

การพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล



นางเนตรรชนี กมลรัตนานันท์

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)

เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR) are the thesis authors' files submitted through the University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต

สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2558

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

DEVELOPMENT OF A TRAINING MODULE OF EVALUATION CAPACITY BUILDING
FOR NURSE TEACHERS

Mrs. Netrachanee Kamonratananun



A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Educational Measurement and
Evaluation

Department of Educational Research and Psychology

Faculty of Education

Chulalongkorn University

Academic Year 2015

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล
โดย	นางเนตรรัชณี กมลรัตนานันท์
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	รองศาสตราจารย์ ดร.กมลวรรณ ตังธนกานนท์

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาคุษฎีบัณฑิต

.....คณบดีคณะครุศาสตร์
(รองศาสตราจารย์ ดร.บัญชา ชลาภิรมย์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(รองศาสตราจารย์ ดร.กมลวรรณ ตังธนกานนท์)

.....กรรมการ
(รองศาสตราจารย์ ดร.โชติกา ภาชีผล)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ณัฐภรณ์ หลาวทอง)

.....กรรมการภายนอกมหาวิทยาลัย
(ผู้ช่วยศาสตราจารย์ ดร.สุวิมล กฤษศยาสา)

เนตรรัชนี กมลรัตนานันท์ : การพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล (DEVELOPMENT OF A TRAINING MODULE OF EVALUATION CAPACITY BUILDING FOR NURSE TEACHERS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร.ศิริเดช สุชีวะ, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: รศ. ดร.กมลวรรณ ดังจนกานนท์, 306 หน้า.

การวิจัยนี้มีวัตถุประสงค์เพื่อ (1) พัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล (2) ศึกษาผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลต่อ ความรู้ ทักษะ และทักษะการประเมินภาคปฏิบัติ (3) ศึกษาผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลต่อความคลาดเคลื่อนจากผู้ประเมิน คือ ความคลาดเคลื่อนจากการกด ปล่อยคะแนน ความคลาดเคลื่อนจากการให้คะแนนแก่กลุ่มตรงกลาง และความคลาดเคลื่อนแบบฮาโล

การวิจัยนี้เป็นการวิจัย และพัฒนา ตัวอย่างในการสำรวจสภาพปัจจุบันของการประเมินภาคปฏิบัติ เป็นอาจารย์จากคณะพยาบาลศาสตร์ใน 4 ภาคของประเทศไทย และจากคณะพยาบาลศาสตร์สวนดุสิต จำนวน 96คน ตัวอย่างในการใช้โมดูล คืออาจารย์พยาบาลจำนวน 17 คน เลือกรูปแบบเจาะจง โดยการจับคู่ และการสุ่มอย่างง่าย เข้ากลุ่มทดลองใช้โมดูล 6 คน และกลุ่มใช้โมดูลจริง 11 คน เครื่องมือที่ใช้เก็บรวบรวมข้อมูลคือ แบบสอบถามสำรวจสภาพปัจจุบัน Rubricการประเมินภาคปฏิบัติ วัสดุทัศนสถานการณ์จำลอง การฝึกปฏิบัติงานของนักศึกษาพยาบาล แบบวัดความรู้ แบบวัดทัศนคติ แบบวัดความพึงพอใจ คู่มือฝึกอบรม และแบบสัมภาษณ์มีโครงสร้าง วิเคราะห์ข้อมูลโดย ร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และ เปรียบเทียบค่าเฉลี่ยคะแนนความรู้ ทัศนคติก่อน และหลังการใช้โมดูลโดยใช้สถิติทดสอบที (paired t-test) และการวิเคราะห์ความแปรปรวนที่มีการวัดซ้ำ (repeated measures ANOVA) วิเคราะห์ความคลาดเคลื่อนจากผู้ประเมินคือ การกด ปล่อยคะแนน การให้คะแนนแก่กลุ่มตรงกลาง และความคลาดเคลื่อนแบบฮาโล ด้วยโมเดลการวิเคราะห์หลายองค์ประกอบของราสช์ (many-facet rasch model) การวิเคราะห์ข้อมูลโดยใช้โปรแกรม SPSS และ FACETS และข้อมูลเชิงคุณภาพใช้การวิเคราะห์เนื้อหา ได้ผลการวิจัยดังนี้

1. ผลการสำรวจสภาพปัจจุบัน พบปัญหาการตัดสินใจผลการฝึกปฏิบัติงานของนักศึกษาพยาบาล จาก 2 ปัจจัยคือ ความเข้าใจในเกณฑ์การประเมินไม่ตรงกัน และความเป็นอัตนัยของผู้ประเมิน
2. การพัฒนาโมดูลได้หน่วยการเรียนรู้ 5 หน่วยคือ (1) ความรู้พื้นฐานด้านการประเมินภาคปฏิบัติ (2) ความรู้เกี่ยวกับความคลาดเคลื่อนจากผู้ประเมิน (3) กรอบของการอ้างอิง (4) การฝึกการสังเกตพฤติกรรมในการประเมิน และ (5) การฝึกการตัดสินใจ
3. ผลของโมดูล ต่อความรู้ ทัศนคติ และทักษะการประเมินภาคปฏิบัติ พบว่าค่าเฉลี่ยคะแนนความรู้ด้านการประเมินภาคปฏิบัติหลังเข้ารับการฝึกอบรมสูงกว่าก่อนเข้ารับการฝึกอบรม อย่างมีนัยสำคัญทางสถิติที่ระดับ .001, ค่าเฉลี่ยคะแนนทัศนคติหลังการใช้โมดูลไม่แตกต่างจากก่อนการใช้โมดูลที่ระดับนัยสำคัญ .05, ผลการวัดทักษะ 3 ครั้งให้ผลการประเมินไม่แตกต่างกันที่ระดับนัยสำคัญ .05
4. ผลการใช้โมดูลต่อความคลาดเคลื่อนจากผู้ประเมิน ก่อน และหลังการฝึกอบรม ทดสอบด้วยสถิติ reliability of separation, separation ratio และ Fixed Chi-square ไม่พบความแตกต่างของการกดคะแนน ปล่อยคะแนน, การให้คะแนนแก่กลุ่มตรงกลาง และความคลาดเคลื่อนแบบฮาโล ที่ระดับนัยสำคัญ .05

ภาควิชา วิชาวิจัยและจิตวิทยาการศึกษา

ลายมือชื่อนิสิต

สาขาวิชา การวัดและประเมินผลการศึกษา

ลายมือชื่อ อ.ที่ปรึกษาหลัก

ปีการศึกษา 2558

ลายมือชื่อ อ.ที่ปรึกษาร่วม

5384458327 : MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION

KEYWORDS: TRAINING MODULE / EVALUATION-CAPACITY BUILDING / NURSE TEACHER

NETRACHANEE KAMONRATANANUN: DEVELOPMENT OF A TRAINING MODULE OF EVALUATION CAPACITY BUILDING FOR NURSE TEACHERS. ADVISOR: ASSOC. PROF.SIRIDEJ SUJIVA, Ph.D., CO-ADVISOR: ASSOC. PROF.KAMONWAN TANGDHANAKANOND, Ph.D., 306 pp.

The purpose of this research were to 1) develop an evaluation capacity building module for nurse teachers in practicum evaluation of student nurses, 2) to investigate the effects of the module on knowledge, attitude and skill in practicum evaluation, and 3) to investigate the effects of the module on rater effects: leniency, severity, central tendency and halo effect.

This research was research and development. Samples for a survey were 96 instructors from the faculty of nursing in 4 regions of Thailand and from Suan Dusit University. The participants consisted of 17 instructors of the faculty of nursing, Suan Dusit University selected by matching the instructors from the same or similar department, then simple random sampling 6 participants to the trial group and others to the actual group alternatively.

Research instruments included (1) a questionnaire for survey on current conditions of practicum evaluation, (2) scoring rubrics, (3) a set of video clips, (4) a training manual, (5) the attitude scales, (6) a test, and (7) structured interview questions. Data analysis were employed by using percentage, mean, standard deviations, paired t-test, repeated measure ANOVA and many-facet rasch measurement. SPSS and FACET program. Results were as follows;

The survey on practicum evaluation problems revealed that the most common problem was grading in practicum because of 2 sources; i e., the raters' subjectivity, and the raters' different perception on the descriptors of the rubrics.

2. The development of the module were consisted of 5 units; (1) basic knowledge in practicum evaluation, (2) rater error training, (3) frame of reference training, (4) behavioral observation training, and (5) decision making training.

3. The effect of the module on knowledge, attitudes and skills in 3 times shown that the mean score of knowledge was higher than that before training at the statistical significant level of .001. However, the mean scores of attitude and skill before the training were not statistically significant difference from those after the training.

4. The effect of the module on rater effects when considering with reliability of separation, separation ratio and Fit mean-square, had no statistically significant difference in leniency, severity, central tendency and halo effects among 3 times.

Department: Educational Research and Psychology Student's Signature

Field of Study: Educational Measurement and Evaluation Advisor's Signature

Evaluation Co-Advisor's Signature

Academic Year: 2015

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้หากก่อให้เกิดประโยชน์ต่อการศึกษาแม้เพียงส่วนใด ขอกราบถวายเป็นสักการะบูชาแด่พระบาทสมเด็จพระจุลจอมเกล้าเจ้าอยู่หัว พระบาทสมเด็จพระมงกุฎเกล้าเจ้าอยู่หัว องค์พระราชทานกำเนิดจุฬาลงกรณ์มหาวิทยาลัย ด้วยความสำนึกในพระมหากรุณาธิคุณอย่างหาที่สุดมิได้

ขอกราบขอบพระคุณรองศาสตราจารย์ ดร. ศิริเดช สุชีวะ และ รองศาสตราจารย์ ดร. กมลวรรณ ตังธนาภานนท์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ให้คำปรึกษา และชี้แนะแนวทางในการศึกษา ตลอดจนเสียสละเวลา และ ให้กำลังใจในการพัฒนา ปรับปรุง และแก้ไข ในทุกขั้นตอนของการวิจัย

ขอกราบขอบพระคุณ ศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี ประธานกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร. โชติกา ภาซีผล ผู้ช่วยศาสตราจารย์ ดร. ณีฎฐภรณ์ หลาวทอง ผู้ช่วยศาสตราจารย์ ดร. สุวิมล กฤษศยาสาส์น ที่ให้ความกรุณาในการเป็นกรรมการสอบวิทยานิพนธ์ ตลอดจนให้คำแนะนำที่เป็นประโยชน์

ขอกราบขอบพระคุณคณาจารย์ภาควิชา วิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยทุกท่าน ที่ให้ความกรุณาในการถ่ายทอดความรู้ ให้ความเมตตาในการให้คำแนะนำ ด้วยความหวังใจด้วยดีเสมอมา

ขอกราบขอบพระคุณผู้ทรงคุณวุฒิ คณะผู้บริหาร และอาจารย์ในสถาบันการศึกษาพยาบาลทุกคน ที่ให้ความอนุเคราะห์ในการพัฒนาการวิจัย รวมถึงคณาจารย์คณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิตที่ให้ความอนุเคราะห์ในการเป็นกลุ่มตัวอย่างในการวิจัย และที่สำคัญยิ่ง ขอกราบขอบพระคุณ อาจารย์ ดร. เบ็ญจา เตากล่ำ คณบดี คณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต ที่ให้ความเมตตาในการชี้แนะแนวทางในการพัฒนาการวิจัย และอนุเคราะห์ในการดำเนินการวิจัยในทุกขั้นตอน

ขอกราบเท้าคุณแม่ สุดใจ กมลรัตนานันท์ ผู้ให้ชีวิต และอยู่เคียงข้างในทุกยามของชีวิตไม่ว่า สุข หรือทุกข์ และคุณพ่อ พันธุ์ศักดิ์ กมลรัตนานันท์ ผู้ล่วงลับ ขอขอบคุณครอบครัว กมลรัตนานันท์ และเพื่อนร่วมรุ่น ผู้เป็นกำลังใจตลอดมา

ท้ายที่สุดนี้ขอขอบพระคุณ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย ในการสนับสนุนทุนวิจัย " ทุน 90 ปี จุฬาลงกรณ์มหาวิทยาลัย" กองทุนรัชดาภิเษกสมโภช ประจำปี 2557

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	1
สารบัญภาพ	1
บทที่ 1.....	1
บทนำ.....	1
ความเป็นมา และความสำคัญของปัญหา.....	1
คำถามวิจัย	4
วัตถุประสงค์ของการวิจัย.....	4
ขอบเขตการวิจัย.....	4
คำจำกัดความที่ใช้ในการวิจัย.....	6
ประโยชน์ที่ได้รับจากการวิจัย	9
บทที่ 2.....	10
เอกสารและงานวิจัยที่เกี่ยวข้อง.....	10
ตอนที่ 1 มโนทัศน์เกี่ยวกับการประเมินการปฏิบัติงาน.....	10
ตอนที่ 2 พยาบาลศาสตร์กับการประเมินภาคปฏิบัติ	39
ตอนที่ 3 มโนทัศน์เกี่ยวกับความคลาดเคลื่อนจากผู้ประเมิน	46
ตอนที่ 4 แนวคิดเกี่ยวกับการเสริมสร้างสมรรถนะการประเมิน.....	73
ตอนที่ 5 การพัฒนาโมดูลการฝึกอบรม.....	82
ตอนที่ 6 กรอบแนวคิดของการวิจัย	100

บทที่ 3.....	103
ขั้นตอนที่ 1 การสำรวจสภาพปัจจุบันของการประเมินภาคปฏิบัติ	105
ขั้นตอนที่ 2 การพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์ พยาบาล	108
ขั้นตอนที่ 3 ทดลองใช้โมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์ พยาบาล	113
ขั้นตอนที่ 4 การประเมินผลการใช้โมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมิน สำหรับอาจารย์พยาบาล	118
ขั้นตอนที่ 5 ปรับปรุงโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์ พยาบาล.....	125
บทที่ 4.....	126
ผลการวิเคราะห์ข้อมูล	126
ตอนที่ 1 ผลการพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์ พยาบาล	126
ตอนที่ 2 ผลการวิเคราะห์ผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมิน สำหรับ อาจารย์พยาบาล ต่อความรู้ ทักษะคติ และทักษะการประเมินภาคปฏิบัติ.....	171
ตอนที่ 3 ผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์ พยาบาล ต่อความคลาดเคลื่อนจากผู้ประเมิน.....	199
บทที่ 5.....	218
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ.....	218
สรุปผลการวิจัย.....	220
ตอนที่ 1 ผลการพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์ พยาบาลในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล.....	220
ตอนที่ 2 ผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์ พยาบาล ต่อความรู้ ทักษะคติ และทักษะการประเมินภาคปฏิบัติ.....	224

ตอนที่ 3 ผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์ พยาบาลต่อความคลาดเคลื่อนจากผู้ประเมิน ดังรายละเอียดต่อไปนี้.....	227
อภิปรายผลการวิจัย.....	230
ข้อเสนอแนะ.....	236
ข้อเสนอแนะสำหรับการนำผลการวิจัยไปใช้.....	236
ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป.....	237
รายการอ้างอิง.....	239
ภาคผนวก.....	253
ภาคผนวก ก.....	254
รายชื่อผู้ทรงคุณวุฒิ.....	254
ภาคผนวก ข.....	258
เครื่องมือที่ใช้ในการวิจัย.....	258
ภาคผนวก ค.....	288
คำสั่งที่ใช้ในการวิเคราะห์คุณภาพการประเมิน.....	288
ผนวก ง.....	292
ผลการวิเคราะห์ข้อมูล.....	292
ภาคผนวก จ.....	304
ภาพกิจกรรมของการพัฒนาโมดูล และการนำโมดูลการฝึกอบรมไปใช้.....	304
ประวัติผู้เขียนวิทยานิพนธ์.....	306

สารบัญตาราง

หน้า

ตารางที่ 1 ตัวอย่างตารางพิมพ์เขียวของแบบทดสอบ (Test specification) ของกิจกรรมการพยาบาล การประเมินสัญญาณชีพ.....	13
ตารางที่ 2 ตัวอย่างรูปรีคแบบภาพรวมของกิจกรรมการพยาบาล การประเมินสัญญาณชีพ	19
ตารางที่ 3 ตัวอย่างรูปรีคแบบแยกส่วน ของกิจกรรมการพยาบาล การประเมินสัญญาณชีพ	20
ตารางที่ 4 เปรียบเทียบข้อดีข้อเสียของดัชนีที่ใช้หาความเที่ยงระหว่างผู้ประเมิน	32
ตารางที่ 5 การเปรียบเทียบรูปแบบของการประเมินคะแนนตามนิยามเชิงปฏิบัติการ	50
ตารางที่ 6 จำนวนผู้ให้ข้อมูลที่ต้องการจำแนกตามสถาบัน และสาขาวิชาที่รับผิดชอบ	106
ตารางที่ 7 ผังข้อสอบในการสร้างแบบวัดความรู้เรื่องการประเมินภาคปฏิบัติ	114
ตารางที่ 8 กำหนดการจัดการฝึกอบรมการเสริมสร้างสมรรถนะการประเมิน	116
ตารางที่ 9 เกณฑ์ในการแปลความหมายระดับคุณภาพของสมรรถนะแต่ละสมรรถนะ	125
ตารางที่ 10 จำนวนผู้ให้ข้อมูลที่ได้ผลตอบกลับจำแนกตามคณะพยาบาลศาสตร์ และสาขาวิชาที่รับผิดชอบ	127
ตารางที่ 11 ข้อมูลทั่วไป.....	128
ตารางที่ 12 ผลการสำรวจปัญหาของการประเมินภาคปฏิบัติ ปัญหาอุปสรรคและแนวทางแก้ไข .	130
ตารางที่ 13 สรุปความคิดเห็นของอาจารย์ผู้สอนภาคปฏิบัติต่อการประเมินภาคปฏิบัติ.....	132
ตารางที่ 14 การสำรวจความต้องการในการใช้รูปรีคการประเมินภาคปฏิบัติ.....	149
ตารางที่ 15 ตัวอย่างรูปรีคการประเมินภาคปฏิบัติ สมรรถนะที่ 1	151
ตารางที่ 16 ความคิดเห็นของผู้ทรงคุณวุฒิภายในฯ ในการเลือกระดับความยากง่ายของกิจกรรมการพยาบาล	152
ตารางที่ 17 การตรวจสอบคุณภาพบทวิดิทัศน์ให้สอดคล้องกับรูปรีคการประเมิน และสถานการณ์.....	154
ตารางที่ 18 การตรวจสอบคุณภาพของรูปรีคที่ใช้ในการประเมินภาคปฏิบัติ	160
ตารางที่ 19 การตรวจสอบคุณภาพของแบบวัดความพึงพอใจต่อโมดูลฯ.....	162

ตารางที่ 20 การตรวจสอบคุณภาพด้านความสอดคล้องระหว่างข้อความกับวัตถุประสงค์ของการวัดของแบบวัดความรู้ด้านการประเมินภาคปฏิบัติ.....	163
ตารางที่ 21 การตรวจสอบความตรงเชิงเนื้อหาของแบบทดสอบความรู้พื้นฐานเกี่ยวกับการประเมินภาคปฏิบัติ.....	166
ตารางที่ 22 การตรวจสอบความตรงเชิงเนื้อหาของแบบประเมินทัศนคติต่อการประเมินภาคปฏิบัติ.....	168
ตารางที่ 23 การตรวจสอบความสัมพันธ์รายข้อของแบบวัดทัศนคติเกี่ยวกับการประเมินภาคปฏิบัติ.....	169
ตารางที่ 24 การตรวจสอบความตรงเชิงเนื้อหาของคู่มือสำหรับอบรมการเสริมสร้างสมรรถนะการประเมิน ด้วยค่า Item Objective Congruence Index (IOC).....	169
ตารางที่ 25 ข้อมูลทั่วไป.....	172
ตารางที่ 26 การเปรียบเทียบค่าเฉลี่ยคะแนนความรู้เกี่ยวกับการประเมินภาคปฏิบัติของอาจารย์พยาบาล ก่อนและหลังการใช้โมดูลฯ.....	172
ตารางที่ 27 การเปรียบเทียบค่าเฉลี่ยของคะแนนทัศนคติต่อการประเมินภาคปฏิบัติของอาจารย์พยาบาล ก่อนและหลังการใช้โมดูลการฝึกอบรมการเสริมสร้างสมรรถนะการประเมิน.....	173
ตารางที่ 28 ผลการวิเคราะห์ความคลาดเคลื่อนของปฏิสัมพันธ์ระหว่างผู้ประเมิน กับเวลาในการวัด 2 ครั้ง (Rater X Time interactions) จากการวัดด้วยวิธีทัศนชุดเดียวกัน.....	175
ตารางที่ 29 ผลการวัดความพึงพอใจของผู้เข้ารับการฝึกอบรมกลุ่มทดลองใช้โมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล.....	176
ตารางที่ 30 ข้อมูลทั่วไป.....	181
ตารางที่ 31 ผลการทดสอบ sphericity และผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยของคะแนนความรู้พื้นฐานด้านการประเมินภาคปฏิบัติ จากการวัดความรู้ 3 ครั้งด้วยแบบทดสอบชุดเดียวกัน.....	182
ตารางที่ 32 ผลการวิเคราะห์ความแปรปรวน และผลการเปรียบเทียบค่าเฉลี่ยรายคู่ของค่าเฉลี่ยของคะแนนความรู้เกี่ยวกับการประเมินภาคปฏิบัติ ระหว่างเวลาของข้อมูลชุดวัดซ้ำด้วยแบบสอบฉบับเดิม.....	183
ตารางที่ 33 ผลการทดสอบ sphericity และผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยของคะแนนทัศนคติเกี่ยวกับการประเมินภาคปฏิบัติ จากการวัดความรู้ 3 ครั้งด้วยวัดฉบับเดียวกัน.....	184

ตารางที่ 34 ผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยของคะแนนทัศนคติต่อการประเมิน ภาคปฏิบัติ จากการวัดซ้ำ 3 ครั้งด้วยแบบวัดฉบับเดียวกัน	185
ตารางที่ 35 ผลการเปรียบเทียบรายคู่ของคะแนนทักษะจากการวัดก่อน และหลังการ.....	189
ตารางที่ 36 ความสอดคล้องระหว่างผู้ประเมิน และความเที่ยงภายในผู้ประเมิน	191
ตารางที่ 37 ผลการวัดความพึงพอใจของผู้เข้ารับการฝึกอบรมกลุ่มใช้โมดูลการฝึกอบรมเพื่อ เสริมสร้าง สมรรถนะการประเมินสำหรับอาจารย์พยาบาล.....	192
ตารางที่ 38 เปรียบเทียบผลการประเมินก่อน และหลังการฝึกอบรม ของฟาเซทผู้ประเมิน	203
ตารางที่ 39 สถิติสรุปรวมในระดับกลุ่มสำหรับฟาเซทผู้ประเมิน	205
ตารางที่ 40 เปรียบเทียบผลการวัดระดับความสามารถของผู้ถูกประเมิน ก่อน และหลังการ ฝึกอบรม.....	209
ตารางที่ 41 สถิติสรุปรวมในระดับกลุ่มสำหรับฟาเซทผู้ถูกประเมิน	211
ตารางที่ 42 เปรียบเทียบผลการวัดก่อน และหลังการใช้โมดูลฝึกอบรมของสมรรถนะที่ใช้วัด.....	214
ตารางที่ 43 สถิติสรุปรวมในระดับกลุ่มสำหรับฟาเซทสมรรถนะ.....	216

สารบัญภาพ

	หน้า
ภาพที่ 1 กรอบแนวคิดของการวิจัย.....	102
ภาพที่ 2 ขั้นตอนการวิจัย.....	104
ภาพที่ 3 ขั้นตอนการสุ่มเลือกชุดวีดิทัศน์.....	110
ภาพที่ 4 ตัวอย่างตอนหนึ่งในวีดิทัศน์ สถานการณ์จำลองเพื่อเสริมสร้างสมรรถนะการประเมิน สำหรับอาจารย์พยาบาล.....	161
ภาพที่ 5 แสดงแผนที่ของตัวแปร หรือฟาเซทที่ศึกษา (Variable map).....	217



บทที่ 1

บทนำ

ความเป็นมา และความสำคัญของปัญหา

พ.ร.บ. การศึกษาแห่งชาติ พ.ศ. 2542 ให้ความสำคัญกับการประเมินตามสภาพจริง เพื่อให้ผู้เรียนสามารถเชื่อมโยงชีวิตจริงกับการเรียนได้อย่างกลมกลืน (2544) การวัด และประเมินผลทางการศึกษาในปัจจุบันจึงมุ่งเน้นการประเมินการปฏิบัติงาน (performance assessment) ซึ่งเป็นการทดสอบการปฏิบัติงานของผู้เรียน โดยวัดกระบวนการปฏิบัติงาน และ หรือผลงานซึ่งเกิดจากการเรียนรู้ที่เน้นการสังเกตพฤติกรรม ที่ไม่ได้เน้นเฉพาะการใช้ทักษะทางร่างกาย แต่เป็นการผสมผสาน ความรู้ความเข้าใจ ความคิดสร้างสรรค์ และความสามารถอื่นเพื่อพัฒนาผู้เรียนให้มีความสมบูรณ์ ด้านร่างกาย จิตใจ และสติปัญญา (เอมอร จังศิริพรปกรณ์, 2550) รวมถึงการศึกษาด้านพยาบาลศาสตร์ ที่ผู้เรียนมีการเรียนรู้ทางทฤษฎี และ สอดแทรกหลักคุณธรรม จริยธรรม และศิลปะในการให้การพยาบาลแบบองค์รวม คือดูแลทั้งร่างกาย จิตใจ อารมณ์สังคม และ จิตวิญญาณ ซึ่งถ่ายทอดผ่านการฝึกปฏิบัติงานบนหอผู้ป่วย เพื่อให้เกิดความรู้ที่เกิดจากประสบการณ์เฉพาะตน และนำไปสู่การสำเร็จเป็นพยาบาลวิชาชีพ โดยกระบวนการพัฒนานักศึกษาพยาบาลนั้นต้องอาศัยกระบวนการของการวัด และประเมินผลภาคปฏิบัติ โดยมีปัจจัยสำคัญคือ คุณภาพการประเมินในการนิเทศงานบนหอผู้ป่วยของอาจารย์ผู้สอนภาคปฏิบัติ

คุณภาพของการประเมินการปฏิบัติงาน ขึ้นอยู่กับหลายปัจจัย คือ 1) คุณภาพของเครื่องมือที่ใช้ประเมิน และ 2) สมรรถนะของผู้ประเมิน (Nitko, 2007; ชวลิต ชูกำแพง, 2551; สุวิมล ว่องวานิช, 2546) ในด้านของคุณภาพของเครื่องมือ พิจารณาทั้งความตรง และความเที่ยง รวมถึงงานที่ใช้ในการประเมิน ซึ่งอาจประเมินกระบวนการ หรือประเมินที่ผลผลิตของงานก็ได้ แต่จะต้องสอดคล้อง หรือตอบคำถามสิ่งที่ต้องการประเมินได้ และเครื่องมือที่ใช้ในการประเมินจะต้องสอดคล้องกับงานด้วย เครื่องมือที่ใช้ในการประเมินผลการเรียนรู้ภาคปฏิบัติของนักศึกษาพยาบาลในปัจจุบัน นิยมใช้แนวคิดพื้นฐานมาจากสมรรถนะทางวิชาชีพซึ่งกำหนดโดยสภาการพยาบาล ร่วมกับแนวคิดตามกรอบมาตรฐานคุณวุฒิระดับอุดมศึกษาแห่งชาติ (Thai Qualifications Framework for Higher Education, TQF:HEd) ที่ว่าด้วยมาตรฐานผลการเรียนรู้ ประกอบด้วย 1) คุณธรรม จริยธรรม 2) ความรู้ 3) ทักษะทางปัญญา

4) ความสัมพันธ์ระหว่างบุคคล และความรับผิดชอบ 5) ทักษะการวิเคราะห์เชิงตัวเลข การสื่อสาร และการใช้เทคโนโลยีสารสนเทศ และ สำหรับสาขาพยาบาลศาสตร์ มีการเพิ่มมาตรฐานคุณวุฒิด้านที่ 6 คือ ทักษะการปฏิบัติทางวิชาชีพสาขาพยาบาลศาสตร์ (สำนักงานคณะกรรมการการอุดมศึกษา, 2556) โดยแต่ละสถาบันมีการประยุกต์ใช้ด้วยแบบฟอร์มเฉพาะแตกต่างกันไป

ศาสตร์ด้านการพยาบาลให้ความสำคัญกับการประเมินภาคปฏิบัติอย่างมากเนื่องจากมีธรรมชาติของงานที่สอดคล้องกับการปฏิบัติงานจริงของพยาบาล และมีข้อดีต่อทั้งผู้เรียน และผู้สอน เช่น ช่วยทำให้การเรียนรู้ตามวัตถุประสงค์การเรียนรู้ที่ค่อนข้างซับซ้อน สามารถวัดประเมินผลได้ชัดเจนขึ้นจากผลงานการปฏิบัติ อย่างไรก็ตามข้อดีของการประเมินภาคปฏิบัติ เกี่ยวกับคุณภาพของการประเมิน คือ มีความเป็นปรนัยต่ำ ซึ่งเป็นผลมาจากความคลาดเคลื่อนจากผู้ประเมิน (rater effect) (Nitko, 2007)

ความคลาดเคลื่อนจากผู้ประเมิน (Rater effect, rater error or rater bias) เป็นปัญหาหลักของการประเมินการปฏิบัติงาน (Reynolds, Livingston, & Willson, 2009) แบ่งเป็น 2 ประเภทหลัก คือ ความคลาดเคลื่อนแบบสุ่ม และความคลาดเคลื่อนแบบเป็นระบบ (Yue, 2011) โดยควรให้ความสำคัญต่อความคลาดเคลื่อนแบบเป็นระบบ เนื่องจากมีผลกระทบต่อภาพรวมของการประเมินทั้งหมดมากกว่าความคลาดเคลื่อนแบบสุ่มที่ไม่ได้เกิดผลกระทบในภาพรวม (Kane, 2011) ความคลาดเคลื่อนแบบ เป็นระบบมีหลายประเภท เช่น ความคลาดเคลื่อนจาก การกด หรือ ปล่อยคะแนน (severity/leniency effect) หรือ ผู้ประเมินบางคนให้คะแนนเกาะกลุ่มตรงกลางทั้งกลุ่ม (central tendency effect) หรือใช้ทัศนคติในด้านบวก หรือลบต่อมิติหนึ่งไปใช้ในการประเมินมิติอื่น ๆ ที่ไม่เกี่ยวข้อง (halo effect) การแก้ปัญหาความคลาดเคลื่อนจากผู้ประเมินที่นิยมใช้ในปัจจุบันคือ การเตรียมความพร้อมของผู้ประเมิน (Donaldson & Gray, 2012) คืออาจารย์พยาบาลผู้สอนภาคปฏิบัติ เป้าหมายสำคัญคือการเสริมสร้างสมรรถนะการประเมิน ทั้งในระดับบุคคล และองค์กร ผลลัพธ์ที่ต้องการให้เกิด 3 ด้านหลักคือ ด้านความรู้ ด้านทัศนคติ และ ด้านทักษะในการประเมินภาคปฏิบัติ (Taut, 2007) การวิจัยครั้งนี้ประยุกต์ใช้โมเดลสหวิทยาการ (Multidisciplinary model) ของ Preskill และ Boyle (2008) เพื่อการเสริมสร้างสมรรถนะการประเมิน มีหัวใจสำคัญที่การมีส่วนร่วมของผู้มีส่วนได้ส่วนเสีย และการประเมินกระแสหลัก (Mainstreaming evaluation) เป็นการสร้างวัฒนธรรมการประเมินภายในองค์กรทำให้เกิดความคุ้นชิน และรู้สึกว่าการประเมินเป็นส่วนหนึ่งของชีวิตการทำงานประจำวัน โดยมีเป้าหมายสำคัญคือ มีสมรรถนะการประเมินอย่างยั่งยืนภายในองค์กร ซึ่งวัดได้จากการเรียนรู้เกี่ยวกับการประเมิน การเก็บรวบรวมข้อมูล การวิเคราะห์ข้อมูล การแปลความหมายข้อมูล และการใช้ข้อมูลการประเมินเพื่อการตัดสินใจ และนำผลการตัดสินใจไปสู่ การลงมือปฏิบัติเพื่อแก้ไข

ปัญหา หรือพัฒนาองค์กร โมเดลทหุวิทยาการมีกลยุทธ์ที่ทำให้เกิดการมีส่วนร่วม ของผู้มีส่วนได้ส่วนเสีย และการสร้างวัฒนธรรมการประเมินอย่างยั่งยืนผ่านการสอน และการเรียนรู้เกี่ยวกับการประเมินจาก 3 ขั้นตอน คือ 1) การเริ่มต้น การออกแบบ การลงมือปฏิบัติ และการประเมินผล และ 2) การสร้างความยั่งยืน และ 3) การเผยแพร่เป็นการเชื่อมโยงระหว่างขั้นตอนที่ 1 และขั้นตอนที่ 2 โดยขั้นตอนที่ 1 การเริ่มต้น การออกแบบ การลงมือปฏิบัติ และการประเมินผล โดยผู้มีส่วนได้ส่วนเสียจะต้องมีความเข้าใจตรงกันเกี่ยวกับการประเมิน และการสร้างสมรรถนะการประเมิน มีแรงจูงใจร่วมกัน และมีความคาดหวังถึงเป้าหมายของความสำเร็จที่ตรงกัน

โมเดลทหุวิทยาการ มีกลยุทธ์ในการดำเนินการให้เลือกใช้ได้ 10 กลยุทธ์ สำหรับการวิจัยครั้งนี้ เลือกใช้ 3 กลยุทธ์ คือ การฝึกอบรม การเข้ามามีส่วนร่วมในกระบวนการประเมิน และการให้คำปรึกษา โดยมีวัตถุประสงค์เพื่อให้เกิดการพัฒนาความรู้ ทักษะและทัศนคติที่ดีต่อการประเมิน ขั้นตอนที่ 2 การสร้างความยั่งยืนให้กับสมรรถนะการประเมิน ซึ่งประกอบด้วยกระบวนการ การปฏิบัติ ระดับนโยบาย และทรัพยากรที่สนับสนุนการประเมินอย่างชัดเจน และขั้นตอนที่ 3 เป็นจุดเชื่อมขั้นตอนที่ 1และขั้นตอนที่ 2 โดยเชื่อว่าเมื่อเกิดขั้นตอนของการเรียนรู้ การสร้างทักษะ และทัศนคติที่ดีต่อการประเมินจากการได้รับแรงจูงใจ และการมีส่วนร่วมแล้วนำไปสู่การสร้าง ความยั่งยืน จากการค้นหา วิเคราะห์ และ ใช้ผลการประเมินเพื่อพัฒนาองค์กรจนนำไปสู่การแลกเปลี่ยนความรู้ ทักษะและทัศนคติสู่หน่วยงานภายนอก (Preskill & Boyle, 2008) การประเมินว่าผู้เข้ารับการฝึกอบรมมีคุณภาพการประเมินในระดับใด มีความคลาดเคลื่อนในการประเมินจากผู้ประเมินมากน้อยเพียงใด ซึ่งสามารถตรวจสอบได้ด้วยการใช้ดัชนีความสอดคล้อง หรือค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างผู้ประเมิน (Intraclass Correlation Coefficient: ICC) หรือ ความเที่ยงระหว่างผู้ประเมิน (Inter Rater Reliability: IRR) และในปัจจุบันพบว่ามีการใช้โมเดลหลายองค์ประกอบของราส์ซ (Many-Facet Rasch Model: MFRM) ในการตรวจสอบความคลาดเคลื่อนจากผู้ประเมิน (Xi & Mollaun, 2011)

จากการศึกษาค้นคว้าเอกสารงานวิจัยที่เกี่ยวข้องกับปัญหา หรือการแก้ปัญหาการศึกษาภาคปฏิบัติในนักศึกษาพยาบาล ในด้านพยาบาลศาสตร์ในประเทศไทยพบน้อยมาก ส่วนใหญ่ เป็นการศึกษาเพื่อพัฒนาคุณภาพของเครื่องมือที่ใช้ในการประเมินภาคปฏิบัติ (จิราภา คงสมฤทธิ์, 2544; ปิยะรัตน์ หยกสุริยันต์, 2549; อนัญญา คูอาริยะกุล, 2554) ซึ่งไม่ตรงกับประเด็นปัญหาที่แท้จริง ในขณะที่ในต่างประเทศมีการศึกษาถึงประเด็นปัญหาเกี่ยวกับการตัดสินใจประเมินทักษะทางวิชาชีพพยาบาล การกำหนดเกรด การใช้สมรรถนะเป็นฐานในการประเมิน และยังพบว่าการกำหนดเกรดในการ

ประเมินผลการเรียนภาคปฏิบัติยังคงเป็นปัญหามาอย่างต่อเนื่องยาวนาน (Chambers, 1998) เนื่องจากต้องอาศัยการตัดสินจากผู้ประเมินซึ่งเผชิญปัญหาในเรื่องความเป็นอัตนัย (subjective) ของผู้ประเมิน ดังนั้นการเสริมสร้างสมรรถนะการประเมินให้แก่อาจารย์พยาบาลที่ทำหน้าที่ประเมินภาคปฏิบัติ นักศึกษาพยาบาล ผ่านโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินจึงเป็นการแก้ปัญหาที่ตรงจุด ด้วยเหตุนี้ผู้วิจัยจึงสนใจศึกษาเพื่อให้ได้แนวทางปฏิบัติที่ชัดเจนในการแก้ปัญหาความคลาดเคลื่อนจากผู้ประเมิน และเพิ่มคุณภาพการประเมินภาคปฏิบัติในนักศึกษาพยาบาล

คำถามวิจัย

1. โมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลควรเป็นอย่างไร
2. ผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลต่อความรู้ ทักษะ และทักษะด้านการประเมินภาคปฏิบัติเป็นอย่างไร
3. ผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลต่อความคลาดเคลื่อนจากผู้ประเมินเป็นอย่างไร

วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล
2. เพื่อศึกษาผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลต่อ ความรู้ ทักษะ และทักษะการประเมินภาคปฏิบัติ
3. เพื่อศึกษาผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลต่อความคลาดเคลื่อนจากผู้ประเมิน

ขอบเขตการวิจัย

1. การพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล ในรูปการฝึกอบรม ทำเฉพาะในอาจารย์พยาบาลที่ทำหน้าที่เป็นผู้สอนภาคปฏิบัติของอาจารย์พยาบาล ในสังกัดสำนักงานคณะกรรมการการอุดมศึกษา
2. โมดูลการฝึกอบรมประกอบด้วย 5 หน่วยการเรียนรู้คือ

- 2.1 ความรู้พื้นฐานเกี่ยวกับการประเมินภาคปฏิบัติ
- 2.2 ความรู้พื้นฐานเกี่ยวกับความคลาดเคลื่อนจากผู้ประเมิน
- 2.3 กรอบของการอ้างอิง
- 2.4 การฝึกการสังเกตพฤติกรรม
- 2.5 การฝึกการตัดสินใจ

3. สมรรถนะการประเมินมี 3 ด้าน ประกอบด้วย

3.1 ด้านความรู้ เกี่ยวกับการประเมินภาคปฏิบัติ ประกอบด้วย ความรู้พื้นฐานด้านการประเมินภาคปฏิบัติ, ความรู้เกี่ยวกับความคลาดเคลื่อนจากผู้ประเมิน, กรอบของการอ้างอิง, หลักของการสังเกตพฤติกรรมในการประเมิน และหลักในการตัดสินใจ

3.2 ทักษะที่เกี่ยวข้องกับการประเมินภาคปฏิบัติ และกระบวนการของการประเมินภาคปฏิบัติ

3.3 ทักษะในการประเมินภาคปฏิบัติเป็นทักษะจากการฝึกการประเมินกับวิธีทัศน์ สถานการณ์จำลองการฝึกปฏิบัติงานของนักศึกษาพยาบาลในรายวิชาปฏิบัติการพยาบาลพื้นฐานของนักศึกษาชั้นปีที่ 2

4. สมรรถนะที่ใช้ในการประเมินภาคปฏิบัติประกอบด้วย 7 สมรรถนะหลักคือ 1) คุณธรรม จริยธรรม 2) ความรู้ 3) ทักษะทางปัญญา 4) ความสัมพันธ์ระหว่างบุคคล และความรับผิดชอบ 5) ทักษะการวิเคราะห์เชิงตัวเลข การสื่อสาร และการใช้เทคโนโลยีสารสนเทศ 6) ทักษะการปฏิบัติทางวิชาชีพสาขาพยาบาลศาสตร์ และ 7) เอกลักษณ์ของมหาวิทยาลัยสวนดุสิต

5. ความคลาดเคลื่อนจากผู้ประเมิน ประกอบด้วยความคลาดเคลื่อนจากการกด ปุ่ม คีย์บอร์ด คัดลอก ความคลาดเคลื่อนจากการให้คะแนนเกาะกลุ่มตรงกลาง และความคลาดเคลื่อนแบบฮาลอ

6. การพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมิน โดยอาศัยกระบวนการคือ

- 1) วิเคราะห์สภาพปัจจุบันของการประเมินภาคปฏิบัติทางวิชาชีพพยาบาลในนักศึกษาพยาบาล
- 2) การออกแบบโมดูลการฝึกอบรมผู้ประเมินภาคปฏิบัติ 3) การทดลองใช้โมดูลการฝึกอบรมภาคปฏิบัติ
- 4) ประเมินผลการนำไปใช้ของโมดูลการฝึกอบรมผู้ประเมินภาคปฏิบัติ และ 5) การปรับปรุงโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินของอาจารย์พยาบาลในการประเมินภาคปฏิบัติในนักศึกษาพยาบาล

7. การประเมินผลของโมดูลการฝึกอบรมเพื่อการเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลเป็นการศึกษาผลของการใช้โมดูลการฝึกอบรม ต่อความรู้ ทักษะ และทักษะ ต่อการประเมินภาคปฏิบัติ ก่อน และหลังการใช้โมดูล ทันที และหลังการใช้โมดูล 3 เดือน

8. การประเมินผลการใช้โมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมิน สำหรับอาจารย์พยาบาลต่อความคลาดเคลื่อนจากผู้ประเมินเป็นการศึกษาผลของการใช้โมดูลต่อความคลาดเคลื่อนจากการกด ปุ่มคั่น การให้คะแนนเกาะกลุ่มตรงกลาง และความคลาดเคลื่อนแบบฮาโล

คำจำกัดความที่ใช้ในการวิจัย

1. โมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมิน สำหรับอาจารย์พยาบาล หมายถึง หน่วยการเรียนรู้ หรือ หน่วยการฝึกอบรมที่ได้รับการพัฒนาจากแนวคิดของ โมเดลพหุวิทยาการของ Preskill และ Boyle โดยใช้กลยุทธ์การฝึกอบรม และการให้คำปรึกษา ประกอบด้วย 5 หน่วยการเรียนรู้คือ

1.1 ความรู้พื้นฐานเกี่ยวกับการประเมินภาคปฏิบัติ มีเนื้อหาครอบคลุม ความหมายของการประเมินภาคปฏิบัติ กระบวนการประเมินภาคปฏิบัติ การออกแบบ การเลือกงาน การสร้างเครื่องมือในการประเมิน การตรวจสอบคุณภาพของเครื่องมือ และการตัดสินผลการประเมิน และการทดสอบความรู้ท้ายบท

1.2 ความรู้พื้นฐานเกี่ยวกับความคลาดเคลื่อนจากผู้ประเมิน มีเนื้อหาครอบคลุม ความหมายของความคลาดเคลื่อนแต่ละประเภท สาเหตุ และวิธีแก้ไข รวมถึงการนำเสนอตัวอย่างของคะแนนที่แสดงถึงความคลาดเคลื่อนประเภทต่าง ๆ และกิจกรรมท้ายบทมีการให้ข้อมูลเพื่อให้ผู้เข้ารับการฝึกอบรมวิเคราะห์เพื่อระบุความคลาดเคลื่อนจากข้อมูลดังกล่าว

1.3 กรอบของการอ้างอิง เป็นการอธิบายถึงระดับคุณภาพของรูปชีวิตที่ใช้ประเมินในแต่ละระดับ และการนำเสนอตัวอย่างวิธีที่ต้นสถานการณัจำลองการฝึกปฏิบัติงานของนักศึกษาพยาบาลที่มีระดับคุณภาพบนมาตรฐานการวัด ระดับ 1 ถึง 4 และเปิดโอกาสให้ผู้เข้ารับการฝึกอบรมอภิปราย

1.4 การฝึกทักษะด้านการสังเกต มีเนื้อหาครอบคลุม ความหมาย ความสำคัญ กระบวนการของการสังเกต กฎของการสังเกต กลยุทธ์ที่ใช้ในการสังเกตพฤติกรรม การทดลองสังเกตพฤติกรรม และนำผลการสังเกตมาอภิปรายร่วมกัน

1.5 ฝึกทักษะการตัดสินใจ มีเนื้อหาครอบคลุม ความสำคัญ ทฤษฎีพื้นฐาน และกระบวนการตัดสินใจตามสัญชาตญาณ กระบวนการตัดสินใจที่ถูกต้อง และการใช้สถานการณ์ในวีดิทัศน์เพื่อฝึกการตัดสินใจ จากนั้นสุ่มสอบถามผู้เข้ารับการฝึกอบรมถึงกระบวนการตัดสินใจ และอภิปรายร่วมกัน

2. สมรรถนะการประเมิน หมายถึงสมรรถนะด้านความรู้ ทักษะ และทักษะของอาจารย์พยาบาล ในการประเมินภาคปฏิบัตินักศึกษาพยาบาล

3. สมรรถนะที่ใช้ในการประเมินทักษะภาคปฏิบัติทางวิชาชีพพยาบาล หมายถึง 7 สมรรถนะหลัก รวม 26 สมรรถนะย่อย คือ

3.1 คุณธรรม จริยธรรม มี 5 สมรรถนะย่อยคือ (1) ความซื่อสัตย์ (2) ควบคุมตนเอง (3) ยอมรับความแตกต่างของความเชื่อ (4) จิตอาสา และ (5) รักษาจรรยาบรรณวิชาชีพ พึงทักสิทธิ์ของผู้ใช้บริการ

3.2 ความรู้ มี 2 สมรรถนะย่อยคือ (1) บอกเหตุผล และหลักการสำคัญของการปฏิบัติกิจกรรมพยาบาลได้ และ (2) มีส่วนร่วมในการอภิปรายแสดงถึงการมีความรู้ความเข้าใจในสาระสำคัญ

3.3 ทักษะทางปัญญา มี 3 สมรรถนะย่อย คือ (1) ระบุจุดแข็ง จุดอ่อน และสิ่งที่ต้องพัฒนาตนเองเกี่ยวกับความรู้ และทักษะการปฏิบัติงาน (2) ค้นคว้าได้ตรงประเด็นจากหลากหลายแหล่งข้อมูล (3) วางแผนการพยาบาลตามกระบวนการพยาบาล และแก้ไขปัญหาได้เหมาะสมกับสถานการณ์

3.4 ความสัมพันธ์ระหว่างบุคคล และความรับผิดชอบ มี 4 สมรรถนะย่อยคือ (1) ปฏิสัมพันธ์ต่อบุคลากรในทีมสุขภาพฯ หรือผู้ใช้บริการ และการปรับตัวต่อวิชาชีพ (2) แสดงถึงความคิดบวกต่อวิชาชีพ และต่อผู้อื่น (3) แสดงตนเป็นสมาชิกที่ดีของกลุ่มตามบทบาทของตน (4) พัฒนาคณะอย่างต่อเนื่อง

3.5 ทักษะการวิเคราะห์เชิงตัวเลข การสื่อสาร และ การใช้เทคโนโลยีสารสนเทศ และ สำหรับสาขาพยาบาลศาสตร์ มี 3 สมรรถนะย่อย คือ (1) การเข้าใจความหมายของตัวเอง และนำข้อมูลตัวเลขมาใช้ได้ถูกต้อง (2) ทักษะการสื่อสารด้านการฟัง พูด อ่าน เขียน ภาษาไทย และภาษาต่างประเทศที่ถูกต้องครบถ้วน (3) การสืบค้นข้อมูลที่เป็นปัจจุบันจากแหล่งต่าง ๆ ทั้งจำนวน เอกสารอ้างอิง แหล่งอ้างอิง ความน่าเชื่อถือของแหล่งอ้างอิง

3.6 ทักษะการปฏิบัติทางวิชาชีพสาขาพยาบาลศาสตร์ มี 5 สมรรถนะย่อย คือ (1) ปฏิบัติการพยาบาลอย่างเป็นองค์รวม โดยใช้กระบวนการพยาบาลที่สอดคล้องกับสถานการณ์ และสามารถสะท้อนถึงการสร้างเสริมสุขภาพ การป้องกัน การรักษาและการฟื้นฟูสภาพ (2) ปฏิบัติการพยาบาลถูกหลักการพยาบาลในกิจกรรมพยาบาล นั้นๆ (3) ปฏิบัติการพยาบาลถูกหลักการป้องกันการแพร่กระจายเชื้อ (4) ปฏิบัติการพยาบาลด้วยความเมตตา กรุณา และเอื้ออาทร และทันเหตุการณ์สอดคล้องกับความต้องการของผู้ใช้บริการ (5) ปฏิบัติการพยาบาลโดยคำนึงถึงความเป็นปัจเจกบุคคล และความหลากหลายทางวัฒนธรรม (6) ปฏิบัติการพยาบาลโดยยึดหลักความปลอดภัย

3.7 เอกลักษณ์ของมหาวิทยาลัยสวนดุสิต มี 3 สมรรถนะคือ (1) แต่งกายถูกระเบียบ สะอาด เรียบร้อย ผมเรียบร้อยไม่ยุ่งเหยิง รองเท้าหุ้มส้นถูกระเบียบสะอาดทำทางสง่าในแต่ละอริยาบท ทั้งขณะนั่ง ยืนหรือเดิน (2) หน้าตายิ้มแย้ม ท่าทางเป็นมิตร พูดจาสำเนียงสุภาพ งดงาม (3) ความประณีต พิถีพิถันในการทำงาน

4. การประเมินภาคปฏิบัติ หมายถึง การประเมินการปฏิบัติงานของนักศึกษาพยาบาลบนหอผู้ป่วย ในรูปวีดิทัศน์สถานการณ์จำลองการขึ้นฝึกปฏิบัติงานของนักศึกษาพยาบาลชั้นปีที่ 2 รายวิชา ปฏิบัติการพยาบาลพื้นฐาน

5. ความคลาดเคลื่อนจากผู้ประเมินหมายถึง ความคลาดเคลื่อนของผลการประเมินที่เกิดจากการที่ผู้ประเมินใช้ความเป็นอัตนัยของผู้ประเมินในการประเมินผลการปฏิบัติงาน ในการวิจัยครั้งนี้ศึกษา 3 ประเภท คือ

5.1 ความคลาดเคลื่อนจากการกด ปล่อยคะแนน มีลักษณะของคะแนนที่กระจุกตัวอยู่ส่วนปลายของมาตรการวัด อาจเป็นปลายมาตรด้านคะแนนน้อย หรือด้านคะแนนมาก

5.2 ความคลาดเคลื่อนจากการให้คะแนนเกาะกลุ่มตรงกลาง มีลักษณะของคะแนนกระจุกตัวอยู่ที่ส่วนกลางของมาตรการวัด

5.3 ความคลาดเคลื่อนแบบฮาโล มีลักษณะคะแนนของทุกมิติการวัดกระจุกตัวอยู่ที่ระดับคะแนนใดคะแนนหนึ่ง

ประโยชน์ที่ได้รับจากการวิจัย

องค์ความรู้ที่ได้จากการศึกษาค้นคว้า และพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลที่เป็นผู้สอนภาคปฏิบัติจะก่อให้เกิดประโยชน์ดังนี้

1. ประโยชน์เชิงวิชาการ

1.1 ได้นำองค์ความรู้ด้านการวัดและประเมินผล ด้านการประเมินภาคปฏิบัติ สู่การปฏิบัติที่เป็นรูปธรรม

1.2 ได้โมดูลการสร้างสมรรถนะการประเมินภาคปฏิบัติในนักศึกษาพยาบาล สำหรับอาจารย์พยาบาลที่เป็นผู้สอนภาคปฏิบัติที่ผ่านการตรวจสอบคุณภาพของเนื้อหา ความต้องการในการใช้ และผลของโมดูลการฝึกอบรม เป็นโมดูลสำเร็จรูปที่สะดวกสามารถนำไปใช้ได้ในวันพฤหัสบดีเดียวกันแต่ต่างบริบทกัน

2. ประโยชน์เชิงปฏิบัติการ

2.1 ได้แนวปฏิบัติที่ดีที่สามารถนำไปใช้ในการแก้ไขปัญหาการประเมินภาคปฏิบัติในนักศึกษาพยาบาลสำหรับอาจารย์พยาบาล

2.2 ได้เครื่องมือ และคู่มือการปฏิบัติที่นำไปสู่การแก้ไขปัญหาในการประเมินภาคปฏิบัติงาน

3. ประโยชน์เชิงนโยบาย

3.1 ได้แนวทางในการวางนโยบายเกี่ยวกับการวัด และประเมินผลการเรียนในภาคปฏิบัติของนักศึกษาพยาบาล

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาแนวคิด ทฤษฎี เอกสารและงานวิจัยที่เกี่ยวข้องกับ การประเมิน ภาควิชาปฏิบัติในนักศึกษาพยาบาลของอาจารย์พยาบาล โดยแบ่งเป็น 5 ตอน ดังนี้

- ตอนที่ 1 มโนทัศน์เกี่ยวกับการประเมินการปฏิบัติงาน
- ตอนที่ 2 การศึกษาด้านพยาบาลศาสตร์กับการประเมินภาควิชาปฏิบัติ
- ตอนที่ 3 มโนทัศน์เกี่ยวกับความคลาดเคลื่อนจากผู้ประเมิน
- ตอนที่ 4 มโนทัศน์เกี่ยวกับการเสริมสร้างสมรรถนะการประเมิน
- ตอนที่ 5 มโนทัศน์เกี่ยวกับการพัฒนาโมดูลการฝึกอบรม
- ตอนที่ 6 กรอบแนวคิดในการวิจัย

ตอนที่ 1 มโนทัศน์เกี่ยวกับการประเมินการปฏิบัติงาน

1. ความหมายของการประเมินการปฏิบัติงาน

ชวลิต ชูกำแพง (2551) เป็นการประเมินองค์ประกอบสำคัญสองส่วนคือ งาน หรือกิจกรรม ที่ครูมอบหมายให้นักเรียนแสดงออกตามคำสั่งเพื่อให้ได้ผลงาน และเกณฑ์ในการให้คะแนนที่นำมา ประกอบในการพิจารณาให้คะแนนตามเกณฑ์ที่สร้างไว้

บุญเชิด ภิญโญนนตพงษ์ (2544) เป็นวิธีการประเมินที่ครูมอบหมายงาน หรือกิจกรรมให้ ผู้เรียนทำ และกำหนดให้ผู้เรียนใช้ความรู้และทักษะหลายๆ ด้าน โดยครูใช้ดุลยพินิจในการใช้เกณฑ์ มาตรฐานคุณภาพที่ได้กำหนดไว้ทำการประเมินว่ามีคุณภาพงานระดับใด เพื่อให้ได้ข้อมูลสารสนเทศ ว่าผู้เรียนเกิดการเรียนรู้มากน้อยเพียงใด

ศิริชัย กาญจนวาสี (2546 ในสุวิมล ว่องวานิช, 2546บรรณาธิการ) เป็นการประเมิน ความสามารถในการปฏิบัติงานของผู้เรียนภายใต้สภาพการณ์ และเงื่อนไขที่สอดคล้องกับสภาพจริง

โดยพิจารณาจากกระบวนการทำ และคุณภาพของงาน เกณฑ์การประเมินอาจสร้างขึ้นจากมิติ ความสำคัญ (rubric) ของสมรรถนะด้านต่าง ๆ ของผลงานนั้น

Hibbard (1996) เป็นชุดของกลยุทธ์ ที่ใช้เพื่อการประยุกต์ใช้ความรู้ ทักษะ อุนิสัยในการ ทำงานผ่านชิ้นงาน ที่มีความหมาย และทำให้เกิดการเข้าถึงการเรียนรู้

Airasian (1991) เป็นการประเมินที่ผู้สอนสังเกต และตัดสินในสองลักษณะคือ 1) กระบวนการทำกิจกรรม หรือทักษะของผู้เรียนผ่านทางกิจกรรม หรือ 2) ผลงาน หรือผลผลิต ซึ่งจะเป็นการประเมินที่มากกว่าการใช้การทดสอบด้วยการตอบคำถามในแบบทดสอบ

Nitko (2007) การแสดงผลงานที่มอบหมายให้ผู้เรียนได้ทำกิจกรรม ซึ่งต้องอาศัยความรู้ และ ทักษะจากหลาย ๆ จุดประสงค์การเรียนรู้โดยใช้เกณฑ์ที่ได้นิยามไว้อย่างชัดเจน เพื่อจะทำให้ทราบ ได้ว่าผู้เรียนมีผลการปฏิบัติในระดับใด โดยการให้ผู้เรียนทำกิจกรรมที่ต้องอาศัยความรู้ ทักษะทาง ปัญญา เนื่องจากมีการใช้ความรู้ที่ได้เรียนมา และวัดทักษะจากการนำความรู้มาสู่การปฏิบัติงานที่ ได้รับมอบหมาย

Palm (2008) ให้ความหมายของการประเมินการปฏิบัติงานที่มีจุดเน้น 2 ส่วนคือ (1) การตอบสนอง (Response-centred) ให้ความสำคัญกับการประเมินกระบวนการ และ (2) การกระตุ้น (Stimulate-centred) ให้ความสำคัญกับการสังเกตผลงานที่ได้ หรือการประเมินผลงาน

การประเมินการปฏิบัติงานจึงหมายถึงการที่ผู้สอนมอบหมายงานที่สอดคล้องกับวัตถุประสงค์ ของการประเมิน เพื่อให้ผู้เรียนนำความรู้ไปสู่การปฏิบัติ การวัดจึงอาจวัดกระบวนการปฏิบัติ หรือ ผลงานที่สำเร็จเป็นชิ้นงานก็ได้ โดยมีประเด็นสำคัญ สามส่วนคือ (1) ผู้สอน มีการมอบหมายงานโดย ตั้งวัตถุประสงค์การเรียนรู้ไว้ล่วงหน้า (2) ผู้เรียนต้องใช้ความรู้ที่ได้รับมาถ่ายทอดสู่การปฏิบัติใน สถานการณ์จริงที่แสดงให้เห็นทักษะการปฏิบัติ และผลงานที่มีลักษณะเฉพาะ และ 3) การประเมินผล การปฏิบัติของผู้เรียนว่ามีคุณภาพเพียงใด ต้องอาศัยดุลยพินิจของผู้สอนในการประเมินคุณภาพทักษะ การปฏิบัติ และผลงานโดยอ้างอิงเกณฑ์มาตรฐานที่ตั้งไว้

2 กระบวนการประเมินการปฏิบัติงาน

กระบวนการประเมินภาคปฏิบัติ ประกอบด้วยขั้นตอน ดังนี้ (Johnson, Penny, & Gordon, 2009; เอมอร์ จังศิริพรพรรณ, 2550 ; บุญเชิด ภิญโญอนันตพงษ์, 2544) (1) การออกแบบการประเมิน (2) เลือกรงานการปฏิบัติให้ชัดเจน (3) ออกแบบกิจกรรมสำหรับใช้ประเมิน (4) กำหนดวิธีวัดทักษะการปฏิบัติ (5) กำหนดเครื่องมือที่ใช้ในการวัดทักษะการปฏิบัติ (6) กำหนดวิธีการประเมินผล และ (7) รายงานผล มีรายละเอียดดังนี้

2.1 ออกแบบการประเมิน (Johnson et al., 2009) โดยกำหนดกรอบของแบบทดสอบ (test framework) และพิมพ์เขียวของแบบทดสอบ (test specifications) เพื่ออธิบายโครงสร้างตัวอย่างสมรรถนะของงานที่ต้องการประเมิน เพื่อให้เนื้อหามีความครอบคลุมตามวัตถุประสงค์ มีขั้นตอนคือ

2.1.1 กำหนดวัตถุประสงค์ เพื่อให้เกิดความชัดเจนถึงเป้าหมาย เพื่อนำผลการประเมินมาอ้างอิงผู้เรียน ให้ทราบถึงความรู้ ทักษะ และความสามารถ ที่ต้องการให้เกิดขึ้นในอนาคต

2.1.2 ให้นิยามกรอบของงาน (framework) กำหนดสมรรถนะ เช่น เนื้อหา ทักษะ กระบวนการ หรือ สมรรถนะที่ตรงกับความต้องการของโครงสร้าง

2.1.3 กำหนดพิมพ์เขียวของแบบทดสอบ (test specifications) บรรยายถึงสมรรถนะของการประเมินที่เหมาะสมต่อกรอบของงานซึ่งควรมีองค์ประกอบ เช่น เนื้อหา และกระบวนการที่จะใช้ประเมิน รูปแบบของแบบทดสอบ จำนวน และชนิดของคำถามในแบบทดสอบ และกระบวนการในเชิงปฏิบัติการเกี่ยวกับโครงสร้าง เพื่อให้เกิดความมั่นใจได้ว่ามีความสอดคล้องกันระหว่างแบบทดสอบกับคะแนนที่อ้างอิงไปถึง มีองค์ประกอบดังนี้

2.1.3.1 ให้นิยามเนื้อหา ประกอบด้วยความรู้ ทักษะ และความสามารถ หรือ สมรรถนะที่ต้องการประเมิน (knowledge, skill, ability: KSAs) โดยพิจารณาจากสมรรถนะของผู้ถูกประเมิน เช่น ระดับการศึกษาของผู้ถูกประเมิน การระบุสาระสำคัญของความรู้ ทักษะ และความสามารถหรือสมรรถนะ (KSAs) ที่ต้องการจึงขึ้นอยู่กับกลุ่มเป้าหมายที่ต้องการประเมิน มีขั้นตอนคือ

1) การกำหนด KSAs ตัวอย่างเช่น KSAs ของพยาบาลเพื่อสอบรับใบประกอบวิชาชีพ ซึ่งเป็นระดับพื้นฐานของวิชาชีพ ย่อมต่างจาก KSAs ของพยาบาลที่ต้องประเมินเพื่อรับวุฒิปริญญาพยาบาล ชำนาญการ

2) ความสำคัญของระดับการประเมินอยู่ในระดับใด เช่นการประเมินในระดับจังหวัด หรือ ระดับประเทศ หรือเป็นระดับวิชาชีพ

- 3) เนื้อหาสาระที่ต้องการประเมิน ครอบคลุมเนื้อหาใดบ้าง
- 4) กระบวนการ หรือทักษะที่ต้องการให้เกิดขณะประเมิน การกระจายของหัวข้อการประเมิน และ หรือทักษะในตารางกำหนดพิมพ์เขียวของแบบทดสอบ (Table specification)
- 5) กำหนดลักษณะของKSAs มีขั้นตอนในการวิเคราะห์กระบวนการประเมิน (core of procedures in practice analysis) ดังนี้ 1) ขอความเห็นผู้เชี่ยวชาญ 2) พัฒนาข้อคำถามเชิงสำรวจ เป็นรายการของชิ้นงาน หรือกิจกรรม 3) เลือกตัวแทนของตัวอย่างในกลุ่มเป้าหมายที่เป็นระดับอาชีพ เพื่อตอบแบบสำรวจ 4) นำผลการให้คะแนนแบบสำรวจของกลุ่มเป้าหมายมาแยกพิจารณาตามสมรรถนะของงาน 5) วิเคราะห์ข้อมูล และตัดสินความสัมพันธ์ที่สำคัญของงาน และ 6) ใช้ผลการวัดเชิงปริมาณของงาน ที่สำคัญนั้นเพื่อการพัฒนาการกำหนดพิมพ์เขียวของแบบทดสอบ โดยการอธิบายสมรรถนะของเนื้อหาที่สำคัญในการประเมินคือ ความรู้ ทักษะ และความสามารถ หรือสมรรถนะที่ต้องการประเมิน (knowledge, skill, ability: KSAs)

2.1.3.2 ทำโครงร่างของข้อคำถาม (Task shell) เป็นรายข้อ โดยกำหนดเป็นลักษณะคำถามที่ต้องประเมิน (Item shell) มีลักษณะเป็นเค้าโครงของแต่ละข้อคำถาม เพื่อให้สามารถสร้างลักษณะของข้อคำถามคล้าย ๆ กันได้ ตัวอย่างของทั้งพิมพ์เขียวของแบบทดสอบ และ โครงร่างของแบบสอบรายข้อ ดังแสดงในตารางที่ 1

ตารางที่ 1 ตัวอย่างตารางพิมพ์เขียวของแบบทดสอบ (Test specification) ของกิจกรรมการพยาบาล การประเมินสัญญาณชีพ

มาตรฐานของเนื้อหา	เปอร์เซ็นต์ของเนื้อหา
1. การเตรียมความพร้อมอุปกรณ์/ ของใช้	10%
2. การเตรียมความพร้อมผู้ป่วย	15%
3. การบอกวัตถุประสงค์ของการประเมินสัญญาณชีพ	15%
4. การปฏิบัติตามขั้นตอนการประเมินสัญญาณชีพ	30%
5. การบันทึกผลการประเมินสัญญาณชีพ และการให้คำแนะนำ	15%
6. การตอบคำถามจากการบูรณาการความรู้สู่การปฏิบัติ	15%

โครงร่างของแบบสอบรายข้อ (Item shell) ของกิจกรรมการพยาบาลการประเมินสัญญาณชีพ เช่น ข้อ 1 การเตรียมความพร้อมอุปกรณ์/ ของใช้ พิจารณาจากความครบถ้วนของ

อุปกรณ์สำคัญที่ใช้ในการประเมินสัญญาณชีพ เช่น ปรอท หูฟัง เครื่องวัดความดันโลหิต และ อุปกรณ์เสริมอื่นๆ เช่น นาฬิกาที่มีเซ็นเซอร์ วัดอัตราการเต้นของหัวใจ

ข้อ 2 การเตรียมความพร้อมผู้ป่วย เป็นการสื่อสารเพื่อแนะนำตนเอง และบอกขั้นตอนการปฏิบัติ

ข้อ 3 การบอกวัตถุประสงค์ของการประเมินสัญญาณชีพ เป็นการสื่อสารที่ชัดเจน เข้าใจง่าย และ ถูกหลักวิชาการ และสอดคล้องกับสถานการณ์จริงของผู้ป่วย

ข้อ 4 การปฏิบัติตามขั้นตอนการประเมินสัญญาณชีพ พิจารณาจากการปฏิบัติตามลำดับขั้นตอนที่สำคัญ บางขั้นตอนที่ไม่มีผลเสียต่อผู้ป่วยอาจปฏิบัติสลับขั้นตอนกันได้ เช่น วัดความดันก่อนการคลำชีพจร หรือนับการหายใจ หรือการทำอย่างหลังก่อนอย่างแรก แต่ขั้นตอนของการวัดปรอทควรทำก่อนกิจกรรมอื่นเนื่องจากจะต้องใช้เวลาในการวัดอย่างน้อย 1 - 3 นาที หากทำกิจกรรมอื่น ก่อนการวัดปรอทจะทำให้เสียเวลาในการรอผลการวัดอุณหภูมิ และรบกวนเวลาพักผ่อนของผู้ป่วยเพิ่มมากขึ้นอีก

ข้อ 5 การบันทึกผลการประเมินสัญญาณชีพ และการให้คำแนะนำ เป็นการลงข้อมูลผลการประเมินสัญญาณชีพในแฟ้มประวัติของผู้ป่วย ต้องมีความชัดเจน อ่านออกง่าย ถูกต้องตรงตามข้อมูลต้นฉบับจากแฟ้มการประเมินสัญญาณชีพ และการให้คำแนะนำเป็นการแปลผลการประเมินสัญญาณชีพ และนำไปแนะนำการปฏิบัติตัวที่จะต้องสอดคล้องกับปัญหาที่เกี่ยวกับสัญญาณชีพของผู้ป่วย

ข้อ 6 การตอบคำถามจากการบูรณาการความรู้สู่การปฏิบัติ คำตอบจะต้องมีความถูกต้องตามหลักวิชาการ และสอดคล้องกับสถานการณ์จริงของผู้ป่วย หรือสถานการณ์สมมติที่มีให้

จากนั้นนำโครงร่างของข้อทดสอบรายข้อ (Item shell) บรรจุเข้าไปในพิมพ์เขียวของแบบทดสอบ (Test specification)

2.2 การเลือกงาน ควรกำหนดให้ชัดเจน ทั้งจำนวนชิ้นงาน ลักษณะการประเมินว่าเป็นรายบุคคลหรือ รายกลุ่ม ประเภทของงานภาคปฏิบัติประกอบด้วย

2.2.1 งานที่มอบหมายตามความต้องการ หรืองานที่ต้องสร้างผลงานขึ้นมา (structure, on-demand tasks) โดยชี้แจงให้ผู้เรียนทราบล่วงหน้าถึงสิ่งที่ต้องลงมือทำ อธิบายถึงผลลัพธ์ที่ต้องการหลังการทำ แจ้งผู้เรียนให้ทราบเมื่อจะประเมิน และเปิดโอกาสให้ผู้เรียนเตรียมความพร้อมก่อนการประเมิน ประเภทของงานที่มอบหมายมีดังนี้

2.2.1.1 งานเขียนตอบ (Paper-pencil task) ที่สะท้อนความคิด เหตุผลในรูปการเขียนบรรยาย และกระบวนการในการแก้ปัญหา อาจอยู่ในรูปชิ้นงานสำเร็จแล้ว เช่น บทความ รายงาน หรือกระบวนการที่ใช้ เช่น อธิบายว่าแก้ไขปัญหานั้นได้อย่างไร

2.2.1.2 งานที่ไม่ใช่งานเขียนตอบ (Task beyond paper-pencil) ต้องอาศัยเครื่องมือ อุปกรณ์ในการปฏิบัติขึ้นอยู่กับลักษณะเนื้อหาวิชา ต้องอาศัยการปฏิบัติจริงให้เห็นกระบวนการ และหรือผลลัพธ์

2.2.2 งานที่เกิดขึ้นตามธรรมชาติ หรืองานที่มีลักษณะธรรมชาติของงานภาคปฏิบัติ (naturally occurring or typical performance tasks) เช่น เกิดขึ้นที่หอผู้ป่วย การประเมินจะต้องอาศัยการสังเกต และประเมินในสถานการณ์ที่เกิดขึ้นจริง ไม่มีการแจ้งผู้เรียนที่กำลังถูกประเมิน และไม่มีการเข้าไปแทรกแซงกระบวนการเรียนรู้ต่างๆ ยกเว้นจะเกิดผลเสียร้ายแรง ข้อเสียคือ ใช้เวลามาก ไม่สามารถควบคุมได้ว่า จะเกิดงานลักษณะใด และจะเกิดเมื่อไหร่ ไม่สามารถบอกได้ว่าผู้เรียนทุกคน จะแสดงพฤติกรรมที่ต้องการประเมินหรือไม่

2.2.3 โครงการระยะยาว (long-term project) ประเภทของโครงการระยะยาวมีดังนี้

2.2.3.1 โครงการที่เป็นงานเดี่ยว พิจารณาจากผลลัพธ์ที่เป็นผลผลิตของผู้เรียน เช่น รายงานเป็นฉบับ หรือการสะสมงานที่ใช้เวลาตลอดภาคการศึกษา

2.2.3.2 โครงการที่เป็นงานกลุ่ม ต้องการให้ผู้เรียนตั้งแต่สองคนขึ้นไปทำงานร่วมกัน เพื่อประเมินการทำงานเป็นทีม และสามารถสร้างสรรค์ผลผลิตที่มีคุณภาพ เป้าหมายทางการเรียนขึ้นอยู่กับลักษณะรายวิชา และระดับของผู้เรียน

2.2.4 แฟ้มสะสมผลงาน (portfolio) เป็นผลงาน ที่เจ้าของประวัติสะสม และคัดเลือก สามารถสะท้อนความเป็นเอกลักษณ์ และเอกลักษณ์ของเจ้าของ แสดงถึงคุณค่าของบุคคลโดยมีเอกสารรับรอง และตรวจสอบได้ มีประโยชน์ในการนำเสนอผลงานของผู้เรียน และแสดงให้เห็นถึงพัฒนาการด้านการศึกษาของผู้เรียน มี 2 ประเภทคือ

2.2.4.1 แฟ้มสะสมผลงานทั้งหมด (working portfolio) ประกอบด้วยรายละเอียดของผลงานตั้งแต่ กระบวนการวางแผน การปฏิบัติ การแก้ปัญหา การแสดงผล และการประเมินทั้งการประเมินจากตนเอง และจากผู้อื่น

2.2.4.2 แฟ้มสะสมผลงานดีเด่น (presentation portfolio) เจ้าของประวัติเป็นผู้คัดสรรเอกสาร หรือผลงานที่แสดงให้เห็นถึงความสามารถพิเศษ ทักษะ และสมรรถนะเด่น ออกมาแนะนำ โดยอาศัยความคิดสร้างสรรค์ของตน

2.2.5 การสาธิต (demonstrations) เป็นงานที่ผู้เรียนได้รับมอบหมายให้แสดงความรู้ และทักษะเพื่อให้ได้ชิ้นงานที่สมบูรณ์

2.2.6 การทดลอง (experiments) เป็นงานที่ได้รับมอบหมายเพื่อให้ผู้เรียนได้วางแผน ลงมือปฏิบัติ และแปลความหมายของผลการทดลองที่ได้ปฏิบัติลงไปจริง ช่วยให้ผู้สอนประเมินได้ว่าผู้เรียนสามารถใช้ทักษะ และวิธีการที่เหมาะสมในการแก้ปัญหาในการทดลองได้เพียงใด

2.2.7 การนำเสนอด้วยปากเปล่า และการแสดงเสมือนจริง (oral presentation and dramatizations) ช่วยให้ผู้เรียนสามารถสื่อสารผ่านภาษาพูดโดยใช้ความรู้ และทักษะทางการพูดในรูปแบบของการสัมภาษณ์ หรือการนำเสนอผลงาน ขึ้นอยู่กับลักษณะวิชาว่าสอดคล้องกับการประเมินด้วยวิธีการใด

2.2.8 สถานการณ์เสมือนจริง และสถานการณ์ที่ออกแบบไว้ (simulations and contrived situations) เป็นเหตุการณ์ หรือสถานการณ์ที่ได้ควบคุม และให้ผู้เรียนแสดงปฏิสัมพันธ์กับคนในบทบาทสมมติ หรือใช้สื่อต่างๆ เช่น ใช้สื่อวีดิทัศน์นำเสนอสถานการณ์ที่อาจเป็นสถานการณ์จริงแล้วตั้งคำถามให้ผู้เรียนแก้ปัญหา เป็นต้น

2.3 ออกแบบกิจกรรมที่จะใช้ในการประเมิน โดยกำหนดตัวบ่งชี้ที่ใช้ในการวัดทักษะการปฏิบัติ (performance indicator) เพื่อกำหนดน้ำหนักความสำคัญของงาน และ ขั้นตอนของการจำแนกกิจกรรม หรือขั้นตอนในการทำงานตามลำดับก่อนหลัง หรืออาจเรียกว่าเป็นการวิเคราะห์งาน (job analysis) ตามลักษณะงาน โดยมีขั้นตอนของการได้มาซึ่งสารสนเทศเกี่ยวกับงาน และการกำหนดน้ำหนักความสำคัญของงานเรียงตามลำดับดังนี้

2.3.1 การรวบรวมรายละเอียดของข้อมูลงานภาคปฏิบัติ

2.3.2 ตรวจสอบความถูกต้องของข้อมูลที่ได้มา

2.3.3 เขียนบรรยายอย่างละเอียดเกี่ยวกับงานภาคปฏิบัติระบุ ความรู้ ความสามารถ และทักษะที่ต้องใช้ในการทำงานภาคปฏิบัติ

2.3.4 ปรับข้อมูลให้มีความทันสมัยอยู่เสมอ

2.4 กำหนดวิธีประเมินทักษะภาคปฏิบัติ มีหลายวิธีดังนี้

2.4.1 การเขียนตอบเป็นวิธีที่เหมาะสมสำหรับการปฏิบัติที่ต้องการวัดความรู้เกี่ยวกับการปฏิบัติก่อนการปฏิบัติจริง เนื่องจาก เป็นงานที่มีความเสี่ยงหรือ มีผลกระทบต่อตนเอง และผู้อื่นเป็นอันมาก เช่น การช่วยฟื้นคืนชีพผู้ใหญ่

2.4.2 การให้ผู้เรียนปฏิบัติงานให้ดูในสถานการณ์จำลอง หรือสถานการณ์จริงที่ผู้เรียนอาจรู้ตัว หรือไม่รู้ว่าถูกประเมินก็ได้ เน้นการสังเกตพฤติกรรมของผู้เรียนในขณะดำเนินการปฏิบัติงาน

2.4.3 การประเมินตัวอย่างของงานที่ได้จากการปฏิบัติจริง (work sample) เป็นวิธีวัดประเมินผลผลิต หรือผลจากการดำเนินงานจนได้ชิ้นงาน ที่อยู่ในรูปแบบต่าง ๆ เช่น รายงาน งานฝีมือ บทประพันธ์ เป็นต้น

2.5 กำหนดเครื่องมือที่ใช้ในการประเมินทักษะภาคปฏิบัติ โดยกำหนดชนิดของเครื่องมือ เช่น แบบสอบภาคปฏิบัติ แบบตรวจรายการ มาตรฐานประมาณค่า เป็นต้น อาจใช้เครื่องมือมากกว่า 1 ประเภท ทั้งนี้ขึ้นอยู่กับลักษณะงาน และวัตถุประสงค์การเรียนรู้

2.6 กำหนดวิธีการประเมินผล การกำกับติดตาม และรายงานผล ข้อมูลที่ได้จากการประเมินทักษะภาคปฏิบัติจะถูกนำมาตัดสินคุณค่า หรือคุณภาพซึ่งอาจเทียบกับกลุ่ม เทียบกับเกณฑ์ที่ผู้สอนตั้งไว้ตั้งแต่แรก หรือเทียบกับตัวผู้เรียนเองแล้วจึงรายงานผล

การวิจัยครั้งนี้ใช้กระบวนการของการออกแบบตามขั้นตอนข้างต้น โดยเลือกใช้งานในการประเมินประเภท สถานการณ์จำลองการขึ้นฝึกปฏิบัติงานของนักศึกษาพยาบาลชั้นปีที่ 2 ในรายวิชาปฏิบัติการพยาบาลพื้นฐาน เนื่องจากเป็นวิชากลาง ที่อาจารย์จากต่างสาขาวิชาสามารถเรียนรู้ร่วมกันได้ โดยสถานการณ์จำลองอยู่ในรูปสื่อวีดิทัศน์

3. เครื่องมือ และเกณฑ์การแปลความหมายของเครื่องมือที่ใช้ในการประเมินภาคปฏิบัติ

3.1 ประเภทของเครื่องมือที่ใช้ในการประเมินภาคปฏิบัติ ใช้ในการประเมินใน 2 ลักษณะ คือ ประเมินกระบวนการ และประเมินผลผลิตของงาน มีหลายประเภทดังนี้

3.1.1 ระเบียบเหตุการณ์ (anecdotal record) เป็นการบันทึกข้อมูลจากการสังเกตอย่างไม่เป็นทางการ ประกอบด้วย ชื่อผู้ถูกสังเกต วัน เวลา สถานที่ ที่สังเกต เหตุการณ์ที่เกิดขึ้นไม่ควรเขียนโดยใช้ความรู้สึกส่วนตัว ข้อดีคือ ไม่ต้องมีรูปแบบตายตัว สามารถบันทึกได้ทุกเวลาทุกสถานการณ์

ผู้สังเกตไม่จำเป็นต้องผ่านการฝึกฝนเป็นพิเศษ นำมาใช้เป็นข้อมูลเชิงคุณภาพร่วมกับแบบประเมินอื่น ๆ ก่อนการตัดสินผลการประเมิน ข้อเสียคือ มีความเป็นปรนัยต่ำ นำไปใช้ในการประเมินสรุปผลไม่ได้ และต้องอาศัยเวลา หากสังเกตไม่ทันจะทำให้เนื้อหาขาดหายไป

3.1.2 แบบตรวจสอบรายการ (checklist) เป็นรายการของพฤติกรรมในการทำกิจกรรมตามขั้นตอนต่างๆ โดยเว้นช่องว่างในช่องรายการ ไว้ให้กาเครื่องหมาย ว่าผู้ถูกประเมินได้ทำหรือไม่ได้ทำรายการนั้น แต่ไม่มีการตัดสินคุณภาพของงาน ข้อดีคือ ประหยัดเวลา ใช้งานง่าย ในหลายสถานการณ์ และ ข้อมูลที่ได้นำไปวิเคราะห์ได้ง่าย ข้อเสียคือ ขาดข้อมูลเชิงคุณภาพ ไม่ได้รายละเอียด

3.1.3 มาตรฐานค่า (rating scale) ใช้ในการประเมินทั้งกระบวนการ และผลผลิต โดยมีรายการพฤติกรรมที่ต้องการประเมิน เหมือนแบบตรวจสอบรายการ (checklist) แต่มีตัวบ่งชี้คุณภาพของระดับการปฏิบัติ และมีช่วงของมาตรวัด ที่มีค่าเป็นตัวเลข หรือระดับคุณภาพของพฤติกรรมให้เลือกตามการตัดสินของผู้ประเมิน ข้อดีคือ ใช้งานง่าย ไม่ใช้เวลามาก ข้อเสียคือ อาจประเมินพฤติกรรมได้บางส่วนทำให้เกิดความคลาดเคลื่อนจากผู้ประเมินได้ง่าย เนื่องจากไม่มีการบรรยายสมรรถนะของแต่ละระดับคุณภาพไว้ ทำให้ผู้ประเมินใช้ความรู้สึกส่วนตัวในการตัดสินผลการประเมินได้ง่าย

3.1.4 การให้คะแนนโดยรูบรีค (scoring rubric) คล้ายมาตรฐานค่าแต่มีคำอธิบายถึงระดับคุณภาพของงาน มาตรวัดอาจมีหลายช่วงเช่นเดียวกับ มาตรฐานค่า (rating scale) เช่น 3 - 10 ช่วง แต่พบว่าช่วงมาตรประเมิน 4 ถึง 6 ระดับมีความเที่ยงระหว่างผู้ประเมิน (inter-rater reliability) สูงอย่างมีนัยสำคัญทางสถิติ จึงควรใช้ช่วงของมาตรประเมินนี้ (Johnson et al., 2009) รูบรีคมีหลายประเภทแบ่งตามเกณฑ์การให้คะแนนได้ดังนี้ (Nitko, 2007)

3.1.4.1 รูบรีคแบบภาพรวม (holistic rubrics) เป็นการประเมินแบบภาพรวมของคุณภาพงาน หรือกระบวนการที่ผู้เรียนตอบทำ ตัวอย่างเช่น การให้คะแนนการประเมินความดันโลหิตแบ่งเป็นระดับ มีการแบ่งช่วงคะแนนเป็น 5 ระดับ และพิจารณาภาพรวมของสมรรถนะทั้งหมดที่ต้องการประเมิน ดังแสดงในตารางที่ 2

ตารางที่ 2 ตัวอย่างรูปรีคแบบภาพรวมของกิจกรรมการพยาบาล การประเมินสัญญาณชีพ

ระดับการประเมิน	คำอธิบายลักษณะของระดับคะแนนที่ควรได้	ค่าคะแนน
ดีมาก	แสดงให้เห็นว่าเข้าใจได้ทั้งหมดเกี่ยวกับวัตถุประสงค์ของการวัดความดันโลหิต และอธิบายได้อย่างถูกต้องแม่นยำเกี่ยวกับรายละเอียด และเตรียมอุปกรณ์ เตรียมผู้ป่วยได้ครบถ้วนแม่นยำ ลงมือปฏิบัติได้อย่างคล่องแคล่ว แม่นยำ เป็นไปตามขั้นตอนในเวลาไม่นาน บันทึกผลการประเมินได้อย่างถูกต้องแม่นยำเป็นระเบียบเรียบร้อย ตอบคำถามได้ถูกต้อง และวิเคราะห์สถานการณ์ที่หลากหลายได้	5
ดี	แสดงให้เห็นว่าเข้าใจได้ทั้งหมดเกี่ยวกับวัตถุประสงค์ของการวัดความดันโลหิต และอธิบายได้อย่างถูกต้องแม่นยำเกี่ยวกับรายละเอียดและอุปกรณ์ เตรียมผู้ป่วยได้ครบถ้วนแม่นยำ ลงมือปฏิบัติได้เป็นไปตามขั้นตอน ใช้เวลาในการปฏิบัติไม่นานมากนัก บันทึกผลการประเมินได้อย่างถูกต้องแม่นยำแต่ขาดความเป็นระเบียบเรียบร้อย ตอบคำถามได้ถูกต้องตามสถานการณ์ แต่ไม่สามารถวิเคราะห์สถานการณ์ที่หลากหลายได้	4
พอใช้	แสดงให้เห็นว่าเข้าใจได้ทั้งหมดเกี่ยวกับวัตถุประสงค์ของการวัดความดันโลหิต และอธิบายได้อย่างถูกต้องในวัตถุประสงค์หลัก แต่ไม่ถูกต้องในวัตถุประสงค์รองหรือที่เกี่ยวกับรายละเอียด และเตรียมอุปกรณ์ เตรียมผู้ป่วยได้เกือบครบถ้วน อาจขาดอุปกรณ์เสริมที่ไม่สำคัญต่อการปฏิบัติ ลงมือปฏิบัติได้เป็นไปตามขั้นตอนไม่ครบถ้วนอาจข้ามขั้นตอน หรือลืมนิติปฏิบัติในขั้นตอนที่ไม่สำคัญ ใช้เวลาในการปฏิบัติค่อนข้างนาน บันทึกผลการประเมินได้ถูกต้องเป็นส่วนใหญ่ ผิดในส่วนที่ไม่สำคัญแต่ขาดความเป็นระเบียบเรียบร้อย ตอบคำถามได้ถูกต้องในสาระสำคัญ	3
ต้องปรับปรุง	แสดงให้เห็นว่าเข้าใจได้บางส่วนเกี่ยวกับวัตถุประสงค์ของการวัดความดันโลหิต และอธิบายได้อย่างถูกต้องในประเด็นรองแต่ไม่สามารถอธิบายในวัตถุประสงค์หลักได้ เตรียมอุปกรณ์ เตรียมผู้ป่วยได้ไม่ครบถ้วน อาจขาดอุปกรณ์เสริมที่ไม่สำคัญต่อการปฏิบัติ ลงมือปฏิบัติได้เป็นไปตามขั้นตอนไม่ครบถ้วนอาจข้ามขั้นตอนหรือลืมนิติปฏิบัติในขั้นตอนที่ไม่สำคัญ ใช้เวลาในการปฏิบัติค่อนข้างนาน บันทึกผลการประเมินได้ถูกต้องเป็นส่วนใหญ่ ผิดในส่วนที่ไม่สำคัญแต่ขาดความเป็นระเบียบเรียบร้อย ตอบคำถามได้ถูกต้องในสาระสำคัญ	2
ไม่ดี	แสดงให้เห็นว่าเข้าใจได้ทั้งหมดเกี่ยวกับวัตถุประสงค์ของการวัดความดันโลหิต และอธิบายได้อย่างถูกต้องในวัตถุประสงค์หลัก แต่ไม่ถูกต้องในวัตถุประสงค์รองหรือที่เกี่ยวกับรายละเอียด และเตรียมอุปกรณ์ เตรียมผู้ป่วยได้เกือบครบถ้วน อาจขาดอุปกรณ์เสริมที่ไม่สำคัญต่อการปฏิบัติ ลงมือปฏิบัติได้เป็นไปตามขั้นตอนไม่ครบถ้วนอาจข้ามขั้นตอนหรือลืมนิติปฏิบัติในขั้นตอนที่ไม่สำคัญ ใช้เวลาในการปฏิบัติค่อนข้างนาน บันทึกผลการประเมินได้ถูกต้องเป็นส่วนใหญ่ ผิดในส่วนที่ไม่สำคัญแต่ขาดความเป็นระเบียบเรียบร้อย	1

3.1.4.2 รูบริคแบบแยกส่วน (analytic rubrics) เป็นการประเมินสมรรถนะแยกตามองค์ประกอบที่เฉพาะลงไป ตัวอย่างเช่น แบบประเมินสัญญาณชีพที่แยกตามสมรรถนะรายองค์ประกอบ 5 องค์ประกอบคือ 1) การเตรียมอุปกรณ์ของใช้ 2) การเตรียมความพร้อมของผู้ป่วย 3) ขั้นตอนการปฏิบัติ 4) การลงบันทึกผลการพยาบาล และการให้คำแนะนำ และ 5) การตอบข้อซักถาม ในที่นี้นำเสนอตัวอย่างการประเมินด้านการเตรียมอุปกรณ์ ดังแสดงในตารางที่ 3

ตารางที่ 3 ตัวอย่างรูบริคแบบแยกส่วน ของกิจกรรมการพยาบาล การประเมินสัญญาณชีพ

ระดับการประเมิน	หัวข้อการประเมิน/คำอธิบายลักษณะของระดับคะแนน	การประมาณค่าคะแนน
ดีมาก	การเตรียมความพร้อมอุปกรณ์/ของใช้ บอกอุปกรณ์ของใช้ที่สำคัญและอุปกรณ์ส่วนเสริมเครื่องใช้หลักได้ครบถ้วน และบอกเหตุผลในการเตรียมอุปกรณ์แต่ละชนิดได้ถูกต้อง เตรียมอุปกรณ์ของใช้ในเวลาที่อันรวดเร็ว ปฏิบัติได้ต่อเนื่องคล่องแคล่ว และเมื่อมีการกำหนดสถานการณ์อื่นๆให้สามารถหยิบเตรียมอุปกรณ์ขึ้นมาใช้ได้อย่างคล่องแคล่วเป็นธรรมชาติ	5
ดี	บอกอุปกรณ์ของใช้ที่สำคัญได้ครบ รวมถึงอุปกรณ์ส่วนเสริมเครื่องใช้หลักอย่างครบถ้วน และสามารถบอกเหตุผลในการเตรียมอุปกรณ์แต่ละอย่างได้ถูกต้อง สอดคล้องกับสถานการณ์ที่กำลังดำเนินอยู่นั้น และ/หรือเตรียมอุปกรณ์ของใช้ได้ในเวลาไม่นาน เมื่อมีการกำหนดสถานการณ์อื่นๆ ให้ ปฏิบัติได้ต่อเนื่อง ค่อนข้างคล่องแคล่ว	4
พอใช้	บอกอุปกรณ์ของใช้ที่สำคัญได้ครบ เช่น เครื่องวัดความดันโลหิตและหูฟัง โปรท และวาสลิน แต่ไม่ได้เตรียมวัสดุอุปกรณ์ของใช้ที่เป็นส่วนเสริมเครื่องใช้หลักเช่นไม่ได้เตรียมสำลี และบอกเหตุผลในการเตรียมอุปกรณ์สำคัญได้แต่ไม่ครบถ้วน และ/หรือ ใช้เวลาครุ่นคิดในการเตรียมอุปกรณ์ ของใช้ มีการปฏิบัติที่ต่อเนื่องแต่ค่อนข้างนาน	3
ต้องปรับปรุง	บอกอุปกรณ์ของใช้ได้เพียง 2-3 ชนิด และ/หรือ บอกเหตุผลของการเตรียมอุปกรณ์แต่ละชนิดได้บ้างเล็กน้อย) ใช้เวลานานกว่าที่ควรจะเป็นในการตัดสินใจปฏิบัติในแต่ละขั้นตอนของการเตรียมอุปกรณ์ และ/หรือไม่มีความคล่องแคล่วในการปฏิบัติ หรือปฏิบัติได้ไม่ต่อเนื่อง หรือปฏิบัติไปเป็นไปตามขั้นตอน ตกหล่นในขั้นตอนสำคัญ	2
ไม่ดี	บอกอุปกรณ์ของใช้ได้น้อยมาก และ/หรือ บอกเหตุผลของการเตรียมอุปกรณ์แต่ละชนิดได้น้อยมาก ใช้เวลานานกว่าที่ควรจะเป็นในการตัดสินใจปฏิบัติในแต่ละขั้นตอนของการเตรียมอุปกรณ์ และ/หรือไม่มีความคล่องแคล่วในการปฏิบัติ หรือปฏิบัติได้ไม่ต่อเนื่องหรือไม่ปฏิบัติ	1

3.1.4.3 Annotated holistic scoring rubric เป็นการผสมผสานระหว่างรูบริคแบบภาพรวม และรูบริคแบบแยกส่วน โดยการใช้รูบริคแบบภาพรวมในการให้คะแนน แต่จะให้ข้อมูลย้อนกลับกับผู้เรียนในลักษณะของรูบริคแบบแยกส่วน โดยระบุถึงจุดเด่น และจุดด้อยของผู้เรียน

3.1.4.4 Generic (General) rubric เป็นการใช้รูบริคเพื่ออธิบายงานที่นำไปประยุกต์ใช้กับชุดของงานที่มอบหมายทั้งหมดซึ่งอาจมีหลายชิ้นงาน เช่นการใช้รูบริคเดียวกันกับการประเมินความสามารถในการเขียนคิดวิเคราะห์ และแก้ไขปัญหาในด้านงานเขียน การแก้ปัญหาทางคณิตศาสตร์ หรือ การปฏิบัติการในห้องทดลอง เป็นต้น

3.1.4.5 Task-specific rubric เป็นรูบริคที่นำไปใช้ในการให้คะแนนงานปฏิบัติที่มีลักษณะเฉพาะของงาน คือนำกรอบการให้คะแนนของ Generic rubrics มาปรับใช้กับงานที่มีลักษณะเฉพาะซึ่งจะต้องมีการดำเนินการประเมินสอดคล้องไปกับ Generic rubrics

การวิจัยในครั้งนี้เลือกใช้เครื่องมือการประเมินคือ รูบริคแบบแยกส่วนเพื่อให้เห็นองค์ประกอบเฉพาะของแต่ละมิติที่ทำการประเมิน เพื่อง่ายต่อการประเมิน

4. การตรวจสอบคุณภาพการประเมินการปฏิบัติงาน

มีการตรวจสอบคุณภาพ 2 ประเด็นสำคัญคือ (1) คุณภาพของเครื่องมือที่ใช้ในการประเมิน โดยพิจารณาจากความตรง และความเที่ยง และ (2) คุณภาพการประเมินของผู้ประเมิน โดยพิจารณาจากความเที่ยง และความสอดคล้องกันระหว่างผู้ประเมิน มีรายละเอียดดังนี้

4.1 ตรวจสอบความตรง ดังนี้ (Nitko, 2007; ชวลิต ชูกำแพง, 2551; สุวิมล ว่องวานิช, 2546)

4.1.1 ความตรงตามเนื้อหาเป็นการตรวจสอบว่ามีเนื้อหาที่ครอบคลุม และเป็นตัวแทนของเนื้อเรื่องที่มุ่งวัดได้

4.1.2 ความตรงตามสภาพ เป็นการตรวจสอบว่าความสามารถในปัจจุบัน หรือพยากรณ์ความสามารถในอนาคตได้หรือไม่ และสิ่งสำคัญที่มีผลต่อความตรงของเครื่องมือในการประเมินการปฏิบัติงานคือ ความสอดคล้องกันของกิจกรรมงานการปฏิบัติที่ได้รับมอบหมายกับวัตถุประสงค์การเรียนรู้ และการใช้เกณฑ์การให้คะแนนความสามารถในการปฏิบัติที่ได้รับการนิยามเป็นอย่างดี (well-defined scoring criteria) จะช่วยเพิ่มความตรงให้กับการประเมิน

4.2 ตรวจสอบความเที่ยง มีวิธีการตรวจสอบดังนี้

4.2.1 การประมาณค่าความเที่ยงในเวลาต่างกัน (Nitko, 2007)

วิธีที่ 1 สอบซ้ำ (test-retest method) โดยนำแบบสอบฉบับเดียวกันไปทดสอบ ผู้สอบกลุ่มเดียวกัน 2 ครั้ง โดยทิ้งระยะระหว่างการสอบครั้งแรก และการสอบครั้งที่สอง และนำคะแนนระหว่างการวัด 2 ครั้งมาคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

วิธีที่ 2 การใช้แบบสอบสมมูลทดสอบในเวลาต่างกัน (alternate forms on different occasions) และนำคะแนนสอบทั้งสองครั้งมาคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

4.2.2 โดยการประมาณค่าความเที่ยงโดยการทดสอบเพียงครั้งเดียว มีหลายวิธีคือ

วิธีที่ 1 การตรวจสอบความเที่ยงของแบบสอบคู่ขนาน (alternate-forms reliability) โดยให้ผู้สอบกลุ่มเดียวกันทำแบบทดสอบ 2 ชุดที่สมมูลกันในการสอบครั้งเดียวกัน แล้วนำคะแนนที่ได้มาคำนวณค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน

วิธีที่ 2 วิธีคำนวณค่าสัมประสิทธิ์แอลฟา (coefficient alpha) เป็นเทคนิคการแบ่งครึ่งข้อสอบเพื่อนำคะแนนทั้งสองส่วนมาคำนวณความสัมพันธ์สำหรับประมาณค่าความเที่ยงแบบความสอดคล้องภายในของแบบสอบนั้น

วิธีที่ 3 วิธีคำนวณค่าสัมประสิทธิ์ความเที่ยงแบ่งครึ่งข้อสอบ (split-halves coefficient) เป็นการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่วัดได้จากการแบ่งครึ่งข้อสอบที่สมมูลกัน จากนั้นจึงคำนวณด้วยสูตรของสเปียร์แมน บราวน์

4.3 โดยการประมาณค่าความเที่ยงของผู้ประเมิน ในการประเมินการปฏิบัติงานจะให้ความสำคัญกับการหาความเที่ยงระหว่างผู้ประเมิน (inter-rater) และความเที่ยงภายในตัวบุคคล (intra-rater) ในการประเมินการปฏิบัติงานเนื่องจากจะต้องใช้ประสบการณ์ และดุลยพินิจของผู้ประเมินสูงในการให้คะแนนงานการปฏิบัติ จึงมีความเป็นปรนัยต่ำ จึงต้องอาศัยเกณฑ์การวัดที่มีความชัดเจน มีการฝึกทักษะและสมรรถนะผู้ประเมินให้มีสมรรถนะใกล้เคียงกัน (ชวลิต ชูกำแพง, 2551; สุวิมล ว่องวานิช, 2546) ดังนั้นการพิจารณาความเที่ยงจึงพิจารณาใน 2 ลักษณะคือ 1) การหาความเที่ยงภายในตัวบุคคล และ 2) การหาความเห็นที่ใกล้เคียงกันระหว่างผู้ประเมิน

5. การตรวจสอบคุณภาพของผู้ประเมิน

คุณภาพของผู้ประเมินที่สำคัญคือ (1) ความเที่ยงระหว่างผู้ประเมิน (inter-rater reliability: IRR) และ (2) ความสอดคล้องระหว่างผู้ประเมิน (inter-rater agreement: RAI) โดยคุณภาพทั้ง 2 ส่วนนี้ มักใช้ในความหมายร่วมกัน หรือใกล้เคียงกัน ซึ่งอาจทำให้เกิดความสับสนได้ เนื่องจากใช้สถิติกลุ่มเดียวกัน แต่แตกต่างกันที่หน่วยของการวิเคราะห์ (unit of analysis) ทั้ง IRR และ IRA ต่างมีการวิเคราะห์ความแปรปรวนรวม (total variance) แต่ต่างนิยามกันโดยความแปรปรวนรวมของ IRA นิยามตามทฤษฎีการทดสอบแบบดั้งเดิม เป็นผลรวมของความแปรปรวนของคะแนนจริง และความแปรปรวนของความคลาดเคลื่อน จากการวัดแบบสุ่ม ซึ่งจะไม่มีส่วนเกี่ยวข้องกับผู้ประเมิน ส่วนความแปรปรวนของ IRR ตามนิยามของ James และคณะ (Liao, Hunt, & Chen, James, 1982; James, Demaree, & Wolf, 1984; 2010) ประกอบด้วย 2 ส่วนคือ (1) เป็นความแปรปรวนของความคลาดเคลื่อนจากการวัดแบบสุ่ม มักมีสาเหตุจากอารมณ์ความรู้สึกของผู้ประเมิน และ (2) คือความแปรปรวนอย่างเป็นระบบ เป็นความแปรปรวนของคะแนนจริง และความแปรปรวนที่สะท้อนความคลาดเคลื่อนจากผู้ประเมิน

IRR มีหน่วยของการวิเคราะห์คือ การเปรียบเทียบความแปรปรวนของคะแนนระหว่างผู้ประเมินที่แตกต่างกัน หรือ สนใจศึกษาความแปรปรวนของคะแนนในผู้ประเมินต่างคนกัน ในการประเมินผู้ถูกประเมินทุกคน (the consistency of scores from different raters across all students) ส่วนIRA ศึกษาความแปรปรวน ของคะแนนภายในผู้ถูกประเมิน หรือ ให้ความสำคัญกับความคงที่ของคะแนนที่ผู้ประเมินต่างคนกันประเมินผู้ถูกประเมินทีละคน (the stability of scores from different raters to each student) (Liao et al., 2010) ความเที่ยงระหว่างผู้ประเมิน เป็นการประเมินเพื่อทราบว่าผู้ประเมินมีความสามารถในการแยกความแตกต่างระหว่างหัวข้อในการประเมิน หรือผู้ถูกประเมินมากน้อยเพียงใด ในขณะที่ความสอดคล้องระหว่างผู้ประเมินเป็นการบอกระดับของการให้คะแนนตรงกัน หรือใกล้เคียงกันของ ผู้ประเมินต่องานชิ้นเดียวกัน หรือผู้ถูกประเมินคนเดียวกัน การหาค่าความเที่ยงระหว่างผู้ประเมินมีหลายวิธี แบ่งเป็นผู้ประเมินสองคน และ ผู้ประเมินมากกว่าสองคน ดังรายละเอียดต่อไปนี้ (Nitko, 2007)

5.1 วิธีการประเมินความเที่ยง หรือความสอดคล้องระหว่างผู้ประเมิน

5.1.1 การคำนวณค่าร้อยละของความเห็นที่ตรงกัน (percentage agreement) มีหลักในการคำนวณง่ายที่สุด โดยนำคะแนนที่เห็นตรงกันหารด้วยจำนวนเหตุการณ์ทั้งหมดที่ต้องสังเกต

โดยไม่ได้นำโอกาสในการให้คะแนนตรงกันโดยบังเอิญมาคำนวณ จึงทำให้เกิดความสอดคล้องในการให้คะแนนที่สูงเกินความเป็นจริง (overestimate the level of agreement) (Hallgren, 2012) ถ้าผู้ประเมิน เห็นด้วยกับคะแนนที่ให้ไปแล้วทุกประเด็นแสดงว่ามีร้อยละของการเห็นด้วย ร้อยละ 100 หากเห็นไม่ตรงกันร้อยละ 100 แสดงว่ามีความเห็นด้วยตรงกันร้อยละ 0 สามารถคำนวณโดยสูตรต่อไปนี้ (Nitko, 2007)

$$PA = [\text{สัดส่วนต่อร้อยละของการตัดสินว่าผ่าน}] + [\text{สัดส่วนต่อร้อยละของการตัดสินว่าไม่ผ่าน}]$$

ตัวอย่างเช่น มีแบบทดสอบ ชุด ก และชุด ข ให้ผู้เรียนจำนวน 25 คน ทำแบบทดสอบ และการผ่านคือ ต้องได้คะแนนร้อยละ 80 ขึ้นไป และพบว่าผลการประเมินด้วยความเห็นชอบของผู้ประเมินทั้งสองคนที่ตรงกัน พบว่าผ่าน 11 คน และไม่ผ่านจำนวน 9 คน สามารถคำนวณค่าร้อยละของความเห็นด้วยได้ดังนี้

$$P = \frac{11}{25} + \frac{9}{25} = \frac{20}{25} = 0.80$$

5.1.2 การคำนวณค่าสัมประสิทธิ์ความสอดคล้องของKendal (Kendall's Coefficient of concordance: KCC) ใช้กับข้อมูลที่มีระดับการวัด เป็น Nominal scale โดยใช้หลักการคำนวณของ Spearman's rank correlation coefficient ในการตรวจสอบความสัมพันธ์ของผลการประเมินระหว่างผู้ประเมิน 2 คน ใช้สถิติ Kendall's tau แต่หากมีผู้ประเมินตั้งแต่สามคนขึ้นไปใช้สถิติ Kendall's W เพื่อพิจารณาว่ามีการให้คะแนนอยู่ในช่วงเดียวกันหรือไม่ (Shih Chieh Liao, Elizabeth A Hunt, & Walter Chen, 2010)

5.1.3 การคำนวณค่าสหสัมพันธ์ของผลการประเมินของผู้ประเมินสองคน (correlation of two scorers' results) อาจเรียกอีกชื่อหนึ่งว่า ความเที่ยงของผู้ให้คะแนน (scorer reliability) หรือ ความเที่ยงระหว่างผู้ประเมินหลายคน (inter-rater reliability) เป็นความเที่ยงที่คาดหวังว่าจะมีความสม่ำเสมอของการให้คะแนนผู้เรียนคนเดียวกันระหว่างผู้ให้คะแนนสองคน หรือมากกว่า 2 คน โดยการคำนวณจากค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน

5.1.4 Rater agreement index (RAI) เป็นการหาผลต่างของคะแนนการประเมินของผู้ประเมินหารด้วยคะแนนที่เป็นไปได้ทั้งหมดลบ 1 และนำผลรวมที่เป็นค่าสัมบูรณ์ มาลบออกจากจำนวนเต็ม 1 ค่า RAI มีค่าอยู่ระหว่าง 0 - 1 หากมีค่าเป็น 1 แสดงว่าผลการประเมินมีความสอดคล้องกันของ

ผู้ประเมินอย่างสมบูรณ์ 100% หรือ หากมีค่าเข้าใกล้ 1 มากแสดงถึงความสอดคล้องของการให้คะแนน ของผู้ประเมินมีสูง การคำนวณมีหลายสถานการณ์ ในที่นี้จะนำเสนอสถานการณ์ที่ใกล้เคียงกับการประเมินภาคปฏิบัติของอาจารย์พยาบาล มีเงื่อนไขของการคำนวณดังนี้ (Bury-stock, Shaw, Laurie, & Chissom, 1996)

กรณีที่ 1 มีมิติการวัดด้านเดียว ผู้ถูกประเมินคนเดียว และผู้ประเมิน 2 คน

$$RAI = 1 - \frac{|R_1 - R_2|}{I - 1}$$

RAI แทน ดัชนีแสดงความสอดคล้องระหว่างผู้ประเมิน

R_1 แทน คะแนนที่ได้จากผู้ประเมินคนที่ 1

R_2 แทน คะแนนที่ได้จากผู้ประเมินคนที่ 2

I แทน คะแนนที่จะเป็นไปได้ทั้งหมด

วิธีการนี้เหมาะสำหรับรายการที่มีมิติการวัดไม่มาก หรือ เป็น รูปรีคแบบภาพรวม

กรณีที่ 2 มีมิติการวัดหลายด้าน ผู้ถูกประเมินคนเดียว และผู้ประเมิน 2 คน คำนวณจากสูตร

$$RAI = 1 - \frac{\sum^K |R_{1k} - R_{2k}|}{K(I - 1)}$$

RAI แทน ดัชนีแสดงความสอดคล้องระหว่างผู้ประเมิน

R_{1k} แทน คะแนนที่ได้จากผู้ประเมินคนที่ 1 ในมิติที่ k (k=1,2,3...k)

R_{2k} แทน คะแนนที่ได้จากผู้ประเมินคนที่ 2 ในมิติที่ k (k=1,2,3...k)

K แทน จำนวนของมิติที่ใช้ประเมินทั้งหมด

I แทน คะแนนที่จะเป็นไปได้ทั้งหมด

กรณีที่ 3 ผู้ประเมินหลายคน ประเมินงานหลายงาน หรือผู้ถูกประเมินหลายคน หลายพฤติกรรม หรืออย่างใดอย่างหนึ่ง ตัวอย่างเช่น มีมิติการวัดด้านเดียว ผู้ถูกประเมินคนเดียว และ ผู้ประเมินหลายคน คำนวณจากสูตร

$$RAI = 1 - \frac{\sum^K |R_m - \bar{R}|}{(M - 1)(I - 1)}$$

RAI แทน ดัชนีแสดงความสอดคล้องระหว่างผู้ประเมิน

R_m แทน คะแนนที่ได้จากผู้ประเมินคนที่ m ($m=1,2,3..m$)

\bar{R} แทน คะแนนเฉลี่ยของผู้ประเมินทั้งหมด

I แทน คะแนนที่จะเป็นไปได้ทั้งหมด

M แทน จำนวนผู้ประเมินทั้งหมด

เงื่อนไขที่ผู้ประเมินหลายคน ทำการประเมินหลายมิติในผู้ถูกประเมินคนเดียว

$$RAI = 1 - \frac{\sum^K \sum^M |R_{mk} - \bar{R}_k|}{K(M-1)(I-1)}$$

เมื่อ R_{mk} แทนการประเมินของผู้ประเมินคนที่ m ต่อมิติการประเมินที่ K มิติ

\bar{R}_k แทน ค่าเฉลี่ยของการประเมินของจำนวนมิติการวัดที่ k นั่นคือ

$$\bar{R}_k = \frac{1}{M} \sum^M R_{mk}$$

$a(R_{mk})$ แทนค่าเฉลี่ยของการประเมินจากผู้ประเมินคนที่ $M-1$ จากผู้ประเมินคนอื่น ๆ

นอกเหนือจาก M นั่นคือ $a(R_{mk}) = \frac{1}{M-1} \sum_{i \neq m} R_{ik}$

5.1.5 การคำนวณค่าสัมประสิทธิ์ แคปปา (Kappa coefficient) ใช้กับข้อมูลระดับ นามบัญญัติ (Nominal scale) คือเห็นด้วยตรงกันทั้งหมดในการให้คะแนนการประเมิน เท่ากับ 1 คะแนน หรือมีความเห็นของผู้ประเมินไม่ตรงกันเท่ากับ 0 เท่านั้น (all or nothing agreement) จึงมีช่วงของความไม่เห็นด้วยแคบ แตกต่างจากค่าที่คำนวณได้จาก intra-class correlation coefficient ที่มีช่วงของความไม่เห็นด้วยกว้างกว่า (magnitude of disagreement) และเหมาะสมกับมาตรการวัดในทุกระดับ ตั้งแต่ระดับอันดับ ระดับช่วงชั้น และระดับอัตราส่วน (Hallgren, 2012) การคำนวณทำได้โดยนำคะแนนที่สังเกตได้มาเทียบกับโอกาสโดยบังเอิญที่ผู้ประเมินทุกคนเห็นตรงกัน มีวิธีคำนวณหลายวิธีดังนี้

กรณีที่ 1 ผู้ประเมินสองคน คำนวณโดยใช้ Cohen's kappa ซึ่งมีแนวคิดพื้นฐานมาจาก ค่าความเที่ยงที่ว่าจะต้องคำนึงถึงความคลาดเคลื่อนอื่นๆ เช่น โอกาสที่จะเกิด โดยเป็นการวัดการเห็นด้วยกับ คะแนนที่ให้กับผู้ถูกประเมิน ระหว่างผู้ประเมินที่สังเกตได้ และโอกาสในการเห็นด้วย ดังแสดงในสมการ

Cohen's kappa = (p agreement - p chance agreement) / (1- p chance agreement)

p agreement แทน ความเห็นที่ผู้ประเมินเห็นตรงกันที่สังเกตได้

p chance agreement แทน โอกาสในการที่ผู้ประเมินเห็นด้วยตรงกัน

กรณีที่ 2 ผู้ประเมินหลายคน การประเมินด้วยค่า Fleiss kappa เป็นการคำนวณโดยใช้หลักการเดียวกับการคำนวณด้วย Cohen's kappa แต่สามารถคำนวณในเงื่อนไขที่ผู้ประเมินมากกว่าสองคนขึ้นไป เมื่อ n = จำนวนของผู้ถูกประเมิน หรือจำนวนชิ้นงานที่ถูกประเมิน k = จำนวนของระดับคะแนนที่ใช้ในการประเมิน และ m = จำนวนผู้ประเมินต่อผู้ถูกประเมินแต่ละคนโดยผู้ประเมิน แต่ละคนอาจประเมินผู้ถูกประเมินทุกคน หรือบางคนก็ได้ การใช้ Fleiss kappa จึงเหมาะกับสถานการณ์ที่มีการสุ่มผู้ประเมินเพื่อทำการประเมินผู้ถูกประเมิน และไม่เหมาะกับการศึกษาที่ ผู้ประเมินทุกคน ทำการประเมินงานทุกงาน (Fully-crossed design) โดยไม่มีการสุ่ม ขยายความจากสูตรเดิมของ Cohen's kappa ดังนี้

$$K_j = 1 - \frac{\sum_{i=1}^k x_{ij} (m - x_{ij})}{mn(m - 1)q_j(1 - q_j)}$$

เมื่อ

K_j แทนจำนวนของระดับคะแนนที่ใช้ในการประเมินที่ได้รับการประเมินที่มาตรคะแนนที่ j

i แทนทุกๆผู้ถูกประเมิน หรืองานประเมินชิ้นที่ $i = 1, 2, \dots, n$

j แทนระดับของการประเมิน $j = 1, 2, \dots, k$

x_{ij} แทนจำนวนของผู้ประเมินที่ให้คะแนนการประเมินที่ระดับ j ต่อผู้ถูกประเมินหรืองาน i

โดยการประมาณค่าความเที่ยงของผู้ประเมิน เป็นวิธีการตรวจสอบความแม่นยำของการประเมิน

q_j แทนความเป็นไปได้ที่พยากรณ์ได้ (expected probabilities)

Fleiss kappa มีความยุ่งยากในการคำนวณจึงทำให้ในระยะแรกการใช้งานไม่เป็นที่แพร่หลาย ต่อมามีการพัฒนาโปรแกรมการคำนวณ เช่น โปรแกรม RECal ในการวิเคราะห์ความเที่ยงระหว่างผู้ประเมินหลายคน จึงทำให้มีความเป็นไปได้ในการใช้งานมากขึ้น (Graham, Milanowski, & Miller, 2012)

5.1.6 Intra-class correlation coefficient (ICC) เป็นสถิติที่ใช้ในการวัดความไม่ต่างกันของความแปรปรวนสำหรับคู่ใดคู่หนึ่งของการวัด หรือ การวัดที่มีจำนวนการวัดมากกว่านั้น

โดยใช้พื้นฐานของการวัดสัดส่วนของความแปรปรวน ที่นิยามว่า objects of measurement เรียกว่า targets ของการวัด อาจเป็นผู้ถูกประเมิน หรืองานที่ถูกประเมิน สถิติพื้นฐานที่ใช้ คือ Classical test theory และ Generalizability theory โดยFisher (1938 cited in McGraw & Wong, 1996) ให้นิยามของคำว่า Class หมายถึง ความสัมพันธ์ระหว่างการวัดภายในชั้น ๆ หนึ่ง เนื่องจากมีหน่วยการวัด และความแปรปรวนเดียวกัน ICCเหมาะกับมาตรการวัดในทุกระดับ ตั้งแต่ระดับอันดับ ระดับช่วงชั้น และ ระดับอัตราส่วน โดยมีเกณฑ์การพิจารณาระดับคุณภาพของงานดังนี้ หากค่าน้อยกว่า 0.40 แสดงว่ามีความสอดคล้องกันระหว่างผู้ประเมินต่ำ 0.40 ถึง 0.59 แสดงว่ามีความสอดคล้องระหว่างผู้ประเมิน ปานกลาง 0.60 ถึง 0.74 แสดงว่ามีความสอดคล้องระหว่างผู้ประเมินสูง และ 0.75 ถึง 1 แสดงว่ามีความสอดคล้องระหว่างผู้ประเมินมากที่สุด (Hallgren, 2012) และเหมาะกับเครื่องมือที่มีระดับการวัดตั้งแต่ 5 ระดับคะแนนขึ้นไป (Graham et al., 2012)

ICC มีแนวคิดพื้นฐานมาจากการศึกษาความเที่ยงระหว่างผู้ประเมินหลาย ๆ คน จากการศึกษาความแปรปรวนในลักษณะของเงื่อนไขการวัดที่เราสนใจ (facet) นับเป็น One- facet generalizability study ใน G study สำหรับนิยามของ ICC คือสิ่งที่ถูกวัดจะต้องได้รับการสุ่ม เช่น ผู้ถูกประเมิน ชิ้นงานที่ใช้ประเมิน โดยถือเป็นแหล่งของความแปรปรวน ผู้ริเริ่มในการใช้ ICC คือ Harris (1913 cited in McGraw & Wong, 1996)ใช้ means square วิเคราะห์ความแปรปรวน ต่อมา Shrout และ Fleiss (1979) ได้นำเสนอโมเดล ที่ใช้ในการประเมินความเที่ยงระหว่างผู้ประเมิน 3 โมเดล โดยกล่าวถึงความเที่ยงระหว่างผู้ประเมินว่าเป็นสัดส่วนของความแปรปรวนของสิ่งที่เราสนใจจะวัด ต่อความแปรปรวนรวม บวกกับความคลาดเคลื่อนจากการวัดโมเดล 3 โมเดลของ Shrout และ Fleiss ประกอบด้วย

โมเดลที่ 1 งานแต่ละงานจะได้รับการประเมินโดยสุ่มจากกลุ่มของผู้ประเมิน k คนต่างกลุ่มกัน (a difference of k judges, randomly selected from a larger population of judges) ทำการประเมินงานที่ใช้ประเมินต่างกัน (Judges are nested in targets or nested design) เขียนเป็นสัญลักษณ์คือ ICC (1,k) โดยอิทธิพล และ ปฏิสัมพันธ์ระหว่างผู้ประเมิน งานการประเมิน และความคลาดเคลื่อนแบบสุ่ม จะไม่ถูกแยกพิจารณา มีโมเดลเชิงเส้นดังนี้

$$x_{ij} = \mu + b_j + w_{ij}$$

จากสมการ

μ แทนค่าเฉลี่ยของคะแนนการประเมินของประชากรทั้งหมด หรือคะแนนที่สังเกตได้

b_j แทน true score ของจำนวนงานการประเมิน j โดยมีค่าเฉลี่ยเท่ากับ 0 และมีความแปรปรวนเท่ากับ σ_{τ^2} และเป็นอิสระจากส่วนประกอบอื่น ๆ ในโมเดล

w_{ij} แทนผลรวมส่วนที่เหลือไม่ได้แยกแยะว่าเป็นความคลาดเคลื่อนจากองค์ประกอบใด แต่เป็นผลรวมของอิทธิพลของผู้ประเมิน ปฏิสัมพันธ์ระหว่างผู้ประเมินและงานการประเมิน และความคลาดเคลื่อนจากการวัด มีค่าเฉลี่ยเท่ากับ 0 และมีความแปรปรวนเท่ากับ σ_w^2 ใช้สถิติวิเคราะห์แบบ One-way random effects

โมเดลที่ 2 ผู้ประเมินคนเดียวกัน หรือชุดเดียวกัน ทำการประเมินงานที่ได้จากการสุ่มที่เหมือนกัน (a random sample of k judges is selected from a larger population, judges rates n targets altogether) เขียนเป็นสัญลักษณ์คือ ICC (2, k) หรือเรียกว่า Two-way random effects ใช้ในการศึกษาความสอดคล้องในการให้คะแนน (degree of agreement) ระหว่างผู้ประเมิน ที่ให้ตัวแปรด้านผู้ประเมินเป็นอิทธิพลแบบสุ่ม (random effects)

โมเดลที่ 3 ไม่มีการสุ่มผู้ประเมิน หรือ งาน ผู้ประเมินคนเดิมทำการประเมินงานทั้งหมด (Fixed facets) (each target is rated by each of the same k judges) เขียนเป็นสัญลักษณ์คือ ICC (3, k) ผู้ประเมินทุกคนประเมินงานประเมินที่เหมือนกัน (the same raters rate each case, these are the only raters) หรือเรียกกว่า fully crossed designs หรือ Two-way mixed effect ซึ่งใช้ในการศึกษาความสม่ำเสมอในการให้คะแนน (Consistency) ระหว่างผู้ประเมิน เนื่องจากไม่ได้ให้ความสำคัญกับตัวแปรความแปรปรวนของผู้ประเมิน ตัวอย่างเช่น ผู้ประเมินสองคนทำการประเมินงานชิ้นเดียวกันเนื่องจากกำหนดให้ผู้ประเมินเป็นอิทธิพลคงที่ (fixed effects)

ทั้งโมเดลที่ 2 และ 3 แตกต่างจากโมเดลที่ 1 คือองค์ประกอบของ w_{ij} มีการแยกระบุองค์ประกอบชัดเจน ไม่นำมารวมกันเช่น โมเดล 1 และมีการประมาณค่าความคลาดเคลื่อนของผู้ประเมิน k ในการประเมินงานการประเมิน n ชุด เขียนเป็นสมการได้ดังนี้

$$x_{ij} = \mu + a_i + b_j + (ab)_{ij} + e_{ij}$$

x_{ij} , μ และ b_j มีนิยามเหมือนในสมการของโมเดลที่ 1 ส่วน a_i เป็นค่าเฉลี่ยของคะแนนการประเมินของผู้ประเมินคนที่ i , $(ab)_{ij}$ เป็นระดับที่ผู้ประเมินคนที่ i ให้คะแนนเบี่ยงเบนไปจาก

แนวโน้มปกติที่ต้นเคยประเมินต่องานการประเมินข้อที่ j และ e_{ij} แทนความคลาดเคลื่อนแบบสุ่มในการประเมินคะแนนของผู้ประเมินคนที่ i งานประเมินข้อที่ j และ b_j มีข้อตกลงเบื้องต้นว่ามีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ σ_{T2} เช่นเดียวกับโมเดลที่ 1 ส่วน e_{ij} แทนความคลาดเคลื่อนที่แยกออกมาเป็นอิสระมีการกระจายของข้อมูลปกติ และมีค่าเฉลี่ยเท่ากับ 0 ความแปรปรวนเท่ากับ σ_{e2} , ส่วนโมเดลที่ 2 ต่างจากโมเดลที่ 3 ที่ข้อตกลงเบื้องต้นของค่า a_i และ $(ab)_{ij}$ ในโมเดลที่ 2 ค่า a_i เป็น ตัวแปรแบบสุ่ม (random variable) ที่มีข้อตกลงว่าค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ σ_{j2} , ในขณะที่โมเดลที่ 3 ตัวแปรนี้เป็นค่าคงที่ (fixed effect) กับข้อตกลงเบื้องต้นที่ค่า Σa_i เท่ากับ 0 ค่าพารามิเตอร์ที่เกี่ยวข้องกับ σ_{j2} , คือค่า θ_{j2} , $= \Sigma a_i^2 / (k - 1)$ ซึ่งทำให้โมเดลที่ 3 ไม่มีการประเมินซ้ำ ๆ กันหลาย ๆ ครั้ง โดยผู้ประเมินแต่ละคนกับงานประเมิน แต่ละชิ้นงาน ทำให้ไม่สามารถประมาณค่าแยกจากกันระหว่างค่า $(ab)_{ij}$ และ e_{ij} ได้ เนื่องจากผู้ประเมิน ทำการประเมินงานเดียวกันทั้งหมด จึงประมาณค่าได้เฉพาะปฏิสัมพันธ์ระหว่างองค์ประกอบที่ต่างชิ้นงานที่ประเมิน สำหรับการประเมินงานการประเมินที่ j เหมือนกัน จึงมีข้อตกลงว่า

$$\sum_{i=1}^k (ab)_{ij} = 0$$

ผลของข้อตกลงนี้ทำให้สรุปได้ว่าองค์ประกอบที่มีปฏิสัมพันธ์สององค์ประกอบใด ๆ ก็ตาม ในกรณีเป็นงานประเมินชิ้นเดียวกัน ความสัมพันธ์ระหว่าง $(ab)_{ij}$ และ $(ab)_{ij}$ มีความสัมพันธ์ทางลบ สำหรับเทคนิคทางสถิติที่ใช้ในการคำนวณนั้น โมเดลที่ 2 คือ Two-way random effects ส่วนโมเดลที่ 3 ใช้ Two-way mixed effect

ข้อควรคำนึงในการใช้ค่า ICC มีดังนี้ (Gisev, Bell, & Chen, 2013; Hallgren, 2012)

1. การเลือกใช้โมเดลแบบ one-way หรือ two-way กรณีเลือกใช้ one-way จะใช้ในการศึกษาความแปรปรวนจากแหล่งเดียวคือ ผู้ถูกประเมินแต่ละคนถูกประเมินด้วยรายการพฤติกรรมที่แตกต่างกัน (unordered observations are nested within objects) ในขณะที่โมเดลแบบ two-way จะใช้ในการศึกษาความแปรปรวนระหว่างผู้ประเมินกับงานที่ใช้ในการประเมิน จึงเหมาะที่จะใช้ในการประเมินประเภท fully crossed designs

2. การเลือกใช้วิธีการในการคำนวณ อาจแบ่งการตรวจสอบคุณภาพเป็น 2 กลุ่มคือการคำนวณโดยเทียบกับเกณฑ์ เป็นการให้คะแนนของผู้ประเมินเทียบกับเกณฑ์มาตรฐานที่ควรจะ

เป็น หากต้องการทราบคุณภาพของผู้ประเมินเทียบกับเกณฑ์มาตรฐาน จึงควรหาความสอดคล้องอย่างสมบูรณ์ (absolute agreement) และหากต้องการเปรียบเทียบผลการให้คะแนนของผู้ประเมินภายในกลุ่มว่ามีคุณภาพการให้คะแนนที่มีความสม่ำเสมอหรือไม่ ควรหาความสอดคล้องในการให้คะแนนของผู้ประเมิน (consistency in the ratings)

3. หน่วยในการวิเคราะห์ (unit of analysis) การพิจารณาหน่วยที่ใช้ในการวิเคราะห์ หากทำการประเมินโดยผู้ประเมินหลายคน การหาคุณภาพความเที่ยงระหว่างผู้ประเมินควรเลือกใช้หน่วยในการวิเคราะห์เป็นแบบ average of ratings ICCs แต่หากงาน หรือผู้ถูกประเมินที่ได้รับการประเมินโดยผู้ประเมินหลายคน และนำผลของความเที่ยงระหว่างผู้ประเมินไปใช้ในการสรุปอ้างอิงกับงานที่ถูกประเมินโดยผู้ประเมินคนเดียว หน่วยของการวิเคราะห์ที่ควรเลือกใช้ควรจะเป็น single-measures ICCs โดยทั่วไปค่า average of ratings ICCs มักที่ค่าสูงกว่า single-measure ICCs หากทั้งสองค่ามีทิศทางที่สวนทางกัน ควรนำเสนอทั้งสองค่า

4. ผู้วิจัยควรระบุว่า จะเลือกใช้โมเดลการวิเคราะห์หรืออิทธิพลแบบสุ่ม หรือแบบคงที่ (random or fixed effects) หากผู้ประเมินถูกสุ่มเลือกกลุ่มประชากรของผู้ประเมินหลายชุด และการประเมินจะถูกนำไปใช้ในการสรุปอ้างอิงกับประชากรกลุ่มนั้นควรเลือกใช้โมเดลแบบ random effects model แต่หากผู้วิจัยไม่ต้องการสรุปอ้างอิงไปยังกลุ่มประชากร แม้ผู้ถูกประเมินอาจได้รับการสุ่ม หรือไม่ก็ตาม แต่ผู้ประเมินไม่ได้รับการสุ่มควรเลือกใช้โมเดลแบบ mixed effects model

สำหรับการวิจัยครั้งนี้เลือกใช้ two-way mixed effect model ในการวิเคราะห์ 2 ส่วนคือ ส่วนที่ 1 เป็นการวิเคราะห์ความสม่ำเสมอในการให้คะแนน (consistency) ระหว่างผู้ประเมิน (ผู้เข้ารับการฝึกอบรม) ทั้งหมด เลือกใช้ หน่วยการวิเคราะห์ แบบ average of rating ICCs และ ส่วนที่ 2 เป็นการวิเคราะห์ความสอดคล้องระหว่างผู้ประเมิน (ผู้เข้ารับการฝึกอบรม) แต่ละคนกับผลการประเมินจากผู้ทรงคุณวุฒิ เลือกใช้หน่วยการวิเคราะห์แบบ single-measures ICCs

5.1.7 การเลือกใช้ดัชนีสำหรับตรวจสอบความเที่ยงระหว่างผู้ประเมิน

วิธีการตรวจสอบความเที่ยงระหว่างผู้ประเมินแต่ละวิธี มีข้อดีข้อเสียแตกต่างกัน จากเอกสารทางวิชาการเกี่ยวกับแนวปฏิบัติในการประเมินภาคปฏิบัติของศูนย์ปฏิรูปทางการศึกษาของอเมริกา (Center for educator compensation reform) ได้เสนอว่า เพื่อให้เกิดความครบถ้วนของข้อมูลที่ได้รับ จึงไม่ควรนำเสนอดัชนีเพียงค่าใดค่าหนึ่งเพียงค่าเดียว (Graham et al., 2012) เนื่องจากแต่ละวิธีมีข้อดีข้อเสียของดัชนีที่ใช้ตรวจสอบความเที่ยงระหว่างผู้ประเมิน ดังแสดงในตารางที่ 4

ตารางที่ 4 เปรียบเทียบข้อดีข้อเสียของดัชนีที่ใช้หาความเที่ยงระหว่างผู้ประเมิน

ชนิดของดัชนี	แนวคิดพื้นฐาน	ข้อดี	ข้อเสีย
Percent absolute agreement	ความถี่ที่ผู้ประเมินให้คะแนนตรงกันมากน้อยเพียงใด	<ul style="list-style-type: none"> - คำนวณง่ายเมื่อจำนวนผู้ประเมินและระดับของมาตราประเมินน้อย - ดีความได้ง่าย - เป็นวิธีที่ดีที่สุดเมื่อผู้ถูกประเมินได้รับคะแนนในระดับเดียวกันมาก ๆ เนื่องจากพิจารณาที่ความถี่ 	<ul style="list-style-type: none"> - คำนวณ และแปลผลได้ยาก ถ้ามีระดับคะแนนหลายระดับ - ไม่ได้นำโอกาสในการให้คะแนนตรงกันโดยบังเอิญมาคิด จึงทำให้เกิดการสอดคล้องกันมากเกินไปจนเกินความเป็นจริง (Overestimate) - ไม่สามารถอธิบายแตกต่างระหว่างระดับความไม่เห็นด้วยมากกว่า 2 ระดับ
Cohen's kappa	ผู้ประเมินให้คะแนนการประเมินตรงกันมากน้อยเพียงใดเมื่อนำโอกาสในการให้คะแนนตรงกันโดยบังเอิญมาพิจารณา	เป็นวิธีการที่ดีในการประมาณความสอดคล้องกันระหว่างผู้ประเมินในการประเมินผู้ถูกประเมินที่แตกต่างกันหลายกลุ่ม	<ul style="list-style-type: none"> - หากมีระดับมาตรการประเมินหลายระดับ จะทำให้ยากต่อการคำนวณ และแปลผล - อาจทำให้ได้ค่าต่ำ หากผู้ประเมินส่วนใหญ่ให้คะแนนเกาะกลุ่มไปด้านปลายมาตรการประเมินด้านสูง หรือต่ำมาก ๆ
Intra-class correlation	สัดส่วนของความแปรปรวนระหว่างผลการปฏิบัติงานของผู้ถูกประเมินกับความคลาดเคลื่อนของผู้ประเมินมีสัดส่วนเป็นอย่างไร	<ul style="list-style-type: none"> - คำนวณได้ง่ายกว่าวิธีการประเมินอื่นเพื่อมีผู้ประเมินหลายคน และระดับของมาตรการประเมินมี 5 ระดับหรือมากกว่า - ผลการคำนวณจะใช้ได้ดีในข้อมูลที่มีระดับการวัดที่มีความต่อเนื่อง (interval, ratio scale) 	<ul style="list-style-type: none"> - จำเป็นต้องอาศัยความเข้าใจในสถิติที่ใช้คำนวณ - อาจทำให้เกิดการวิเคราะห์ผิดพลาดหากมีความแปรปรวนระหว่างผู้ถูกประเมินต่ำ

ที่มา: Graham, Milanowski, & Miller, 2012

5.2 การศึกษาวิจัยที่เกี่ยวกับความเที่ยงระหว่างผู้ประเมิน

โดยทั่วไปการนำเสนอผลการตรวจสอบคุณภาพของผู้ประเมิน มักนำเสนอค่าความสอดคล้องระหว่างผู้ประเมิน หรือ ความเที่ยงระหว่างผู้ประเมิน ค่าใดค่าหนึ่ง บางงานวิจัยนำเสนอมากกว่า 1 ค่า จากการสังเคราะห์งานวิจัย ของ Flinn, Braham และ Nair (2015) ตั้งแต่ปี ค.ศ. 1993 ถึง 2010 โดยสืบค้นข้อมูลบนฐานข้อมูลออนไลน์ ที่มีการศึกษาในเรื่อง Inter-rater reliability หรือ test-retest reliability โดยคัดกรองงานวิจัยที่มีคุณภาพตามเกณฑ์ที่ตั้งไว้ ผลการสังเคราะห์ได้งานวิจัยจำนวน 18 เรื่อง พบว่าส่วนใหญ่ศึกษาในอเมริกา ตัวอย่างมีจำนวนระหว่าง 152 ถึง 550 พบว่ามีการใช้ percentage agreement ในการวิเคราะห์ความเที่ยงระหว่างผู้ประเมินในงานวิจัยหลายเรื่อง แต่มีข้อด้อยอย่างมากเนื่องจากไม่ได้มีโอกาสในการให้คะแนนตรงกันโดยบังเอิญมาคำนวณ จึงมีการใช้วิธีการอื่นร่วมด้วย เช่น Kendall's concordance coefficient, Percentage agreement, Cohen's kappa, Intraclass coefficient(ICC), Person correlations, spearman rho, ANOVA และ t-test ส่วนใหญ่ (10 ใน 18 เรื่อง) รายงานผล ความเที่ยงระหว่างผู้ประเมิน 2 ค่า เช่น ค่าpercentage agreement กับค่าICC, ค่าpercentage agreement กับค่าCohen's kappa หรือ percentage agreement กับค่า Person correlations เป็นต้น ตัวอย่างเช่น งานวิจัยต่อไปนี้

Wang และคณะ (2014) ทำการประเมินความเที่ยงระหว่างผู้ประเมิน คือ พยาบาลวิชาชีพที่มีความเชี่ยวชาญสูง ในการประเมินภาวะเสี่ยงต่อการเกิดแผลกดทับ จำนวน 6 คน ประเมินความเสี่ยง ต่อการเกิดแผลกดทับในคนไข้จำนวน 23 คน โดยแต่ละคนแยกกันประเมิน โดยไม่ได้มีการปรึกษารือกัน ใช้เครื่องมือในการประเมิน 3 ชนิดคือ มาตรการประเมินของBraden, Norton และของ Waterlow ค่าความเที่ยงระหว่างผู้ประเมินที่นำมาใช้คือ ICC พบว่ามาตรการประเมินของทั้งสามชนิดไม่แตกต่างกันมาก คือ Braden มีค่า ICC อยู่ระหว่าง 0.603 (95% CI: 0.435-0.770) ถึง 0.964 (95% CI: 0.936-0.982) ของ Norton มีค่า ICC อยู่ระหว่าง 0.595 (95% CI: 0.426-0.764) ถึง 0.975 (95% CI: 0.955-0.988) และของ Waterlow มีค่า ICC อยู่ระหว่าง 0.592 (95% CI: 0.422- 0.762) ถึง 0.990 (95% CI: 0.982- 0.995)

Niveau, Lacasa, Berclaz และ Germond (2015) ศึกษาความเที่ยงระหว่างผู้ประเมิน ในการใช้เครื่องมือในการประเมินการล่งละเมิดทางเพศชนิด Criteria-based content analysis (CBCA) ซึ่งมีทั้งสิ้น 19 หัวข้อการประเมิน ใช้ในการวิเคราะห์คำให้การของเด็ก เพื่อตีความทางกฎหมายเมื่ออยู่ในระหว่างการพิจารณาคดีของศาลในอเมริกา ศึกษาในตัวอย่าง ผู้ประเมิน 3 คนซึ่งเป็นนักจิตวิทยา และมีความรู้ความเข้าใจในกระบวนการสัมภาษณ์ และเครื่องมือ ที่ใช้ในการประเมิน

เป็นอย่างดี ทำการประเมินด้วยเครื่องมือ CBCA กับวิธีทัศนกรรมการสัมภาษณ์โดยตำรวจ กับเด็กที่ถูกล่วงละเมิดทางเพศ ผลการศึกษาไม่พบความแตกต่างของค่าเฉลี่ยที่ได้จากการประเมินของผู้ประเมินความเที่ยงระหว่างผู้ประเมินตรวจสอบด้วยค่า Pearson's correlation coefficients 3 คู่ได้ค่า 0.71, 0.73 และ 0.81, $p < .01$ ตามลำดับ และค่า Kendall's w coefficient ได้ค่า 0.85, $\text{Khi}^2 = 238.32$ ด้วยค่า $p < .01$ ซึ่งเป็นความเที่ยงระหว่างผู้ประเมิน และตรวจสอบความสอดคล้องระหว่างผู้ประเมิน เป็นรายข้อด้วยค่า percentage agreement ค่าอยู่ระหว่างร้อยละ 48 ถึง 94 และ Kendall's w coefficient ค่าอยู่ระหว่าง 0.40 ถึง 0.79

Ekegren, Hart และ Gabbe (2016) ศึกษาความเที่ยงระหว่างผู้ประเมินสองกลุ่ม คือ ผู้เชี่ยวชาญกับผู้ประเมินจำนวน 16 คน และ ระหว่างผู้ประเมินด้วยกันเอง ผู้ประเมินจำนวน 16 คน คือ เจ้าหน้าที่ ที่ปฏิบัติงานในการประเมินระดับความพิการที่กำลังจะเกิดขึ้นในผู้ป่วยที่ได้รับอุบัติเหตุ เช่น พยาบาล นักกายภาพบำบัด หรือนักจิตวิทยา ด้วยเครื่องมือคือ Extended Glasgow outcome scale (GOS-E) ทำการศึกษากฎนิศึกษาบทสัมภาษณ์ของผู้ป่วยที่บาดเจ็บจำนวน 15 ราย แล้วประเมินออนไลน์ ผลการศึกษาความเที่ยงระหว่างผู้ประเมินด้วย GOS-E นำเสนอ 2 ค่า คือ percentage agreement และ Cohen's kappa ผลของ percentage agreement ระหว่างผู้ประเมิน 16 คน เท่ากับร้อยละ 63 ถึงร้อยละ 100 ส่วนผลที่มีความสอดคล้องระหว่างผู้ประเมิน ร้อยละ 100 เป็นมาตรการประเมินส่วนปลายทั้งสองด้านคือ การประเมินที่ได้คะแนนน้อยที่สุด และการประเมินที่ได้คะแนนดีที่สุด ส่วนมาตรการประเมินกลางๆ จะได้ผลของ percentage น้อยลงมา และความเที่ยงระหว่างผู้เชี่ยวชาญกับผู้ประเมินทั้ง 16 คน พบค่าอยู่ระหว่างร้อยละ 73 ถึงร้อยละ 100 และ ค่า Kappa อยู่ระหว่าง 0.65 ถึง 1 คือมีความเที่ยงระหว่างผู้ประเมินอยู่ในระดับดีถึงระดับดีมาก

5.3 ปัจจัยที่มีผลต่อความตรง และความเที่ยงของการประเมินการปฏิบัติงาน (Nitko, 2007)

5.3.1 ความสอดคล้องของงานที่มอบหมาย และเครื่องมือที่ใช้ในการประเมิน เช่น หากเป้าหมายการเรียนรู้คือเพื่อให้ผู้เรียนสามารถเตรียมยาฉีด การประเมินการปฏิบัติงานก็ควรเลือกงานให้เหมาะสม เช่น ให้ผู้เรียนสาธิตการผสมยาฉีดได้จริง และสามารถบอกสมรรถนะของเทคนิคการผสมยาฉีดว่ามีคุณภาพระดับใด แทนการเขียนบทความเรียงเรียงวิธีการคำนวณ และการผสมยาฉีด

5.3.2 ความสอดคล้องของงานที่มอบหมายกับการใช้กระบวนการคิดที่สอดคล้องกับหลักสูตร งานที่ได้รับมอบหมายจะต้องใช้กระบวนการคิดที่มีความเฉพาะเจาะจงตามเนื้อหาของ

หลักสูตร ทันทีที่ผู้เรียนลงมือปฏิบัติงานตามที่ได้รับมอบหมายผู้สอนสามารถประเมินกระบวนการคิดที่สอดคล้องกับหลักสูตรได้

5.3.3 ความหลากหลายของวิธีการประเมิน ควรใช้วิธีการประเมินการปฏิบัติงานให้หลากหลายวิธีเพื่อให้สามารถตอบสนองต่อมาตรฐานการศึกษาระดับชาติ และตอบสนองต่อเป้าหมายการเรียนรู้ในหลักสูตรเพื่อให้ได้การประเมินผลการเรียนรู้ที่มีความครอบคลุมมากที่สุด

5.3.4 การมีส่วนร่วมในการสร้างเกณฑ์ของผู้ใช้เกณฑ์ ควรให้ผู้สอนหลายๆคนช่วยกันสร้างเกณฑ์การให้คะแนน

5.3.5 จำนวนผู้ประเมิน ควรใช้ผู้ประเมินมากกว่า 1 คนในการประเมินการปฏิบัติงาน

5.3.6 การทำความเข้าใจให้ตรงกันระหว่างผู้ประเมิน โดยแจ้งถึงเกณฑ์ที่จะใช้ในการประเมินให้ผู้เรียนเข้าใจ

5.3.7 การขจัดปัจจัยที่ไม่สำคัญต่อวัตถุประสงค์ของการประเมินทิ้งไปเพื่อที่จะได้ผลการประเมินที่สะท้อนความสามารถที่แท้จริงของผู้เรียน

5.3.8 การฝึกทักษะและสมรรถนะของผู้ประเมินให้ใกล้เคียงกัน และสุ่มตรวจสอบผู้ประเมินเป็นครั้งคราว

6. การตัดสินผลการประเมินทักษะการปฏิบัติงาน

ประกอบด้วยขั้นตอนสำคัญดังนี้ (สุมิตล ว่องวาณิช, 2546)

6.1 กำหนดสมรรถนะของการปฏิบัติงาน (ชิ้นงาน) มีจุดเน้นที่ทักษะภาคปฏิบัติตามวัตถุประสงค์ของธรรมชาติรายวิชานั้น ๆ ที่อาจารย์รวมเอาความรู้ภาคทฤษฎีกับทักษะภาคปฏิบัติเข้าด้วยกัน และนำมาสู่การสร้างเครื่องมือที่มุ่งเน้นทักษะการปฏิบัติที่มีพื้นฐานมาจากความรู้ภาคทฤษฎี มีการกำหนดส่วนประกอบของงานที่ต้องวัด และกำหนดน้ำหนักความสำคัญของชิ้นงานให้ชัดเจน

6.2 รวมข้อมูลที่ได้จากการวัดผลจากงานที่ทำแต่ละชิ้นเข้าด้วยกัน โดยทั่วไปต้องมีการปรับคะแนนดิบตามน้ำหนักความสำคัญของงาน มีวิธีการปรับคะแนนดิบหลายวิธีดังนี้

6.2.1 กรณีคะแนนเต็มของแต่ละชิ้นงานเท่ากันให้คุณคะแนนดิบด้วยน้ำหนักความสำคัญของงานนั้นแล้วจึงรวมคะแนน เช่น

ชิ้นงาน	คะแนนเต็ม	คะแนนที่ได้	น้ำหนักงาน	ผลการปรับคะแนน
การทดสอบภาคปฏิบัติ	50	30	2	$30 \times 2 = 60$
รายงาน	50	30	3	$30 \times 3 = 90$
การปฏิบัติงาน	50	40	3	$40 \times 3 = 120$
				คะแนนรวม = 270

6.2.2 กรณีคะแนนเต็มของแต่ละชิ้นงานไม่เท่ากัน แปลงคะแนนดิบให้อยู่ฐานเดียวกัน เช่น แปลงให้เป็นฐาน 100 แล้วจึงคูณด้วยน้ำหนักความสำคัญของงานแต่ละชิ้น เช่น

ชิ้นงาน	คะแนนเต็ม	คะแนนที่ได้	แปลงเป็นฐาน 100	น้ำหนักงาน	คะแนนหลังการแปลง
การทดสอบภาคปฏิบัติ	30	20	67	2	$67 \times 2 = 134$
รายงาน	40	30	75	3	$75 \times 3 = 225$
การปฏิบัติงาน	50	40	80	3	$80 \times 3 = 240$
				คะแนนรวม = 599	

6.2.3 วิธีแปลงคะแนนให้เป็นคะแนนมาตรฐาน เป็นการแปลงคะแนนดิบให้มีฐานเดียวกัน โดยไม่คำนึงถึงส่วนเบี่ยงเบนมาตรฐานของคะแนนแต่ละชิ้นงาน เช่น คะแนน Z หรือคะแนน T หรือคะแนน สเตโนน แล้วจึงนำน้ำหนักความสำคัญมาคูณ และรวมผลคูณเป็นคะแนนรวม ตัวอย่างของการใช้คะแนนสเตโนน จากสูตร $stanine = 5 + 2Z$

ชิ้นงาน	คะแนนเต็ม	คะแนนที่ได้	สเตโนน	น้ำหนักงาน	ผลคูณ
การทดสอบภาคปฏิบัติ	30	20	3	2	6
รายงาน	40	30	5	3	15
การปฏิบัติงาน	50	40	7	3	21
				คะแนนรวม = 42	

6.3 กำหนดกรอบที่ต้องการอ้างอิงผลการตัดเกรด (frame of reference) ใช้เพื่อให้ทราบว่าระดับคุณภาพของงานภาคปฏิบัติที่จะใช้ในการตัดสินเกรดของผู้เรียนอยู่ในระดับใด มี 3 ประเภทคือ

6.3.1 การเปรียบเทียบกับระดับความสามารถภายในกลุ่ม หรือแบบอิงกลุ่ม (norm-referenced) หมายถึงการเปรียบเทียบระดับคุณภาพของงานของผู้เรียนภายในกลุ่มผู้เรียนมีจุดประสงค์เพื่อจำแนกผู้เรียน

นอกจากกัน เหมาะกับการประเมินผลสรุป (summative assessment) การปรับคะแนนจุดตัดว่าจะให้ตกหรือผ่านอาจปรับขึ้นหรือลงโดยอาศัยค่าความคลาดเคลื่อนมาตรฐาน (standard error of the measure) วิธีนี้มีข้อดีคือ ทำได้ง่าย สะดวกต่อผู้ใช้ ข้อเสียคือผลการประเมินขึ้นอยู่กับความสามารถของคนในกลุ่ม หากกลุ่มผู้เรียนส่วนใหญ่มีระดับความสามารถต่ำผู้เรียนที่มีระดับความสามารถปานกลางอาจมีระดับคุณภาพการปฏิบัติดีเมื่อเทียบกับผู้เรียนในกลุ่มที่มีระดับความสามารถต่ำมาก ๆ แต่ระดับคุณภาพนั้นอาจไม่ถึงระดับคุณภาพตามมาตรฐานก็ได้ รวมถึงไม่สามารถระบุถึงจุดบกพร่องของผู้เรียนได้

6.3.2 การเปรียบเทียบระดับความสามารถกับเกณฑ์มาตรฐาน หรือแบบอิงเกณฑ์ (criterion-referenced) เป็นการเปรียบเทียบระดับความสามารถกับเกณฑ์มาตรฐานที่กำหนด เหมาะกับการประเมินความก้าวหน้า (formative assessment) เป็นการให้ข้อมูลที่ละเอียด ทำให้ทราบว่าผู้เรียนและผู้สอนควรปรับปรุงจุดบกพร่องที่ใด เป็นการให้ข้อมูลป้อนกลับที่ดี ข้อเสียของวิธีนี้คือ ยากที่จะหาจุดตัด หรือเกณฑ์มาตรฐานที่จะจัดจำแนกผู้ที่มีทักษะกับไม่มีทักษะออกจากกันได้อย่างเหมาะสม และยากที่จะแปลความหมายของเกณฑ์ที่กำหนด

6.3.3 การเปรียบเทียบระดับความสามารถกับการพัฒนาตนเองหรือแบบอิงตนเอง (self-referenced) เป็นการเปรียบเทียบความสามารถในการปฏิบัติงานของผู้เรียนกับทักษะความสามารถเดิมที่มีอยู่ของตนเอง ผลการประเมินขึ้นอยู่กับระดับของพัฒนาการในตัวผู้ถูกประเมิน ข้อดีคือช่วยกระตุ้นความตื่นตัวของผู้เรียน ให้มีความตั้งใจปฏิบัติงานมากขึ้น เกรดที่ได้ขึ้นอยู่กับความสามารถของตนเอง ไม่เน้นการแข่งขันระหว่างสมาชิกในกลุ่ม ข้อเสียคือไม่มีการให้ข้อมูลย้อนกลับว่ามีทักษะการทำงานตามเกณฑ์หรือไม่

6.4 การให้เกรดจะขึ้นอยู่กับธรรมชาติของการกำหนดกรอบที่ต้องการอ้างอิง แยกได้ดังนี้

6.4.1 การให้เกรดโดยใช้หลักการประเมินผลแบบอิงกลุ่มจะเน้นการกระจายของกลุ่มเป็นสำคัญ โดยการกระจายเป็นโค้งปกติ ผลการตัดเกรดเป็นระดับต่าง ๆ จึงขึ้นอยู่กับคะแนนเฉลี่ยของกลุ่มกับส่วนเบี่ยงเบนมาตรฐาน แต่หากการกระจายไม่เป็นโค้งปกติควรใช้คะแนนมัธยฐานกับขนาดของส่วนเบี่ยงเบนมาตรฐานที่กำหนด

6.4.2 การให้เกรดโดยใช้หลักการประเมินผลแบบอิงเกณฑ์ ระดับของเกรดที่ได้จะขึ้นอยู่กับคุณภาพของการปฏิบัติที่เป็นไปตามคุณภาพของงานที่กำหนดขึ้นเป็นเกณฑ์ ตัวอย่างเช่น

- A = ผลการปฏิบัติงานดีเยี่ยม มีทักษะการปฏิบัติงานถูกต้อง มีความคิดสร้างสรรค์คุณภาพของงานที่ผลิตสวยงาม
- B= คุณภาพของงานดี มีทักษะการปฏิบัติงานถูกต้อง คุณภาพของงานสวยงาม แต่ยังไม่ได้แสดงความคิดสร้างสรรค์ที่เป็นของตนเองเท่าที่ควร
- C= คุณภาพของงานเป็นที่พอใจ มีทักษะการปฏิบัติงานถูกต้อง คุณภาพของงานพอใช้
- D= คุณภาพของงานเป็นที่พอใจน้อย ยังต้องปรับปรุง การปฏิบัติงานยังมีผิดพลาด คุณภาพของงานยังไม่ดีเท่าที่ควร
- F= คุณภาพของงานไม่เป็นที่พอใจ ยอมรับไม่ได้ การปฏิบัติงานผิดพลาด ผลงานยังใช้ไม่ได้

หรืออาจนำคะแนนรวมของผู้เรียนมาเปรียบเทียบตามเกณฑ์ที่กำหนด เช่น

คะแนนรวม	เกรด
90% ขึ้นไป	A
80%-89%	B
70%-79%	C
60%-69%	D
ต่ำกว่า 60%	F

การประเมินผลโดยการตัดเกรดแบบอิงเกณฑ์มักไม่รายงานผลโดยการแสดงผลของเกรดเพียงอย่างเดียว แต่ควรระบุวัตถุประสงค์ที่วัดประกอบด้วย ลักษณะการรายงานนี้เรียกว่า การรายงานความก้าวหน้า

6.4.3 การให้เกรดโดยใช้หลักการประเมินอิงตนเอง ทำโดยติดตามเปรียบเทียบพัฒนาการของผู้เรียนตั้งแต่เริ่มต้นจนถึงสิ้นสุดการเรียนการสอนเพื่ออัตราการเจริญเติบโต (growth rate) ใช้เพื่อกระตุ้นแรงจูงใจของผู้เรียนในการประเมินความก้าวหน้ามากกว่าที่จะใช้ในการประเมินสรุปรวม

ตอนที่ 2 พยาบาลศาสตร์กับการประเมินภาคปฏิบัติ

1. หลักสูตรพยาบาลศาสตร์

ข้อบังคับสภาการพยาบาลว่าด้วยการให้ความเห็นชอบหลักสูตรการศึกษาระดับปริญญาพยาบาล และการผดุงครรภ์ ระดับวิชาชีพ พ.ศ. 2554 (สภาการพยาบาล, 2554) ให้ความหมายของหลักสูตรว่า หมายถึงหลักสูตรการศึกษาระดับปริญญาพยาบาลและการผดุงครรภ์ ระดับปริญญาตรี หรือประกาศนียบัตร เทียบเท่าปริญญาตรี หลักสูตรต้องมีจำนวนหน่วยกิตรวมไม่น้อยกว่า 140 หน่วยกิต และไม่เกิน 150 หน่วยกิต โดยมีสัดส่วนจำนวนหน่วยกิตของแต่ละหมวดวิชา ดังนี้

1.1 หมวดวิชาศึกษาทั่วไป จำนวนหน่วยกิต ไม่น้อยกว่า 30 หน่วยกิต

1.2 หมวดวิชาเฉพาะ จำนวนหน่วยกิต ไม่น้อยกว่า 104 หน่วยกิต จำแนกได้ ดังนี้

1.2.1 กลุ่มวิชาพื้นฐานวิชาชีพ จำนวนหน่วยกิต ไม่น้อยกว่า 28 หน่วยกิต

1.2.2 กลุ่มวิชาชีพ จำนวนหน่วยกิตไม่น้อยกว่า 76 หน่วยกิต โดยจำแนกเป็นรายวิชา ภาคทฤษฎี และการเรียนในห้องปฏิบัติการทางการพยาบาล ไม่น้อยกว่า 51 หน่วยกิต และ รายวิชา ภาคปฏิบัติในสถานการณ์จริง ไม่น้อยกว่า 25 หน่วยกิต โดย 1 หน่วยกิตภาคปฏิบัติในสถานการณ์จริง ต้องจัดการเรียนการสอน ไม่ต่ำกว่า 60 ชั่วโมง ทั้งนี้รายวิชาภาคทฤษฎี และภาคปฏิบัติดังกล่าว จะต้องจัดให้มีกลุ่มรายวิชาการพยาบาลมารดาทารก และการผดุงครรภ์ และการวางแผนครอบครัว ไม่น้อยกว่า 12 หน่วยกิต หมวดวิชาเลือกเสรี จำนวนหน่วยกิตไม่น้อยกว่า 6 หน่วยกิต การกำหนด รายวิชาและจำนวนหน่วยกิต

2. รายละเอียดของหลักสูตรวิชากลุ่มวิชาชีพ

รายวิชาตามข้อบังคับของสภาการพยาบาล มีรายละเอียด ดังนี้

2.1 รายวิชาเฉพาะสาขาวิชาชีพการพยาบาลและการผดุงครรภ์ โดยต้องมีเนื้อหาสาระครอบคลุมทั้ง ภาคทฤษฎี การเรียนในห้องปฏิบัติการ และภาคปฏิบัติทางการพยาบาลระดับวิชาชีพ ที่ปฏิบัติต่อมนุษย์ทุกวัย ทั้งภาวะสุขภาพปกติและเจ็บป่วย ได้แก่ การประเมินภาวะสุขภาพ มโนคติและทฤษฎีทางการพยาบาล สารสนเทศทางการพยาบาล หลักการ และกระบวนการวิจัยทางการพยาบาล การบริหารการพยาบาลและการบริหารองค์กรสุขภาพ

2.2 การพยาบาลผู้ใหญ่ทั้งในภาวะวิกฤติและเรื้อรัง (รวมอายุรศาสตร์ ศัลยศาสตร์และนรีเวชวิทยา)

2.3 การพยาบาลเด็กและวัยรุ่น

- 2.4 การพยาบาลจิตเวช และสุขภาพจิต
- 2.5 การพยาบาลอนามัยชุมชน และการรักษาโรคเบื้องต้น
- 2.6 การพยาบาลมารดา และทารก การผดุงครรภ์ และการวางแผนครอบครัว
- 2.7 การพยาบาลผู้สูงอายุ
- 2.8 กฎหมายและจรรยาบรรณวิชาชีพ

ตัวอย่างหลักสูตรที่สอดคล้องกับข้อบังคับสภาการพยาบาล เช่น หลักสูตรพยาบาลศาสตรบัณฑิต คณะพยาบาลศาสตร์ มหาวิทยาลัยราชภัฏสวนดุสิต ปรับปรุง พ.ศ. 2554 ชื่อหลักสูตรพยาบาลศาสตรบัณฑิต (พย.บ.) จำนวน 141 หน่วยกิต เป็นหลักสูตรระดับปริญญาตรี 4 ปี ตามมาตรฐานคุณวุฒิระดับปริญญาตรี สาขาพยาบาลศาสตร์ พ.ศ. 2552 ผ่านการปรับปรุงจากหลักสูตรพยาบาลศาสตรบัณฑิต โดยมี คณะกรรมการกลั่นกรองหลักสูตรพิจารณา คณะกรรมการสภาวิชาการพิจารณาเห็นชอบหลักสูตร สภามหาวิทยาลัยอนุมัติหลักสูตร และสภาการพยาบาลให้ความเห็นชอบหลักสูตร เริ่มเปิดสอน ภาคการศึกษา ที่ 1 ปีการศึกษา 2555 โดยมีโครงสร้างหลักสูตรดังนี้ (คณะพยาบาลศาสตร์, 2554)

1. หมวดวิชาการศึกษาทั่วไป 30 หน่วยกิต ประกอบด้วย
 - 1.1 กลุ่มวิชาสังคมศาสตร์
 - 1.2 กลุ่มวิชามนุษยศาสตร์
 - 1.3 กลุ่มวิชาภาษา
 - 1.4 กลุ่มวิชาวิทยาศาสตร์และ คณิตศาสตร์ 12 หน่วยกิต
2. หมวดวิชาเฉพาะ 105 หน่วยกิต
 - 2.1 กลุ่มพื้นฐานวิชาชีพ
 - 2.2 กลุ่มวิชาชีพ
3. หมวดวิชาเลือกเสรีไม่น้อยกว่า 6 หน่วยกิต

หมวดวิชาเฉพาะกลุ่มวิชาชีพประกอบด้วย ภาคทฤษฎี/ทดลองและภาคปฏิบัติรวม 77 หน่วยกิต เมื่อเทียบกับข้อบังคับสภาการพยาบาลว่าด้วยการให้ความเห็นชอบหลักสูตรการศึกษาระดับปริญญาตรี การพยาบาล และการผดุงครรภ์ ระดับวิชาชีพ พ.ศ. 2554 จำนวนหน่วยกิตไม่น้อยกว่า 76 หน่วยกิต โดย 1 หน่วยกิตภาคปฏิบัติจะต้องจัดการเรียนการสอน ไม่นต่ำกว่า 60 ชั่วโมง/ภาคการศึกษา เมื่อเทียบเวลาของวิชาในกลุ่มวิชาชีพจึงเป็นสัดส่วนหน่วยกิต มากเกินกว่าครึ่งหนึ่งของจำนวนหน่วยกิต ทั้งหมดในหลักสูตร จึงเห็นได้ว่าการศึกษาระดับปริญญาตรีให้ความสำคัญกับการฝึกปฏิบัติงานเป็นอย่างมาก

การศึกษาภาคปฏิบัติของนักศึกษาพยาบาลเริ่มจากเมื่อศึกษาในภาคทฤษฎี แล้วจึงนำความรู้ที่ได้รับไปสู่การปฏิบัติ โดยเริ่มตั้งแต่การชมวีดิทัศน์ปฏิบัติการพยาบาลด้านต่าง ๆ หรือการสาธิตของอาจารย์ ในสถานการณ์จำลอง หรือในห้องปฏิบัติการทางการพยาบาล จากนั้นจึงให้นักศึกษาฝึกปฏิบัติการพยาบาลกับสถานการณ์จำลอง จนเกิดทักษะในระดับหนึ่ง จึงมีการทดสอบผ่านการสอบภาคปฏิบัติในการปฏิบัติกิจกรรมการพยาบาล แล้วจึงขึ้นฝึกปฏิบัติงานในสถานการณ์จริงบนหอผู้ป่วย ภายใต้การดูแลของอาจารย์ผู้สอนภาคปฏิบัติ

3. มโนทัศน์ หรือทฤษฎีพื้นฐานที่ใช้ในการประเมินภาคปฏิบัติสำหรับนักศึกษาพยาบาล

มโนทัศน์ที่เป็นที่ยอมรับ และใช้อย่างแพร่หลายในการประเมินภาคปฏิบัติทั้งในสถานศึกษาพยาบาลศาสตร์ ไปจนถึงการประเมินการปฏิบัติงานของพยาบาลประจำการในปัจจุบัน คือการใช้สมรรถนะเป็นฐานในการประเมินเนื่องจากมีความเป็นรูปธรรมมากในการอธิบายผลการประเมินที่ได้ (Calman, Watson, Norman, Redfern, & Murrells, 2002)

ทฤษฎีความเชี่ยวชาญด้านการพยาบาลของ Benner (Benner's theory of nursing expertise) (Benner, 1984 cited in Calman et. al, 2002) เป็นแนวคิดของการพัฒนาความเชี่ยวชาญในวิชาชีพพยาบาลจากผู้ที่ยังอ่อนประสบการณ์ (novice) ไปถึงพยาบาลผู้เชี่ยวชาญ (expert) ผ่านกระบวนการ 5 ขั้นตอน คือ 1) ขั้นเริ่มต้น (beginner) ผู้เรียน เรียนรู้ผ่านผู้สอน ซึ่งผู้เรียนจะได้รับความรู้ ข้อเท็จจริง สมรรถนะและการปฏิบัติเฉพาะด้าน และนำความรู้และทักษะไปใช้ในสถานการณ์นั้น ๆ โดยตรง ไม่มีความยืดหยุ่นหรือการพลิกแพลงไปใช้ในสถานการณ์อื่น 2) หลังผ่านการเรียนรู้ของผู้เริ่มต้นแล้วจะมีการพัฒนาไปสู่ขั้น ผู้เริ่มต้นเข้าสู่ระดับสูง (advanced beginner level) ในขั้นนี้ผู้เรียนแต่ละคนเริ่มมีความยืดหยุ่นที่จะใช้และปรับความรู้ และทักษะเข้ากับสถานการณ์อื่น ๆ มากขึ้น แต่ยังยึดติดอยู่กับเนื้อหาของทฤษฎี 3) ขั้นของผู้มีสมรรถนะ (competence stage) ผู้เรียนสามารถสร้างงานที่มีลำดับลดหลั่นของสำคัญของงาน และมีการวางแผนระยะยาวได้ ซึ่งเป็นการเพิ่มระดับของความสามารถ แม้แผนงานจะคงมีความยืดหยุ่นน้อย ยึดติดตรงตามทฤษฎี แต่มีการวิเคราะห์และมีความละเอียดขึ้น 4) ความสามารถ (proficiency stage) เป็นขั้นที่มองภาพกว้างมากกว่าที่จะเป็นการมองภาพส่วนปลีกย่อย ที่ไม่มีความเกี่ยวเนื่องกัน และมองเห็นสมรรถนะบางอย่างที่มีความโดดเด่นแตกต่างจากผู้ที่ยังเรียนขั้นต้น รวมถึงสามารถเข้าใจปัญหา และแก้ปัญหาตามสถานการณ์ได้อย่างดี 5) ขั้นผู้เชี่ยวชาญเป็นผู้ที่มีความสามารถไม่เพียงแต่เข้าใจในงานแต่ยังสามารถตัดสินใจที่จะทำอะไร

เป็นลำดับต่อไป ด้วยความเชี่ยวชาญชำนาญ และเป็นไปโดยธรรมชาติของการตอบสนองที่ผ่านการ ตกผลึกจากความรู้ และประสบการณ์ที่ผ่านมาสามารถตัดสินใจ และแก้ไขปัญหาจากสถานการณ์ ทัวไปได้อย่างมีวิจารณญาณ ไปจนถึงในสถานการณ์ที่ตนเองไม่เคยพบเห็นมาก่อน (Campitelli & Gobet, 2011) ซึ่งสอดคล้องกับการใช้มาตรฐานค่าของเกณฑ์ที่ใช้อ้างอิงในการประเมินในคลินิกของ Bondy สำหรับในประเทศไทย มีการแบ่งระดับของสมรรถนะตามกรอบมาตรฐานคุณวุฒิระดับอุดมศึกษา แห่งชาติ (Thai Qualifications Framework for Higher Education, TQF:HEd) ที่ว่าด้วยมาตรฐานผล การเรียนรู้ ประกอบด้วย 1) คุณธรรม จริยธรรม 2) ความรู้ 3) ทักษะทางปัญญา 4) ความสัมพันธ์ระหว่าง บุคคล และความรับผิดชอบ 5) ทักษะการวิเคราะห์เชิงตัวเลข การสื่อสาร และการใช้เทคโนโลยีสารสนเทศ และสำหรับการพยาบาลศึกษามีการเพิ่มมาตรฐานคุณวุฒิอีกด้านหนึ่งคือ 6) ทักษะการปฏิบัติทางวิชาชีพ สาขาพยาบาลศาสตร์ (สำนักงานคณะกรรมการการอุดมศึกษา, 2556) ดังรายละเอียดต่อไปนี้

1. คุณธรรม จริยธรรม

1.1 มีความรู้ ความเข้าใจในหลักศาสนา หลักจริยธรรม และจรรยาบรรณวิชาชีพ ตลอดจน สิทธิมนุษยชน สิทธิเด็ก สิทธิผู้บริโภคน สิทธิผู้ป่วย ตลอดจนสิทธิของผู้ประกอบวิชาชีพการ พยาบาล ที่มีความสำคัญต่อการปฏิบัติการพยาบาล

1.2 สามารถแยกแยะความถูกต้อง ความดี และความชั่วได้

1.3 เคารพในคุณค่าและศักดิ์ศรีของความเป็นมนุษย์

1.4 มีความรับผิดชอบต่อการกระทำของตนเอง

1.5 มีระเบียบวินัย และซื่อสัตย์

1.6 ปฏิบัติตามจรรยาบรรณวิชาชีพ และมีความสามารถจัดการกับปัญหาจริยธรรมใน การดำรง ชีพ และในการปฏิบัติงานในวิชาชีพการพยาบาล

1.7 เป็นแบบอย่างที่ดีต่อผู้อื่นทั้งในการดำรงตนและการปฏิบัติงาน

1.8 ส่งเสริมให้ผู้ป่วย/ผู้ใช้บริการได้รับรู้ และเข้าใจสิทธิของตนเองเพื่อปกป้องสิทธิ ของตนเอง ที่จะถูกละเมิด

2. ความรู้

2.1 มีความรู้ และความเข้าใจในสาระสำคัญของศาสตร์ที่เป็นพื้นฐานชีวิต และพื้นฐานทางวิทยาศาสตร์สุขภาพ ที่ครอบคลุมทั้งวิทยาศาสตร์ มนุษยศาสตร์ สังคมศาสตร์ กฎหมายและการปกครองระบอบประชาธิปไตย

2.2 มีความรู้ และความเข้าใจในสาระสำคัญของศาสตร์ทางวิชาชีพการพยาบาล ระบบสุขภาพ และปัจจัยที่มีผลต่อการเปลี่ยนแปลงของสังคมและต่อระบบสุขภาพ

2.3 มีความรู้ และความเข้าใจในสาระสำคัญของกระบวนการพยาบาล และการนำไปใช้

2.4 มีความรู้ และความเข้าใจในสาระสำคัญของกระบวนการแสวงหาความรู้ การจัดการความรู้ กระบวนการวิจัย กระบวนการบริหารและการจัดการองค์กร

2.5 มีความรู้ และความเข้าใจในสาระสำคัญเกี่ยวกับเทคโนโลยีสารสนเทศทางการพยาบาล และระบบจำแนกข้อมูลทางการพยาบาล

2.6 มีความรู้ ความเข้าใจใน วัฒนธรรม สถานการณ์ที่เปลี่ยนแปลงของประเทศ และสังคมโลกที่มีผลกระทบต่อภาวะสุขภาพและประชาชน

3. ทักษะทางปัญญา

3.1 ตระหนักรู้ในศักยภาพและสิ่งที่เป็นจุดอ่อนของตน เพื่อพัฒนาตนเองให้มีความสามารถเพิ่มมากขึ้น สามารถนำไปสู่การปฏิบัติการพยาบาล การสอน การแสวงหาความรู้ที่มีประสิทธิภาพ และการเป็นผู้นำที่เข้มแข็ง

3.2 สามารถสืบค้น และวิเคราะห์ข้อมูลจากแหล่งข้อมูลที่หลากหลาย

3.3 สามารถนำข้อมูล และหลักฐานไปใช้ในการอ้างอิง และแก้ไขปัญหาอย่างมีวิจารณญาณ

3.4 สามารถคิดวิเคราะห์อย่างเป็นระบบ โดยใช้องค์ความรู้ทางวิชาชีพและที่เกี่ยวข้อง รวมทั้ง ใช้ประสบการณ์เป็นฐาน เพื่อให้เกิดผลลัพธ์ที่ปลอดภัยและมีคุณภาพ ในการให้บริการ การพยาบาล

3.5 สามารถใช้กระบวนการทางวิทยาศาสตร์ ทางการวิจัย และนวัตกรรมที่เหมาะสม ในการแก้ไขปัญหา

3.6 สามารถพัฒนาวิธีการแก้ไขปัญหาที่มีประสิทธิภาพสอดคล้องกับสถานการณ์ และบริบททางสุขภาพที่เปลี่ยนไป

4. ความสัมพันธ์ระหว่างบุคคล และความรับผิดชอบ

4.1 มีความสามารถในการปรับตัวเชิงวิชาชีพและมีปฏิสัมพันธ์อย่างสร้างสรรค์กับผู้ใช้บริการ ผู้ร่วมงาน และผู้บังคับบัญชา

4.2 สามารถทำงานเป็นทีมในบทบาทผู้นำและสมาชิกทีม ในทีมการพยาบาล ทีมสุขภาพ และทีมในชุมชนของระบบบริการสาธารณสุขทุกระดับและในบริบทหรือสถานการณ์ที่แตกต่างกัน

4.3 สามารถแสดงออกซึ่งภาวะผู้นำในการผลักดันให้เกิดการเปลี่ยนแปลงที่ดีในองค์กร ในสถานการณ์ที่หลากหลาย และสถานการณ์เฉพาะหน้า

4.4 มีความรับผิดชอบต่อหน้าที่ ต่อสังคม และรับผิดชอบต่อในการ พัฒนาตนเอง วิชาชีพ องค์กรและสังคมอย่างต่อเนื่อง

5. ทักษะการวิเคราะห์เชิงตัวเลข การสื่อสาร และการใช้เทคโนโลยีสารสนเทศ

5.1 สามารถประยุกต์ใช้หลักตรรกะ คณิตศาสตร์และสถิติ ในการพยาบาลอย่างเหมาะสม

5.2 สามารถแปลงข้อมูลให้เป็นข่าวสารที่มีคุณภาพ รวมทั้งสามารถอ่านวิเคราะห์ และถ่ายทอดข้อมูลข่าวสารแก่ผู้อื่นได้อย่างเข้าใจ

5.3 สามารถสื่อสารภาษาไทยได้อย่างมีประสิทธิภาพทั้งการพูด การฟัง การอ่าน การเขียนและการนำเสนอ รวมทั้งสามารถอ่านวารสาร และตำราภาษาอังกฤษอย่างเข้าใจ

5.4 สามารถใช้ไมโครคอมพิวเตอร์พื้นฐานที่จำเป็น

5.5 สามารถเลือกและใช้รูปแบบการนำเสนอสารสนเทศ ตลอดจนใช้เทคโนโลยีสารสนเทศ และการสื่อสารได้อย่างมีประสิทธิภาพ และเหมาะสมกับสถานการณ์

6. ทักษะการปฏิบัติทางวิชาชีพ

6.1 สามารถปฏิบัติทักษะการพยาบาลอย่างเป็นองค์รวมโดยประยุกต์ใช้ศาสตร์ และศิลปะทางการพยาบาล รวมทั้งใช้กระบวนการพยาบาล หลักฐานเชิงประจักษ์ และการสื่อสารเชิงบำบัดในการพยาบาลบุคคล ครอบครัว และชุมชน

6.2 สามารถปฏิบัติกร การสร้างเสริมสุขภาพ การป้องกันโรค การรักษาพยาบาล การบำบัด และการบรรเทาอาการ และการฟื้นฟูสุขภาพ แก่ผู้ใช้บริการทุกภาวะสุขภาพและทุกช่วงวัยรวมทั้งการผดุงครรภ์ ในทุกระดับของสถานบริการสุขภาพ ตามพระราชบัญญัติวิชาชีพการพยาบาลและการผดุงครรภ์ (พ.ศ. 2528) และที่แก้ไขเพิ่มเติมโดยพระราชบัญญัติวิชาชีพ การพยาบาลและการผดุงครรภ์ (ฉบับที่ 2) พ.ศ. 2540

6.3 สามารถปฏิบัติการพยาบาลด้วยความเมตตา กรุณา และเอื้ออาทร โดยยึดมั่นในคุณธรรม จริยธรรม กฎหมาย และสิทธิของผู้ป่วย

6.4 สามารถปฏิบัติการพยาบาลโดยคำนึงถึงความเป็นปัจเจกบุคคล และความหลากหลายทางวัฒนธรรม

6.5 แสดงภาวะผู้นำในการปฏิบัติงาน สามารถบริหารทีมการพยาบาล ทีมสหสาขาวิชาชีพ และการทำงานในชุมชน ในหน่วยบริการสุขภาพชุมชน

สำหรับการวิจัยครั้งนี้ได้ประยุกต์ใช้ทฤษฎีความเชี่ยวชาญด้านการพยาบาลของ Bender ร่วมกับกรอบมาตรฐานคุณวุฒิระดับอุดมศึกษาแห่งชาติ และเอกลักษณ์ของมหาวิทยาลัยสวนดุสิต คือ มีบุคลิกภาพที่ดี มีความประณีต และพิถีพิถันในการทำงาน มีทักษะด้านคอมพิวเตอร์ และเทคโนโลยีสารสนเทศมาใช้ในการประเมินทักษะการปฏิบัติงาน

4. ปัญหาที่พบในการประเมินการฝึกปฏิบัติงานของนักศึกษาพยาบาล

ปัญหาที่พบในการประเมินการฝึกปฏิบัติงานของนักศึกษาล้วนคล้ายคลึงกับปัญหาในการประเมินการปฏิบัติงานทั่วไป ดังรายละเอียดต่อไปนี้ (Gallagher, Smith, & Ousey, 2012)

4.1 เครื่องมือที่ใช้ พบว่าแต่ละสถานศึกษาด้านการพยาบาลในประเทศอังกฤษใช้สมรรถนะแตกต่างกันไป และใช้ทฤษฎีพื้นฐานของการประเมินแตกต่างกันไปดังนั้นในการเปรียบเทียบของแต่ละสถาบันจึงเป็นเรื่องยาก (Calman et al., 2002) โดยเฉพาะการทำความเข้าใจกับเกณฑ์หรือสมรรถนะของแต่ละบุคคล ซึ่งอาจสร้างความสับสนให้ผู้ใช้เครื่องมือ ผลการประเมิน มีการใช้เกณฑ์ในการกำหนดผลการเรียนภาคปฏิบัติแตกต่างกันไป บางสถาบันอาจใช้เกณฑ์ผ่านกับไม่ผ่าน บางสถาบันให้เป็นการกำหนดเกรดเป็น ระดับ ปัญหาที่พบเช่น มีภาวะเกรดเฟ้อ เมื่อเทียบกับเกรดในวิชาทฤษฎีพบว่าจำนวนผู้ได้เกรดสูงๆในรายวิชาปฏิบัติมีมากกว่าในรายวิชาทฤษฎีมากถึง 5 เท่า

4.2 ผู้ประเมิน อาจไม่เข้าใจวัตถุประสงค์ของการฝึกทักษะอย่างแท้จริง ไม่ได้รับการเตรียมการเป็นผู้ประเมิน และที่สำคัญคือปัญหาความลำเอียงจากผู้ประเมิน ซึ่งเป็นปัญหาของความเป็นปรนัยของผู้ประเมินต่ำ และความคลาดเคลื่อนหรือ ความลำเอียงของผู้ประเมินก็จะยังคงเป็นปัญหาต่อไป แนวทางในการแก้ไขปัญหามีงานวิจัยต่างๆออกมารองรับคือ การฝึกผู้ประเมิน การทำความเข้าใจและการแปลค่าของคำศัพท์ทางวิชาการ การประเมินให้แม่นยำ การกำหนดเกรดที่เที่ยงตรง และการเขียนอุบัติการณ์มาสนับสนุนเหตุผลในการกำหนดเกรดของนักศึกษาแต่ละคน (Donaldson & Gray, 2012)

5. แนวทางในการแก้ปัญหาการประเมินการฝึกปฏิบัติงานของนักศึกษาพยาบาล

พบแนวปฏิบัติในการแก้ปัญหาการประเมินภาคปฏิบัติของนักศึกษาพยาบาลดังนี้ (Donaldson & Gray, 2012)

5.1 การพัฒนาเครื่องมือในการทดสอบทักษะปฏิบัติ ควรใช้รูบรีค การใช้รูบรีคในการประเมินระหว่างกลาง รวมถึงการประเมินสรุปรวม

5.2 การใช้วิธีการประเมิน หลายวิธี ไม่ควรใช้วิธีใดวิธีหนึ่งวิธีเดียว

5.3 ควรมีการฝึกอบรมและให้ความรู้ใหม่ๆแก่ผู้ประเมิน รวมถึงการประเมิน และกำกับ ติดตามการนำผลการประเมินไปใช้

ข้อแนะนำที่สำคัญในการแก้ปัญหาการประเมินการปฏิบัติในสาขาพยาบาลศาสตร์ คือการแก้ปัญหาที่การเตรียมผู้ประเมิน และเพื่อให้เกิดความเข้าใจถึงความสำคัญของการเตรียมผู้ประเมิน จึงจะกล่าวถึง ความลำเอียง หรือความคลาดเคลื่อนจากการประเมินอันเกิดจากผู้ประเมินในลำดับต่อไป

ตอนที่ 3 มโนทัศน์เกี่ยวกับความคลาดเคลื่อนจากผู้ประเมิน

ความคลาดเคลื่อนจากผู้ประเมิน มีสาเหตุมาจากการใช้ความเป็นอัตนัยของผู้ประเมินในการประเมิน (rater error , rater bias or rater effect) มีรายละเอียดดังนี้

1. ประเภทของความคลาดเคลื่อน

ความคลาดเคลื่อนจากผู้ประเมิน (rater effect) มีหลายประเภทจำแนกได้ดังนี้ (Reynolds et al., 2009)

1.1 ความคลาดเคลื่อนฮาโล (Halo effect) สาเหตุเกิดจากความรูสึกประทับใจเกี่ยวกับกลุ่มคนนั้นๆ ต่อมิติใดมิติหนึ่ง และมีผลต่อการประเมินมิติอื่นๆ หรือ การนำความประทับใจในสมรรถนะที่ดีด้านหนึ่งมาส่งผลกับการตัดสินใจผลการประเมินสมรรถนะด้านอื่น ๆ ที่ไม่เกี่ยวข้องกัน มีลักษณะดังนี้

1.1.1 เป็นแนวโน้มที่จะยึดติดอยู่กับมิติเกี่ยวกับสมรรถนะของตัวผู้ถูกประเมินมากกว่าที่จะยึดระดับที่แตกต่างกันของมิติที่ต้องประเมินจริง (Borman, 1979; Saal, Downey, & Lahey, 1980)

1.1.2 เกิดจากการไม่สามารถแยกมิติต่าง ๆ ให้ความชัดเจนได้ (Saal et al., 1980)

1.1.3 เป็นแนวโน้มที่จะจัดวางตำแหน่งของผู้ถูกประเมินในระดับเดียวกันโดยใช้มิติการวัดที่แตกต่างกัน (Bernardin & Walter, 1977)

1.1.4 เป็นแนวโน้มที่ผลการประเมินจะได้รับอิทธิพลจากสมรรถนะเอกมิติด้านบวกหรือด้านลบที่ผู้ประเมินนำมาใช้ในการประเมิน โดยเป็นมิติที่ไม่เกี่ยวข้องกับมิติที่ต้องการประเมิน (Reynolds et al., 2009)

1.2 ความคลาดเคลื่อนจากการกด ปล่อยคะแนน (severity /Leniency effect) มักเกิดขึ้นเมื่อผู้ประเมินไม่มีความรู้เพียงพอในเรื่องที่จะประเมิน ทำให้ผู้ประเมินสร้างกลไกการทดแทนโดยให้คะแนนสูงหรือต่ำเกินไปทั้งกลุ่ม จนเป็นผลให้เกิดปรากฏการณ์ที่เรียกว่า ceiling effect และ floor effort หรือผู้ประเมินบางคนประเมินในระดับคะแนนกลาง ๆ ทั้งกลุ่ม (central tendency effect)

จากการศึกษาเอกสารงานวิจัยพบว่าความคลาดเคลื่อนแบบกด ปล่อยคะแนนมีสมรรถนะการประเมินเหมือนกันแต่ทิศทางตรงข้ามกัน มักทำการศึกษาไปพร้อม ๆ กันเนื่องจากใช้ดัชนีชี้วัดตัวเดียวกัน โดยมีลักษณะดังนี้

1.2.1 เป็นแนวโน้มที่จะให้คะแนนการประเมินสูงกว่า หรือต่ำกว่าในผู้ประเมินแต่ละคนมากกว่าที่จะเป็นการประเมินจากพฤติกรรมของผู้ถูกประเมิน (Saal & Landy, 1977 cited in Saal et al., 1980)

1.2.2 เป็นชุดการตอบสนองของผู้ประเมินที่มีสมรรถนะของความยาก ง่ายในการให้คะแนนที่มีความสม่ำเสมอในการให้คะแนนยากกว่า หรือง่ายกว่าโดยใช้เกณฑ์นอกเหนือจากระดับความสามารถที่ได้กำหนดไว้ในเกณฑ์ (Decotiis, 1977 cited in Saal et al., 1980)

1.2.3 เป็นการประเมินคะแนนที่ห่างจากค่าเฉลี่ยของกลุ่มผู้ประเมินคนอื่นๆไปในทิศทางสูงกว่า หรือต่ำกว่า (Saal et al., 1980)

1.3 ความไม่แม่นยำในการวัด (inaccuracy) หมายถึงความคลาดเคลื่อนแบบสุ่มที่เกิดจากความแตกต่างระหว่างคะแนนการประเมินตามศักยภาพผู้ถูกประเมิน และคะแนนศักยภาพที่แท้จริงของผู้ประเมิน (Yue, 2011)

1.4 ความคลาดเคลื่อนจากการให้คะแนนที่มีแนวโน้มคะแนนเกาะกลุ่มตรงกลาง (central tendency effect) และแบบจำกัดช่วงในการให้คะแนน (restriction of range) ซึ่งเป็นการจำกัดช่วงของการให้คะแนนเกาะกลุ่มอยู่ด้านใดด้านหนึ่งของมาตรการประเมินอาจเป็นการจำกัดอยู่ในกลุ่มคะแนนสูง จะเป็นการปล่อยคะแนน ถ้าจำกัดในช่วงคะแนนต่ำจะเป็นการกดคะแนน และถ้าจำกัดคะแนนอยู่ตรงกลางจะเป็นการให้คะแนนที่ไม่สามารถแยกคุณภาพการประเมินได้ว่าดี หรือไม่ดี เนื่องจากมีคุณภาพกลาง ๆ ทั้งหมด สาเหตุมักเกิดจากผู้ประเมินไม่มีความรู้เพียงพอจนไม่สามารถ

แยกหรือจำแนกได้ว่าถูกผู้ประเมินคนใดมีสมรรถนะเหนือกว่าหรือด้อยกว่าผู้ถูกประเมินคนอื่น ๆ อย่างไร มีสมรรถนะดังนี้

1.4.1 เป็นการให้คะแนนของผู้ประเมินที่ไม่มีทิศทางไปในทางกด หรือปล่อยคะแนน เป็นไปในทางกลางๆ (Saal et al., 1980)

1.4.2 เป็นแนวโน้มที่จะประเมินผู้ถูกระเมินทุกคนอยู่ใกล้เคียงค่าเฉลี่ยของทั้งกลุ่ม (Korman, 1971 cited in Saal et al., 1980)

1.5 ความคลาดเคลื่อนจากการรับรู้ของผู้ประเมินว่าผู้ถูกประเมินมีสมรรถนะที่คล้ายผู้ประเมิน (similar-to-me effect) ซึ่งจะทำให้ผู้ประเมินลำเอียงเพราะรู้สึกชอบ หรือพึงพอใจต่อตัวผู้ถูกประเมิน ซึ่งจะเกี่ยวข้องกับภูมิหลังของผู้ประเมินโดยตรง ทำให้ผู้ประเมินให้คะแนนผู้ถูกประเมินคนนั้น สูงกว่าคนอื่น ๆ

1.6 ลำดับของผลจากการประเมิน (order effects) คือการเปลี่ยนแปลงของคะแนนการประเมินที่เกิดขึ้นเมื่อผู้ประเมินใช้เกณฑ์การประเมินอย่างเคร่งครัดในช่วงแรกของการประเมิน แต่เมื่อเวลาผ่านไปจะประเมินโดยใช้เกณฑ์หย่อนยานลง หรือลืมนำเกณฑ์นั้นทำให้เกิดผลกระทบกับความเที่ยงที่เรียกว่า rater drift หรือ reliability decay

1.7 ความคลาดเคลื่อนแบบขัดแย้ง (contrast effects) เป็นแนวโน้มของผู้ประเมินที่จะเปรียบเทียบผู้ถูกประเมินกับตัวผู้ประเมิน (Murray, 1938 cited in Saal et al., 1980)

1.8 ความคลาดเคลื่อนที่เกิดจากความรู้สึกประทับใจเมื่อแรกพบ (first impression error) เกิดจากความรู้สึกประทับใจของผู้ประเมิน ทำให้ ให้คะแนนสูง โดยไม่ได้พิจารณาผลงาน หรือการปฏิบัติ

1.9 ความคลาดเคลื่อนที่เกิดจากการประเมินให้คะแนนสูง เนื่องจากมีสัมพันธภาพส่วนตัวกับผู้ถูกประเมิน (friendship bias) โดยไม่ได้พิจารณาผลงาน หรือการปฏิบัติ ลักษณะคะแนนที่ออกมาเหมือน similar-to-me effect หรือ first impress error แต่ต่างที่สาเหตุ

1.10 ความคลาดเคลื่อนของผู้ประเมินที่เกิดจากการประเมินที่ผู้ประเมินให้คะแนนการประเมินในมิติ หรือสมรรถนะที่คล้ายกับในมิติ หรือสมรรถนะที่ผู้ประเมินคาดหวังไว้ในใจ (logical error) โดยให้ผลการประเมินไปในทิศทางเดียวกัน คล้ายความคลาดเคลื่อนแบบฮาโล แต่ต่างที่ฮาโลเป็นการนำมิติที่ต่างกันมาตัดสินภายในคน คนเดียวกัน ในขณะที่ logical error ยึดกับมิติแทนบุคคล เช่น เมื่อผู้ประเมินคิดว่ามิติใดมิติหนึ่งมีสมรรถนะเกี่ยวเนื่องกับอีกมิติหนึ่ง จึงประเมินทั้งสองมิติให้มีผลสอดคล้องเป็นไปในทิศทางเดียวกัน

1.11 ความคลาดเคลื่อนที่เกิดจากการใช้ผลการปฏิบัติงานช่วงเวลาที่ใกล้เคียงกับระยะเวลาที่จะสิ้นสุดการประเมินมาใช้ (recency) โดยไม่ได้พิจารณาตั้งแต่ระยะแรก ระยะกลาง และระยะสุดท้ายของช่วงเวลาที่ใช้ประเมิน ตัวอย่างเช่น การประเมินภาคปฏิบัติงานของนักศึกษาที่ขึ้นฝึกในระยะเวลา 4 สัปดาห์ แต่ผู้ประเมินนำผลการปฏิบัติงานที่ใกล้เคียงกับเวลาปัจจุบันคือสัปดาห์ที่ 4 มาใช้ในการประเมิน เป็นต้น

แม้จะมีความคลาดเคลื่อนหลายประเภทที่เกิดจากผู้ประเมิน แต่ความคลาดเคลื่อนที่สำคัญ และได้รับการศึกษาอย่างต่อเนื่อง คือ ความคลาดเคลื่อนประเภท กด ปล่อยคะแนน ให้คะแนนเกาะกลุ่มตรงกลาง และความคลาดเคลื่อนแบบฮาโล (severity/leniency effect, central tendency effect, halo effect) เนื่องจากเป็นความคลาดเคลื่อนอย่างเป็นระบบ (systematic error) ซึ่งจะมีผลต่อคะแนนของผู้ถูกประเมินเป็นภาพรวมของกลุ่ม ในขณะที่ความคลาดเคลื่อนแบบสุ่มจะเกิดผลกระทบต่อทำให้คะแนนผู้ถูกประเมินบางคนเท่านั้น (Kane, 2011) เอกสารงานวิจัยส่วนใหญ่พบการศึกษาความคลาดเคลื่อนแบบฮาโล (halo effect) และ แบบกด ปล่อยคะแนน (leniency or severity) (T. R. Athey & R. M. McIntyre, 1987; Bernadin & Buckley, 1981; Bernardin, 1978; Bernardin & Walter, 1977; Borman, 1979; Feldman, Lazzara, Vanderbilt, & DiazGranados, 2012; Ivancevich, 1979; Knoch, Read, & von Randow, 2007; Latham, 1975; Murphy, Jako, & Anhalt, 1993; E. D. Pulakos, 1984; Saal et al., 1980; Wolfe & McVay, 2010)

2. ดัชนีที่ใช้ในการศึกษาความคลาดเคลื่อนจากผู้ประเมิน

ในการวินิจฉัยว่าการประเมินนั้นมีคุณภาพหรือไม่ จะต้องอาศัยดัชนีที่ใช้บ่งชี้เพื่อระบุถึงสมรรถนะที่มีคุณภาพของการประเมินซึ่งอาจพิจารณาได้จาก ความคลาดเคลื่อนจากผู้ประเมินลดลง และความสอดคล้องกัน หรือความเที่ยงระหว่างผู้ประเมินสูง โดยอาศัยดัชนีดังแสดงในตารางที่ 5

ตารางที่ 5 การเปรียบเทียบรูปแบบของการประเมินคะแนนตามนิยามเชิงปฏิบัติการ

เกณฑ์ที่ใช้ในการประเมินคุณภาพ	นิยามเชิงปฏิบัติการของดัชนีที่ใช้เป็นเกณฑ์
ความคลาดเคลื่อนแบบฮาโล	<ul style="list-style-type: none"> - Dimension intercorrelations - Principle- component analysis - Standard deviations - Rater x Ratee interaction
ความคลาดเคลื่อนแบบปล่อย/ กัดคะแนน	<ul style="list-style-type: none"> - Mean dimension ratings - Rater main effect - Skewness
ความคลาดเคลื่อนแบบเกาะกลุ่มคะแนน ตรงกลาง/การให้คะแนนแบบจำกัดช่วง คะแนน	<ul style="list-style-type: none"> - Mean dimension ratings - Standard deviations - Kurtosis - Ratee main effect
ความเที่ยงระหว่างผู้ประเมิน	<ul style="list-style-type: none"> - Standard deviations - Pearson product-moment correl - Rater x Ratee interaction

ที่มา: Saal, Downey & Lahey, 1980

3. การแก้ปัญหาความคลาดเคลื่อนจากผู้ประเมิน

แนวทางในการแก้ปัญหาความคลาดเคลื่อนจากผู้ประเมิน มีดังนี้ (Rudner, 1992)

3.1 ทำความเข้าใจกับคุณลักษณะของความคลาดเคลื่อนเพื่อหลีกเลี่ยงการเกิดความคลาดเคลื่อนนั้น เช่น ทำความเข้าใจกับ 1) คุณลักษณะทางจิตวิทยาของผู้ประเมิน (psychological traits) เช่น แรงจูงใจ ความวิตกกังวล ความเข้มของแรงจูงใจ ความสำเร็จ ความเชื่อในสมรรถนะแห่งตน และอารมณ์ 2) คุณลักษณะด้านบุคลิกภาพ 3) ความเชื่อเดิมของผู้ประเมิน 4) ความเป็นมาของผู้ประเมิน เช่น เพศ เชื้อชาติ 5) บทบาทการประเมินที่ได้รับมอบหมาย 6) รูปแบบทางสังคมของผู้ประเมิน 7) ประสบการณ์เดิมในการให้คะแนนการประเมิน 8) การได้รับการฝึกอบรมการเป็นผู้ประเมิน (Yue, 2011)

3.2 การฝึกอบรมผู้ประเมิน เป็นการเพิ่มความเที่ยงจากการประเมิน มีงานวิจัยจำนวนมากรองรับว่าเป็นวิธีการแก้ปัญหาความคลาดเคลื่อนจากผู้ประเมินได้ดีที่สุดในขณะนี้ การศึกษาเกี่ยวกับ

การฝึกอบรม ตั้งแต่อดีตถึงปัจจุบัน แสดงให้เห็นถึงประโยชน์ของการฝึกอบรมผู้ประเมินว่าสามารถลดผลกระทบจากผู้ประเมินได้ (Johnson et al., 2009)

3.3 การใช้เทคนิคทางสถิติ โดยศึกษาความแตกต่างระหว่างค่าเฉลี่ยของคะแนนจากผู้ประเมิน และค่าเฉลี่ยของคะแนนจริงจากผลการประเมินทั้งหมดเรียกว่า ความคลาดเคลื่อนจากผู้ประเมิน (rater effect) มีวิธีในการปรับความลำเอียงนี้โดยใช้วิธีการทางสถิติ ดังนี้ (Raymond & Houston, 1990)

3.3.1 การถดถอยยกกำลังสองน้อยที่สุด (Least squares regression) เป็นหนึ่งในวิธีที่เก่าแก่ที่สุดของสถิติยุคใหม่ และเป็นรากฐานของสถิติสมัยใหม่อื่นๆ (Abdi, 2003) ซึ่งเทคนิควิธียกกำลังสองน้อยที่สุดที่เป็นที่นิยมอย่างมาก ในการประมาณค่าพารามิเตอร์ และปรับข้อมูลให้เหมาะสมต่อการทำหน้าที่ของข้อมูลในชุดข้อมูล และใช้ในการบอกสมรรถนะที่เหมาะสมของการประมาณค่า

3.3.2 การยกกำลังสองน้อยที่สุดถ่วงน้ำหนัก (Weighted least squares: WLS) การประมาณค่าพารามิเตอร์ด้วยวิธียกกำลังสองน้อยที่สุด จะเป็นการประมาณค่าของคะแนนจริงแบบไม่มีความคลาดเคลื่อน หรือให้ความคลาดเคลื่อนเป็นศูนย์ อย่างไรก็ตาม หากการประเมินคะแนนมีความสม่ำเสมอในระหว่างผู้ประเมินแล้ว ก็จะฝ่าฝืนข้อตกลงเบื้องต้นของการวิเคราะห์ถดถอยโดยปกติ คือ ความแปรปรวนของความคลาดเคลื่อนที่มีค่าเท่ากันระหว่างผู้ถูกประเมินทุกคน และผู้ประเมินทุกคน ผลก็คือ ความแปรปรวนของการประมาณค่าพารามิเตอร์จะเกิดการขยายตัวขึ้น (inflate)

3.3.3 การใช้รหัสโมเดลเป็นโมเดลพารามิเตอร์เดียวที่ใช้ในข้อสอบแบบ dichotomous ที่กล่าวว่า โอกาสของผู้ทำข้อสอบจะตอบถูกขึ้นอยู่กับความสามารถและความยากของข้อสอบ (Wright & Stone, 1979 cited in Raymond & Houston, 1990) โมเดลของรหัส หรือโมเดล 2 พารามิเตอร์ และ 3 พารามิเตอร์ ที่ระบุว่าความสัมพันธ์ระหว่างความสามารถของบุคคล และโอกาสในการตอบคำถามได้ถูกต้องอยู่ในรูปของฟังก์ชัน โลจิสติก ในขณะที่โมเดล 2 พารามิเตอร์ ยอมให้เกิดความแปรปรวนของความชันรายข้อ และ โมเดล 3 พารามิเตอร์ ยอมให้เกิดความแปรปรวนของความชัน และเส้นโค้งที่เข้าหาค่าที่ไม่สิ้นสุด โมเดลของรหัสมีข้อตกลงเบื้องต้นว่า ความชันจะเท่ากัน และค่าต่ำสุดของเส้นโค้งคือค่าศูนย์ การขยายแนวคิดของโมเดลของรหัสไปสู่การนำไปใช้จริงในประเมินโดย Wright และ Master (1982 cited in Raymond & Houston, 1990) โดยการให้ความสำคัญกับโอกาสของบุคคลในการได้รับการประเมินให้อยู่ในระดับขั้นที่ $m+1$ มากกว่าที่จะอยู่ในระดับขั้นที่ m และเพิ่มพารามิเตอร์อีก 1 พารามิเตอร์ นอกเหนือจากความสามารถของบุคคล และความยากของการประเมิน คือ พารามิเตอร์ที่มีความเกี่ยวข้องกับความแตกต่างระหว่าง

ระดับชั้นของมาตรฐานค่าที่มีค่าใกล้เคียงกัน (threshold parameter) สำหรับมาตรฐานค่าที่อยู่ในระดับชั้น M จะมีพารามิเตอร์รวมของ $M - 1$ พารามิเตอร์เหล่านี้มีข้อตกลงเบื้องต้นว่าเหมือนกันในผู้ประเมินทุกคน ซึ่งเขียนสมการได้ดังนี้

$$\phi_{ijm} = \frac{\exp[\alpha_i - (\beta_j + \tau_m)]}{1 + \exp[\alpha_i - (\beta_j + \tau_m)]},$$

$i = 1$ ถึง n ; $j = 1$ ถึง p ; $m = 1$ ถึง $(M-1)$; (2)

เมื่อ

ϕ_{ijm} หมายถึง โอกาสของผู้ถูกประเมินคนที่ i จะถูกประเมินด้วยผู้ประเมิน j ด้วยการประเมินในกลุ่มคะแนน $m+1$ กับกลุ่มคะแนน m

α_i หมายถึง ระดับความสามารถของผู้ถูกประเมิน i

β_j หมายถึง ดัชนีความคลาดเคลื่อนของผู้ประเมิน j

τ_m หมายถึง ความยากที่สัมพันธ์กับระดับชั้นความยาก $m+1$ กับระดับชั้นความยาก m เรียกว่าเป็นพารามิเตอร์ระดับชั้นความยาก (threshold parameter)

จากสมการที่ 2 นี้ เมื่อมีการประมาณค่า α_i , β_j และ τ_m จึงมีความเป็นไปได้ในการพยากรณ์พารามิเตอร์ของทั้งผู้ถูกประเมินและผู้ประเมิน ซึ่งอธิบายด้วย curvilinear relationship เรียกว่า ฟังก์ชันโลจิส ระหว่างการประเมินที่สังเกตได้ กับผลการปฏิบัติงานจริง การเปลี่ยนรูปของโลจิส จะช่วยในการลด ความคลาดเคลื่อนแบบคะแนนต่ำทั้งกลุ่ม หรือคะแนนสูงทั้งกลุ่ม (ceiling effect or floor effect) ในการประเมินจากการสังเกตโดยการยึดส่วนทางการกระจายของคะแนน ส่วนการแปรผลจะนำเสนอในรูปของหน่วยทางคณิตศาสตร์ ที่เรียกว่า โลจิท (logit) หรือ odd ratio

การประมาณค่าพารามิเตอร์ในโมเดลทำได้โดยการประมาณค่าซ้ำๆ ด้วยวิธี maximum likelihood algorithm เพื่อให้เกิดการปรับการประมาณค่าที่เป็นไปได้สูงสุด พร้อมกันระหว่างบุคคลและผู้ประเมิน โดยเริ่มต้นการประมาณค่าของ ความสามารถของผู้ประเมิน ความลำเอียงของผู้ประเมิน และระดับชั้นของความยากในโมเดล และนำมาเปรียบเทียบกับประเมินจริง และผลต่างระหว่างการประเมินจริง และการประเมินโดยใช้โมเดลพยากรณ์ และทำการประมาณค่าซ้ำ ๆ เพื่อลดความคลาดเคลื่อนให้น้อยลง โมเดลการประมาณค่าของราส์ซเมื่อมีการพิจารณาองค์ประกอบหลาย ๆ องค์ประกอบจึงเรียกว่า การประเมินหลายองค์ประกอบ (Multifaceted rating designs) ซึ่งใช้ในการประเมินในระดับมหภาคอย่างแพร่หลายในปัจจุบัน (Iramaneerat &

Yudkowsky, 2007; Lumley & McNamara, 1995; Myford & Wolfe, 2000; Sudweeks, Reeve, & Bradshaw, 2004)

3.3.4 การประมาณค่าพารามิเตอร์โดยใช้วิธี E - M algorithm เป็นการวิเคราะห์ที่มีพื้นฐานของการวิเคราะห์ถดถอยพหุคูณ โดยใช้พื้นฐานของการประมาณด้วยวิธี maximum likelihood ใช้กับข้อมูลหลายตัวแปร โดยประมาณค่าเฉลี่ย ความแปรปรวนและ ความแปรปรวนร่วม

4. โมเดลหลายองค์ประกอบของราส์ช (Many Facet Rasch Model: MFRM)

4.1 ความเป็นมาของโมเดลหลายองค์ประกอบของราส์ช

Linacre (1989 cited in Linacre, 1994) ได้นำแนวคิดของ Andrich และ Masters (cited in Linacre, 1994) ในการขยายขอบเขตการใช้โมเดลของราส์ชที่มีพื้นฐานของแนวคิดมาจากทฤษฎีการตอบสนองข้อสอบจากแบบตรวจให้คะแนน 2 ค่า (Dichotomous) ที่มีองค์ประกอบ (Facet) ของการวิเคราะห์ ผลจากการสังเกตปฏิสัมพันธ์ระหว่าง 2 องค์ประกอบคือ ข้อสอบและผู้สอบ ไปสู่อองค์ประกอบที่ 3 คือ ผู้ประเมิน หรือมากกว่า 2 องค์ประกอบในหลายสถานการณ์ของการทดสอบ ไปถึงสถานการณ์การทดสอบที่มีหลายองค์ประกอบ สู่แบบตรวจให้คะแนนมากกว่า 2 ค่า (Polytomous) โดย Polytomous IRT model มีข้อจำกัดของการนำไปใช้คือ ความสัมพันธ์ระหว่างความสามารถของผู้ตอบ กับโอกาสของการเลือกตอบแต่ละรายการคำตอบ เป็นความสัมพันธ์ที่ไม่ใช่เชิงเส้นตรง (ศิริชัย กาญจนวาสี, 2550) การนำโมเดลของ Rasch ไปใช้เป็นการลดข้อจำกัดของความสัมพันธ์ ไปสู่ความสัมพันธ์เชิงเส้น (additive linear model) โดยเพิ่มพารามิเตอร์ การกคคะแนนของผู้ประเมินเข้าไปในโมเดล การทำความเข้าใจกับราส์ชโมเดลที่ง่ายที่สุดคือ การศึกษาตัวอย่างของ Wright และ Linacre (1995 cited in Wright & Mok, 2004) โดยยกตัวอย่างการกระโดดข้ามรั้วของนักกระโดดรั้ว 2 คน คนที่ 1 แทนสัญลักษณ์ B_m และคนที่ 2 แทนด้วยสัญลักษณ์ B_n กระโดดข้ามรั้วสูงระดับ i ผลของการกระโดดอาจให้ผลดังนี้ $1 =$ กระโดดผ่าน $0 =$ กระโดดไม่ผ่าน และ $-1 =$ ไม่ได้สังเกตว่าโดดผ่านหรือไม่ หากนำเสนอคะแนนจากการสังเกต จะทำให้เกิดการขาดหายของข้อมูลที่ไม่ทันสังเกต แต่เมื่อแทนค่าด้วย โอกาสในการเกิดเหตุการณ์ ในรูปแมทริกซ์ พบว่าจะไม่มีข้อมูลที่หายไป เมื่อแทนค่า X_{mi} ด้วยผลการโดดของผู้กระโดดคนแรกที่ระดับความสูง i และแทนค่า X_{ni} ด้วยผลการโดดของผู้กระโดดคนที่สองที่ระดับความสูง i ทั้ง X_{mi} และ X_{ni} อาจได้คะแนน 0 กรณีกระโดดไม่ข้าม หรือ 1 ถ้ากระโดดข้าม ดังนั้นโอกาสในการเกิดผลการโดดของทั้ง

สองคนจึงเป็น 2×2 หรือความเป็นไปได้ 4 ผลลัพธ์ คือ คนแรกกระโดดผ่าน คนที่สองกระโดดไม่ผ่าน หรือสลับกัน หรือกระโดดผ่านทั้งคู่ หรือกระโดดไม่ผ่านทั้งคู่ ข้อมูลที่ให้ผลลัพธ์ในการกระโดดเหมือนกันไม่ได้มีประโยชน์ในการหาความสัมพันธ์ระหว่างความแข็งแรงของแต่ละคน นั่นคือ หากกระโดดไม่ผ่านเหมือนกัน หรือกระโดดผ่านเหมือนกันจะทำให้ข้อมูลที่ได้ไม่มีประโยชน์ที่จะบอกได้ว่าคนไหนแข็งแรงกว่ากัน มีเพียงข้อมูล 2 ส่วนที่เหลือคือ คนแรกกระโดดผ่านคนที่สองกระโดดไม่ผ่าน หรือในทางกลับกัน ข้อมูล 2 ส่วนหลังเป็นประโยชน์มากกว่าที่จะบอกได้ว่าใครแข็งแรงกว่ากัน เขียนผลเป็นแมทริกซ์ ได้ดังนี้ (Wright & Mok, 2004)

	คนที่ 2 กระโดดผ่าน $X_{ni} = 1$	คนที่ 2 กระโดดไม่ผ่าน $X_{ni} = 0$
คนที่ 1 กระโดดผ่าน $X_{mi} = 1$	N11, Pni, Pmi	N10, Pmi (1-Pni)
คนที่ 1 กระโดดไม่ผ่าน $X_{mi} = 0$	N01, (1-Pmi) Pni	N00, (1-Pmi) (1-Pni)

เมื่อ

N11 แทน จำนวนครั้งที่ทั้งคนที่ 1 และคนที่ 2 กระโดดผ่านรั้ว (เป็นข้อมูลที่ไม่มีประโยชน์)

N10 แทน จำนวนครั้งที่คนที่ 1 กระโดดผ่าน คนที่ 2 ไม่ผ่าน

N01 แทน จำนวนครั้งที่คนที่ 2 กระโดดผ่าน คนที่ 1 ไม่ผ่าน

N00 แทน จำนวนครั้งที่ทั้งสองกระโดดไม่ผ่านรั้วทั้งคู่ (เป็นข้อมูลที่ไม่มีประโยชน์)

Pni แทน โอกาสที่คนที่ 2 กระโดดผ่านรั้ว i และ $1-Pni$ แทนโอกาสที่คนที่ 2 กระโดดไม่ผ่าน

Pmi แทน โอกาสที่คนที่ 1 กระโดดผ่านรั้ว i และ $1-Pmi$ แทนโอกาสที่คนที่ 1 กระโดดไม่ผ่าน

หากสมมติสถานการณ์การวัดโดยให้ N10 แทนจำนวนครั้งที่คนที่ 1 กระโดดผ่าน แต่คนที่ 2 กระโดดไม่ผ่าน และ N01 แทนจำนวนครั้งที่คนที่ 2 กระโดดผ่าน แต่คนที่ 1 กระโดดไม่ผ่าน ดังนั้นโอกาสของผลลัพธ์ของการกระโดดจึงเป็นสัดส่วนของจำนวนครั้งที่ของ $N10 / N01$ ซึ่งจะได้สารสนเทศมากกว่าผลต่างระหว่าง $N10 - N01$ คือสามารถบอกได้ว่าคนที่ 1 เก่งกว่าคนที่ 2 เป็นก็เท่าซึ่งเป็นค่าที่นำเสนอให้เห็นภาพความคงที่ของข้อมูลมากขึ้น ดังนำเสนอในแมทริกซ์ของสถานการณ์การวัดต่อไปนี้

		สถานการณ์การวัด			
		ก	ข	ค	ง
คนที่ 1 กระโดดผ่านคนที่ 2 ไม่ผ่าน	N10	9	90	900	504
คนที่ 1 กระโดดไม่ผ่านคนที่ 2 ผ่าน	N01	1	10	100	496
ผลต่างของการแข่งขัน(di)	N10- N01	8	80	800	8
สัดส่วน (Ratio)	N10/ N01	9	9	9	\cong $1 \cong$ \cong

เมื่อสถานการณ์ ก, ข และ ค เป็นสถานการณ์ที่คนที่ 1 กระโดดได้ผ่านมากกว่าคนที่ 2 9 เท่า เป็นความคงที่ในสัดส่วน แต่ไม่คงที่ในผลต่าง สถานการณ์ ง เป็นกรณีที่คนที่ 1 และคนที่ 2 มีผลการกระโดดที่ใกล้เคียงกัน ดังเห็นได้จากสัดส่วนของ N10/ N01 ส่วนค่า N10 - N01 สะท้อนภาพที่ไม่ค่อยตรงกับความเป็นจริงนัก เนื่องจากหากพิจารณาค่าผลต่างนี้ คล้ายจะมีนัยว่าคนที่ 1 เก่งกว่าคนที่ 2 หากอยู่ในสถานการณ์ ก ที่มีสัดส่วนการกระโดดผ่านของ N10/ N01 เท่ากับ 9 จากตัวอย่างนี้แสดงให้เห็นว่าสัดส่วนของ N10/ N01 จะทำให้ได้ข้อมูลซ้ำ ๆ กันว่าคนที่ 1 มีความแข็งแรงกระโดดข้ามรั้วขณะคนที่ 2 เป็นก็เท่าไปเรื่อย ๆ เพื่อให้ง่ายต่อการตีความจึงมีการนำเสนอโมเดลของโอกาสในการเกิดสัดส่วนของข้อมูลซ้ำ ๆ กันนี้ เป็นประโยชน์สัญลักษณ์ได้ดังนี้

$$\frac{N_{10}}{N_{01}} \approx \frac{P_{mi}(1-P_{ni})}{P_{ni}(1-P_{mi})}$$

ซึ่งจะทำให้คนที่ 1 และคนที่ 2 มีความสัมพันธ์ที่มีความหมายมากขึ้นในสถานการณ์การกระโดดข้ามรั้วที่ระดับความสูงของรั้วที่แตกต่างกันออกไป และเพื่อให้เกิดความเป็นรูปธรรม และเป็นข้อมูลที่สามารคาคัดคะแนนได้ถึงผลของการเปรียบเทียบโดยไม่ได้ยึดติดกับระดับการวัดที่แตกต่างกัน i หรือ j โอกาสในการเกิดเหตุการณ์สามารถเขียนเป็นประโยชน์สัญลักษณ์ทางคณิตศาสตร์ได้ดังนี้

$$\frac{P_{mi}(1 - P_{ni})}{P_{ni}(1 - P_{mi})} \equiv \frac{P_{mj}(1 - P_{nj})}{P_{nj}(1 - P_{mj})}$$

สำหรับระดับการวัดที่แตกต่างกันทั้งหมดของ i หรือ j

$$\left(\frac{P_{ni}}{(1 - P_{ni})}\right) \equiv \left(\frac{P_{nj}}{(1 - P_{nj})}\right) \left(\frac{(1 - P_{mj})}{P_{mj}}\right) \left(\frac{P_{mi}}{(1 - P_{mi})}\right),$$

แสดงให้เห็นความสัมพันธ์ที่ระดับการวัดที่แตกต่างกัน (ต่างระดับสถานการณ์ของการวัดในงานลักษณะเดียวกัน) แต่ยังคงมีความสัมพันธ์ระหว่างกันคงที่ระหว่างคนที่ m และคนที่ n (เป้าหมายของการวัด) ซึ่งสามารถสรุปอ้างอิงความสัมพันธ์ในเชิงเปรียบเทียบได้ หาก คนที่ 0 และระดับการวัดที่ 0 และกำหนดให้โอกาสในการเกิด P_{00} เท่ากับ 0.5 เมื่อเทียบระดับการวัด 2 ระดับ ระหว่างระดับการวัดที่ 0 กับระดับการวัดที่ i ในคนที่ n เขียนเป็นประโยคสัญลักษณ์ทางคณิตศาสตร์ได้เป็น

$$\left(\frac{P_{ni}}{1-P_{ni}}\right) \equiv \left(\frac{P_{no}}{1-P_{no}}\right) \left(\frac{P_{oi}}{1-P_{oi}}\right) \left(\frac{(1-P_{oo})}{P_{oo}}\right) = \left(\frac{P_{no}}{1-P_{no}}\right) \left(\frac{P_{oi}}{1-P_{oi}}\right), \quad (1)$$

$$\text{นั่นคือ } \left(\frac{P_{ni}}{1-P_{ni}}\right) \equiv f(n) \times g(i) = \frac{b_n}{d_i}$$

$$\text{และ } f(n) = b_n, \quad g(i) = 1/d_i$$

$$b_n = \frac{P_{no}}{(1-P_{no})}, \quad 1/d_i = \frac{P_{oi}}{(1-P_{oi})}$$

สัดส่วนระหว่างโอกาสในการเกิดระดับการวัดที่ 0 ในคนที่ n กับโอกาสในการเกิดระดับการวัดที่ระดับ 0 กับระดับ i เรียกว่า odd ratio มีค่าตั้งแต่ 0 ถึง infinity ขึ้นอยู่กับคนที่ n หรือระดับการวัดที่ i กับกรอบของการอ้างอิงที่ตกลงร่วมกัน การทำให้สัดส่วนอยู่ในรูป log odd หรือเรียกว่า logit scale เป็นการแปลงข้อมูลเพื่อให้เกิดสมการบวกเชิงเส้น (addictive linear model) จะทำให้ได้ข้อมูลที่จะบอกได้ว่าสัดส่วนระหว่างเหตุการณ์เกิดได้มากน้อยแค่ไหน เมื่อนำมาคำนวณด้วย logarithm ในสมการที่(1) จะได้ผลดังนี้

$$\ln\left(\frac{P_{ni}}{1-P_{ni}}\right) \equiv \ln\left(\frac{P_{no}}{1-P_{no}}\right) + \ln\left(\frac{P_{oi}}{1-P_{oi}}\right) \equiv \ln(b_n) - \ln(d_i)$$

$$\text{ซึ่ง } \ln\left(\frac{P_{ni}}{1-P_{ni}}\right) \equiv B_n - D_i \quad \text{หรือ} \quad \log\left(\frac{P_{ni}}{1-P_{ni}}\right) \equiv B_n - D_i \quad (2)$$

$$\text{หรือ } P_{ni} = \frac{\exp(B_n - D_i)}{1 + \exp(B_n - D_i)} \quad (3)$$

สมการที่ 2 และ 3 เป็น รัสซโม่เดลสำหรับข้อมูล dichotomous เมื่อ P_{ni} แทนโอกาสของคน ที่ n ที่มีความสามารถ B_n ทำเป้าหมายในการวัดที่ i สำเร็จด้วยระดับความยากที่ D_i โดยผลต่างระหว่างผลที่คาดว่าจะเกิด (expectation) หรือคะแนนที่คาดว่าจะได้จากการวัด กับผลที่ได้จากการสังเกต หรือคะแนนจริงที่ได้จากการวัด หากมีค่าต่างกันมากเท่าใดแสดงว่ามีความคลาดเคลื่อนจากการวัดมากเท่านั้น เขียนเป็นสมการคือ $Y_{ni} = X_{ni} - P_{ni}$ โดยหาได้จากความแปรปรวนของ X_{ni} $P_{ni}(1 - P_{ni})$ การประมาณค่าความแตกต่างระหว่างผลต่างนี้ ทดสอบด้วย χ^2 (Chi-square

test) ด้วยค่า degree of freedom ของแต่ละค่าของ X_{ni} ซึ่งเป็นสมการที่แทนค่าด้วยคะแนนดิบ แต่เมื่อแปลงเข้าสู่สมการมาตรฐานจะได้ สมการคือ $z_{ni} = (x_{ni} - P_{ni})/\sqrt{P_{ni}(1 - P_{ni})}$ ค่า standard residual ที่เป็นบวกแสดงถึงคะแนนที่สังเกตได้มีค่าสูงกว่าคะแนนที่คาดหวัง หากค่าเป็นลบแสดงว่าคะแนนที่สังเกตได้ต่ำกว่าคะแนนที่คาดหวัง เมื่อแปลงคะแนนดิบของผลต่างเป็นคะแนนมาตรฐานและนำมายกกำลังสอง จะเรียกว่า outfit mean square และ infit mean square ของแต่ละคน แต่ละเป้าหมายของการวัด และแต่ละ threshold ของการประมาณค่า โดย outfit mean square เป็นการย่อยส่วนของ out-lier sensitive mean square residual goodness of fit statistic ที่เป็นการคำนวณแบบไม่ได้ถ่วงน้ำหนัก (unweighted of the fit statistic) ทำการวัดค่าเฉลี่ยของข้อมูลที่ไม่สอดคล้องกันระหว่างคะแนนที่สังเกตได้ กับค่าคะแนนที่คาดหวังโดยนำผลรวมยกกำลังสองของค่าเฉลี่ยของผลต่างทั้งหมดที่เกิดขึ้นกับคน n ทุกคนในการวัดแต่ละ i ของการวัด ค่าที่ได้จาก outfit mean square มักมีค่ามากเกินไปจนเกินความเป็นจริงหากแบบทดสอบง่ายเกินไป หรือยากเกินไปจนทำให้เกิดความแตกต่างระหว่างค่าที่ไม่คาดหวังกับค่าที่เป็นจริง ดังนั้นเพื่อหลีกเลี่ยงปรากฏการณ์นี้ ควรเลือกใช้ infit mean square โดยทำการถ่วงน้ำหนักค่าเฉลี่ยยกกำลังสองของความคลาดเคลื่อนที่เหลือ (weight average mean square of residuals) จึงทำให้ปรากฏการณ์ที่เกิดความแตกต่างกันอย่างมากระหว่างค่าที่คาดหวังกับค่าคะแนนจริงไม่แตกต่างกันมาก

จากตัวอย่างราศีชโมเดลที่มีเงื่อนไขของการวัดที่เรียกว่าฟาเซท (facet) สองฟาเซทคือ เป้าหมายของการวัดคนที่ n (target or object) และงานที่ใช้ในการวัด (agent) i แต่เนื่องจากเงื่อนไขของการวัดมีเงื่อนไขที่น่าสนใจที่จำเป็นต้องศึกษา จึงมีการขยายขอบเขตของ ฟาเซทที่ศึกษาออกไปเป็นหลายๆฟาเซท นอกเหนือจาก 2 ฟาเซทดังกล่าวเช่น ผู้ตัดสิน (judge) หรือผู้ประเมิน (rater)

4.2 โมเดลหลายองค์ประกอบของราศีช สามารถเพิ่มเงื่อนไขการวัด หลายเงื่อนไข เขียนสมการเพิ่มเติมจากสมการที่ (2) ได้ดังนี้ 4.2.1 โมเดลพื้นฐาน ทางคณิตศาสตร์ ที่เรียกว่า rating scale model ที่กล่าวถึงโอกาสของผู้ถูกประเมินที่จะได้รับการประเมินด้วยคะแนน k มากกว่า $k-1$ บนสมรรถนะการวัด (trait) i โดยผู้ประเมิน j ขึ้นอยู่กับอิทธิพลเชิงบวกของความยากของสมรรถนะที่วัด (additive effects of the difficulty of the trait) D_i , การกดคะแนนของผู้ประเมิน C_j , ระดับของผลงานของผู้ถูกประเมิน (ratee) B_n และความยากของกลุ่มมาตรวัดที่ k เมื่อเปรียบเทียบกับกลุ่มมาตรวัดที่ $k-1$ ในทุกสมรรถนะการวัด และผู้ประเมินทุกคน F_k

$$\log\left(\frac{P_{nij k}}{P_{nij k-1}}\right) = B_n - D_i - C_j - F_k \quad (4)$$

เมื่อให้ $P_{nij k}$ แทน ความน่าจะเป็นของผู้ถูกประเมิน n จะถูกให้คะแนนในหัวข้อที่ i โดยผู้ประเมินคนที่ j ให้คะแนนการประเมิน k

$P_{nij k-1}$ แทน ความน่าจะเป็นของผู้ถูกประเมิน n จะถูกให้คะแนนในหัวข้อที่ i โดยผู้ประเมินคนที่ j ให้คะแนนการประเมิน $k-1$

B_n แทน ค่าความสามารถของผู้ถูกประเมิน n

D_i แทน ความยากของการประเมินรายการที่ i

C_j แทน การกดคะแนน (ความเข้มงวดในการให้คะแนน) ของผู้ประเมิน j

F_k แทน ความยากของการประเมินบนกลุ่มคะแนน $k-1$ ถึงกลุ่มคะแนน k เมื่อ $k=1, M$

4.2.2 โมเดลที่เน้นมาตรการวัดที่ผู้ประเมิน (Judge-scale model) เป็นสถานการณ์ที่ผู้ประเมิน (judge or rater) แต่ละคน ใช้มาตรในการประเมินในการตีความหมายการให้คะแนนของตนเอง ดังสมการ

$$\log\left(\frac{P_{nij k}}{P_{nij k-1}}\right) = B_n - D_i - C_j - F_{jk} \quad (5)$$

เมื่อ F_{jk} แทน ความยากของขั้นของกลุ่มคะแนน (category) จาก $k - 1$ ถึงกลุ่มคะแนน k ของผู้ประเมิน j , ส่วนค่าพารามิเตอร์อื่น ๆ นิยามตามสมการที่ 4

4.2.3 โมเดลที่เน้นมาตรการตามรายการประเมิน (Item-scale model) เป็นสถานการณ์ที่รายการประเมินแต่ละรายการมีโครงสร้างในมาตรการวัดตามมาตรการวัดของรายการนั้น ตามสมการต่อไปนี้

$$\log\left(\frac{P_{nij k}}{P_{nij k-1}}\right) = B_n - D_i - C_j - F_{ik} \quad (6)$$

เมื่อ F_{ik} แทน ความยากของขั้นของกลุ่ม (category) คะแนนจากระดับ $k - 1$ ถึงกลุ่มคะแนน k ของมาตรการวัดที่มีลักษณะเด่นของรายการประเมิน i และ $k = 1, M_i$ ส่วนค่าพารามิเตอร์อื่น ๆ นิยามตามสมการที่ 4

4.2.4 โมเดลที่มี 4 ฟาเซท (Four- facet model) เป็นสถานการณ์ที่รายการประเมินแต่ละรายการถูกจัดรูปแบบให้นำไปใช้กับแต่ละจำนวนของงาน

$$\log\left(\frac{P_{nij k}}{P_{nij k-1}}\right) = B_n - A_m - D_i - C_j - F_k \quad (7)$$

เมื่อ A_m แทน ความยากของงาน m และค่าพารามิเตอร์อื่นๆนิยามตามสมการที่ 4

สำหรับการวิจัยครั้งนี้ใช้ rating scale model ที่มี 3 ฟาเซท คือ ผู้ประเมิน (rater) ผู้ถูกประเมิน (ratee) และ มาตรการประเมิน (item) โดยเน้นฟาเซทของผู้ประเมิน ใช้ในการศึกษาสมรรถนะในการให้คะแนน แบบกต ปล่อยคะแนน เน้นฟาเซทของผู้ถูกประเมินในการประเมินสมรรถนะการให้คะแนนแบบเกาะกลุ่มตรงกลาง เน้นฟาเซทของรายการประเมินในการประเมินสมรรถนะของการเกิดความคลาดเคลื่อนแบบฮาโล และการใช้โมเดลที่มี 4 ฟาเซท คือ ประเมิน (rater) ผู้ถูกประเมิน (ratee) มาตรการประเมิน (item) และช่วงเวลาที่ใช้ในการประเมิน 3 ช่วงเวลา (time) เพื่อศึกษาปฏิสัมพันธ์ระหว่างสมรรถนะของการให้คะแนนของผู้ประเมินคือ การกต ปล่อยคะแนน หรือให้คะแนนเกาะกลุ่มตรงกลาง ตามช่วงเวลาที่เปลี่ยนไป

4.3 ดัชนีสำคัญที่ใช้โมเดลหลายองค์ประกอบของรราส์ซ 4 ดัชนีสำคัญดังนี้ (Wolfe & Dobria, 2008)

4.3.1 การประมาณค่าพารามิเตอร์ สำหรับผู้ประเมิน ผู้ถูกประเมิน แต่ละคน และองค์ประกอบต่างๆในแต่ละฟาเซทที่สนใจศึกษา ซึ่งจะอยู่ในรูปของคะแนนที่สังเกตได้ คะแนนที่ได้จากการคาดหวังจากโมเดลของฟาเซท และองค์ประกอบต่างๆ ในฟาเซท

4.3.2 ค่าเฉลี่ยที่เหมาะสม (fair average) เป็นการแปลงค่าคะแนนดิบไปสู่หน่วยการวัดที่สามารถระบุระดับของข้อมูลที่ขาดหายไป หรือการกระจายของข้อมูลที่ก่อให้เกิดความไม่สมบูรณ์ของข้อมูลดิบ เปรียบเทียบระหว่างบุคคลกับองค์ประกอบย่อยๆ ภายในฟาเซท

ทั้งการประมาณค่าพารามิเตอร์ และค่าเฉลี่ยที่เหมาะสมเป็นค่าที่สะท้อนระดับของบุคคล หรือองค์ประกอบที่ต้องการศึกษาว่ามีค่าแตกต่างจากค่าที่คาดหวังด้วยโมเดลหลายองค์ประกอบของรราส์ซมากน้อยเพียงใด โดยการประมาณค่าพารามิเตอร์ เป็นการแปลงข้อมูลให้อยู่ในรูปหน่วยการวัดที่เรียกว่า โลจิต หรือ log-odd ที่มีการปรับค่าความแตกต่างระหว่างระดับของความคลาดเคลื่อนของผู้ประเมิน และ/หรือความแตกต่างที่เกิดขึ้น และค่าเฉลี่ยที่เหมาะสมของบุคคลเปลี่ยนรูปของหน่วยการวัดแบบโลจิตไปสู่คะแนนดิบ ซึ่งจะมีประโยชน์ในกรณีที่ต้องการเปรียบเทียบองค์ประกอบในฟาเซทที่อาจไม่ได้มีการเก็บข้อมูลจริงได้ครบถ้วนทุกส่วน เช่น ผู้ประเมินทุกคนไม่ได้ประเมินงาน หรือผู้ถูกประเมินทุกคน (fully crossed design) แต่ต้องการเปรียบเทียบความ

คลาดเคลื่อนระหว่างผู้ประเมิน ซึ่งจะทำได้จากกระบวนการประมาณค่าและ การหาค่าเฉลี่ยที่เหมาะสมนี้ ซึ่งถ้าเหล่านี้จะปรากฏในตารางผลการวิเคราะห์ของโมดูล FACETS ซึ่งจะใช้ในการวิจัยครั้งนี้

4.3.3 สถิติทดสอบความสอดคล้องของโมเดล (Fit statistic) เป็นระดับความสอดคล้องของข้อมูลเชิงประจักษ์กับค่าที่คาดหวังจากการพยากรณ์การให้คะแนนของผู้ตรวจ จากโมเดลหลายองค์ประกอบของราส์ซ โดยนิยมใช้ค่าส่วนเหลือมาตรฐาน (Standardized residual) เมื่อนำมายกกำลังสอง จึงเรียกว่า mean-square fit statistics (Eckes, 2009) ค่าที่เป็นไปได้คือ $-\infty$ ถึง $+\infty$ ค่าที่เป็นลบแสดงถึงทิศทางของคะแนนที่มีความสม่ำเสมอของคะแนนจากการสังเกตได้กับคะแนนที่คาดหวังจากโมเดลมากเกินไป (overly consistent with the expected values) ค่าที่เป็นบวกแสดงถึงทิศทางของคะแนนที่มีความสม่ำเสมอของคะแนนจากการสังเกตได้ต่ำกว่าที่คาดหวัง (overly inconsistent with the expected values) และคะแนนที่เข้าใกล้ศูนย์แสดงว่าทิศทางของคะแนนมีปริมาณความแปรปรวนเหมาะสมใกล้เคียงกับค่าที่คาดหวัง (Wolfe & Dobria, 2008) ในการใช้ค่านี้ควรมีค่าไม่เกิน 1 หากเกินแสดงว่าการประเมินไม่ปฏิบัติตามที่คาดหวังที่ควรจะเป็น ค่ามาตรฐานส่วนที่เหลือที่ 1 แสดงว่าเป็น 1SD ห่างจากค่าที่คาดหวังที่ศูนย์ การแปลค่า แปลตามองค์ประกอบที่กำลังพิจารณา คือ 1) สถิติ INFIT เป็นค่าเฉลี่ยยกกำลังสองของส่วนเหลือมาตรฐานแบบถ่วงน้ำหนัก คือ การนำค่าเฉลี่ยขององค์ประกอบทุกองค์ประกอบที่ใช้ในโมเดลที่เราากำลังศึกษามาถ่วงน้ำหนักเฉลี่ยโดยผู้ถูกประเมินทุกคน สมรรถนะการประเมินทุกสมรรถนะที่ได้รับการประเมินด้วย ผู้ประเมินทุกคน คะแนนการประเมินที่ต่ำสุด และสูงสุดจะถูกถ่วงน้ำหนักให้มีน้ำหนักน้อยลงมา (Myford & Wolfe, 2003) และ 2) สถิติ OUTFIT เป็นดัชนีของค่าเฉลี่ยกำลังสองของส่วนเหลือมาตรฐานยกกำลังสอง ของผู้ถูกประเมินทุกคน (all ratee) สมรรถนะทุกสมรรถนะ (all trait) ที่ได้รับการประเมินด้วย ผู้ประเมินทุกคน (all rater) มาคำนวณโดยตรง ไม่ถ่วงน้ำหนักเฉลี่ยระหว่างค่าคะแนนการประเมินที่มีคะแนนสูงสุด และต่ำสุด เป็นค่าเฉลี่ยของคะแนนที่วัดได้จริง ที่แสดงความแตกต่างระหว่างคะแนนสังเกตได้กับค่าที่คาดหวัง หากเข้าใกล้ค่าที่หวังของดัชนี INFIT และ OUTFIT แสดงว่าค่าสังเกตได้กับค่าที่คาดหวังใกล้เคียงกัน หากค่าเป็นลบ หรือต่ำกว่า 1 แสดงว่าค่าสังเกตได้มีค่าเข้าใกล้ค่าที่คาดหวังมากกว่าที่โมเดลทำนายไว้ (overfit) และถ้าค่าเป็นบวกและมากกว่า 1 แสดงว่าค่าที่สังเกตได้อยู่ห่างจากค่าที่คาดหวังมากกว่าที่โมเดลทำนาย (misfit) แต่ Wright และ Linacre (Wright & Linacre, 1994) แนะนำให้ใช้ช่วงของดัชนีที่ 0.6-1.4 แต่ทั้งนี้ต้องระวังการแปลผลในข้อมูลที่มีการกระจายของค่าเฉลี่ยกำลังสองมากผิดปกติด้วย ควรนำระดับการกดคะแนน และระดับผลการปฏิบัติงานของ

ผู้ถูกประเมินมาพิจารณาพร้อมด้วย สถานการณ์ที่ดีในการใช้หลักเกณฑ์นี้ในการแปลผลควรใช้ในสถานการณ์ที่ ช่วงของระดับการกด ปล่อยคะแนนควรเป็นกึ่งหนึ่งของระดับความสามารถในการปฏิบัติงาน หรือผลงานของผู้ถูกประเมิน การกำหนดช่วงของการแปลผลอาจพิจารณาตามระดับความสำคัญของผลการประเมิน หากเป็น ผลการประเมินที่มีผลกระทบสูง เช่น ผลการประเมินที่ใช้ในระดับชาติ ควรกำหนดช่วงของค่าดัชนีที่ 0.8 ถึง 1.2 หากระดับการประเมินไม่มีผลกระทบในวงกว้างอาจยืดหยุ่นให้ช่วงของค่าดัชนีที่ 0.5 ถึง 2.0 ในการวิจัยครั้งนี้ใช้เกณฑ์การแปลผลของค่าที่คาดหวังของ Fit mean-square ที่ 0.6 - 1.4 ซึ่ง ไม่เข้มงวดเกินไป หรือ ยืดหยุ่นเกินไป

4.3.4 ค่าพารามิเตอร์ที่ใช้ในการประมาณค่าปฏิสัมพันธ์ความคลาดเคลื่อน (Interaction bias) ระหว่าง 2 ฟาเซท คือ บุคคลกับฟาเซทที่ประกอบอยู่ในโมเดลที่ต้องการศึกษา ซึ่งมีที่มาจากค่าเบี่ยงเบนของ องค์ประกอบของคู่ฟาเซทที่ต้องการศึกษากับ ที่เบี่ยงเบนไปจากค่าเฉลี่ยจากการประมาณค่าพารามิเตอร์ โมเดลที่ใช้คือ

$$\log\left(\frac{P_{nk}}{P_{nk-1}}\right) = B_n - D_i - C_j - F_k - C_{ij} \quad (8)$$

เมื่อให้ P_{nk} แทนความน่าจะเป็นของผู้ถูกประเมิน n จะถูกให้คะแนนการประเมิน k P_{nk-1} แทนความน่าจะเป็นของผู้ถูกประเมิน n จะถูกให้คะแนนในหัวข้อที่ i โดยผู้ประเมินคนที่ j ให้คะแนนการประเมิน $k-1$

B_n แทน ค่าความสามารถของผู้ถูกประเมิน n

D_i แทน ค่าความยากของกระทางการประเมินที่ i

C_j แทน การกด ปล่อยคะแนนของผู้ประเมินคนที่ j

F_k แทน ความยากของขั้นของกลุ่มคะแนน (category) k

C_{ij} แทน ผลต่างของสองฟาเซทที่ต้องการศึกษาปฏิสัมพันธ์ระหว่างฟาเซท ว่าเบี่ยงเบนไปจากค่าเฉลี่ยจากการประมาณค่าพารามิเตอร์ ที่อยู่ในรูปฟาเซทที่ต้องการศึกษาเช่น DIF (Differential item function) หรือ DRF (Differential rater function)

วัตถุประสงค์ในการศึกษาค่าปฏิสัมพันธ์ความคลาดเคลื่อนคือ (Linacre, 2014)

1. ใช้ในการวินิจฉัย misfit โดยพิจารณาจาก residual ที่แยกเป็นส่วนๆ ของคู่ขององค์ประกอบของแต่ละฟาเซท เช่น residual ระหว่างคู่ของผู้ประเมินกับกระทางในรายการประเมิน (raterX item) จากนั้นจึง

แปลงค่านี้ไปสู่โลจิทการวัด ประมาณค่าขนาดที่ไม่ได้คาดหวัง และความมีนัยสำคัญทางสถิติของภาวะ misfit ซึ่งเป็นความคลาดเคลื่อนอย่างเป็นระบบ

2. ใช้ในการศึกษาความตรงในการวัด เป็นการศึกษาความคลาดเคลื่อนที่มีขนาดเล็กแต่เป็นระบบ ของกระแทกที่ใช้ในการวัด หรือศึกษาผู้ประเมินคนหนึ่งต่อกลุ่มของผู้ประเมิน ซึ่งการศึกษาด้วย fit mean-square ทัวไปไม่สามารถศึกษาถึงความต้องการส่วนนี้ได้ เนื่องจากการศึกษาปฏิสัมพันธ์ เฉพาะพาเซทย่อยรายคู่ของแต่ละพาเซท ที่จะทำได้ข้อมูลถูกต้องเหมาะสม

3. ใช้ในการศึกษาอิทธิพลของความคลาดเคลื่อน (effect of bias) Myford และ Wolfe (2008) ได้กล่าวถึงการศึกษาไว้ 2 ขั้นตอนคือ ขั้นตอนที่ 1 ทำการประมาณค่าพารามิเตอร์ทุกองค์ประกอบในทุกพาเซท ยกเว้นพาเซทของปฏิสัมพันธ์รายคู่ที่ต้องการศึกษาขั้นตอนที่ 2 กำหนดให้พาเซทและองค์ประกอบของพาเซทในขั้นตอนที่ 1 เป็นค่าคงที่ และทำการประมาณค่าปฏิสัมพันธ์ของพาเซทรายคู่ที่สนใจศึกษา และค่า standard error เนื่องจากมีการวัดขนาดของความคลาดเคลื่อน และ standard error รวมถึงมีการทดสอบนัยสำคัญทางสถิติด้วยสถิติทดสอบ t ทำให้ความคลาดเคลื่อนสามารถวัดออกเป็นหน่วยการวัดเช่นเดียวกับองค์ประกอบอื่นๆของพาเซท คือ หน่วยเป็นโลจิท การทดสอบนี้มีเงื่อนไขว่า สถิติทดสอบ t ที่ mean = 0 และ S.D = 1 หากนำความคลาดเคลื่อนออกแล้ว ทดสอบความมีนัยสำคัญของขนาดของความคลาดเคลื่อนว่าไม่มีความคลาดเคลื่อนใด นอกเหนือจากความคลาดเคลื่อนจากการวัด แสดงให้เห็นว่ามี misfit เกิดขึ้นเล็กน้อยเพียงใด ภายหลังจากขจัด ความคลาดเคลื่อนออก หากพาเซทหลักที่สนใจศึกษา เช่น ผู้ประเมินกับ ผู้ถูกประเมิน เมื่อทดสอบด้วยสถิติทดสอบ t พบว่ามีนัยสำคัญทางสถิติแปลผลการวิเคราะห์ได้ว่า ผู้ประเมินคนใดคนหนึ่งมีความคลาดเคลื่อนจากการวัดกับกลุ่มผู้ถูกประเมิน หรือ ผู้ประเมินคนนั้น มีการกด ปล่อย คะแนนแตกต่างกันในกลุ่มผู้ถูกประเมิน

สำหรับการวิจัยครั้งนี้เป็นการศึกษาการทำหน้าที่ต่างกันของผู้ประเมินในช่วงเวลาต่างกัน 3 ช่วงเวลา (differential rater function over time) คือ ก่อนการฝึกอบรม หลังการฝึกอบรมทันที และ 3 เดือนหลังการฝึกอบรม

4. ใช้ในการศึกษาความแปรปรวนเป็นส่วนๆ ของความคลาดเคลื่อนที่ไม่สามารถอธิบายได้ ขนาดความคลาดเคลื่อนในรูปหน่วยการวัดเป็นโลจิท สามารถนำมาประมาณค่าปริมาณของความคลาดเคลื่อนแบบระบบในรูปของความแปรปรวนของความคลาดเคลื่อน (RMSE: Root mean-square error)

4.4 การวิเคราะห์คุณภาพการประเมินด้วยโมเดลหลายองค์ประกอบของราส์ซ

โมเดลหลายองค์ประกอบของราส์ซสามารถนำมาใช้ในการอธิบายคุณภาพของการวัดและประเมิน ทั้งในระดับกลุ่ม (group level) ไปจนถึงระดับบุคคล (individual level)

4.4.1 ประโยชน์ของการนำไปใช้ในระดับกลุ่ม

ประโยชน์ในการอธิบายผลการวิเคราะห์ในระดับกลุ่มซึ่งจะทำให้เห็นคุณภาพการประเมินที่เป็นระบบ เช่น

4.4.1.1 ผู้ประเมินในกลุ่มผู้ประเมินทำการวัด และประเมินในลักษณะกคคะแนนแตกต่างกันอย่างไร

4.4.1.2 ผู้ประเมินสามารถแยกระดับความสามารถของผู้ถูกประเมินตามสมรรถนะ (Trait) ที่ใช้ประเมินได้อย่างมีประสิทธิภาพหรือไม่

4.4.1.3 ผู้ถูกประเมินแสดงให้เห็นได้ถึงความแตกต่างอย่างมีนัยสำคัญทางสถิติในแต่ละระดับของสมรรถนะที่ใช้วัดหรือไม่

4.4.1.4 ผู้ประเมินมีประสิทธิภาพเพียงใดที่จะแยกสมรรถนะ (trait) ของแต่ละสมรรถนะได้ หรือ สมรรถนะที่ใช้ประเมินสามารถสะท้อนให้เห็นถึงความแตกต่างของระดับความยากอย่างมีนัยสำคัญหรือไม่

4.4.2 สถิติที่ใช้ในการทดสอบในการวิเคราะห์ระดับกลุ่ม (Myford & Wolfe, 2003) เช่น Separation statistic, Fixed effects Chi-square tests, Summary fit statistics

4.4.2.1 สถิติทดสอบแบบ fit (Fit mean square) เป็นระดับของความสอดคล้องกันระหว่างคะแนนที่คาดคะเนจากการประมาณค่าในโมเดล กับค่าที่สังเกตได้จากฟาเซทที่ต้องการศึกษา เช่น ผู้ประเมิน ผู้ถูกประเมิน หรือสมรรถนะที่ต้องการประเมิน ตัวอย่างเช่น สถิติทดสอบที่ใช้กับฟาเซทผู้ประเมิน (rater) ทำการประเมินผู้ถูกประเมินบนสมรรถนะนั้น จากสมการที่ 9

$$VE_{nij} = \sum_{k=0}^m (k - E_{nij})^2 P_{nij} \quad (9)$$

เมื่อ VE_{nij} แทนความแปรปรวนของการสังเกตที่อยู่ในรูป square root การนำไปใช้ในระดับบุคคล

K แทน จำนวนครั้งของดัชนีที่แทนกลุ่มของมาตรการประเมินแต่ละครั้งของการประเมิน

m แทน จำนวนกลุ่มของมาตรการประเมิน

$$E_{nij} = \sum_{k=0}^m KP_{nijk}$$

เมื่อ E_{nij} แทน การประเมินที่คาดคะเนด้วยการประมาณค่าพารามิเตอร์ของราล์ซ

หากพบค่า Standardized absolute residual มากกว่า 2 แสดงถึงความแตกต่างระหว่างคะแนนที่สังเกตได้กับคะแนนที่ได้จากการประมาณค่าด้วยโมเดล และแสดงว่าเป็น 2 SD ที่เบี่ยงเบนไปจากค่าที่คาดหวังที่ค่า 0 จะสะท้อนถึงความคลาดเคลื่อนของการวัด การยกกำลังสองของค่า Standardized absolute residual จะเป็นการเฉลี่ยโดยรวมของค่าพารามิเตอร์ของผู้ถูกประเมิน สมรรถนะที่ต้องการประเมินเพื่อให้ได้ค่าทางสถิติสำหรับผู้ประเมินแต่ละคน การหาค่าเฉลี่ยนี้ทำได้สองวิธีคือ แบบไม่ถ่วงน้ำหนัก (Unweight fit mean-square) และแบบถ่วงน้ำหนัก (Weight fit mean-square)

1) Unweight fit mean-square หรือ Outfit mean-square (rater outfit) เป็นวิธีหาค่าเฉลี่ยแบบง่ายของค่ายกกำลังสองของ standardized residuals ของผู้ประเมิน (rater) ต่อผู้ถูกประเมิน (ratee) และสมรรถนะที่ประเมิน (trait) ค่านี้มีความไวสูง ต่อโอกาสการประเมินที่ไม่ตรงกับคะแนนที่ประมาณการจากโมเดล เนื่องมาจากผู้ประเมินคนใดคนหนึ่งที่มีความไม่คงที่ในการประเมิน หรือ มีความไวสูงต่อค่าคะแนนที่ออกนอกกลุ่มของคะแนนที่คาดหวัง (outfit is short for outlier-sensitive fit statistic)

$$MS_U = \frac{\sum_{n=1}^N \sum_{i=1}^I Z_{Rnij}^2}{NI} \quad (10)$$

เมื่อ

n แทนจำนวนของผู้ถูกประเมิน และ

i แทนจำนวนของสมรรถนะที่ประเมิน

2) Weight fit mean-square หรือ Infit mean-square (rater infit) เป็นการถ่วงน้ำหนักเฉลี่ย standardized residual ยกกำลังสอง ของผู้ถูกประเมินและ สมรรถนะทั้งหมดโดยผู้ประเมิน ดังนั้นการประเมินของผู้ประเมินที่ให้คะแนนบนมาตรฐานการวัดที่สูงสุด และต่ำสุดจะถูกนำมาถ่วงน้ำหนักให้มีน้ำหนักน้อยลง หรือเป็นการประมาณค่าความคงเส้นคงวาในการให้คะแนนของผู้ประเมิน ทำให้มีความไวต่ำต่อการให้คะแนนของผู้ประเมินคนใดคนหนึ่งที่ให้คะแนนไม่คงที่ซึ่งตรงข้ามกับ Outfit mean-square แต่ทำให้มีความไวต่อผลรวมของการประเมินที่ไม่ตรงกับค่าประมาณค่าของโมเดล จึงทำให้ Infit มีความสำคัญมากกว่า Outfit (Eckes, 2009) มีสมการดังนี้

$$MS_W = \frac{\sum_{n=1}^N \sum_{i=1}^I Z_{Rnij}^2 V_{Enij}}{\sum_{n=1}^N \sum_{i=1}^I V_{Enij}} \quad (11)$$

ทั้ง infit และ outfit mean -square มีค่าตั้งแต่ 0 ถึง infinity แต่ค่าที่คาดหวังคือ 1 ดังนั้นค่าที่เข้าใกล้ 1 แสดงว่าค่าที่สังเกตได้ใกล้เคียงกับค่าที่ได้จากการประมาณค่าด้วยโมเดลของราส์ซ หากค่าน้อยกว่า 1 แสดงว่าคะแนนที่สังเกตได้เข้าใกล้ค่าที่คาดหวังมากกว่าที่โมเดลของราส์ซพยากรณ์ ถ้าค่ามากกว่า 1 แสดงว่าคะแนนที่สังเกตได้ห่างจากค่าที่คาดหวังมากกว่าที่โมเดลของราส์ซพยากรณ์ เนื่องจาก infit เป็นดัชนีที่มีที่มาจากกรถ่วงน้ำหนักค่าเฉลี่ยในองค์ประกอบทั้งหมดในฟาเซทที่ต้องการศึกษา และมีความไวต่อค่าที่อยู่นอกเหนือจากค่าที่คาดหวังโดยโมเดล โดยทั่วไปจึงให้ความสำคัญกับดัชนี Infit มากกว่า Outfit (Linacre, 2014; Myford & Wolfe, 2003) สำหรับการแปลผลการวิเคราะห์ Wright และ Linacre (1994) แนะนำให้ใช้ช่วงของค่าของดัชนีอยู่ที่ 0.6 ถึง 1.4 ทั้งนี้ขึ้นอยู่กับสถานการณ์แต่ละสถานการณ์ การแปลความหมายของค่าดัชนีนี้ว่าดีหรือไม่ การแปลค่าจะต้องคำนึงถึงระดับการกตคะแนนในการให้คะแนนของผู้ประเมินร่วมกับระดับผลงานของผู้ถูกประเมิน ดังนั้นในการแปลค่าดัชนีนี้จึงไม่สามารถสรุปด้วยค่า ค่าเดียวแต่ต้องพิจารณาปัจจัยอื่น ๆ ที่กล่าวไว้ด้วย โดยระดับของการกต ปล่อยคะแนนควรประมาณครึ่งหนึ่งของคุณภาพงานของผู้ถูกประเมิน (ratee performance) หากเป็นการวัดในระดับมหภาค หรือมีผลกระทบในระดับสูง เช่นการวัดระดับชาติ ควรมีการกำหนดช่วงของค่าดัชนี ระหว่าง 0.8 ถึง 1.2 หากเป็นการวัดที่ไม่มีความสำคัญมาก หรือไม่ใช้ระดับมหภาค อาจกำหนดช่วงดัชนีระหว่าง 0.5 ถึง 2 โดยทั่วไป ค่า mean-square ที่มากกว่า 1 จะถือว่าปัญหาในการวัดมากกว่าค่านี้น้อยกว่า 1 สำหรับการวิจัยครั้งนี้ใช้ช่วงของค่าดัชนีนี้ระหว่าง 0.6 ถึง 1.4

4.4.2.2 สถิติทดสอบ Fixed effect chi-square เป็นผลการวิเคราะห์จากสถิติ Chi-square เพื่อทดสอบสมมติฐานของอิทธิพลที่ถูกกำหนดให้คงที่ (fixed effect) ของอิทธิพลของฟาเซททั้งหมดที่ร่วมกันกำหนดค่าพารามิเตอร์ และค่าความคลาดเคลื่อนจากการวัด แยกการวิเคราะห์ตามฟาเซท โดยค่าดัชนีนี้มีความไวต่อขนาดของตัวอย่างมากนั่นคือหากตัวอย่างมากจะพบว่าสถิติทดสอบมีนัยสำคัญทางสถิติแม้ความแปรปรวนระหว่างองค์ประกอบในฟาเซทที่ศึกษาอยู่มีค่าน้อยก็ตาม ตัวอย่างการใช้สถิติทดสอบนี้ เช่น การทดสอบสำหรับฟาเซทของผู้ประเมิน (rater) ใช้ทดสอบสมมติฐานที่ว่าผู้ประเมินทุกคน (all rater) ทำการประเมินผู้ถูกประเมินในระดับคุณภาพของการกตคะแนนระดับเดียวกัน หลังการวิเคราะห์ค่าความคลาดเคลื่อนจากการวัดแล้ว (the same level of severity)

การทดสอบสมมติฐานของฟาเซทของผู้ถูกประเมิน ที่ว่าผู้ถูกประเมินทุกคน (all ratee) มีระดับคุณภาพงานที่ใช้ประเมินในระดับเดียวกัน ดังสมการที่ 12 (Eckes, 2009)

$$\chi^2 = \frac{\sum (W_i \times D_i^2) - (\sum W_i \times D_i)^2 / \sum W_i}{J-1} \quad (12)$$

degree of freedom = L - 1

D แทน ความยากของสมรรถนะ

L แทน จำนวนของสมรรถนะที่ต้องการวัด และ

$$W_i = \frac{1}{SE_i^2} \quad \text{เมื่อ } i = 1, L$$

เมื่อต้องการทดสอบฟาเซทของผู้ประเมิน แทนที่ D_i ด้วย C_i และแทนที่ degree of freedom ด้วย $J-1$ หรือเมื่อต้องการทดสอบฟาเซทของผู้ถูกประเมิน แทนที่ D_i ด้วย B_n และแทนที่ degree of freedom ด้วย $N-1$ การนำเสนอผลการวิเคราะห์ควรนำเสนอค่า ไค สแควร์ และ degree of freedom หากค่า ไค สแควร์ มีนัยสำคัญทางสถิติแสดงว่า มีองค์ประกอบอย่างน้อย 2 องค์ประกอบในฟาเซทมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ

4.4.2.3 ความเชื่อมั่นแยกส่วน (The separation of reliability index) เป็นดัชนีที่ใช้วัดตัวแปรแฝง ให้ค่าการวัดในแต่ละองค์ประกอบว่ามีการกระจาย หรือแยกกันตลอดช่วงบนมาตรวัดหรือไม่ สถิติ Separation (G) เป็นสัดส่วนของส่วนเบี่ยงเบนมาตรฐานของคะแนนจริง (true standard deviation) ของฟาเซทที่ต้องการศึกษาต่อค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของฟาเซทนั้นจากสมการ

$$G = \frac{\text{True SD}}{\text{RMSE}}$$

เมื่อ True SD เป็นส่วนเบี่ยงเบนมาตรฐานของฟาเซทที่ต้องการศึกษา

RMSE เป็น root mean-square of standard error ของฟาเซทที่ต้องการศึกษา

ตัวอย่างของการใช้ ความเชื่อมั่นแยกส่วนในการทดสอบสมรรถนะของงาน (trait) ที่ว่าทุกสมรรถนะของงานมีระดับความยากอยู่ในระดับเดียวกัน (the same level of performance) หลังการวิเคราะห์ค่าความคลาดเคลื่อนจากการวัดแล้ว โดยมีการแยกการแปลค่าความเชื่อมั่นตามฟาเซทที่ต้องการศึกษาดังนี้ 1) ความเชื่อมั่นแยกส่วนของผู้สอบ (ratee separation of reliability) ฟาเซทที่ต้องการศึกษาคือ ผู้สอบ ความเชื่อมั่นแยกส่วนของผู้สอบจึงเป็นอัตราส่วนความแปรปรวนของคะแนนที่สังเกตได้กับคะแนนจริงในการประมาณค่าความสามารถของผู้สอบโดยแปลผลเป็นร้อยละของความแปรปรวนถ้ามีค่ามากแสดงว่ามีความน่าเชื่อถือได้ของคะแนนของผู้สอบได้มาจากข้อสอบ

และผู้ตรวจ 2) ความเชื่อมั่นแยกส่วนของผู้ประเมิน (rater separation of reliability) แสดงถึงความแปรปรวนของการให้คะแนนที่ไม่ต้องการให้เกิดระหว่างผู้ตรวจที่มีการกตคะแนนเป็นดัชนีแสดงความแตกต่างในเรื่องการกตคะแนนของผู้ประเมิน ถ้าค่าสูงแสดงว่ามีความแตกต่างกันมากในการประเมินคะแนนของผู้ประเมิน 3) ความเชื่อมั่นแยกส่วนของสมรรถนะ หรือข้อสอบ (trait separation of reliability) เป็นดัชนีที่แสดงถึงความสามารถในการจำแนกความยากของข้อสอบว่าต่างกันหรือไม่ ถ้ามีค่าสูงแสดงสมรรถนะของข้อสอบมีความแตกต่างกัน สรุปการแปลผลได้ว่า แปลผลต่างกันตามลักษณะองค์ประกอบที่พิจารณา เช่นองค์ประกอบของผู้ประเมิน (rater) ค่าที่เข้าใกล้ศูนย์แสดงว่าผู้ประเมินมีระดับการกต ปล่อยคะแนนใกล้เคียงกัน สำหรับองค์ประกอบของผู้ถูกประเมิน (ratee) ค่าที่เข้าใกล้ศูนย์ แสดงว่ากลุ่มผู้ถูกประเมินถูกประเมินได้รับการประเมินคะแนนใกล้เคียงกันทำให้เกิดความคลาดเคลื่อนแบบเกาะกลุ่มตรงกลางได้ (central tendency effect)

4.4.3 การใช้งานในระดับบุคคล (individual level) สามารถอธิบายปรากฏการณ์ที่พบโดยอาศัยข้อมูลจากโลจิทที่วัดได้ว่ามีระดับการกต ปล่อยคะแนนคือ fit mean-square, standard error หรือความถี่ของเหตุการณ์ว่ามีมากน้อยเพียงใดเมื่อเทียบเป็นรายบุคคล หากพิจารณาในด้านของปฏิสัมพันธ์ของความคลาดเคลื่อน (interaction bias) โลจิทที่ประมาณได้คือ ขนาดของความคลาดเคลื่อน (bias size) ค่า standard error และค่า Z - score จะสะท้อนว่าระดับของการประเมินเฉพาะคู่ที่เราสนใจศึกษา มีค่าเบี่ยงเบนไปจากค่าที่คาดหวังด้วยโมเดล หรือทำให้เห็นปัญหาที่ชัดเจนขึ้น เช่นเมื่อทราบว่ามีการกตคะแนนซึ่งเกิดขึ้นในการประเมิน เป็นการประเมินระดับกลุ่ม ควรมีการระบุที่เกิดจากผู้ประเมินรายใดที่มีรูปแบบการประเมินแบบกตคะแนน หรือปล่อยคะแนน เป็นการประเมินคุณภาพการประเมินระดับบุคคล ซึ่งพิจารณาจากฟิเชทของผู้ประเมิน (rater) ผู้ถูกประเมิน(ratee) สมรรถนะที่จะประเมิน (trait) สามารถอธิบายในระดับบุคคลได้ดังนี้

4.4.3.1 หากพบว่าระดับการกต ปล่อยคะแนนเกิดขึ้นในกลุ่มผู้ประเมิน MFRMอธิบายได้ว่าเกิดการกต หรือปล่อยคะแนนกับผู้ประเมินคนใด

4.4.3.2 สามารถค้นพบความคลาดเคลื่อนจากผู้ประเมินที่เกิดจากการที่ผู้ประเมินคนใดคนหนึ่งมีรูปแบบการให้คะแนนการประเมินผู้ถูกประเมินคนหนึ่ง แตกต่างจากผู้ประเมินคนอื่นๆ อย่างเป็นระบบ

4.4.3.3 สามารถค้นหาผู้ประเมินที่ไม่สามารถแยกแยะความสามารถของผู้ถูกประเมินได้ หรือให้คะแนนผู้ถูกประเมินในระดับคะแนนใกล้เคียงกันหมดจนไม่สามารถแยกความแตกต่างของความสามารถของผู้ถูกประเมินได้

4.5 ข้อดีของโมเดลหลายองค์ประกอบของราส์ชในด้านการนำไปใช้ มีการนำMFRMมาใช้ในการศึกษาความคลาดเคลื่อนอย่างเป็นระบบในการประเมินการปฏิบัติในงานวิจัยหลายเรื่อง (Myford & Wolfe, 2000) ทำให้ทราบคุณภาพของผู้ถูกประเมินผู้ประเมิน และงานที่ใช้ในการประเมิน สอดคล้องกับค่าที่ได้ทำนายใน โมเดล มากน้อยเพียงใดซึ่งจะสะท้อนคุณภาพของการประเมิน และมีข้อดีเหนือกว่า CTT และ G-Theory สามารถสรุปได้ดังนี้ (Basturk, 2008; Linacre, 1994; Sudweeks et al., 2004)

4.5.1 การวัดโดยใช้โมเดล Rasch ได้นำความคลาดเคลื่อนแต่ละแหล่งมาใช้ในการวิเคราะห์ ทั้งในลักษณะของกลุ่ม ความยากรายข้อ และรูปแบบของการตัดสินใจการกด หรือปล่อยคะแนน ซึ่งจะทำให้ได้ข้อมูลของคะแนนที่เป็นมาตรวัดที่มีความสัมพันธ์เชิงเส้นตรงทำให้วิเคราะห์โดยสถิติทั่วไปได้ง่าย และทำให้การแปลความหมายชัดเจนในเชิงกลุ่มทั้งในรูปของการกดคะแนน หรือความยากง่ายของงานที่มอบหมาย

4.5.2 การตรวจสอบรายข้อ หรือการตัดสินใจบนพื้นฐานของโมเดล Rasch จะไม่มีอิทธิพลของความหลากหลายของตัวอย่างจากการวัด ดังนั้นการสรุปอ้างอิงที่ตรงตามจริงก็สามารถกระทำได้นอกเหนือจากตัวอย่างในขณะนั้น ทำให้เกิดผลดีคือทำให้สามารถตัดสินศักยภาพของการทำงานของกลุ่มได้ แม้ข้อมูลจะขาดหายไป ตัวอย่างเช่น กลุ่มอาจไม่ได้ทำงานการประเมินทั้งหมด หรือบางกลุ่มถูกประเมินโดยผู้ประเมินบางคน

4.5.3 วิธีการของ Rasch ที่จะทำให้โมเดลมีความสอดคล้อง (fit procedure) ถูกนำมาใช้เพื่อให้ได้รูปแบบการตอบสนองที่ไม่ได้คาดหมายมาก่อน เพื่อให้เป็นประโยชน์ต่อการขยายขอบเขตของสิ่งที่ได้จากการประเมินในระดับกลุ่มแต่ละกลุ่ม งานแต่ละงานหรือ ผู้ประเมินแต่ละคน ซึ่งกระบวนการในลักษณะนี้จะไม่เป็นไปตามรูปแบบของโมเดลอย่างตายตัว โดยMFRM จะทำให้ได้มาตรการวัดที่มีระดับอันตรายภาคขึ้น ที่เป็นคะแนนเดียวขึ้นไป และเรียกคะแนนนี้จะถูกเรียกว่าเป็น โลจิต ซึ่งทำให้สามารถคำนวณค่าทางสถิติได้โดยการบวก ลบ คูณ หารได้

4.5.4 เนื่องจาก G-theory ใช้คะแนนดิบในการวิเคราะห์ความเที่ยง โดยไม่ผ่านการปรับคะแนนให้เป็นมาตรฐานก่อนการวิเคราะห์ และการประมาณค่าส่วนประกอบของความแปรปรวนจะเป็นการประมาณค่าการแจกแจงของผู้ถูกประเมิน ซึ่งจะเป็นภาพกว้างของทั้งกลุ่มผู้ถูกประเมิน

ความคลาดเคลื่อนจากการวัดเป็นการพิจารณาจาก interaction effect และความคลาดเคลื่อนอย่างสุ่มหรืออาจสรุปได้ว่าส่วนประกอบของความแปรปรวนประกอบด้วยสาม แหล่งรวมกันคือ main effect, interaction effect และ random error แต่ Rasch Model มีการนำคะแนนของผู้ถูกประเมินแต่ละคน มาพิจารณา และปรับในแต่ละงานที่มอบหมาย หรือกระทั่งข้อสอบรายข้อ และผู้ประเมินรายคน ก่อนทำการวิเคราะห์จึงทำให้ได้สารสนเทศที่ลงรายละเอียดในรายบุคคลมากกว่าเป็นภาพรวม เช่นใน G-theory จากเหตุผลดังกล่าวจึงทำให้ ความเที่ยงใน Rasch Model จะมีค่าสูงกว่า G-theory

4.5.5 การตัดสินใจให้ผู้ถูกประเมินผ่านหรือไม่ผ่านการประเมินนั้น ใน G- theory จะยึดหลักของคะแนนดิบเพียงอย่างเดียว โดยจะให้สารสนเทศเพียงแค่ว่าระดับที่ผ่านในภาพรวมของผู้ถูกประเมินทั้งหมดในทุกคะแนนดิบที่เป็นไปได้ที่จะให้ผ่าน หรือตก มีประโยชน์ในภาพกว้าง เช่น จะเลือกงานประเมินกี่ชิ้นงาน จำนวนผู้ประเมินกี่คน ต่อผู้ถูกประเมินกี่คน รูปแบบการประเมินควรเป็นอย่างไรจึงจะให้ผลการประเมินที่มีคุณภาพสูงสุด เป็นต้น ในขณะที่ โมเดล Rasch สามารถลงรายละเอียดในผู้ประเมินแต่ละคน ผู้ถูกประเมินแต่ละคน และงานที่มอบหมายในแต่ละงานทำให้ได้สารสนเทศมากกว่า และสอดคล้องกับความเป็นจริงในการตัดสินใจให้ผ่านหรือตก

4.6 การนำโมเดลหลายองค์ประกอบของราส์ชไปใช้ในการประเมินภาคปฏิบัติ

4.6.1 ในต่างประเทศ จากการทบทวนตำรา เอกสารทางวิชาการ และงานวิจัยต่าง ๆ พบว่ามี การศึกษาถึงประโยชน์ของโมเดล นำไปใช้ในการตรวจสอบคุณภาพการประเมินภาคปฏิบัติ ที่ชัดเจนในหลาย สาขาวิชาทางการศึกษา เช่น ด้านภาษาศาสตร์ และด้านวิทยาศาสตร์สุขภาพ หรือด้านการฝึกอบรม สิ่งที่ทำให้โมเดลหลายองค์ประกอบของราส์ชที่มีความซับซ้อนของข้อมูลการคำนวณซ้ำๆ และการเชื่อมโยงกันหลายระดับของข้อมูลมีความเป็นไปได้ในทางปฏิบัติคือ การมีโมเดลสำเร็จรูปเพื่อใช้ในการวิเคราะห์ โมเดลที่มีใช้ในขณะนี้เช่น FACETS, Conquest, T-Rasch, Bigsteps (MS-DOS Windows) การวิจัยส่วนใหญ่ใช้ FACETS ในการวิเคราะห์ ดังตัวอย่างต่อไปนี้

4.6.1.1 ด้านการศึกษา จากการศึกษาของ Basturk (2008) ใช้ MFRM ในการประเมินการ นำเสนอผลงานด้วย PowerPoint ในระดับอุดมศึกษา มีวัตถุประสงค์เพื่อศึกษา ความสามารถของกลุ่มในการ นำเสนอ, ความยากของงาน และการกด ปล่อยคะแนนของผู้ประเมิน ในการประเมินผลการนำเสนอผลงาน ในรายวิชาหนึ่ง ของนักศึกษา 30 คนที่ได้รับการสุ่มจากนักศึกษาจำนวน 5 ชั้นเรียน ให้นำเสนองานหน้าชั้น โดยมีผู้ให้คะแนนการนำเสนอ 7 คน ใช้รูบริก เป็นแบบประเมิน ผลการศึกษาพบว่าผู้ประเมินส่วนใหญ่ให้ คะแนนสอดคล้องกับการคาดการณ์โมเดล ไม่มีผู้ประเมินคนใดให้คะแนนไม่สอดคล้องกับการคาดการณ์

ของโมเดล Separation index สามารถแยกแยะความสามารถในการนำเสนอได้อย่างสมเหตุสมผล และผู้ประเมินทุกคนให้คะแนนในลักษณะกตคะแนนอย่างสม่ำเสมอ สถิติ Infit และ Outfit สามารถบอกระดับความยากของแต่ละหัวข้อการประเมินได้

Güler (2014) ทำการวิเคราะห์แบบทดสอบแบบปลายเปิดมีเนื้อหาเกี่ยวกับสถิติพื้นฐาน จำนวน 10 ข้อ ในนักศึกษาชั้นปีที่ 3 จำนวน 55 คน ตรวจสอบแบบทดสอบผู้ประเมิน 3 คนที่มีความเชี่ยวชาญด้านสถิติ มีองค์ประกอบที่สนใจศึกษา 3 องค์ประกอบคือ ผู้สอบ ผู้ตรวจ และกระทงคำถาม โดยการวิเคราะห์ด้วยโมเดลหลายองค์ประกอบของราส์ซใช้สถิติทดสอบ Infit, Outfit mean square, Separation index และ Reliability coefficient ด้วยโปรแกรม FACETS ผลการศึกษาพบว่าผู้สอบ และกระทงคำถาม มีความเที่ยงอยู่ในระดับสูง พบความเที่ยงระหว่างผู้ประเมินในระดับสูงเช่นกัน

4.6.1.2 ด้านภาษาศาสตร์ เช่น Schaefer (2008) ศึกษาความคลาดเคลื่อนจาก ผู้ประเมิน คือ ครูช่วยสอนภาษาอังกฤษที่ใช้ภาษาอังกฤษเป็นภาษาแรก และอยู่ในประเทศญี่ปุ่น จำนวน 40 คน ในการตรวจงานเขียนบทความภาษาอังกฤษใน หัวข้อการเขียนเดียวกัน ของผู้สอบที่ใช้ภาษาอังกฤษเป็นภาษาที่สอง จำนวน 40 ฉบับ ใช้รูปบริบทที่มีมาตรการประเมิน 6 ระดับ 6 ด้านวิเคราะห์คุณภาพ ด้วยโมเดลหลายองค์ประกอบของราส์ซ และโมดูล FACETS มีการฝึกอบรมผู้ประเมินประมาณ 30 นาที ในหัวข้อ frame of reference ผลการศึกษาพบว่า ผู้ประเมินส่วนใหญ่มีการปล่อยคะแนน บางหัวข้อ มีการกตคะแนนบางหัวข้อ การตรวจสอบปฏิสัมพันธ์ระหว่างผู้ประเมิน และผู้ถูกประเมินพบว่า ผู้ประเมินมีแนวโน้มที่จะกต หรือปล่อยคะแนน ในกลุ่มผู้สอบที่มีความสามารถสูง มากกว่าผู้สอบที่มีความสามารถต่ำ สรุปว่า MFRM สามารถวิเคราะห์องค์ประกอบต่าง ๆ ที่ต้องการศึกษาได้เป็นอย่างดี

4.6.1.3 ด้านวิทยาศาสตร์สุขภาพ มีวัตถุประสงค์ในการวิจัยสองประเด็นหลัก คือ เพื่อพัฒนาคุณภาพการเรียนการสอนภาคปฏิบัติ และเพื่อตรวจสอบคุณภาพการปฏิบัติงานทางคลินิก เช่น งานวิจัยที่มีวัตถุประสงค์เพื่อพัฒนาคุณภาพการเรียนการสอนภาคปฏิบัติ เช่น Iramaneerat และ Myford (Iramaneerat & Myford, 2006) ศึกษาความคลาดเคลื่อนจากผู้ประเมินในการประเมินการปฏิบัติงานของแพทย์ประจำบ้านแผนกศัลยกรรม ในแผนกศัลยกรรม 13 แผนก ที่ฝึกปฏิบัติงานเป็นเวลา 4 สัปดาห์ โดยผู้ประเมินจำนวน 13 คน เป็นแพทย์ประจำแผนก ที่แพทย์ประจำบ้านไปฝึกปฏิบัติงาน ทำการประเมินแพทย์ประจำบ้านต่างคนกัน ในสัปดาห์สุดท้ายของการฝึก จากนั้นทำการศึกษาความคลาดเคลื่อนจากผู้ประเมิน 3 ประเภทคือ ความคลาดเคลื่อนแบบกต ปล่อยคะแนน ความคง

เส้นคงวาของการประเมิน และ ความคลาดเคลื่อนจากการให้คะแนนในช่วงแคบ ๆ โดยใช้ MFRM ในการวิเคราะห์ ผลการศึกษาพบว่า MFRM ช่วยให้เกิดความชัดเจนในการศึกษาความคลาดเคลื่อนแบบ กัด ปล่อยคะแนน ความคงเส้นคงวาของการประเมิน และ ความคลาดเคลื่อนจากการให้คะแนน ในช่วงแคบ ๆ ได้

การศึกษาของ Iramaneerat, Yudkowsky, Myford และ Downing (2008) ทำการกำกับติดตามคุณภาพของการทดสอบทางคลินิก OSCE (Objective structured clinical examination) ศึกษาความคลาดเคลื่อนจากผู้ประเมินโดยใช้ G-theory และ MFRM ในการประเมิน ทักษะด้านการติดต่อสื่อสารที่เป็นส่วนหนึ่งในการทดสอบ OSCE ศึกษาในแพทย์ประจำบ้านจำนวน 79 คน ของมหาวิทยาลัย Midwestern ในอเมริกา ทักษะการสื่อสารมี 6 ด้าน จำนวน 18 ข้อ และแบบ ประเมินแบบมาตรฐานค่าระดับ 5 คะแนนผู้ประเมินเป็นคนไข้เสมือนจริง (Standardized Patients: SPs) จำนวน 6 สถานการณ์จำลอง ใช้ MFRM ในการศึกษา 3 พาเซท คือ ความสามารถ ของแพทย์ประจำบ้าน, ความเข้มงวดในการประเมินของคนไข้เสมือนจริง และระดับความยากของ งาน วิเคราะห์ค่า Chi-square เพื่อพิสูจน์ว่ามีความแตกต่างระหว่างระดับของการกัด ปล่อยคะแนน ระหว่างผู้ประเมินหรือไม่ ผลการศึกษาพบว่า ผู้ประเมินคนที่ 1, 2, 4 และ 6 มีความสม่ำเสมอของการ ให้คะแนน ในขณะที่คนที่ 3 และคนที่ 5 ให้คะแนนไม่มีความสม่ำเสมอ และไม่พบการให้คะแนน ในช่วงแคบ ๆ แสดงว่าผู้ประเมินไม่จำเป็นต้องได้รับการฝึกอบรมเพิ่มเติมในเรื่องนี้ ซึ่งทำให้ทราบ ลักษณะการให้คะแนนของผู้ประเมินแต่ละคน และทราบแนวทางการปรับปรุงแก้ไข

Robert, Rothnie, zoanetti และ Crossley (2010) ทำการศึกษาสมรรถนะ ทางจิตวิทยาของแบบประเมิน “Multiple mini-interview” ซึ่งเป็นเครื่องมือใช้เพื่อการทดสอบ OSCE สำหรับนักศึกษาแพทย์ หรือนักศึกษาทันตแพทย์ทั้งในแคนาดา และออสเตรเลีย ประกอบด้วย ข้อคำถาม 39 ข้อ มี 4 ระดับคะแนนเป็นแบบมาตรฐานค่า ทำการสัมภาษณ์ตัวอย่างจำนวน 686 คน ใช้ผู้สัมภาษณ์ 207คน ผลการศึกษาพบว่า MFRM สามารถใช้ในการศึกษา การกัด ปล่อยคะแนนทั้ง ในระดับกลุ่ม และระดับบุคคลได้ สอดคล้องกับการศึกษาของ Dowell, Lynch, Till, Kumwenda และ Husbands (2012) และการศึกษาของ Sebok, Luu และ Klinger (2014) ที่ทำการตรวจสอบ สมรรถนะทางจิตวิทยาของเครื่องมือชนิดนี้ เพื่อนำมาใช้ในการเรียนการสอนของนักศึกษาแพทย์ โดย ใช้ MFRM ในการวิเคราะห์ และพบว่า MFRM มีประโยชน์ในการตรวจสอบคุณภาพของเครื่องมือที่ใช้ ในการประเมินภาคปฏิบัติได้จริง

ตัวอย่างการศึกษาเพื่อตรวจสอบคุณภาพการปฏิบัติงานทางคลินิก ตัวอย่าง เช่น การศึกษาของ Chapman, Henggeler, Halliday-Boykins และ Cunningham (2008) ศึกษาสมรรถนะทางจิตวิทยาในการประเมินสำหรับนักจิตวิทยาบำบัดในวัยรุ่นที่ใช้สารเสพติด คือ การจัดการ กับเหตุการณ์ที่อาจเกิดขึ้นเฉพาะหน้า เนื่องจากจะต้องรวบรวมข้อมูลหลายด้านเพื่อนำมาประเมินสภาพของวัยรุ่นด้วยแบบประเมินมาตรฐานที่เป็นมาตรฐานประมาณค่า 5 ระดับคะแนน ที่มีหัวข้อการประเมินการจัดการกับเหตุการณ์เฉพาะหน้ามี 2 ประเด็นคือ การกำกับติดตามที่ชัดเจน (distinct monitoring: MON) กับพฤติกรรมทางปัญญา (cognitive behavior: CB) โดยใช้สถิติ Fit mean-square ในการตรวจสอบความสอดคล้องระหว่างคะแนนที่สังเกตได้และ คะแนนที่คาดการณ์ด้วยโมเดล และ Separation reliability ผลการศึกษาจากค่าสถิติ Fit mean-square และ Separation reliability ทำให้เข้าใจถึงระดับความยากของแต่ละหัวข้อ และช่วงของการประเมิน

การศึกษาของ Forjaz และคณะ (2015) ศึกษาเพื่อตรวจสอบคุณภาพของเครื่องมือที่พัฒนาขึ้นมาใหม่เพื่อใช้ในการประเมินความเครียดจากโรคพากินสัน มีมิติการวัด 3 ด้าน คือ 1) ความเครียดแบบคงที่ 2) แบบเป็นบางครั้ง และ 3) แบบหลีกเลี่ยงความเครียด ศึกษาภาคตัดขวางในคนไข้พากินสัน จำนวน 362 คน จากศูนย์การรักษาพากินสันระดับชาติ หลายศูนย์โดยให้ทั้งผู้ป่วย และแพทย์ ต่างฝ่ายต่างทำการประเมินภาวะเครียดของคนไข้ และใช้ MFRM ในการประเมิน และกำจัดค่าความคลาดเคลื่อนจากผู้ประเมิน ผลการศึกษาพบว่า เครื่องมือในการประเมินความเครียดสำหรับผู้ป่วยพากินสันนี้มีความสอดคล้องกับที่โมเดลคาดหวัง

4.6.1.4 การศึกษาเกี่ยวกับผลของการฝึกอบรมต่อความคลาดเคลื่อนของการประเมินพบว่า ศึกษาหลายกลุ่มประชากร จากอดีตถึงปัจจุบันผลการศึกษาที่มีความชัดเจน จนสามารถกล่าวได้ว่าการฝึกอบรมช่วยลดความคลาดเคลื่อนจากผู้ประเมินได้ และโมเดลหลายองค์ประกอบของราล์ซมีบทบาทในการพิสูจน์สำคัญในการพิสูจน์ข้อเท็จจริงนี้ซึ่งจะได้กล่าวถึงในหัวข้อการพัฒนาโมดูลการฝึกอบรมในลำดับต่อไป

4.6.2 การนำ MFRM มาใช้ในประเทศไทย พบว่ามีการนำไปใช้ไม่มากนัก และวัตถุประสงค์ของการศึกษามีประเด็นการศึกษาแตกต่างกัน ตัวอย่างเช่น การศึกษาของน้ำผึ้ง อินทะเนตร, อดิชา นัยพัฒน์, ผจงจิต อินทสุวรรณ และ พีรศักดิ์โสภณ (2555) ศึกษาสมรรถนะของคะแนนแบบทดสอบปลายเปิดวิชาคณิตศาสตร์ โดยใช้โมเดลการสุรปู้อ้างอิง และโมเดลหลายองค์ประกอบของราล์ซ ที่มีจำนวนผู้ตรวจต่างกัน คือ 2 คน 3 คน และ 4 คน และมีรูปแบบการตรวจต่างกันคือ ผู้ตรวจทุกคน ตรวจข้อสอบทุกข้อของผู้สอบ

ทุกคน ผู้ตรวจทุกคนตรวจสอบข้อสอบทุกข้อของผู้สอบบางคน และผู้ตรวจตรวจสอบข้อสอบบางข้อของผู้สอบทุกคน ตัวอย่างคือ นักเรียนชั้นมัธยมศึกษาปีที่ 4 จำนวน 180 คน ผลการศึกษาพบว่า เมื่อใช้โมเดลหลายองค์ประกอบของราล์ชวิเคราะห์ความแปรปรวนขององค์ประกอบ พบว่าความแปรปรวนของผู้สอบมีค่าสูงที่สุด รองลงมาเป็นความแปรปรวนของข้อสอบ และความแปรปรวนของผู้ตรวจ ความเชื่อมั่นแยกส่วนของผู้สอบในรูปแบบการตรวจที่ 1 มีค่าสูงที่สุดในทุกเงื่อนไขของจำนวนผู้ตรวจ และมีค่าสูงขึ้นเมื่อจำนวนผู้ตรวจมากขึ้น คะแนนความสามารถในทุกเงื่อนไขการวัดที่แตกต่างกันมีความตรงตามสภาพสูง และแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

การศึกษาของบุชวอร์ธ แสปลีม (2556) ศึกษาความคลาดเคลื่อนจากผู้ตรวจคือ การกวด ปล่อยคะแนน ความลำเอียงด้านเพศ และการทำหน้าที่ต่างกันของผู้ตรวจเมื่อเวลาผ่านไป ในการให้คะแนนความสามารถในการเขียนเรียงความ ของนักเรียนชั้นประถมศึกษาปีที่ 3 จำนวน 50 คน ผู้ตรวจเป็นนักศึกษา ครุศาสตร์ที่เรียนวิชาเอกภาษาไทย จากมหาวิทยาลัยราชภัฏ 3 แห่ง จำนวน 130 คน รูปแบบของเกณฑ์การให้คะแนนที่แตกต่างกัน 3 แบบคือ 1) รูบริคให้คะแนนแบบแยกส่วนประกอบ 2) รวมองค์ประกอบ และ 3) แบบผสมผสาน ผลการวิจัยพบว่าการใช้โมเดลหลายองค์ประกอบของราล์ช ช่วยในการวิเคราะห์ได้ว่า รูบริคการให้คะแนนแบบใดได้ผลการตรวจ จากผู้ตรวจที่มีคลาดเคลื่อนแบบกวด ปล่อยคะแนนในระดับใด และสามารถเปรียบเทียบได้ระหว่างชนิดของรูบริค และช่วยในการวิเคราะห์ปฏิสัมพันธ์ระหว่างเพศกับความลำเอียงของผู้ตรวจได้เป็นอย่างดี

ตอนที่ 4 แนวคิดเกี่ยวกับการเสริมสร้างสมรรถนะการประเมิน

1. ความเป็นมา

ในช่วงต้นทศวรรษที่ 21 ได้มีความตื่นตัวอย่างเป็นประวัติการณ์ในวิชาชีพการประเมิน เนื่องจาก มีการนำเสนอแนวคิดของการประเมินแนวใหม่ ที่เน้นการมีส่วนร่วม (participatory) แบบร่วมแรงร่วมใจ (collaborative) และนำผู้มีส่วนได้ส่วนเสีย (stakeholder) มาใช้ในการประเมินอย่างแพร่หลาย และปรากฏการณ์สำคัญเกิดขึ้นในปี ค.ศ. 2000 เมื่อ James Sanders นายกสมาคมการประเมินแห่งอเมริกา (American Evaluation Association: AEA) ได้นำเสนอหัวข้อหลักของการประชุมวิชาการคือ การเสริมสร้างสมรรถนะการประเมิน (Evaluation capacity building) และการประเมินกระแสหลัก (Mainstreaming evaluation) ในการประชุมปีต่อมา ในปี ค.ศ. 2006 มีการ

จัดตั้งองค์กรการเสริมสร้างสมรรถนะด้านการเรียนรู้ และด้านการประเมิน (Preskill & Boyle, 2008) โดยแนวคิดหลักของการเสริมสร้างสมรรถนะการประเมิน เป็นการสร้างศักยภาพในการประเมินในระดับบุคคล และองค์กรเพื่อการพัฒนาตนเอง พึ่งพาตนเอง บนพื้นฐานของการมีส่วนร่วม และความร่วมแรงร่วมใจของคนในองค์กรเพื่อให้เกิดการสร้างสังคมแห่งการประเมิน นำการประเมินเข้ามาใช้ในวิถีการทำงานประจำ ซึ่งจะเป็นการพัฒนาที่ยั่งยืน รายละเอียดของความหมาย และแนวคิดหลักของการเสริมสร้างสมรรถนะการประเมินมีดังต่อไปนี้

2. ความหมาย

นักวิชาการด้านการประเมินให้ความหมายของการเสริมสร้างสมรรถนะการประเมิน ไว้หลายท่าน แต่นิยามที่ได้รับการอ้างอิงถึงมากที่สุดคือ นิยามโดย Stockdill, Baizerman และ Compton (2002 cited in Preskill & Boyle, 2008) ที่กล่าวว่า “การเสริมสร้างสมรรถนะการประเมินเป็นความตั้งใจในการทำงานเพื่อสร้างสรรค์ และคงไว้ซึ่งกระบวนการขององค์กรอย่างต่อเนื่อง ซึ่งทำให้เกิดการประเมินที่มีคุณภาพ และการนำไปใช้จนเป็นงานประจำ”

โปรแกรมการพัฒนาการประเมินของมหาวิทยาลัย Wisconsin ได้ระบุความหมายของการเสริมสร้างสมรรถนะการประเมินในโปรแกรมว่า เป็นกิจกรรม และกระบวนการที่ช่วยสร้างสรรค์ สร้างความเข้มแข็ง และความยั่งยืนของสมรรถนะในการประเมินในระดับบุคคล และชุมชน (Program development and evaluation, 2008)

Preskill และ Boyle (2008) ได้ศึกษาแนวคิดเกี่ยวกับการเสริมสร้างสมรรถนะการประเมิน และสรุปกรอบแนวคิดครอบคลุมเนื้อหาเพื่อให้ง่ายต่อการนำไปประยุกต์ใช้ในรูปของโมเดลพหุวิทยาการ (Multidisciplinary model) ที่พิจารณาปัจจัยที่มีอิทธิพลต่อกิจกรรมการเสริมสร้างสมรรถนะการประเมิน เช่น ปัจจัยนำเข้า การออกแบบ การนำไปใช้ และผลกระทบของกระบวนการเสริมสร้างสมรรถนะการประเมิน และความยั่งยืนของกระบวนการ โดยได้นิยามการเสริมสร้างสมรรถนะการประเมินว่าเป็นการสอน และการเรียนรู้กลยุทธ์ด้านการออกแบบ และนำไปใช้เพื่อช่วยให้บุคคล กลุ่มคน และองค์กรได้เรียนรู้การสร้างการปฏิบัติการประเมิน อย่างมีประสิทธิภาพ มีประโยชน์ และอย่างมืออาชีพ เป้าหมายสูงสุดของการเสริมสร้างสมรรถนะการประเมินอยู่ที่การปฏิบัติการประเมินอย่างยั่งยืน

3. แนวคิดของการเสริมสร้างสมรรถนะการประเมิน

3.1 แนวคิดหลักของการเสริมสร้างสมรรถนะการประเมิน (toward conceptualizing ECB) ตามการให้นิยามของโดย Stock ill, Baizerman และ Compton (2009) มีกรอบแนวคิดที่ประกอบด้วยองค์ประกอบต่อไปนี้ 1) บริบทของการพึ่งพา (context-dependent) 2) ระบบการลงมือปฏิบัติด้วยความมุ่งเท (intentional action system) และ 3) กระบวนการ รายละเอียดของทั้งสามองค์ประกอบ คือ

3.1.1 บริบทของการพึ่งพา (context- dependent) ความหมายของบริบทของการพึ่งพา คือ ความเป็นจริงของแต่ละองค์กรว่าจะมีบริบทที่มีความเฉพาะอย่างไร

3.1.2 ระบบปฏิบัติการด้วยความมุ่งเท (intentional action system) หมายถึง การเสริมสร้างสมรรถนะการประเมินที่มีขึ้นในองค์กร หรือโมดูลหรือบางหน่วยงานโดยมีความร่วมมือกับหน่วยงานอื่นในรูปของพันธมิตร และความสัมพันธ์ในเชิงนโยบาย เมื่อเวลาผ่านไปทำให้พวกเขาสามารถคงความยั่งยืนของระบบไว้ได้ การจัดการด้านการเสริมสร้างสมรรถนะการประเมินจะต้องเป็นหน่วยงานที่เป็นผู้นำ และทำงานเพื่อการพัฒนากระบวนการปฏิบัติที่ยั่งยืน

3.1.3 กระบวนการระบบ หมายถึงความพยายามในภาพรวมทั้งหมดที่จะสร้างสรรค์ หรือคงไว้ซึ่งระบบปฏิบัติการ การสร้างสมรรถนะการประเมิน จุดเน้นอยู่ที่การพัฒนา และการคงความยั่งยืนของหน่วยการประเมิน ซึ่งปัจจัยสำคัญที่สุดอยู่ที่การได้รับการสนับสนุนจากผู้บริหารระดับสูงสุดขององค์กร

3.2 กรอบแนวคิดการเสริมสร้างสมรรถนะการประเมินในองค์กรที่มีความซับซ้อน (Taylor-Powell & Boyd, 2008) การขยายกรอบแนวคิดการเสริมสร้างสมรรถนะการประเมินไปในองค์กรที่มีความซับซ้อนเนื่องจากต้องการเพิ่มประสิทธิภาพ และความรับผิดชอบ เป้าหมายหลักคือ การสร้างความเข้มแข็งและความยั่งยืนให้กับสมรรถนะขององค์กรและเพื่อ 1) ออกแบบ ลงมือปฏิบัติ และจัดการกับโครงการประเมินอย่างมีประสิทธิภาพ 2) เข้าถึง สร้าง และใช้ความรู้และทักษะการประเมิน 3) เพื่อให้เกิดความองอาจของจิตวิญญาณของการเรียนรู้ การปรับปรุง และความรับผิดชอบต่อองค์กร และ 4) สร้างความตระหนัก และการสนับสนุนโมดูลการประเมิน และ การประเมินตนเอง ให้เป็นกลยุทธ์ในการพัฒนาผลงาน โดยมีกรอบแนวคิดในการดำเนินการเสริมสร้างสมรรถนะการประเมินในองค์กรที่มีความซับซ้อนประกอบด้วยองค์ประกอบ 3 องค์ประกอบ คือ 1) การพัฒนาระดับวิชาชีพ 2) การสนับสนุน และทรัพยากร และ 3) สิ่งแวดล้อมขององค์กร ดังรายละเอียดต่อไปนี้

3.2.1 การพัฒนาระดับวิชาชีพ (professional development) เป็นการสร้างความรู้ ความเชื่อ และทักษะของบุคคลในการประเมิน ในระดับต่างๆ ของวิชาชีพในองค์กร แตกต่างกันใน ระดับของแรงจูงใจ และ ระดับของการฝึกอบรม เป้าหมายเพื่อการพัฒนาความรู้ ทักษะ และทัศนคติ และพฤติกรรมในระดับบุคคล ไปสู่ระดับทีมงาน หรือองค์กร ในการสร้างหรือใช้การประเมินในภาพ ของกลุ่มมีกลยุทธ์ในการดำเนินการพัฒนา หลายกลยุทธ์ เช่น การฝึกอบรม (training) การประเมิน การประเมินวัตุดิบ ทรัพย์สินขององค์กร สถานะทางการเงิน เทคโนโลยี และเวลา

3.2.3 สิ่งแวดล้อมขององค์กร สิ่งแวดล้อมขององค์กรที่เหมาะสมต่อการเสริมสร้างสมรรถนะ การประเมินประกอบด้วย

3.2.3.1 ภาวะผู้นำ หัวหน้าเป็นผู้มีบทบาทสำคัญในการสนับสนุน และสำรวจการ สนับสนุนภายในองค์กร ผู้นำจะต้องเข้าใจ และแสดงออกถึงเป้าหมาย และคุณค่าของการเสริมสร้าง สมรรถนะการประเมินให้ผู้อื่นทราบได้ เป็นผู้จัดระดับความคาดหวังในการประเมิน สนับสนุนการใช้ ผลการประเมิน และแจ้งผลการนำผลการประเมินไปใช้ สนับสนุนการใช้กระบวนการประเมิน ทั้ง อย่างเป็นทางการ และไม่เป็นทางการ และเป็นผู้ร่วมรับผิดชอบในการเสริมสร้างสมรรถนะการ ประเมิน และค้นหาหนทางในการบูรณาการการประเมินเข้ากับวิถีการทำงานประจำในองค์กร

3.2.3.2 อุปสงค์ แสดงให้เห็นความรับผิดชอบอย่างสูงต่อแหล่งทุนที่ได้รับ การช่วยให้ เกิดการยอมรับการประเมินในหน่วยงาน ส่วนใหญ่มักใช้กลยุทธ์ของการกล่าวถึงความสำเร็จของ โครงการ เพื่อโฆษณาประชาสัมพันธ์ความสำเร็จของโครงการ การประเมินที่ตอบข้อสงสัย และการ เรียนรู้อาจสวนทางกับการรายงานผลการประเมินเพื่อทำให้ผู้ให้เงินทุนสนับสนุนพึงพอใจ ดังนั้น ผู้ประเมินโครงการควรหาแหล่งทุนจากหลายๆ แหล่งไม่ควรผูกติดกับแหล่งทุนเดียวเพื่อความเป็น อิสระในการตอบสนองความต้องการที่แท้จริง

3.2.3.3 สิ่งจูงใจ หรือรางวัลในการประเมิน เป็นการสร้างแรงจูงใจของบุคคลให้เข้าสู่ กระบวนการของการประเมิน ทั้งนี้ขึ้นอยู่กับสัญญาตาณญาณภายในตัวบุคคล ประสบการณ์ในอดีตที่ เกี่ยวกับการประเมิน และการรับรู้คุณค่าของประสบการณ์การประเมินว่าจะมีผลให้บุคคลเกิด แรงจูงใจในการเข้าสู่กระบวนการเสริมสร้างสมรรถนะการประเมินอย่างไร บางสถานการณ์การ ประเมินเป็นส่วนหนึ่งของงานประจำ แต่บางสถานการณ์การประเมินเป็นงานชั่วคราวมีกำหนด ระยะเวลาสิ้นสุด และจะเริ่มการประเมินใหม่เมื่อมีความต้องการการประเมินจากภายนอกหน่วยงาน

3.2.3.4 โครงสร้าง มีสมรรถนะของโครงสร้าง 3 สมรรถนะที่ปรากฏชัดเจนว่าจะช่วยให้เกิดการสร้างสมรรถนะในองค์กรขนาดใหญ่ 1) ความสำเร็จของการเสริมสร้างสมรรถนะการประเมินขึ้นอยู่กับโครงสร้างของการสื่อสารที่ช่วยให้ได้รับข้อมูลทั้งในแนบราบ และแนวตั้ง อย่างทั่วถึง มีทิศทาง การไหลเวียนของข้อมูลชัดเจนขึ้นได้โดยอาศัยการใช้เทคโนโลยีเพิ่มขึ้น 2) โครงสร้างของทีมงานของโมดูล การจัดตั้งทีมงาน การช่วยเหลือในการกำจัดอุปสรรค การเสาะแสวงหาอย่างร่วมมือ การแก้ปัญหา กลุ่ม และการสังเคราะห์ รวมถึงโครงสร้างการสนับสนุนจากเพื่อน โครงสร้างการเรียนรู้ กลุ่มให้คำปรึกษาทางการประเมิน และโครงสร้างของระบบพี่เลี้ยงและการให้คำปรึกษา โครงสร้างเหล่านี้มีส่วนช่วยในการเสริมสร้างสมรรถนะการประเมิน 3) ระบบการจัดการข้อมูล ทั้งการช่วยในการสร้าง การจัดการ และการใช้ข้อมูลรวมเข้าด้วยกันเป็นคลังสำหรับกำหนดการเก็บรวบรวมข้อมูลตามความต้องการ เช่น อยู่ในรูป ระบบประมวลข้อมูลจากฐานข้อมูล แม่แบบของข้อมูลสำหรับการใช้ และการสื่อสาร และการประมวลการกำกับติดตามคุณภาพข้อมูล และการแลกเปลี่ยนบทเรียนที่ได้เรียนรู้

3.2.3.5 นโยบาย และกระบวนการ การประเมินอาจเขียนในรูปการวิเคราะห์งาน และการทบทวนผลการปฏิบัติงานประจำปี อาจเห็นนโยบายการประเมินในรูปแนวปฏิบัติสำหรับ โครงการระยะสั้น สำหรับการเลื่อนตำแหน่ง สำหรับทุนการศึกษา สำหรับการได้รับทุน หรือการทำ สัญญาโครงการงานวิจัย การได้รับรางวัลทุนจากภายในหน่วยงาน และความคาดหวังของทีมงาน และกลุ่มทำงาน อย่างไรก็ตามโดยธรรมชาติของระดับการประเมิน และความคาดหวังต่อการประเมินอาจ ไม่สามารถระบุได้ชัดเจน ผู้ทำหน้าที่ระบุความคาดหวังควรเป็นเจ้าของพื้นที่ หรือผู้ที่เกี่ยวข้องกับหน่วยงานที่จะดำเนินโครงการ อาจทำได้โดยการ แจกแบบสอบถามในช่วงท้ายของการดำเนินโครงการ

3.3 กระบวนการในการเสริมสร้างสมรรถนะการประเมิน

มี 10 ขั้นตอน อาจใช้กระบวนการนี้ปรับเปลี่ยนตามสถานการณ์ได้โดยไม่จำเป็นต้อง ยึดตามลำดับของกระบวนการทั้งหมด (Lancaster University, 2013)

3.3.1 วงจรการประเมิน (Evaluation cycle) เป็นขั้นตอนของการสร้าง การนำไปใช้และการ ทบทวนแผนการดำเนินงาน รวมถึงการทบทวนการปฏิบัติกิจกรรมในอนาคต

3.3.2 การเตรียมการประเมิน (Evaluation preparation) เป็นการวิเคราะห์สถานการณ์ ของการคงอยู่ของการประเมิน และการวางแผนการวิเคราะห์ข้อมูลที่จะถูกเก็บรวบรวม

3.3.3 การวางแผนการประเมิน (Evaluation planning) เป็นกรอบของงานที่ใช้สำหรับการ วางแผนการประเมิน เพื่อช่วยในการตัดสินใจ

3.3.4 ตัวบ่งชี้ ผลจากการประเมิน (Evaluation impact indicators) เป็นระดับของผลกระทบ และผลของกระบวนการ ร่วมกับตัวบ่งชี้ด้าน ความสามารถ กระบวนการ และผลลัพธ์

3.3.5 มุมมองด้านการประเมิน (Evaluation perspectives) เป็นเค้าโครงของวิถีทางการร่วมมือสำหรับการประเมินและ เหตุผลสำหรับการประเมินทั้งจากภายในและภายนอกองค์กร

3.3.6 ความสามารถในการปฏิบัติการประเมิน (Evaluation practicalities) เป็นการลงมือเก็บรวบรวมข้อมูล การแลกเปลี่ยนข้อมูล และเป็นเรื่องของระเบียบวิธีการทางการวิจัย และด้านจริยธรรม

3.3.7 การรวบรวมข้อมูลการประเมิน (Evaluation data collection) ข้อมูลจากผู้มีส่วนร่วม ความก้าวหน้า และ ความตั้งใจ ข้อมูลเหล่านี้จะนำไปใช้ในเชิงข้อมูลคุณภาพอย่างไร

3.3.8 การวิเคราะห์ข้อมูลจากการประเมิน (Evaluation data analysis) เป็นจุดเชื่อมระหว่างข้อมูลเชิงปริมาณ และเชิงคุณภาพ และเป็นการวิเคราะห์ข้อมูลเชิงคุณภาพ

3.3.9 การเผยแพร่ผลการประเมิน (Evaluation dissemination) เป็นช่องทางของการแลกเปลี่ยนข้อมูลที่พบจากการประเมินกับบุคคลอื่นๆ

3.3.10 หลักฐานเชิงประจักษ์ด้านการประเมิน (Evaluation Evidence) แผนงานการประเมิน และข้อมูลที่ได้จากการประเมินนำไปสู่การทำหลักฐานเชิงประจักษ์

4. โมเดลที่ใช้ในการเสริมสร้างสมรรถนะการประเมิน

โมเดลที่ได้รับความนิยมนำมาใช้ในการเสริมสร้างสมรรถนะการประเมินคือ โมเดลพหุวิทยาการของ Preskill และ Boyley (Preskill & Boyle, 2008) ไม่ว่าจะการสร้างการประเมินจะอยู่ในระดับการประเมินตนเองในองค์กรหรือ ขนาดใหญ่ที่มีความซับซ้อน ย่อมมีกลยุทธ์ในการสร้างสมรรถนะการประเมิน เพื่อใช้ในกระบวนการเสริมสร้างสมรรถนะการประเมิน ระดับบุคคลและองค์กร มีกลยุทธ์ต่างๆ ดังนี้ (Graig, 2007 ; Huffman, Thomas, & Lawrenz, 2008 ; Preskill & Boyle, 2008)

4.1 การฝึกอบรมเพียงอย่างเดียว (training only) เป็นวิธีที่ได้รับความนิยมมาก เหมาะสำหรับผู้จัดทำโครงการ การเสริมสร้างสมรรถนะการประเมินที่ต้องการนำเสนอการประเมิน ในระยะเริ่มต้น เพื่อให้เข้าใจเค้าโครงของหลักสูตร เป้าหมายของโครงการอยู่ที่การแนะนำการประเมินเข้าสู่องค์กร ช่องทางในการอบรม เช่น ให้ข้อมูลแบบเผชิญหน้า การให้ข้อมูลสดทางระบบออนไลน์ หรือการสื่อสารภายในเครือข่าย หรือระหว่างเครือข่ายสารสนเทศ เมื่อเปรียบเทียบกับวิธีการอื่น วิธีนี้เป็นวิธีที่ประหยัดที่สุด

4.2 การฝึกอบรม และกลุ่มการสอนงาน (training and group coaching) มีรูปแบบคล้ายรูปแบบแรก แต่เนื้อหามีความเข้มข้นมากขึ้น และมีปฏิสัมพันธ์ระหว่างวิทยากร หรือผู้ฝึกผู้ประเมินกับผู้เข้ารับการอบรมมากขึ้น โดยเพิ่มส่วนของการสอนงาน และให้คำปรึกษาด้านการประเมินเพื่อให้ผู้อบรมเข้าใจตรงกัน อาจใช้การถามตอบในช่วงท้ายของการอบรมซึ่งเป็นปฏิสัมพันธ์ระดับต่ำสุดของการสอนงาน อาจมีรูปแบบเผชิญหน้ากับผู้อบรม หรือการโต้ตอบผ่านระบบออนไลน์ในเวลานั้นหรือผ่านระบบการบริการเครือข่ายที่อนุญาตให้สมาชิกส่งข้อมูลทางระบบออนไลน์ได้ รูปแบบการอบรมนี้จะทำให้ได้เนื้อหาของการฝึกอบรมตามความต้องการของผู้เข้ารับการฝึกอบรม และได้สร้างชุมชนของการประเมินผ่านกลุ่มผู้เข้ารับการอบรมที่มีความสนใจในเรื่องเดียวกัน

4.3 การฝึกอบรม และการสอนงานเป็นรายบุคคล (training and individual coaching) ให้ความสำคัญกับการสอนงานเป็นรายบุคคลไปพร้อมกับการมีส่วนร่วม และให้ความสำคัญกับเนื้อหาพื้นฐานที่จะทำให้กลุ่มสามารถสร้างงานร่วมกับผู้สอนงานได้ การสอนงานอาจมีขึ้นตลอดระยะเวลาการฝึกอบรม หรืออาจจำกัดช่วงเวลา เช่น ช่วงแรกของการฝึกอบรม หรือระหว่างการฝึกอบรม และช่วงกลางของการฝึกอบรม

4.4 การแลกเปลี่ยนความรับผิดชอบต่อการส่งข้อมูลการประเมิน (shared responsibility for evaluation deliverables) ผู้ให้คำปรึกษา และผู้ที่ไม่มีส่วนได้ส่วนเสียอยู่ในสัมพันธภาพแบบร่วมมือกัน โดยมีสมรรถนะของการแลกเปลี่ยนความรับผิดชอบสำหรับการส่งออกข้อมูลการประเมินเป็นส่วนใหญ่ ตัวอย่างเช่น ผู้ให้คำปรึกษาอาจพัฒนาเครื่องมือ และมีบทบาทสำคัญในการวิเคราะห์ข้อมูล ในขณะที่ผู้ที่ไม่มีส่วนได้ส่วนเสียมีหน้าที่หลักในการเก็บรวบรวมข้อมูล ทั้งสองอาจทำงานด้วยกันในการร่าง และนำเสนอผลรายงานการประเมิน เป้าหมายของความร่วมมือ คือ การส่งผ่านความรู้ ซึ่งทำให้รูปแบบที่แตกต่างจากรูปแบบที่เคยมีมา และการสรุปผลการประเมินสำหรับตนเองเท่านั้น รูปแบบนี้ไม่เหมาะกับสถานการณ์ที่มีผู้อบรมจำนวนมาก หากเลือกใช้วิธีการนี้ควรใช้เทคนิคการใช้ทีมเป็นฐาน จะช่วยแก้ปัญหาได้ดี

4.5 ระบบการประเมินด้วยการฝึกอบรมแบบพร้อมใช้ (turnkey evaluation system with training) รูปแบบนี้มีส่วนคล้ายกับรูปแบบที่ 4 คือมีจุดเน้นที่การให้คำปรึกษา ผู้ให้คำปรึกษาเป็นผู้ทำหน้าที่หลัก โดยทำหน้าที่เพียงรายงานผลการออกแบบการให้คำปรึกษา หรือการบริหารจัดการเครื่องมือในการออกแบบการให้คำปรึกษา ผู้ให้คำปรึกษาเป็นผู้สร้างระบบการประเมิน และเป็นผู้ฝึกผู้เข้ารับการอบรมในการควบคุมระบบอย่างอิสระ และให้ใช้ความรู้เรื่องการประเมินน้อยกว่าวิธีที่ใช้

ความร่วมมือ ความรู้ที่ถ่ายทอดให้จะเน้นไปที่การลงมือปฏิบัติการประเมินมากกว่าการออกแบบการประเมิน รูปแบบการประเมินไม่ได้มุ่งเน้นการเสริมพลังอำนาจ แต่เน้นประสิทธิผล ผลงาน และการนำไปใช้เนื่องจากผู้ให้คำปรึกษาเป็นผู้ออกแบบระบบทั้งหมด จึงทำให้ผู้เข้ารับการอบรมไม่จำเป็นต้องเรียนรู้ระบบการทำงานของการประเมิน

4.6 การให้การช่วยเหลือด้านเทคนิค (technical assistance) เป็นการเรียนรู้โดยผู้เชี่ยวชาญภายนอก เกี่ยวกับการใช้เครื่องมือ การให้คำแนะนำในการวางแผนการประเมิน การสำรวจ และการลงมือปฏิบัติการสำรวจ รวมถึงการวิเคราะห์ข้อมูล ผู้ประเมินจำเป็นต้องเรียนรู้เทคนิคที่ซับซ้อนขึ้นไปของการประเมิน และขยายขอบเขตของความรู้ออกไปสู่การประเมินในสถานการณ์จริง

4.7 การใช้ตัวเร่งปฏิกิริยาให้เกิดการเปลี่ยนแปลงเร็วขึ้น (catalyst-for -change) โดย Taut (2007) กล่าวว่าควรเลือกใช้วิธีนี้ในบริบทของการประเมินตนเองระหว่างผู้ประเมินกันเองในขณะดำเนินการเสริมสร้างสมรรถนะการประเมิน ซึ่งหมายถึงการที่บุคคลที่อยู่ในตำแหน่งผู้นำทำหน้าที่ช่วยให้เกิดการเปลี่ยนแปลงอย่างสำคัญในเรื่อง 1) ความรู้ และทักษะของสมาชิกคนอื่น ๆ 2) เป็นตัวเร่งให้เกิดองค์กรที่มีการปฏิบัติการประเมินเป็นกระแสหลัก โดยผู้นำมีอำนาจในการจูงใจให้มีการใช้แหล่งสนับสนุนเพื่อเปลี่ยนองค์กรของการปฏิบัติการประเมิน และเพื่อตอบสนองความต้องการด้านโครงสร้างพื้นฐานที่ใช้สำหรับโมดูลการประเมิน เพื่อให้ได้องค์กรกระแสหลักของการประเมิน

4.8 การอบรมเชิงปฏิบัติการ (The workshop/ institute approach) เป็นอีกวิธีหนึ่ง ที่ใช้ในการเสริมสร้างสมรรถนะการประเมิน ส่วนใหญ่มักจัดในรูปการประชุมประจำปี โดยให้ผู้เชี่ยวชาญเป็นผู้ให้ความรู้ ในหัวข้อที่หลากหลายประเด็นของการประเมิน การอบรมเชิงปฏิบัติการของสถาบันนั้นๆ มุ่งเน้นการช่วยเหลือบุคคลให้เรียนรู้ทักษะเฉพาะ หรือทำความเข้าใจแนวคิดของวิธีการแต่ละวิธีการให้ดีขึ้น การจัดการอบรมเชิงปฏิบัติการจะช่วยให้เกิดการพัฒนาในระดับบุคคลคือพัฒนาศักยภาพการประเมิน ในเรื่องใหม่ๆในช่วงเวลาสั้น

4.9 วิธีความร่วมมือ (The collaborative immersion approach) มีแนวคิดพื้นฐานมาจากทฤษฎีทางจิตวิทยา social- constructivist ที่มองว่าสังคมช่วยให้คนเกิดการเรียนรู้จากกระบวนการทางสังคมได้ ร่วมกับกระบวนการพื้นฐานของวิธีการอบรมเชิงปฏิบัติการของสถาบัน โดยช่วยสร้างความรู้ และทักษะในการประเมินในระดับบุคคล ในขณะที่มีการเสริมสร้างสมรรถนะขององค์กรไปพร้อมๆ กัน เป็นการสร้างวิธีในการพัฒนาสมรรถนะการประเมิน ทั้งในระดับบุคคล และองค์กร

โดยเป็นการเสริมสร้างสมรรถนะการประเมินเป็นขั้น ๆ จากระดับบุคคลที่มีระดับการประเมินง่าย ๆ ไปถึงซับซ้อน และในทางกลับกันจากระดับซับซ้อนคือระดับองค์กรลงมาถึงบุคคล

4.10 วิธีการฝึกงาน (The Fellowship/ Internship approach) เป็นการเสริมสร้างสมรรถนะอีกขั้นหนึ่งผ่านการประเมินในสถานการณ์จริง โดยมีผู้เชี่ยวชาญให้คำปรึกษาแนะนำ ผลของการฝึกงานมักได้รับผลลัพธ์ที่ดีมาก ทำให้ผู้ฝึกงานได้รับทักษะการประเมินที่มีความซับซ้อนในสถานการณ์จริง

5. ผลลัพธ์ของการเสริมสร้างสมรรถนะการประเมิน

ผลลัพธ์ของการเสริมสร้างสมรรถนะการประเมินมีทั้งในระดับบุคคล และระดับองค์กร คือ

5.1 ผลลัพธ์ในระดับบุคคล จากการศึกษาวิจัยพบว่าการเสริมสร้างสมรรถนะการประเมินทำให้เกิดผลลัพธ์ด้านสติปัญญา (cognitive outcome) ด้านพฤติกรรม (behavioral outcome) และด้านอารมณ์ความรู้สึก (affective outcome) ดังนี้ (Taut, 2007)

5.1.1 ผลลัพธ์ด้านสติปัญญา (cognitive outcome) เป็นการเพิ่มความรู้ด้านการประเมิน และความคิดในเรื่องการประเมินที่รวมเข้าเป็นส่วนหนึ่งของการปฏิบัติงานประจำวันของวิชาชีพ

5.1.2 ผลลัพธ์ด้านพฤติกรรม (behavioral outcome) เป็นการเพิ่มทักษะที่บูรณาการการประเมินเข้ากับการปฏิบัติงานประจำวัน

5.1.3 ผลลัพธ์ด้านจิตใจ หรือ อารมณ์ (affective outcome) เป็นการเปลี่ยนแปลงในด้านบวกในด้านทัศนคติต่อการประเมิน

5.2 ผลลัพธ์ระดับองค์กรการเสริมสร้างสมรรถนะการประเมินช่วยสนับสนุนการใช้การประเมินในทางปฏิบัติ และสร้างกระแสหลักของการประเมินให้การประเมินเป็นส่วนหนึ่งของงานประจำ (Cousins, Linnemann, & Tucker, 2008 ; Garcia-Iriarte, Suarez-Balcazar, Taylor-Ritzler, & Luna, 2010)

5.3 ผลลัพธ์ในระหว่างองค์กร ทำให้เกิดการเชื่อมต่อทางสาธารณะ และระหว่างบุคคลผ่านทางเครือข่าย กระบวนการ หรือผ่านหุ้นส่วน

5.4 ผลลัพธ์ในระดับสังคมทำให้เกิดกระบวนการคิดอย่างสังคมแห่งการประเมิน รวมถึงองค์กรด้านวิชาชีพ และสาธารณะ

ตอนที่ 5 การพัฒนาโมดูลการฝึกอบรม

1. ความหมายของโมดูล (Cambridge dictionary online, 2016)

ความหมายโมดูลโดยทั่วไปที่ไม่เฉพาะเจาะจงถึงสาขาวิชาใด ๆ โมดูลหมายถึงหนึ่งในชุดของส่วนประกอบใด ๆ ที่สามารถแยกส่วนออกมาได้ และเมื่อนำมารวมกันจะทำให้เกิดความสมบูรณ์ของชุดนั้น ๆ

ความหมายของโมดูลในมุมมองด้านการศึกษามีหมายถึง หน่วยการเรียนรู้ในหลักสูตร ที่เมื่อรวมกันแล้วจะเกิดความสมบูรณ์ แต่ก็สามารถแยกออกมาได้เป็นอิสระ ตัวอย่างเช่น โมดูลที่ใช้ในการฝึกอบรมเพื่อเป็นแนวปฏิบัติสำหรับผู้ปฏิบัติงานในการวางแผนให้เป็นผู้ประสบความสำเร็จในงานภาคปฏิบัติ ใช้สำหรับผู้บริหารโรงเรียน ของหน่วยงาน National center on response to intervention ของรัฐบาลสหรัฐอเมริกา (2012) มีหน่วยการเรียนรู้ในโมดูลประกอบด้วย (1) Power point การนำเสนอ (2) บันทึกย่อสาระสำคัญของวิทยากรผู้บรรยายในงานนำเสนอ (3) คู่มือการฝึกอบรมที่ต้องใช้ร่วมกับ Power point การนำเสนอ (4) กระดาษสำหรับสรุปสิ่งที่เรียนรู้และนำมาแนะนำเสนอในกลุ่ม (5) วิดีทัศน์เกี่ยวกับผู้เชี่ยวชาญ หรือผู้ที่มีประสบการณ์ในการปฏิบัติการณ์นั้น

สำหรับโมดูลการฝึกอบรมในการวิจัยครั้งนี้ จึงหมายถึง หน่วยการเรียนรู้ย่อย จำนวน 5 หน่วย คือ (1) ความรู้พื้นฐานเกี่ยวกับการประเมินภาคปฏิบัติ (2) การฝึกอบรมเกี่ยวกับความคลาดเคลื่อนจากผู้ประเมิน (3) การฝึกทักษะการสังเกต (4) การฝึกอบรมกรอบของการอ้างอิง และ (5) การฝึกทักษะการตัดสินใจ

2. ทฤษฎีพื้นฐานของการฝึกอบรมผู้ประเมิน

การฝึกอบรมผู้ประเมินสามารถอธิบายได้ด้วยทฤษฎีการเรียนรู้ทางสังคม 3 ทฤษฎีหลักคือ ทฤษฎีการเรียนรู้ของ Bloom (Bloom's taxonomy of education objects) ทฤษฎีระดับของกระบวนการ (Levels-of-processing theory) และทฤษฎีการเกื้อหนุนทางสังคม (Social facilitation theory) โดยการฝึกอบรมผู้ประเมินเป็นตัวอย่างหนึ่งของโครงสร้างของกลุ่มสถานการณ์ในการเรียนรู้ ผลของกระบวนการการฝึกอบรมจึงมีพื้นฐานเดียวกับทฤษฎีการเรียนรู้ของ Bloom ที่ใช้ในกระบวนการทศนังทางการสอน โมเดลทั่วไปสำหรับกระบวนการสอนมีระดับของการเรียนรู้ด้านพุทธิพิสัย เรียงลำดับการเรียนรู้จากความรู้ในระดับง่าย ระดับพื้นฐานไปจนถึงระดับที่ซับซ้อน (Bloom, Engelhard, Walker, Furst, & Krathwohl, 1956 cited in Athey & McIntyre, 1987) โดยการฝึกอบรมเป็นกระบวนการหนึ่งของการเรียนรู้ด้านพุทธิพิสัยที่มีระดับการเรียนรู้ที่ค่อนข้างซับซ้อนกว่าการเรียนรู้ด้วยความรู้พื้นฐานง่าย ๆ ต้องอาศัยสื่อใน

การกระตุ้นการเรียนรู้ จึงถือได้ว่าเป็นการเรียนรู้ขั้นสูงสุดของกระบวนการเรียนรู้ สื่อที่ใช้ เช่น วีดิทัศน์ ที่มีระดับของข้อมูลที่ต้องการนำเสนอในโมดูลการฝึกอบรม โดยสิ่งที่ต้องการให้เกิดขึ้นในกระบวนการเรียนรู้ คือ ความแม่นยำในการประเมิน ส่วนทฤษฎีระดับของกระบวนการ (Crain, & Lockhart, 1972 cited in Athey, & McIntyre, 1987) จะทำให้เข้าใจความแตกต่างของรูปแบบการฝึกอบรม ผู้ประเมินที่อาจมีผลต่อการเรียนรู้เนื้อหาของการฝึกอบรม โดยให้นิยามความจำเป็นว่าเป็นการทำหน้าที่ของกระบวนการรับรู้ในส่วนลึกที่จำเป็นต้องจำในเนื้อหาที่ได้รับ ซึ่งกระบวนการรับรู้ข้อมูลนี้จะมี ความเกี่ยวข้องช่วยให้นักเรียนด้านพุทธิพิสัยมีความลุ่มลึกมากขึ้น และยังคงอยู่ในความทรงจำได้นาน เนื่องจากถูกจัดอยู่ในช่องทางของความจำที่มีความโดดเด่น โดยการฝึกอบรมเพื่อให้เกิดความแม่นยำในการประเมิน เนื่องจากจะทำให้เกิดความชัดเจนของความเชื่อมโยงระหว่างมิติของงานที่จะประเมิน กับพฤติกรรมจริงที่กำลังประเมิน รูปแบบการประเมินที่แตกต่างกันอาจทำให้ความเชื่อมโยงระหว่าง มิติของงาน และพฤติกรรมจริงแตกต่างกันจึงทำให้รูปแบบการประเมินแต่ละรูปแบบให้ผลการ ประเมินแตกต่างกัน

ทฤษฎีการเกื้อหนุนทางสังคมของ Zajonx (1965 cited in Athey, & McIntyre, 1987) กล่าวว่า การปรากฏของสิ่งต่าง ๆ ในสถานการณ์การเรียนรู้ หรือสถานการณ์การทำงานจะเพิ่มระดับสิ่งเร้า และในทางกลับกันจะมีผลต่องานในสถานการณ์นั้น เมื่อพิจารณาในมุมของการฝึกอบรม มีสมมติฐานทาง ทฤษฎีที่ว่าความจำเป็นจะได้รับอิทธิพลจากขนาดของกลุ่มในขณะที่เข้ากระบวนการฝึกอบรมผู้ประเมิน จึงมี การศึกษาถึงขนาดของกลุ่มที่มีอิทธิพลต่อการเรียนรู้ในกระบวนการฝึกอบรม ผลการศึกษาพบว่าขนาดของ กลุ่มเล็ก ๆ เช่น 3 - 5 คน จะได้ประสิทธิผลของการฝึกอบรมมากกว่ากลุ่มขนาดใหญ่ เช่น จำนวน 15 คนขึ้นไป ซึ่งทั้งสามทฤษฎีพื้นฐานนี้จะทำให้การพัฒนาโมดูลการฝึกอบรมผู้ประเมินมีความชัดเจนขึ้น และสามารถ อธิบายเชิงตรรกะได้ถึงเหตุผลในการดำเนินการในแต่ละกิจกรรมได้เป็นอย่างดี

3. การพัฒนาโมดูลการฝึกอบรม

การพัฒนาโมดูลควรประกอบด้วยขั้นตอนการพัฒนาอย่างเป็นระบบ 3 ขั้นตอนคือ (Bachman & Palmer, 1996 cited in Jang, 2010; Steve & Terry, 2009)

3.1 ขั้นวางแผน (Planning stage) เป็นขั้นวิเคราะห์สมรรถนะของรูปแบบการทดสอบในการฝึกการประเมิน และความต้องการของผู้มีส่วนได้ส่วนเสียกับการฝึกอบรม มีขั้นตอนคือ

3.1.1 โดยการสำรวจความต้องการรูปแบบในการฝึกอบรมจากผู้มีส่วนได้ส่วนเสีย การหาข้อมูลจากงานวิจัย และผู้เชี่ยวชาญจะช่วยในการตัดสินใจว่าควรเลือกใช้วิธีการใดในการฝึกอบรม สำหรับการวิจัยครั้งนี้มีการวางแผนโดยนำข้อมูลจากขั้นตอนการสำรวจสภาพปัจจุบัน จากผู้ให้ข้อมูลคืออาจารย์พยาบาลจาก 5 สถาบัน มีการสำรวจความต้องการในการเลือกใช้โมดูลมาตรฐานที่ใช้ในการฝึกอบรม

3.1.2 กำหนดวัตถุประสงค์ให้ชัดเจน ผู้พัฒนาโมดูลจะต้องทราบให้ชัดเจนว่าโมดูลที่พัฒนาขึ้นมีวัตถุประสงค์หลักเพื่อใคร และเพื่ออะไร เช่นการวิจัยครั้งนี้มีกลุ่มเป้าหมายที่ต้องการคืออาจารย์พยาบาลที่ทำหน้าที่ประเมินการฝึกปฏิบัติงานนักศึกษาพยาบาลในระดับปริญญาตรี เพื่อต้องการให้เกิดสมรรถนะการประเมิน 3 ด้านคือ ด้านความรู้ ทักษะ และทักษะ เป็นต้น

3.1.3 กำหนดโครงสร้างของวัตถุประสงค์การเรียนรู้ และผลลัพธ์ที่ต้องการให้เกิด

3.1.4 ออกแบบวิธีการ และกระบวนการในการฝึกอบรม ทั้งนี้จะต้องรู้จักผู้เข้ารับการอบรมในระดับหนึ่ง นั่นคือต้องรู้ว่าผู้เข้ารับการฝึกอบรมมีภูมิหลังอย่างไร มีความรู้ความชำนาญด้านใด เพื่อจะได้จัดกิจกรรมในระหว่างการฝึกอบรมได้เหมาะสม สำหรับการวิจัยครั้งนี้ผู้เข้ารับการฝึกอบรมเป็นอาจารย์พยาบาลที่มีประสบการณ์ในการปฏิบัติงานบนหอผู้ป่วยมาแล้ว แต่มีประสบการณ์ในการประเมินนักศึกษาพยาบาลแตกต่างกัน การออกแบบการฝึกอบรมจึงมีลักษณะเป็นการเรียนรู้แบบผู้ใหญ่คือ ให้ความรู้พื้นฐานในโมดูลนั้น ๆ และตามด้วยการฝึกทักษะในโมดูล

3.2 ขั้นกำกับติดตาม (Monitoring stage) เป็นขั้นที่เกิดขึ้นในระหว่างการฝึกอบรมกำลังดำเนินอยู่ เป็นการเปิดโอกาสให้ผู้อบรมได้แลกเปลี่ยนความคิดเห็นหรือสิ่งที่ผู้อบรมเคลือบแคลงสงสัย เพื่อให้ตามกลุ่มทัน หรือบรรลุถึงข้อตกลงร่วมกันกับกลุ่ม เป้าหมายสำคัญของการกำกับติดตามคือการฝึกฝนการประเมินจนทำให้ผลการประเมินของผู้อบรมใกล้เคียงกับผลการประเมินที่มีมาตรฐาน ซึ่งเป็นตัวเชื่อมระหว่างความรู้ที่ผู้อบรมได้รับ และการนำไปใช้ โดย Myford และ Wolfe (2009) ได้ทำการศึกษาการกำกับติดตามผลการทำหน้าที่ต่างกันของผู้ประเมินเมื่อเวลาต่างกัน (differential rater functioning over time: DRIFT) เพื่อใช้ในการค้นหาความแม่นยำ และประเภทของมาตรวัดที่แตกต่างกัน และได้สรุปผลการกำกับติดตามดังนี้

3.2.1 แบ่งผู้ประเมินออกเป็นช่วงเวลา ใช้หลักการบริหารจัดการว่าระยะเวลาที่ผู้ประเมินจะทำหน้าที่ประเมินได้ความแม่นยำควรมีระยะเวลาในการทำหน้าที่นานเท่าใด เช่น หากการทดสอบ

กินระยะเวลา 4 - 5 ชั่วโมงอาจสลับกันทำหน้าที่ และหยุดเป็นชั่วโมงพักหนึ่งชั่วโมง หรือหากมีระยะเวลาหลายวันอาจยืดระยะเวลาในการประเมินนานออกไป

3.2.2 เลือกใช้กลยุทธ์ที่สามารถค้นหา และวัดการเปลี่ยนแปลงในพฤติกรรมของผู้ประเมินได้ คือ 1) เปรียบเทียบผลการประเมินของผู้ประเมินแต่ละคนโดยให้นักวิเคราะห์ทำการเปรียบเทียบคะแนน และหาคะแนนมาตรฐานเพื่อเทียบเคียง (baseline comparison) ในการประเมินในภาพรวม 2) เปรียบเทียบกับสิ่งที่อยู่ใกล้ เช่น เทียบกับคะแนน ณ ช่วงเวลาที่ใกล้เคียงกัน ในการวิจัยครั้งนี้เลือกใช้การเทียบเคียงกับผลการประเมินจากผู้ทรงคุณวุฒิ

3.2.3 เลือกกรอบของการอ้างอิงเพื่อบรรยายผลของDRIFT คือการใช้เมทริกซ์ของการประมาณค่าพารามิเตอร์จะใช้เมื่อผู้ประเมินสองคนให้คะแนนไปต่างทิศทางกันอย่างมาก มีทั้ง internal frame of reference และ external frame of reference

3.2.4 เลือกใช้การตั้งค่าพารามิเตอร์ภายนอกเมื่อต้องใช้ผลการประเมิน ซึ่งเป็นวิธีการทางสถิติ

3.3 ขั้นประเมิน (Assessing stage) เป็นขั้นของการประเมินว่าโมดูลการฝึกอบรมนี้เหมาะสม มีประโยชน์หรือไม่เพียงใด โดยการประเมินความก้าวหน้าในการประเมินของผู้อบรมก็เป็นตัวบ่งชี้หนึ่งของการประเมินคุณภาพการฝึกอบรม ซึ่ง (Borman, 1979) ได้จัดกลุ่มเกณฑ์ที่ใช้ในการประเมินโมดูลการฝึกอบรมผู้ประเมินไว้ 3 ประเภท คือ 1) พฤติกรรมการประเมิน (rating behavior) เช่น การกดคะแนนปล่อยคะแนน การให้คะแนนเกาะกลุ่มตรงกลาง ความคลาดเคลื่อนแบบฮาโล 2) เกณฑ์ที่เกี่ยวข้องกับสมรรถนะที่สำคัญของการประเมิน เช่น ความตรงเชิงสอดคล้อง (convergent validity) หรือความตรงเชิงจำแนก (discriminant validity) และค่าความเที่ยงระหว่างผู้ประเมิน (inter rater reliability) 3) เกณฑ์ที่รวมองค์ประกอบที่หลากหลายของความแม่นยำ แต่พบว่า การศึกษาวิจัยส่วนใหญ่นิยมศึกษาเกณฑ์ในข้อที่ 1 เพียง 1 หรือ 2 ค่า ส่วนในปัจจุบันเกณฑ์ที่นิยมใช้ เช่น การทดสอบผลการใช้อบรมโดยใช้ดัชนีคือ ICC หรือ IRR ซึ่งเป็นเกณฑ์ที่ 2 และใช้ MFRM ในการตรวจสอบ rater effect ซึ่งเป็นเกณฑ์ที่ 1 (Xi & Mollaun, 2011)

นอกจากประเมินคุณภาพด้านการประเมินของผู้ประเมินแล้ว ยังมีการศึกษาถึงผลการเรียนรู้จากการฝึกอบรมอื่น ๆ เช่น แนวคิดพื้นฐานของโมเดลการเปลี่ยนแปลงด้าน ความรู้ ทักษะ และ ทักษะ (Knowledge attitude practice Model) ซึ่งมีจุดเริ่มต้นมาจาก Diffusion of innovation theory ของ Rogers (1962 cited in Chien-Yun, Wan-Fei, Yu-Hsi, & Chia-Hung, 2011) ที่มีพื้นฐานมาจากทฤษฎีการเรียนรู้ของ Albert Bandura โดยโมเดลนี้ช่วยให้เข้าใจถึงความสัมพันธ์

ระหว่างปัจจัย 3 ปัจจัยที่มีผลต่อกัน นั่นคือเมื่อได้รับความรู้ใหม่ๆ แล้ว จะเกิดการเรียนรู้ และนำมาผสมผสานกับความรู้ ประสบการณ์เดิม และความเชื่อที่มีอยู่เดิม แล้วจึงมีการเปลี่ยนแปลงของความเชื่อ และนำไปสู่การเปลี่ยนแปลงพฤติกรรมที่แสดงออกมา โดยทัศนคติมีอิทธิพลต่อพฤติกรรมที่แสดงออก ปัจจัยด้านการเปลี่ยนแปลงของความรู้ ทักษะและทัศนคติเป็นปัจจัยเริ่มต้นของการปรับเปลี่ยนพฤติกรรม สิ่งที่ต้องคำนึงถึงในการประเมินการเรียนรู้คือ

3.3.1 ควรใช้กลุ่มควบคุมในการประเมินโมดูลการฝึกอบรม หากเป็นไปได้เนื่องจากจะได้เป็นการพิสูจน์การเปลี่ยนแปลง ด้านความรู้ ทักษะ และ ทัศนคติว่าเป็นผลมาจากโมดูลการฝึกอบรมหรือไม่

3.3.2 ประเมินความรู้ ทักษะ และ/หรือ ทัศนคติ ทั้งก่อน และหลังการฝึกอบรม

3.3.3 ใช้แบบทดสอบที่เป็นข้อเขียนเพื่อวัดความรู้ และทัศนคติ ก่อน และหลังเข้ารับการฝึกอบรม

3.3.4 ใช้งานภาคปฏิบัติเพื่อเป็นแบบวัดทักษะ ในการวิจัยครั้งนี้ใช้งานภาคปฏิบัติ คือวิธีทัศนสถานการณ์จำลองการฝึกปฏิบัติงานบนหอผู้ป่วยของนักศึกษาพยาบาล โดยมอบหมายให้ ผู้เข้ารับการฝึกอบรม ฝึกประเมินผลการฝึกภาคปฏิบัติของนักศึกษาพยาบาล

3.3.5 เก็บข้อมูลจากที่กล่าวมาข้างต้นให้ได้ร้อยละร้อยของผู้เข้ารับการฝึกอบรม

3.3.6 ใช้ผลการประเมินเพื่อประโยชน์ในการดำเนินต่อไปอย่างเหมาะสม หากผลการฝึกอบรมไม่บรรลุเป้าหมาย ผู้เข้ารับการฝึกอบรมไม่มีการพัฒนาตนเองใน 3 ด้านที่กล่าวมานี้ ควรหันมาพิจารณาผู้ถ่ายทอดความรู้ว่ามีจุดบกพร่องใด ที่ทำให้ไม่บรรลุผลตามเป้าหมาย มากกว่าการมองความบกพร่องในผู้เข้ารับการฝึกอบรม เช่น ไม่สนใจในสิ่งที่ถ่ายทอดให้ หรือเรียนรู้ช้า

4. การดำเนินการฝึกอบรมผู้ประเมิน

การฝึกอบรมผู้ประเมินมีความสำคัญ และจำเป็นต่อการประเมินการปฏิบัติ เนื่องจากทำให้ผลการประเมินมีความเที่ยงมากขึ้น และเป็นการปรับความเที่ยงของผู้ประเมิน เพื่อให้ผู้ประเมินสามารถประเมินโดยใช้มาตรฐานเดียวกัน มีความเข้าใจตรงกันให้มากที่สุด (Johnson, et al., 2009) ขั้นตอนในกระบวนการ การฝึกอบรมประกอบด้วย

4.1 การเลือกผู้ประเมิน และผู้ฝึกผู้ประเมิน (Selecting raters and trainers) ผู้ประเมินที่มีคุณสมบัติในการเป็นผู้ประเมินที่ดีจะทำให้การประเมินประสบความสำเร็จ และคงไว้ซึ่งความคงที่

และความเที่ยงในการให้คะแนน การเลือกผู้ประเมินจึงมีอิทธิพลอย่างมากต่อคะแนน การเลือกผู้ประเมินมีหลักที่ควรคำนึงถึง ดังนี้ (Rudner, 1992)

4.1.1 ตัวแปรด้านประชากร เช่น เชื้อชาติ อายุ ระดับการศึกษา อาชีพ ความชำนาญ และความสนใจ

4.1.2 การคัดเลือกควรจะได้รับ การเห็นชอบในการเป็นตัวแทนจากกลุ่มที่สนใจในเรื่องที่จะประเมิน

4.1.3 ควรเข้ามาทำการประเมินในกลุ่มย่อย อาจแบ่งกลุ่มกันทำงาน โดยแบ่งเป็นกลุ่มเล็ก ๆ เนื่องจากกลุ่มใหญ่จะทำให้การตัดสินใจทำได้ยาก และเมื่อมีการประเมิน และนำผลการประเมินมาเปรียบเทียบข้ามกลุ่มเพื่อให้เกิดความเที่ยง หรือเพื่อตรวจสอบความตรงข้ามกลุ่ม ซึ่งเป็น การตรวจสอบคุณภาพของผู้ประเมินคนนั้นด้วย

ผู้ประเมินที่ดีควรมีคุณสมบัติดังนี้ (Johnson et al., 2009)

- 1) มีความสามารถในการทำความเข้าใจกับเกณฑ์ที่สำคัญ
 - 2) สามารถปฏิบัติงานในสภาพแวดล้อมที่มีกรอบชัดเจนในการทำงาน
 - 3) ปฏิบัติตามคำแนะนำหรือขั้นตอนในการปฏิบัติงาน
 - 4) ตระหนักถึง และเฝ้าระวังความคลาดเคลื่อนที่เกิดจากการตัดสินใจของตนเอง
 - 5) พยายามกำจัดอคติ และมองข้ามประเด็นที่ไม่ใช่ประเด็นหลักในการประเมิน
 - 6) เข้าใจเนื้อหาหลักของการประเมินเป็นอย่างดี และทราบข้อมูลพื้นฐานของ ผู้ถูกประเมิน
- นอกจากนี้ผู้ประเมินควรมีสมรรถนะดังนี้ (Johnson et al., 2009)

- 1) มีความสามารถด้านการจัดการอย่างเป็นระบบ
- 2) มีทักษะในการใช้กลยุทธ์ทางการอบรม และทักษะในการสร้างและดำรงรักษาไว้ซึ่ง

พลวัตรกลุ่มอย่างมีอาชีพ

3) มีคุณสมบัติของการเป็นผู้ประเมินที่ดีในทุกสมรรถนะ คือ มีคุณสมบัติของการมีทักษะ การสื่อสารและการจัดการที่ดี และมีความรอบรู้ในศาสตร์ที่จะต้องทำการประเมินเป็นอย่างดี

4.2 ประมาณการณ์จำนวนผู้ประเมิน ควรศึกษาจากประวัตินผลของการประเมินในเรื่องนั้น ๆ หากค่าเฉลี่ยของเวลาในการประเมิน 1 คน ต่องาน 1 ชิ้น จากผู้ประเมินที่มีประสบการณ์กับผู้ประเมินที่ไม่มีประสบการณ์ ดังนั้นจำนวนของผู้ประเมินจึงขึ้นอยู่กับสัดส่วนของผู้ประเมินที่มีประสบการณ์กับผู้ประเมินที่อ่อนประสบการณ์รวมถึงการคำนึงถึงเวลาที่ใช้ตรวจด้วย ตัวอย่าง เช่นการตรวจงาน

เขียนที่พบว่าหากมีการใช้เกณฑ์แบบ holistic จะใช้เวลาประมาณ 30 วินาทีต่อเรื่อง ในขณะที่ analytic ใช้เวลาประมาณ 1 - 4 นาทีต่อเรื่อง โดยทั่วไปจะใช้ผู้ประเมิน 2 คนในการประเมิน

4.3 ลงมือฝึกอบรม กรอบของการฝึกอบรมโดยทั่วไปจะประกอบด้วยขั้นตอนดังนี้

(Johnson et al., 2009)

4.3.1 การทบทวนงานการประเมินที่ต้องการประเมิน โดยพิจารณาสมรรถนะของงานเหล่านั้น

4.3.2 ให้แนวปฏิบัติในการให้คะแนน เพื่อศึกษาให้เกิดความเข้าใจที่ตรงกันเกี่ยวกับเกณฑ์ในการให้คะแนน

4.3.3 การเทียบเคียงกับงานที่จะประเมินที่มีมาตรฐานที่ดี โดยการศึกษาผลงานการประเมินที่มีคุณภาพงานที่ดี

4.3.4 ลงมือฝึกการให้คะแนนจากผลงานที่ต้องการประเมินโดยงานที่ใช้ประเมินนี้ เป็นงานที่ผ่านการประเมินโดยผู้เชี่ยวชาญ

4.3.5 นำผลการให้คะแนนของผู้ฝึกอบรมมาเทียบเคียงกับคะแนนมาตรฐานที่ให้โดยผู้เชี่ยวชาญ จากนั้นจึงมีการอภิปรายถึงผลการประเมินของผู้เข้ารับการฝึกอบรมเมื่อเทียบเคียงกับผลการประเมินของผู้เชี่ยวชาญ และมีการให้ข้อมูลย้อนกลับแก่ผู้เข้ารับการฝึกอบรมด้วย

4.3.6 ฝึกการประเมินต่อไปเรื่อย ๆ จนเมื่อใกล้สิ้นสุดการฝึกอบรม จะมีการตรวจสอบความแม่นยำของการประเมินอีกครั้ง โดยผู้ประเมินจะผ่านเกณฑ์การเป็นผู้ประเมิน ที่มีคุณภาพจะต้องมีการประเมินที่สอดคล้องกับคณะกรรมการผู้เชี่ยวชาญประมาณร้อยละ 60 ถึง 80 การฝึกอบรมจึงสิ้นสุดลง

4.4 ความรู้ที่ผู้ประเมินจำเป็นต้องทราบเกี่ยวกับสมรรถนะของการให้คะแนนประกอบด้วย

4.4.1 วัตถุประสงค์ของการประเมิน มีความสำคัญโดยตรงต่อกระบวนการฝึกอบรม การประเมิน และเป็นจุดเริ่มต้นของกระบวนการ ก่อนจะพิจารณาเลื่อนระดับคุณภาพของงานที่ประเมิน เพราะความเข้าใจในวัตถุประสงค์ของการประเมินจะช่วยให้ผู้ประเมินทราบถึงบทบาทที่ชัดเจนของตน

4.4.2 เกณฑ์ในการประเมิน โดยทั่วไปจะใช้เกณฑ์การประเมินแบบรูบริกที่ระบุสมรรถนะของผลงานที่พึงประสงค์ และเกณฑ์คุณภาพที่เรียงลำดับคุณภาพของงานตั้งแต่ไม่ผ่านไปจนถึงผ่านตามมาตรฐาน และผ่านตามเกณฑ์ไปจนถึงผ่านในระดับดีเยี่ยม โดยคุณภาพตามลำดับในรูบริกจำเป็นต้องนำมาเทียบกับเกณฑ์ตามมาตรฐานของงานที่ต้องการประเมินชนิดนั้น ๆ จึงจะทำให้เข้าใจในระดับคุณภาพของรูบริกได้ เนื่องจากจะทำให้ผู้ประเมินตัดสินใจได้ว่างานการประเมินนั้นผ่านหรือไม่ผ่านเกณฑ์มาตรฐาน เมื่อ

ผู้ประเมินเข้าใจในเกณฑ์การให้คะแนนในรูบรีค แล้วจะต้องได้รับการฝึกอบรมในเรื่อง ”กฎของการตัดสินใจ”

4.4.3 ความรู้พื้นฐานที่ควรทราบ การประเมินการปฏิบัติหลายอย่างที่จำเป็นจะต้องใช้ ผู้ประเมินที่มีความรู้ความเชี่ยวชาญในเรื่องที่จะประเมิน อาจกำหนดคุณสมบัตินี้ก่อนฝึกอบรม ซึ่งแท้จริงแล้ว การเป็นผู้ชำนาญในสาขานั้นไม่ได้เป็นเครื่องรับประกันความชำนาญที่แท้จริง ยกตัวอย่าง เช่น ครูสอนวิทยาศาสตร์อาจเชี่ยวชาญฟิสิกส์ แต่ไม่เชี่ยวชาญเคมี ดังนั้นการฝึกอบรม ผู้ประเมินจึงควร ฟูพื้นฐานความรู้พื้นฐานที่จำเป็นในการประเมินโดยการทบทวนความรู้ หรือจัดทำเอกสารที่ช่วยสนับสนุนให้ผู้เข้ารับการประเมินได้เห็นจุดสำคัญ และตัดสินได้ว่าผลงานของผู้ถูกประเมินที่คาดหวัง เป็นอย่างไร การทบทวนความรู้พื้นฐานนี้รวมถึงการทบทวนการใช้เกณฑ์ในรูบรีค การทบทวนเกณฑ์ มาตรฐานและการให้คำจำกัดความอื่น ๆ และหลังจากนั้นจึงฝึกการให้คะแนน เรียนรู้การให้คะแนนที่ถูกต้อง และอภิปรายสมรรถนะของงานที่สอดคล้องกับคะแนนที่ถูกต้อง และเนื้อหาของกรอบความรู้ พื้นฐานนั้น

4.5 กระบวนการให้คะแนนมีกระบวนการคือ

4.5.1 ให้คะแนนตามกรอบของกระบวนการที่ได้ทบทวนในขั้นตอนที่ 3 เริ่มจากผู้ประเมิน มองภาพรวมของงานที่ต้องประเมิน และยึดมั่นในหน้าที่ของการตัดสินการประเมิน จากนั้นผู้ประเมิน พิจารณาจุดแข็งจุดอ่อนของงานที่กำลังประเมิน บันทึกย่อไว้ และเริ่มให้คะแนนในขั้นตอนนี้ ขั้นตอนสุดท้ายคือการตัดสินให้คะแนนทั้งหมด

4.5.2 ตั้งคำถามให้มากเกี่ยวกับการตัดสินให้คะแนน เช่นคำถามที่ว่า จะใช้มาตรฐานประมาณค่า เท่าไรในรูบรีค ตัวอย่าง เช่น หากใช้ รูบรีค 4 ระดับคุณภาพ ผู้ประเมินควรทำตามขั้นตอนต่อไปนี้ ในแต่ละมิติ ของการประเมินถามตัวเองว่า “สมรรถนะของผลงานของผู้ถูกประเมินใดที่ใกล้เคียงกับคำอธิบายในรูบรีคมากที่สุดสำหรับคะแนนที่สูงสุด 2 ระดับ และต่ำสุด 2 ระดับ” ถ้าการตัดสินแรกนี้ ผู้ประเมินตัดสินให้สมรรถนะของ ผลงานมีสมรรถนะที่สามารถอธิบายได้ด้วยเกณฑ์จากรูบรีคตั้งแต่ครั้งหนึ่งขึ้นไปของเกณฑ์ ต่อไปให้ถามตนเองต่อ ว่า “สมรรถนะของผลงานนี้ใกล้เคียงกับคำอธิบายในรูบรีคในระดับคะแนนสูงถึงระดับคะแนนสูงสุดหรือไม่” และในทางกลับกันในกรณีคุณภาพผลงานอยู่ในระดับต่ำก็ตั้งคำถามเช่นเดิมแต่คนละทิศทาง การตั้งคำถามเป็น ลำดับเช่นนี้มีประโยชน์สำหรับผู้ประเมินที่ไม่มีประสบการณ์

4.5.3 การเพิ่มคะแนน เป็นการเพิ่มความสำคัญของการประเมินหลังจากได้ทราบชัดเจน ถึงระดับของคะแนนที่จะให้ในงานการประเมินทางการศึกษา ผู้ประเมินสามารถเพิ่มหรือลดคะแนน

ซึ่งจะทำให้ผลการปฏิบัติสูงขึ้น หรือต่ำกว่าคะแนนการประเมินตามรูปแบบปกติเพียงเล็กน้อย เช่น ถ้าผลการประเมินอยู่ในระดับคุณภาพที่สูงกว่าเกณฑ์ปกติผู้ประเมินสามารถเพิ่มคะแนนให้ หรือหักคะแนนลงเมื่อไม่ได้เกณฑ์การประเมินตามคุณภาพปกติ

4.5.4 การกำกับติดตามตนเอง ผู้ประเมินควรได้รับความรู้เกี่ยวกับการกำกับติดตามตนเอง และทราบว่าต้องกำกับติดตามอย่างไรรอบคอบ เพื่อให้เกิดความแม่นยำในการประเมินมากที่สุด โดยการทบทวนเกณฑ์มาตรฐาน หรือทำการปรับค่าความเที่ยงตรงในการให้คะแนนเพื่อป้องกันการประเมินที่ให้ผลต่างกัน ในผู้ที่มีความสามารถใกล้เคียงกัน เนื่องจากการกำกับติดตามตนเองต้องใช้เวลา และจะทำให้เกิดปัญหาความลำเอียงเมื่อต้องตรวจงานการประเมินจำนวนมาก จึงควรเพิ่มขึ้นตอนนี้ในกรณีต้องคัดเลือกคุณสมบัติของผู้ประเมินเพื่อทำการจ้างงาน

4.5.5 การประเมินในสถานที่จริง ในขั้นตอนนี้อาจไม่จำเป็นต้องฝึกปฏิบัติทั้งนี้ขึ้นอยู่กับลักษณะการประเมิน เช่นการประเมินที่ผู้ประเมินฝึกประเมินจนเกิดความแม่นยำแล้วแต่ไม่เคยได้ฝึกประเมินในแหล่งประเมินจริง และไม่เคยประเมินในโครงสร้างการประเมินระดับสูง ซึ่งมักเป็นการประเมินระดับกลุ่ม หากผู้ประเมินทำการประเมินตามรูปแบบการประเมินออนไลน์จำเป็นต้องฝึกประเมินจริงเพื่อฝึกการสื่อสารที่ชัดเจน และสมบูรณ์ครบถ้วนผ่านทางจดหมายอิเล็กทรอนิกส์ แทนการสื่อสารโดยตรงกับผู้ถูกประเมิน

5. โมเดลมาตรฐานที่ใช้ในการอบรมผู้ประเมิน

การอบรมผู้ประเมินเริ่มถูกนำมาใช้เพื่อแก้ปัญหาความคลาดเคลื่อนจากผู้ประเมินในช่วงปลายทศวรรษที่ 80 โดยมีวิธีการฝึกอบรมสรุปได้พอสังเขปดังนี้

5.1 การฝึกอบรมความคลาดเคลื่อนจากผู้ประเมิน(rater error training: RET) เป็นวิธีการแรก ที่มีการนำมาใช้ในการฝึกอบรมผู้ประเมิน พัฒนาขึ้นโดย Latham และคณะ (1975 cited in Pulakos, 1984) โดยมีวัตถุประสงค์เพื่อให้เกิดความแม่นยำในการประเมิน ลดการเกิดความคลาดเคลื่อนจากผู้ประเมินที่เกิดจากการ กด ปลดอยคะแนน และ ความคลาดเคลื่อนแบบฮาโล เนื้อหาของการอบรมจะเกี่ยวกับชนิดของความคลาดเคลื่อนจากการประเมินที่เกิดจากผู้ประเมิน และ วิธีการในการแก้ปัญหาความคลาดเคลื่อนในแต่ละชนิด และต่อมาได้ขยายกรอบความคิดไปสู่ การฝึกอบรมเรื่องความแม่นยำในการประเมิน นั่นคือถ้าผลการประเมินแสดงให้เห็นว่ามีความสอดคล้องกันสูงระหว่างผู้ประเมิน ช่วงของคะแนนไม่กระจุกอยู่บนมาตรวัดด้านใดด้านหนึ่งย่อมแสดงว่าการประเมินนั้นมีความแม่นยำ

5.2 การฝึกอบรมกรอบของการอ้างอิง (Frame of reference: FOR) เป็นวิธีการฝึกอบรมที่มีหลายมิติของการประเมิน ทำให้ผู้ประเมินสามารถแลกเปลี่ยน และใช้กรอบแนวคิดเดียวกันในการประเมิน มีความเห็นความเข้าใจที่ตรงกันในเกณฑ์ ที่ใช้ในการประเมินสิ่งเดียวกัน ผู้ประเมินที่มีความคิดเห็นส่วนตัวที่ไม่สอดคล้องกับกลุ่มควรได้รับตรวจสอบ ดังนั้นผลการประเมินของแต่ละคนจะต้องถูกนำมาพิจารณาเพื่อปรับเปลี่ยนการรับรู้ที่ตรงกันให้มากที่สุด หากพบว่าผู้ประเมินคนใดมีการรับรู้ที่แตกต่างจากกลุ่ม ควรนำเข้าสู่กระบวนการของการฝึกอบรมเชิงปฏิบัติการเพื่อปรับการรับรู้ให้ตรงกับผู้ประเมินคนอื่น ๆ โดยมีขั้นตอนดังนี้

5.2.1 รับฟังคำอธิบายงาน และจากนั้นจึงอภิปรายถึงหน้าที่ และสมรรถนะของงานที่ต้องประเมินว่าต้องมีคุณลักษณะอย่างไรบ้าง

5.2.2 ผู้เข้ารับการฝึกอบรมจะได้รับงานการปฏิบัติที่มีระดับความสำคัญจำนวน 3 ชิ้นงาน ชิ้นงานสะท้อนให้เห็นผู้ถูกประเมินที่มีลักษณะโดดเด่น อีกชิ้นงานหนึ่งเป็นชิ้นงานที่มีคุณภาพปานกลาง และชิ้นงานที่ 3 เป็นชิ้นงานที่ต่ำกว่ามาตรฐาน แต่ละตัวอย่างของงานจะมีคำอธิบายสมรรถนะ ทำให้มีความชัดเจนของสมรรถนะของงานที่มีความโดดเด่น งานที่มีลักษณะกลางๆ และงานที่มีลักษณะต่ำกว่ามาตรฐานทั่วไปซึ่งจะอยู่ในรูปมาตรฐานประมาณค่าที่อิงเกณฑ์ที่ได้จากผู้เข้ารับการฝึกอบรม และจากผู้ฝึกผู้ประเมิน ผู้เข้ารับการอบรมแต่ละคนประเมินแต่ละงานโดยยึดตามเครื่องมือที่มีมาตรฐานวัด

5.2.3 จากนั้นผู้เข้าอบรมจะต้องตัดสินใจให้คะแนน ว่าอยู่บนมาตรฐานค่าระดับใด

5.2.4 ผู้ฝึกผู้ประเมินจะให้ข้อมูลย้อนกลับโดย แจ้งผลการประเมินที่ควรจะเป็นใน่งงานให้ผู้เข้ารับการฝึกอบรมทราบซึ่งจะเป็นข้อมูลแบบอิงกลุ่ม และบอกเหตุผลประกอบด้วยว่าผลการประเมินนั้นปกติ หรือแตกต่างจากกลุ่ม โดยการอภิปรายจะให้ความสำคัญที่ความแตกต่างระหว่างการประมาณค่าที่ถูกต้อง และการประมาณค่าที่แตกต่างจากเกณฑ์การประเมินที่ควรจะเป็น และการให้นำหนักความสำคัญของพฤติกรรมบางพฤติกรรมที่มีลักษณะเฉพาะ การฝึกอบรมมุ่งหวังให้ผู้เข้ารับการฝึกอบรม แยกแยะพฤติกรรมสำคัญที่ควรนำมาประกอบการตัดสินใจในการประเมิน กับพฤติกรรมที่ไม่มีความสำคัญต่อการประเมิน

5.2.5 จากนั้นเข้ากลุ่มอภิปรายผลการประเมิน นำเสนอมุมมองที่แตกต่างจากมาตรฐานวิธีการนี้ทำให้เกิดการเพิ่มขึ้นของความสอดคล้องระหว่างผู้ประเมินทั้งในด้านการประเมินงานเป็นชิ้นงาน และการประเมินตามสถานการณ์จริง

5.3 การฝึกอบรมความแปรปรวนของผู้ประเมิน (rater variability training: RVT) มีจุดประสงค์เพื่อเพิ่มความแม่นยำให้การประเมิน เป็นการผสมผสานระหว่างวิธี ฝึกอบรมแบบ กรอบของการอ้างอิงและการอบรมความคลาดเคลื่อนจากผู้ประเมินเชื่อว่าจะช่วยเพิ่มความแม่นยำในการประเมินได้

5.4 การฝึกอบรมมิติของผลงาน (Performance dimension training: PDT) เป็นการอบรมเพื่อสร้างความคุ้นเคยในมิติที่ต้องการประเมิน ถึงคุณลักษณะของแต่ละมิติของผลงาน

5.5 การฝึกอบรมการสังเกตพฤติกรรม (Behavioral observation training: BOT) เป็นการฝึกเทคนิคการสังเกต โดยให้ความสำคัญกับการรวบรวมข้อมูลจากการสังเกตเพื่อตัดสินประเมิน เมื่อผู้ประเมินทำการสังเกต และตัดสินใจให้คะแนนการประเมิน

5.6 การฝึกการตัดสินใจ Decision making training (DMT) เป็นวิธีที่ช่วยให้เข้าใจในธรรมชาติของกระบวนการในการตัดสินใจ หรือคาดการณ์ได้ถึงผลการประเมินที่จะได้ และสามารถหลีกเลี่ยงการตัดสินใจที่จะทำให้เกิดความลำเอียง หรือ ความคลาดเคลื่อนจากการประเมินได้ ทฤษฎีที่ใช้เป็นแนวคิดพื้นฐานในการประเมินภาคปฏิบัติที่เกี่ยวข้องกับการตัดสินใจในการประเมินภาคปฏิบัติหลายๆ ทฤษฎี นิยมใช้แนวคิดพื้นฐานมาจากทฤษฎีทางจิตวิทยา Social and cognitive psychology มีแนวคิดพื้นฐานมาจากทฤษฎีการเรียนรู้ของ Albert Bandura (Bandura, 1991) ซึ่งเชื่อว่าการเรียนรู้จะต้องอาศัยการสังเกต และจากการมีปฏิสัมพันธ์กับสิ่งแวดล้อม ในขณะที่เดียวกันการเรียนรู้ก็ได้รับอิทธิพลมาจากความคิด ความเชื่อ และการแปลความหมายของสิ่งแวดล้อมของผู้เรียน ซึ่งทฤษฎีของการตัดสินใจงานภาคปฏิบัติ (Theory of performance judgment) ที่มีพื้นฐานมาจากทฤษฎีการเรียนรู้ของ Bandura หลายทฤษฎี สามารถอธิบายในเชิงการประเมินภาคปฏิบัติ ซึ่งแสดงให้เห็นถึงธรรมชาติ หรือ สัจชาตญาณของการตัดสินใจ (Intuitive decision-making) กระบวนการตัดสินใจตามหลักการที่ถูกต้องควรประกอบด้วย (Neck & Barnard, 1996)

5.6.1 การสังเกต และเก็บรวบรวมข้อมูลกระบวนการ หรืองานที่ใช้ในการประเมินภาคปฏิบัติ โดยผู้ประเมินทุกคนจะต้องสังเกตพฤติกรรมสำคัญที่จำเป็นสำหรับงานภาคปฏิบัติ นั้น ๆ ให้ตรงกันซึ่งทราบได้จากเกณฑ์ หรือพฤติกรรมที่แบ่งแต่ละระดับคุณภาพ เพื่อให้เกิดความสอดคล้องกันในการประเมิน และเพื่อลดความคลาดเคลื่อนจากการที่ผู้ประเมินใช้เกณฑ์ต่างชนิดกันในการประเมิน

5.6.2 เลือกพฤติกรรมที่สำคัญ หรืองานสำคัญที่จะใช้ในการประเมิน คือการคัดกรองข้อมูลจากขั้นตอนการสังเกต นำข้อมูลที่สำคัญต่อการประเมินตามเกณฑ์ที่ตั้งไว้มาพิจารณา

5.6.3 จำแนกประเภทของสิ่งที่กำลังทำการตัดสินใจและประเมินคร่าว ๆ หรือการให้น้ำหนักกับพฤติกรรมสำคัญ หรืองานภาคปฏิบัติที่คัดกรองมานั้น และ ทำการประเมินระดับคุณภาพคร่าว ๆ ปัญหาจะเกิดขึ้นเมื่อผู้ประเมินให้น้ำหนักพฤติกรรมแตกต่างกัน ตัวอย่างเช่น ผู้ประเมินคนหนึ่งให้น้ำหนักพฤติกรรมที่เกิดขึ้นในเวลาใกล้เคียงกับเวลาที่สิ้นสุดการประเมิน (Recency error) ในขณะที่ผู้ประเมินอีกคนหนึ่งใช้ข้อมูลจากความประทับใจในครั้งแรกที่ได้รู้จักผู้ถูกประเมินมาให้น้ำหนัก (First impression error) ในขั้นตอนนี้ ผู้ประเมินทุกคนจะต้องรับรู้ในสิ่งเดียวกันว่าพฤติกรรมสำคัญหรืองานภาคปฏิบัติที่คัดกรองมานั้นจะต้องได้รับน้ำหนักเท่าๆ กัน หรือใกล้เคียงกันให้มากที่สุดโดยพิจารณาตลอดระยะเวลาของกระบวนการปฏิบัติ หรือ ชี้นงานภาคปฏิบัติ

5.6.4 การบูรณาการคือการนำข้อมูลจากการสังเกต คือ ความประทับใจในอดีตและข้อมูลที่เป็นปัจจุบันมาเทียบกับค่ามาตรฐานของงานภาคปฏิบัติที่ควรจะเป็น หากข้อมูลที่ได้มีความสอดคล้องกับมาตรฐานก็จะนำไปสู่การตัดสินใจ แต่หากข้อมูลที่ได้ในปัจจุบันขัดแย้งกับข้อมูลในอดีตผู้ประเมินจะพยายามปรับเปลี่ยนการรับรู้ต่อข้อมูลที่ได้ใหม่ให้สอดคล้องกับข้อมูลความประทับใจเดิม หากมีสถานการณ์แวดล้อมบีบบังคับให้ต้องปรับเปลี่ยน ก็จะมีการปรับเปลี่ยนการรับรู้เดิม หรือความประทับใจในอดีตตามข้อมูลใหม่ แล้วจึงนำไปเทียบกับมาตรฐานของงานเพื่อการตัดสินใจ ในการตัดสินใจควรใช้การควบคุมตนเอง (Self-regulation) เพื่อให้การประเมินมีความแม่นยำ และมียุติธรรมมากที่สุด และช่วยให้ใช้ระยะเวลาในการตัดสินใจเร็วขึ้น โดยใช้กระบวนการดังนี้ ขั้นตอนที่ 1 การพูดกับตัวเอง (Internal dialogue or self-talk) ขั้นตอน ที่ 2 การสร้างภาพในใจ (Mental images) และ ขั้นตอนที่ 3 การสร้างความเชื่อและการสร้างสมมติฐาน (Beliefs and assumption)

6. การศึกษาวิจัยที่เกี่ยวข้องกับโมดูลมาตรฐานที่ใช้ในการฝึกอบรม

โมดูลการฝึกอบรมผู้ประเมินประกอบด้วยโมดูลมาตรฐานที่นิยมใช้เช่น Rater error training (RET), Frame of reference training (FOR), behavioral observation training (BOT) และ decision making training (DMT) ซึ่งแต่ละวิธีผ่านการศึกษาถึงประสิทธิผลมาเป็นลำดับตั้งงานวิจัยต่อไปนี้

6.1 วิธีการอบรมเรื่องความคลาดเคลื่อนจากผู้ประเมิน (RET) วิธีนี้นิยมใช้ในการอบรมเพื่อลดความคลาดเคลื่อนจากการประเมินแบบปล่อยคะแนน และแบบฮาโล และเป็นวิธีฝึกอบรมที่ใช้มาตั้งแต่เริ่มมีการนำการเตรียมผู้ประเมินมาใช้ ผู้ริเริ่มคือ Latham, Wexley และ Pursell (1975) ได้ทำการศึกษาเพื่อเปรียบเทียบผลของการฝึกอบรมผู้ประเมิน 2 วิธี คือ แบบการอบรมเชิงปฏิบัติการ

ด้วยการชมวีดิทัศน์ และการอภิปรายกลุ่มในเรื่อง ความคลาดเคลื่อนจากผู้ประเมิน 4 ประเภทคือ contrast effect, halo effect, similarity และ first impressions ในผู้จัดการจำนวน 60 คน แบ่งตัวอย่างเป็น 3 กลุ่มคือ กลุ่มที่ได้รับการอบรมด้วยวิธีที่ 1 กลุ่มที่ 2 ได้รับการอบรมวิธีที่ 2 และกลุ่มควบคุมไม่ได้รับการอบรมใดๆ หลังสิ้นเสร็จการอบรม 6 เดือนจึงให้ตัวอย่างทั้ง 3 กลุ่มทำการประเมินการปฏิบัติงาน โดยใช้สถิติทดสอบที่ ผลการศึกษาพบว่ากลุ่มที่ได้รับการอบรมเชิงปฏิบัติการ ไม่พบความคลาดเคลื่อนจากผู้ประเมินที่มีนัยสำคัญทางสถิติ ในขณะที่กลุ่มอภิปรายกลุ่มพบความคลาดเคลื่อนแบบ first impression และกลุ่มควบคุมพบความคลาดเคลื่อนแบบ contrast effect, halo effect และ similarity ในขณะที่ Borman (1979) ทำการศึกษาผลของการฝึกอบรมระยะสั้น ต่อการลดความคลาดเคลื่อนแบบฮาโลเป็นการบรรยายลักษณะของความคลาดเคลื่อนแบบฮาโล มีระยะเวลา 5-6 นาที และนำเสนอตัวอย่างการให้คะแนนแบบที่มีความคลาดเคลื่อนแบบฮาโล ช่วยลดความคลาดเคลื่อนแบบฮาโลได้ และการศึกษาอื่น ๆ ที่มีการใช้ RET เป็นส่วนหนึ่งในการเตรียมผู้ประเมิน และให้ผลการศึกษาไปในทิศทางเดียวกัน (Pulakos, 1978; Pulakos, 1984; Thornton III & Zorich, 1980; Warmke, 1979; Weigle, 1998)

การศึกษาของ Woehr และ Huffcutt (1994) ทำการสังเคราะห์งานวิจัยที่เกี่ยวกับวิธีการฝึกอบรมแบบ RET, PDT, FOR และ BOT ต่อประสิทธิภาพในการลดความคลาดเคลื่อนจากผู้ประเมิน โดยทำการศึกษางานวิจัยที่มีคุณภาพดีตั้งแต่ปี ค.ศ. 1949 ถึง 1992 โดยคัดเลือกได้จำนวน 29 เรื่อง ผลการสังเคราะห์พบว่า RET เป็นวิธีการฝึกอบรมที่ถูกนำมาใช้มากที่สุด จำนวน 28 เรื่อง รองลงมาคือ FOR จำนวน 15 เรื่อง PDT 8 เรื่อง และ BOT 4 เรื่อง วิธีฝึกอบรมแบบ RET เพียงวิธีเดียวช่วยลดการเกิดความคลาดเคลื่อนแบบฮาโลด้วยขนาดอิทธิพลขนาดปานกลาง แต่มีผลต่อการลดความคลาดเคลื่อนแบบกต ปล่อยคะแนนได้น้อย และมีผลต่อการเพิ่มความแม่นยำจากการประเมิน ไม่ได้ลดความแม่นยำจากการประเมินตามที่เคยเข้าใจกัน

วิธี RET ยังเป็นวิธีมาตรฐานที่การฝึกอบรมผู้ประเมินจนถึงปัจจุบัน (Feldman et al., 2012) อย่างไรก็ตามปัจจุบันพบว่างานวิจัยน้อยมากที่ทำการศึกษาเรื่องผลของ RET ต่อการลดความคลาดเคลื่อนจากผู้ประเมิน ทั้งนี้อาจเนื่องจากการศึกษาจนเกิดองค์ความรู้ที่ชัดเจนแล้ว โดยสาขาที่มีการพัฒนารูปแบบการฝึกอบรมผู้ประเมินเพื่อเพิ่มคุณภาพการประเมินอย่างต่อเนื่อง และชัดเจน เช่น สาขาแพทยศาสตร์ (Holmboe & Hawkins, 2008 ; Kobak, Lipsitz, & F., 2003) มีรูปแบบการฝึกอบรมมาตรฐานที่นิยมใช้คือ RET, FOR, BOT และ PDT (Eppich et al., 2015; Holmboe & Hawkins, 2008)

6.2 Frame of reference เริ่มต้นจากการนำเสนอวิธีการฝึกอบรมโดย Bernardin และ Buckley (1981) ได้เขียนบทความเรื่อง “Strategies in Rater training” เนื้อหาเกี่ยวกับ กลยุทธ์ในการฝึกอบรมการประเมิน ว่าไม่มีหลักฐานพอที่จะมาสนับสนุนความเชื่อที่ว่า ผู้ประเมินที่ได้รับการฝึกอบรม (training rater) จะทำให้มีการเปลี่ยนแปลงให้มีความแม่นยำ หรือความตรงเพิ่มขึ้น การฝึกอบรมอาจช่วยได้ในบางสถานการณ์เฉพาะนั้น ๆ จึงนำเสนอวิธีการใหม่ที่ใช้ในการอบรมการประเมินซึ่งจะทำให้ได้ผลดีขึ้น คือ

- 1) กระบวนการจดบันทึก (diary-keeping) เพื่อเพิ่มทักษะในการสังเกต
- 2) กรอบของการอ้างอิง (frame of reference) เพื่อทำให้ผู้ประเมินเข้าใจเกณฑ์การประเมินตรงกัน และสอดคล้องกัน
- 3) การฝึกทักษะการประเมินเพื่อให้เกิดความชำนาญ

การศึกษาของ Athey และ McIntyre (1987) ศึกษาพบว่าวิธีการฝึกอบรมแบบ FOR เป็นวิธีที่จำได้ง่าย และความจำเกี่ยวกับวิธีการนี้ยังคงอยู่นานที่สุดเมื่อเทียบกับวิธีการฝึกอบรมอื่น และการศึกษาของนักวิชาการหลายท่านเพื่อยืนยันประสิทธิผลของ FOR ในการเพิ่มความแม่นยำในการประเมิน การสังเคราะห์งานวิจัยที่เกี่ยวกับผลของการฝึกอบรม 4 วิธีคือ RET, PDT, FOR และ BOT ของ Woehr และ Huffcutt (1994) ต่อประสิทธิภาพในการลดความคลาดเคลื่อนจากผู้ประเมินการศึกษของผู้ประเมิน ในส่วนของวิธีฝึกอบรมแบบ FOR เป็นวิธีการเพิ่มความแม่นยำในการประเมินได้มากที่สุด และ Woehr (1994) ได้ศึกษาประสิทธิผลของ FOR ที่ทำให้เกิดความแม่นยำในการประเมินว่าเกิดขึ้นเนื่องจากสามารถจำได้ง่ายกว่า (better recall for individual behavior) การศึกษาของ Noonan และ Sulsky (2001) ทำการเปรียบเทียบวิธีการฝึกอบรมแบบ FOR กับ BOT ในการเพิ่มความแม่นยำในการประเมิน โดยแบ่งตัวอย่างออกเป็น 4 กลุ่มคือ กลุ่มที่ฝึกอบรมด้วยวิธี FOR อย่างเดียว วิธีBOT อย่างเดียว วิธีทั้งสองวิธี และกลุ่มที่ไม่ได้รับการฝึกอบรมใดๆเลย ผลการศึกษาพบว่าวิธีการFOR วิธีเดียว และ วิธีFORร่วมกับวิธี BOT ต่างมีผลต่อความแม่นยำของการประเมินอย่างมีนัยสำคัญทางสถิติ และไม่พบความแตกต่างของวิธี FORวิธีเดียว กับวิธี FORร่วมกับวิธี BOT ต่อความแม่นยำในการประเมินที่ระดับนัยสำคัญ .05 เมื่อตรวจสอบจนเกิดความชัดเจนขององค์ความรู้เกี่ยวกับประโยชน์ของ FOR แล้ว ต่อมาการศึกษาวิจัยได้ศึกษาเชิงลึกถึงประสิทธิผลต่อการเพิ่มความแม่นยำในการประเมิน ในด้านอื่นๆ เช่น การศึกษาความยั่งยืน หรือคงทนของวิธีการนี้เทียบกับวิธีการอื่น เช่น การศึกษาของ Roch และ O'Sullivan (2003) เพื่อตรวจสอบสมรรถนะของFOR ว่าสามารถช่วยในการฟื้นความจำเกี่ยวกับพฤติกรรมที่สังเกตจากการประเมินได้หรือไม่ สามารถเพิ่มประสิทธิภาพในการเรียนรู้ในวิธีการฝึกอบรมแบบ BOT ได้ดีเพียงใด และผลของการฝึกอบรมสามารถคงอยู่ได้นานเพียงใดเมื่อเวลาผ่านไป

ไป โดยทำการศึกษาในตัวอย่างนักศึกษามหาวิทยาลัยจำนวน 64 ผลการศึกษาพบว่า ทั้ง FOR วิธีเดียว หรือ FOR ร่วมกับ BOT ช่วยเพิ่มความแม่นยำในการประเมินเมื่อเทียบกับกลุ่มควบคุม และช่วยในการทบทวนความจำเกี่ยวกับพฤติกรรมที่ผู้ประเมินควรสังเกต วิธีการฝึกอบรมแบบ BOT ร่วมกับ FOR ไม่ได้เพิ่มคุณภาพของความจำที่สามารถทบทวนได้

ตัวอย่างของการศึกษาสมรรถนะเชิงลึกของ FOR เช่น Gorman และ Rentsch (2009) ได้ศึกษาประสิทธิภาพของ FOR ต่อการสร้าง schema เนื่องจากแนวคิดที่ว่า Schema เป็นการเรียนรู้จากประสบการณ์เดิมนำไปใช้กับสถานการณ์ใหม่เพื่อสร้างความรู้ใหม่ จนเกิดความเชี่ยวชาญ ซึ่งต้องอาศัยการบูรณาการ schema อย่างเป็นระบบ การฝึกอบรม FOR เป็นส่วนหนึ่งในการสร้าง schema ตัวอย่างประกอบด้วยนักศึกษาระดับปริญญาตรีของมหาวิทยาลัย Public south eastern จำนวน 144 คนที่ลงทะเบียนเข้าร่วมโครงการวิจัยนี้ นักศึกษาได้รับแจ้งวัตถุประสงค์ของการศึกษา มีการทดสอบการวัดความสามารถในการเรียนรู้ด้วย schema ต่อด้วยการฝึกอบรม FOR ในกลุ่มทดลอง และวิธีการฝึกอบรมอื่น ๆ ในกลุ่มควบคุม และทำการวัดความสามารถในการเรียนรู้ หลังการฝึกอบรม ผลการศึกษาไม่พบความแตกต่างระหว่างสองกลุ่ม ในด้านอายุ เพศ เชื้อชาติ ผลการเรียนรู้ หรือความแตกต่างของคะแนนการวัดความสามารถในการสร้างschema ในกลุ่มทดลอง และกลุ่มควบคุม

Roch, Woehr, Mishra และ Kieszczynska (2012) ได้สังเคราะห์งานวิจัยเกี่ยวกับการฝึกอบรมผู้ประเมินในช่วงปี ค.ศ. 1996 ถึง 2008 จำนวน 90 เรื่อง และคัดเลือกงานวิจัยที่มีคุณภาพ หรือมี effect size ในระดับที่ยอมรับได้ตามเกณฑ์ที่ใช้ในการสังเคราะห์ จนได้งานวิจัยจำนวน 36 เรื่องพบว่าส่วนใหญ่มีวัตถุประสงค์ของการศึกษาเพื่อตรวจสอบประโยชน์ของ FOR ต่อคุณภาพการประเมิน ซึ่งสรุปได้ว่าทำให้เกิดความแม่นยำในการประเมินได้ชัดเจน แต่มีงานวิจัยน้อยมาก ที่ทำการศึกษาประสิทธิภาพของ FOR เมื่อช่วงเวลาผ่านไปหลังการฝึกอบรม จึงทำให้ไม่สามารถสรุปถึงประสิทธิผลนี้เมื่อเวลาผ่านไปได้ ส่วนใหญ่ศึกษาร่วมกับความคลาดเคลื่อนจากการประเมินอื่นๆ และพบว่า FOR ไม่ใช่ประเด็นที่เป็นเรื่องใหม่อีกต่อไป

การฝึกอบรมด้วยวิธีการฝึกอบรมเพื่อเตรียมความพร้อมของผู้ประเมินเพียงวิธีเดียวได้ถูกนำมาใช้ในการเพิ่มคุณภาพการประเมินอย่างแพร่หลาย อาจเนื่องจากประหยัดเวลา และได้ผลสัมฤทธิ์ชัดเจน ตัวอย่างเช่น การศึกษาของ George และคณะ (2013) ทำการฝึกอบรมผู้ประเมินคือ ศัลยแพทย์ประจำโรงเรียนแพทย์ ทั้งสิ้นจำนวน 44 คนแบ่งเป็นสองกลุ่ม อบรมด้วยวิธี FOR ระยะเวลาการฝึกอบรมกลุ่มแรก 1 ชั่วโมง และ กลุ่มที่สอง 4 ชั่วโมง เพื่อทำการประเมินแพทย์ประจำบ้าน (Resident)

ในทักษะด้านศัลยกรรมทั่วไป โดยการชมวิดีโอที่บันทึกไว้ขณะทำการผ่าตัดจริงจำนวน 10 ชุด และประเมินด้วยเครื่องมือประเมินแบบ มาตรฐานประมาณค่า 4 ระดับ ผลการวิจัยคือหลังการฝึกอบรม ผู้ประเมินมีร้อยละของความสอดคล้องระหว่างผู้ประเมิน (percentage agreement) ในระดับสูงคือ กลุ่มที่มีระยะเวลาในการฝึกอบรม 1 ชั่วโมงและ 4 ชั่วโมงคิดเป็นร้อยละ 88 และ 80 ตามลำดับ และระยะเวลาในการฝึกอบรมที่แตกต่างกันคือ 1 ชั่วโมง และ 4 ชั่วโมง ไม่มีผลต่อความสอดคล้องระหว่างผู้ประเมิน หรือการศึกษาของ Raczyski, Cohen, Engelhard และ Lu (2015) ศึกษาเปรียบเทียบประสิทธิผลของวิธี FOR 2 วิธีคือ อบรมแบบกลุ่ม (Classroom training) และอบรมแบบเรียนรู้ด้วยตนเองแบบออนไลน์ (Self-paced training) ต่อความแม่นยำในการประเมินในการประเมินขนาดใหญ่ของการประเมินงานเขียน ของนักเรียนเกรด 10 ตัวอย่างเป็นผู้ตรวจงานเขียนจำนวน 66 คนจากกลุ่มผู้ตรวจทั้งสิ้น 90 กว่าคน แบ่งผู้ตรวจเป็นสองกลุ่ม กลุ่มแรกทำการประเมินงานในลักษณะจัดอบรมเป็นกลุ่ม ณ ศูนย์การฝึกอบรม กลุ่มที่ 2 ทำการฝึกอบรมด้วยตนเองผ่านระบบออนไลน์ โดยเนื้อหาของการฝึกอบรมเหมือนกันทำการประเมินงานที่ผู้เชี่ยวชาญประเมินมาแล้วจำนวน 50 ชุด ผลการศึกษาพบว่าในกลุ่มที่ฝึกอบรมด้วยตนเองในระบบออนไลน์ ให้ความแม่นยำในการประเมินมากกว่ากลุ่มที่ฝึกอบรมเป็นกลุ่ม จึงสรุปได้ว่าวิธีการฝึกอบรมแบบ FOR ยังอยู่ในกระแสความสนใจในการศึกษาวิจัยอยู่ในปัจจุบัน

6.3 BOT (Behavioral observation training) มีตัวอย่างการวิจัย เช่น การศึกษาของ Woehr และ Huffcutt (1994) ที่ทำการสังเคราะห์งานวิจัยที่เกี่ยวกับวิธีการฝึกอบรมแบบ RET, PDT, FOR และ BOT ต่อประสิทธิผลในการลดความคลาดเคลื่อนจากผู้ประเมิน ตั้งแต่ปี ค.ศ. 1949 ถึง 1992 ที่ได้กล่าวไปแล้วนั้น ในส่วนของการฝึกอบรมแบบ BOT พบงานวิจัยที่มีคุณภาพ และสามารถคำนวณขนาดอิทธิพลได้ 4 เรื่อง และพบว่า BOT เป็นวิธีที่ใช้ในการศึกษาน้อยที่สุด อย่างไรก็ตามกลับมีขนาดอิทธิพลในระดับปานกลาง ถึงสูงต่อความแม่นยำในการประเมิน และต่อความแม่นยำในการสังเกต งานวิจัยหลังปี ค.ศ. 1992 เช่น การศึกษาของ Noonan และ Sulsky (2001) ทำการเปรียบเทียบวิธีการฝึกอบรมแบบ FOR กับ BOT ในการเพิ่มความแม่นยำในการประเมิน โดยแบ่งตัวอย่างออกเป็น 4 กลุ่มคือ กลุ่มที่ฝึกอบรมด้วยวิธี FOR อย่างเดียว วิธี BOT อย่างเดียว ทั้งสองวิธี และกลุ่มที่ไม่ได้รับการฝึกอบรมใดๆเลย ผลการศึกษาพบว่าวิธีการFOR วิธีเดียว และ วิธี FOR ร่วมกับวิธี BOT ต่างมีผลต่อความแม่นยำของการประเมินอย่างมีนัยสำคัญทางสถิติ และไม่พบความแตกต่างของวิธี FOR วิธีเดียว กับวิธี FORร่วมกับวิธี BOT ต่อความแม่นยำในการประเมินที่ระดับนัยสำคัญ .05

ต่อมา Holmboe (2008) ได้เสนอบทความทางวิชาการ เกี่ยวกับแนวปฏิบัติในการประเมินสมรรถนะ การปฏิบัติงานของนักศึกษาแพทย์ โดยได้กำหนดให้วิธี BOT เป็นหนึ่งในวิธีมาตรฐานที่ใช้ในการฝึกอบรม ผู้ประเมิน และการศึกษาของ Eppich และคณะ (2015) ซึ่งมีเป้าหมายเพื่อพัฒนาคุณภาพการประเมิน ภาควิชาของแพทย์ มีวัตถุประสงค์เพื่อ 1) พัฒนา โมดูลการฝึกอบรมผู้ประเมินเพื่อประเมินทักษะด้านการ ทำงานเป็นทีม และทักษะการติดต่อสื่อสาร 2) เพื่อศึกษาความเที่ยงระหว่าง ผู้ประเมิน โดยใช้วิธีทัศน สถานการณ์จำลอง ในการประเมินผู้ป่วยของแพทย์ ในขณะที่ฝึกทักษะการประเมิน ที่มีระดับคุณภาพ ของทักษะที่ระดับต่ำ ปานกลาง และสูง และ 3) เพื่อศึกษาบทบาทของความเที่ยงระหว่างผู้ประเมิน จากการศึกษา การสังเกตพฤติกรรมกรรมการประเมิน ผลการศึกษาพบว่า 1) การพัฒนาโมดูลการฝึกอบรมใช้แนวมาตรฐาน ของการประเมินที่มีความสำคัญระดับสูง (High- stakes assessments) ระยะเวลาทั้งสิ้น 1 วัน ผู้ประเมิน ประกอบด้วยผู้เชี่ยวชาญ 9 ท่าน มีภูมิหลังเป็นกุมารแพทย์ 5 ท่าน พยาบาลผู้เชี่ยวชาญด้านเดียวกัน 3 ท่าน และนักวิจัยที่เกี่ยวข้องกับด้านบริการสาธารณสุข อีก 1 ท่าน แบ่งช่วงการฝึกอบรมเป็น 3 ช่วงคือ ช่วงที่ 1 ทำการอบรมเรื่อง RET, PDT, FOR และ BOT เป็นระยะเวลา 4 ชั่วโมง ช่วงที่ 2 ผู้ประเมิน ทำการศึกษาคู่มือการฝึกอบรม ทำการสังเกตพฤติกรรมที่ต้องการประเมินจากวีดิทัศน์ ซึ่งเป็นการ ทดลองประเมิน จำนวน 2 ชุดโดยผู้ประเมินแต่ละคนแยกกันประเมินพฤติกรรมในวีดิทัศน์ และสามารถปรึกษาทีมงานเกี่ยวกับพฤติกรรมที่ต้องสังเกตในการประเมินทักษะได้ ใช้ระยะเวลาการ ฝึกอบรม 4 ชั่วโมง ช่วงที่ 3 ในเวลา 2 ชั่วโมง โดยนำผลการประเมินที่ทำให้ช่วงที่ 2 รวมถึงวีดิทัศน์ ทั้งหมด มาประเมินจริง รวมทั้งสิ้น 42 สถานการณ์ ทำการประเมิน 4 ช่วงเวลาคือ 1) ในการฝึกอบรม ช่วงที่ 2 และ 2) ในการฝึกอบรมช่วงที่ 3 3) 4 สัปดาห์หลังการฝึกอบรม สุ่มเลือกวีดิทัศน์ 8 สถานการณ์ ทำการประเมิน และ 4) 16 สัปดาห์ประเมินวีดิทัศน์ทั้งหมด 42 สถานการณ์ ผลการศึกษาพบว่าสามารถ พัฒนาโมดูลการฝึกอบรม และคู่มือการฝึกอบรม และศึกษาความเที่ยงระหว่างผู้ประเมินพบว่าหลังการ ฝึกอบรมอยู่ในระดับสูง และมีความยั่งยืนของสมรรถนะแม้เวลาผ่านไป 16 สัปดาห์หลังการฝึกอบรม

6.4 DMT (Decision making training) กระบวนการตัดสินใจประเมินเป็นขั้นตอนสุดท้ายที่ สำคัญในการกำหนดคะแนนการประเมินให้แก่กระบวนการ หรือผลงานนั้นๆ และในระยะแรก การฝึกอบรมการตัดสินใจของผู้ประเมินถูกนำมาใช้โดย Kahneman และ Tversky (1981, Tversky & Kahneman, 1974, 1978 cited in Hedge & Kavanagh, 1988) ใช้กลยุทธ์ด้านปัญญา (Cognitive strategy) เพื่อให้เกิดความแม่นยำในการประเมิน ตัวอย่างเช่น การศึกษาของ Hedge และ Kavanagh (1988) ทำการศึกษาผลของการฝึกอบรมด้วยวิธี RET, BOT และ DMT ต่อการลดความคลาดเคลื่อนจากการ

กต ปล่อยคะแนน ความคลาดเคลื่อนแบบฮาล์ และ การให้คะแนนในช่วงคะแนนแคบ ๆ (Range restriction) และความแม่นยำในการประเมิน ตัวอย่างจำนวน 52คน ถูกสุ่มเข้ากลุ่มทดลองจำนวน 4 กลุ่ม คือ กลุ่มที่อบรมด้วยวิธี RET จำนวน 14 คน กลุ่มอบรมด้วยวิธี BOT จำนวน 12 คน กลุ่มที่ 3 อบรมด้วยวิธี DMT จำนวน 13 คน และกลุ่มควบคุม 13 คน การฝึกอบรมใช้เวลาเฉลี่ยประมาณ 3 ชั่วโมงครึ่ง ทำการประเมินก่อนฝึกอบรม หลังการฝึกอบรมทันที และ 1 สัปดาห์หลังการฝึกอบรม ผลการศึกษาพบว่าผลการประเมินก่อน และหลังการประเมินมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ การฝึกอบรมวิธี BOT และ DMT ช่วยลดความคลาดเคลื่อนจากการ กต ปล่อยคะแนน ยกเว้นวิธี RET ที่พบว่าทำให้ความคลาดเคลื่อนแบบกต ปล่อยคะแนนลดลงเล็กน้อย แต่ไม่มีนัยสำคัญทางสถิติ วิธี RET และ BOT ช่วยลดความคลาดเคลื่อนแบบฮาล์ วิธี BOT ช่วยลดความคลาดเคลื่อนจากการให้คะแนนช่วงแคบ ๆ และลดความคลาดเคลื่อนแบบฮาล์ และ วิธี RET และ DMT ช่วยเพิ่มความแม่นยำในการประเมิน

สำหรับการศึกษาวิจัยที่เกี่ยวกับการฝึกอบรมในต่างประเทศในปัจจุบัน พบว่ามีการศึกษาถึงผลของวิธี FOR ในตัวอย่างเฉพาะ ที่ต้องการนำประโยชน์ของวิธีการฝึกอบรมนี้ไปใช้ประโยชน์ได้จริง ยังคงได้รับความสนใจจากนักวิจัย(Raczynski et al., 2015) และยังสามารถขยายขอบเขตการศึกษาไปสู่การศึกษาการฝึกอบรมในระบบออนไลน์ รวมไปถึงการเปรียบเทียบระหว่างผู้ประเมินที่เป็นคนจริง กับผู้ประเมินระบบอิเล็กทรอนิกส์ (e-rater) ซึ่งในปัจจุบันมีการนำ e-rater มาใช้งานจริงในการประเมินงานเขียนของการทดสอบทางภาษาอังกฤษ (Knoch et al., 2007; Wang & von Davier, 2014)

ในประเทศไทยการวิจัยเกี่ยวกับการฝึกอบรมผู้ประเมินเพื่อลดความคลาดเคลื่อนจากผู้ประเมิน พบได้น้อยมาก รวมถึงการศึกษาวิจัยที่เกี่ยวกับวิธีการฝึกอบรมผู้ประเมินในสาขาพยาบาลศาสตร์ด้วยเช่นกัน ดังนั้นการศึกษาผลของการฝึกอบรมต่อความคลาดเคลื่อนจากผู้ประเมินจึงเป็นสิ่งจำเป็น และ จากการศึกษาเอกสารงานวิจัยเกี่ยวกับผลของการฝึกอบรมด้วยวิธีต่าง ๆ จึงสรุปได้ว่าวิธีการเหล่านี้สามารถลดความคลาดเคลื่อนจากผู้ประเมิน การวิจัยในครั้งนี้จึงได้นำวิธีการฝึกอบรมในเรื่อง RET, FOR, BOT และ DMT มาใช้เป็นส่วนประกอบในโมดูลการฝึกอบรม เพื่อช่วยให้คุณภาพของการฝึกอบรมดีขึ้น (Wolfe & McVay, 2010)

ตอนที่ 6 กรอบแนวคิดของการวิจัย

จากการค้นคว้าเอกสาร แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้องกับมโนทัศน์การประเมิน การปฏิบัติงาน การประเมินภาคปฏิบัติ การพัฒนาโมดูล และ มโนทัศน์เกี่ยวกับการเสริมสร้างสมรรถนะการประเมิน สามารถนำมาประยุกต์ใช้ในการวิจัยครั้งนี้ดังนี้

1. การวิจัยครั้งนี้ได้บูรณาการ มโนทัศน์ด้านการประเมินการปฏิบัติงาน โดยการนำกระบวนการทัศน์ของการประเมินการปฏิบัติงาน ที่มีการออกแบบการประเมินให้สอดคล้องกับวัตถุประสงค์การเรียนรู้ การออกแบบเครื่องมือเช่น รุบริค ซึ่งเป็นเครื่องมือที่มีคุณภาพสูง เนื่องจากมีคำอธิบายระดับคุณภาพของงานอย่างละเอียด และนำเกณฑ์การประเมินตามเกณฑ์ของกรอบมาตรฐานคุณวุฒิระดับอุดมศึกษาแห่งชาติ (Thai Qualifications Framework for Higher Education, TQF:HEd) และ ด้านเอกลักษณ์ของมหาวิทยาลัยสวนดุสิต มาเป็นเกณฑ์ในการสร้างเครื่องมือ และการมอบหมายงานที่สอดคล้องกับธรรมชาติของการประเมินในวิชาชีพพยาบาล คือ วิดีทัศน์สถานการณ์จำลองการขึ้นฝึกปฏิบัติงานของนักศึกษาพยาบาล รวมถึงการตรวจสอบคุณภาพของเครื่องมือ ที่สำคัญคือการตรวจสอบความเที่ยงระหว่างผู้ประเมิน ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างผู้ประเมิน (Intra-class correlation coefficient) ร่วมกับดัชนีความสอดคล้องระหว่างผู้ประเมิน (Rater agreement index) และการตรวจสอบคุณภาพการประเมินภาคปฏิบัติ ที่สำคัญ คือ ความคลาดเคลื่อนจากผู้ประเมิน ประกอบด้วย ความคลาดเคลื่อนจากการกด ปุ่มยคะแนน จากการให้คะแนนเกาะกลุ่มตรงกลาง และความคลาดเคลื่อนแบบฮาโล ด้วยการวิเคราะห์ด้วยโมเดลหลายองค์ประกอบของราล์ซ

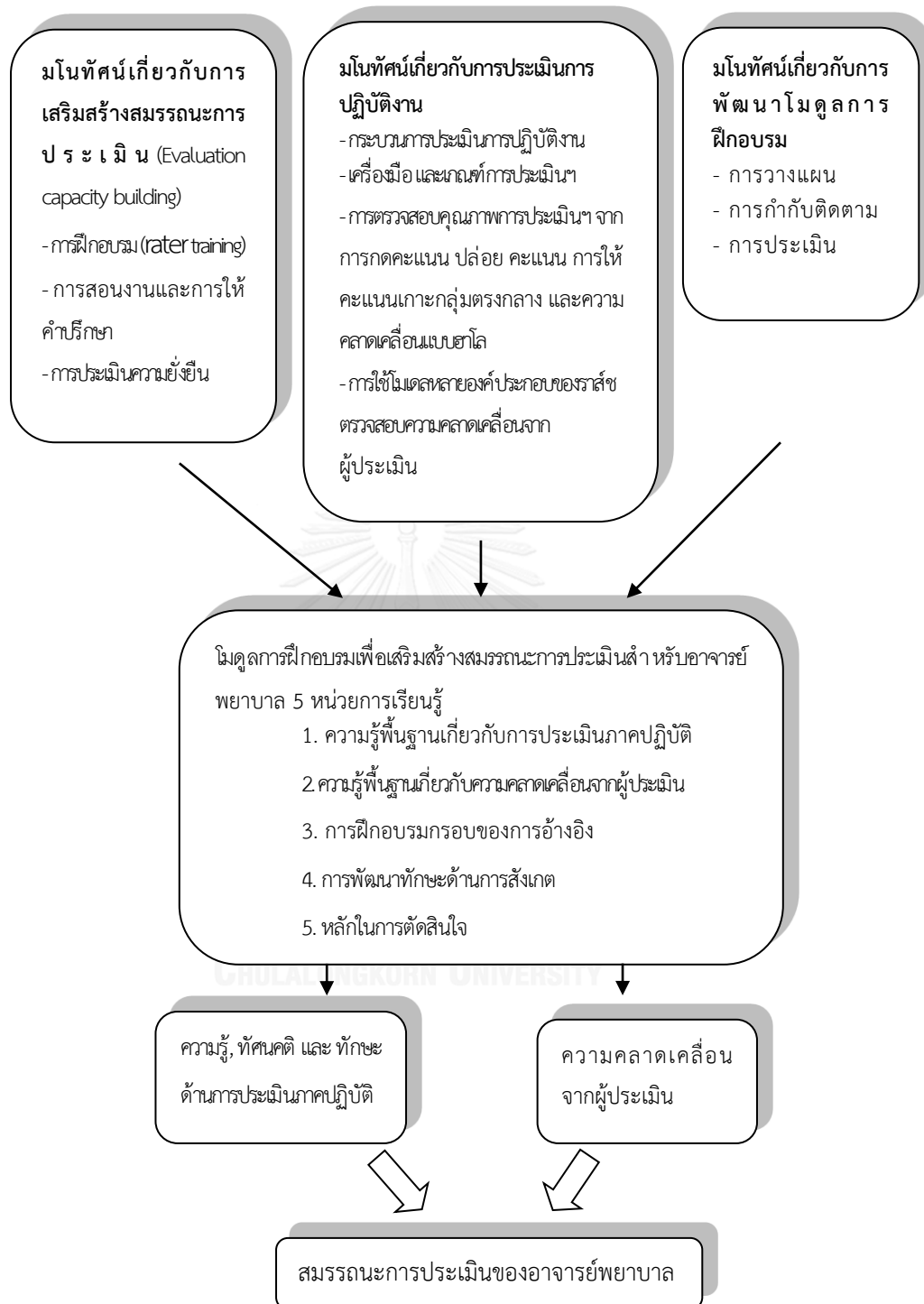
2. การนำมโนทัศน์ของการเสริมสร้างสมรรถนะการประเมินของ โมเดลพหุวิทยาการมาประยุกต์ใช้ โดยการมีส่วนร่วมของผู้มีส่วนได้ส่วนเสียคือ อาจารย์ และผู้บริหาร ของคณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต ในทุกระยะของการวิจัย นับตั้งแต่การร่วมแสดงความคิดเห็นในการพัฒนาโมดูลการฝึกอบรม ตลอดจนการพัฒนาเครื่องมือที่ใช้ประกอบในโมดูลการฝึกอบรม การมีส่วนร่วมในการใช้โมดูล มีส่วนร่วมในการประเมิน และหลังสิ้นสุดการวิจัย ได้วางแผนการมีส่วนร่วมในการใช้ผลการวิจัย ตามนโยบายของผู้บริหารฝ่ายวิชาการที่คาดหวังที่จะนำผลการวิจัยครั้งนี้ไปใช้ให้เกิดประโยชน์ต่อองค์กร

3. การพัฒนาโมดูลการฝึกอบรม ใช้แนวคิดในการพัฒนาโมดูลจากทฤษฎีการเรียนรู้ทางสังคม 3 ทฤษฎีหลักคือ ทฤษฎีการเรียนรู้ของ Bloom (Bloom's taxonomy of education objects) ทฤษฎีระดับของกระบวนการ (Levels-of-processing theory) และทฤษฎีการเกี่ยวพัน

ทางสังคม (Social facilitation theory) ซึ่งต้องการให้เกิดการเรียนรู้ใน 3 ด้านคือ ด้านพุทธิพิสัย จิตพิสัย และทักษะพิสัย โดยอาศัยกระบวนการของการสำรวจสภาพปัจจุบันของการประเมินภาคปฏิบัติ และนำข้อมูลเหล่านี้ร่วมกับ องค์ความรู้ที่เกี่ยวข้อง มาออกแบบวิธีการ และกระบวนการฝึกอบรม ออกแบบเนื้อหา กำกับติดตามผลของการฝึกอบรม ในที่นี้คือกำกับติดตามทักษะการประเมินของ อาจารย์พยาบาล รวมถึงการประเมินผลการใช้โมดูลการฝึกอบรม

4. ความคาดหวังต่อการใช้โมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมิน สำหรับ อาจารย์พยาบาล คือ การเพิ่มพูนสมรรถนะด้านความรู้เกี่ยวกับการประเมินภาคปฏิบัติ ทักษะด้านการประเมินภาคปฏิบัติ และคาดหวังว่าจะให้เกิดทัศนคติที่ดีต่อการประเมินภาคปฏิบัติ และสิ่งที่สะท้อนคุณภาพการประเมินภาคปฏิบัติเป็นอย่างดีอีกประการหนึ่งคือ การลดความคลาดเคลื่อนจากผู้ประเมิน ดังแสดงในภาพที่ 1





ภาพที่ 1 กรอบแนวคิดของการวิจัย

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ผู้วิจัยใช้ระเบียบการวิจัย และพัฒนา (Research and Development) มีวัตถุประสงค์เพื่อ 1) พัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล 2) เพื่อศึกษาผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล ต่อความรู้ ทักษะ และทักษะการประเมินภาคปฏิบัติ 3) เพื่อศึกษาผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลต่อความคลาดเคลื่อนจากผู้ประเมิน แบ่งขั้นตอนของการศึกษาออกเป็น 5 ขั้นตอนดังนี้

ขั้นตอนที่ 1 การสำรวจสภาพปัจจุบันของการประเมินภาคปฏิบัติ

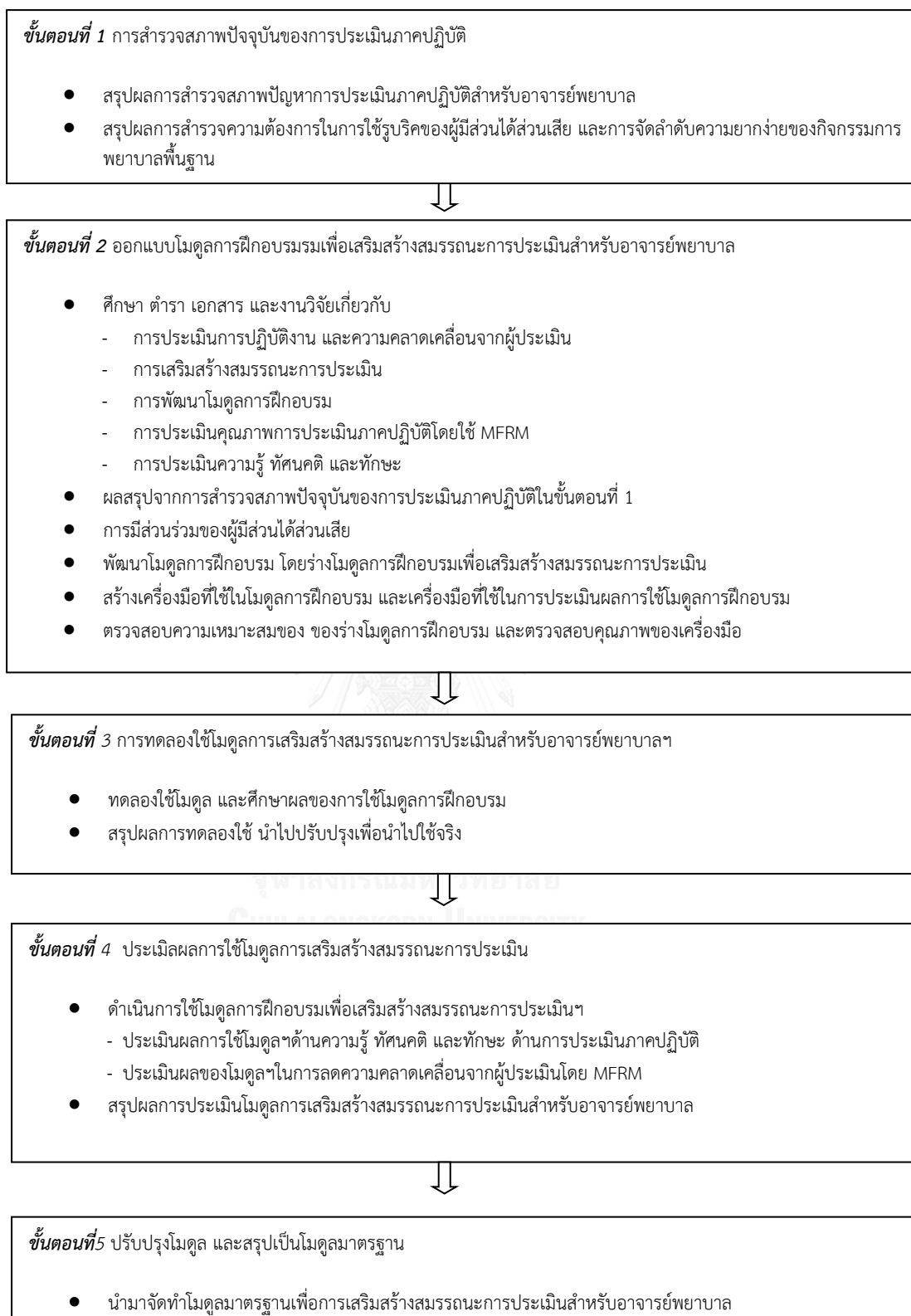
ขั้นตอนที่ 2 การพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับ
อาจารย์พยาบาล

ขั้นตอนที่ 3 ทดลองใช้โมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับ
อาจารย์พยาบาล

ขั้นตอนที่ 4 การประเมินผลการใช้โมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมิน
สำหรับอาจารย์พยาบาล

ขั้นตอนที่ 5 ปรับปรุงโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับ
อาจารย์พยาบาลในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล และสรุปเป็น
โมดูลมาตรฐาน

ลำดับของขั้นตอนทั้ง 5 และขั้นตอนย่อย นำเสนอในภาพที่ 2



ขั้นตอนที่ 1 การสำรวจสภาพปัจจุบันของการประเมินภาคปฏิบัติ

สำรวจสภาพปัจจุบันของการประเมินภาคปฏิบัติ 2 ส่วนคือ 1) สำรวจสภาพปัญหาการประเมินภาคปฏิบัติ ประกอบด้วย กลุ่มผู้ให้ข้อมูล 2 กลุ่มคือ ผู้ให้ข้อมูลกลุ่มที่ 1 อาจารย์พยาบาลผู้ทำหน้าที่สอนภาคปฏิบัติในนักศึกษาพยาบาลในระดับปริญญาตรี และ ผู้ให้ข้อมูลกลุ่มที่ 2 ผู้บริหารฝ่ายวิชาการ และ 2) สำรวจความต้องการในการใช้ระบบการประเมินภาคปฏิบัติ และการจัดลำดับความยากง่ายของกิจกรรมการพยาบาลพื้นฐาน ในผู้มีส่วนได้ส่วนเสีย คือผู้ทรงคุณวุฒิภายในคณะพยาบาลศาสตร์ รายละเอียด ประกอบด้วย ประชากร กลุ่มผู้ให้ข้อมูล เครื่องมือ ขั้นตอนการดำเนินงาน และการวิเคราะห์ข้อมูล ดังต่อไปนี้

1. สำรวจปัญหาการประเมินภาคปฏิบัติ

ประชากร ประกอบด้วยอาจารย์พยาบาลในสถาบันการศึกษาพยาบาลสังกัดสำนักงานคณะกรรมการการอุดมศึกษา 20 แห่ง ยกเว้นมหาวิทยาลัยสุโขทัยธรรมมาธิราช ซึ่งเป็นหลักสูตรต่อเนื่อง จำนวนอาจารย์พยาบาลทั้งหมด 1383 คน

กลุ่มผู้ให้ข้อมูลกลุ่มที่ 1 อาจารย์พยาบาลสังกัดสำนักงานคณะกรรมการ การอุดมศึกษา จากทั้ง 4 ภาคของประเทศไทย และอาจารย์พยาบาลสังกัดคณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต รวม 5 สถาบัน

ผู้วิจัยดำเนินการสุ่มตัวอย่างดังนี้

1) สุ่มอย่างง่าย เลือกสถาบันการศึกษาพยาบาล จากทั้ง 4 ภาคของประเทศไทย ได้สถาบันการศึกษาจำนวน 4 แห่ง เลือกแบบเจาะจง คณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต เป็นตัวอย่างในการสำรวจ เนื่องจากเป็นสถาบันที่จะนำโมดูลนี้ไปใช้ในการวิจัยขั้นตอนที่ 3 และขั้นตอนที่ 4

2) คำนวณขนาดตัวอย่างด้วยแนวคิดของ Cohen (1988 cited in Faul, Erdfelder, Lang, & Buchner, 2007) โดยใช้ โปรแกรม G* Power 3.1 เลือกการวิเคราะห์อำนาจการทดสอบก่อนการวิจัย (prior power analysis) เป็นการประมาณค่าอำนาจทดสอบจากขนาดตัวอย่างที่ใช้ในการวิจัย และ กำหนดค่าอำนาจการทดสอบตามที่ผู้วิจัยต้องการ และเป็นวิธีที่นักวิจัยส่วนใหญ่นิยมใช้ (นงลักษณ์ วิรัชชัย, 2555) กำหนดค่าพารามิเตอร์ดังนี้ 1) ขนาดอิทธิพลปานกลาง = .30, อำนาจการทดสอบ (power of test) $(1 - \beta) = .95$ และระดับนัยสำคัญ α ที่ .05 (Cohen, 1988) จำนวนตัวอย่างที่ได้คือ 96 คน

3) ในแต่ละสถาบันการศึกษาพยาบาล ผู้วิจัยใช้การสุ่มตัวอย่างแบบกำหนดโควตา (Quota sampling) เพื่อให้ได้ตัวอย่างที่กระจายในทุกสาขาวิชา รายละเอียดดังแสดงในตารางที่ 6

ตารางที่ 6 จำนวนผู้ให้ข้อมูลที่ต้องการจำแนกตามสถาบัน และสาขาวิชาที่รับผิดชอบ

คณะพยาบาล มหาวิทยาลัย	จำนวน	จำนวนผู้ให้ข้อมูลจำแนกตามสาขาวิชา					
		การพยาบาลมารดา ทารก และผดุงครรภ์	การพยาบาลเด็ก และวัยรุ่น	การพยาบาลผู้ใหญ่ และผู้สูงอายุ	การพยาบาล สุขภาพจิต และจิตเวช	การพยาบาลชุมชน และ การรณรงค์โรคเบื้องต้น	
เชียงใหม่	43	7	4	21	4	7	
ขอนแก่น	31	7	5	13	3	3	
ราชภัฏสวนสุนันทา	12	2	2	6	1	1	
ราชภัฏสุราษฎร์ธานี	4	1	1	2	0	0	
สวนดุสิต	11	2	2	5	1	1	
รวม	96	18	13	41	9	15	

เครื่องมือการวิจัย ได้แก่แบบสอบถามสภาพปัญหาของการประเมินภาคปฏิบัติ เป็นข้อความชนิดเลือกตอบหลายตัวเลือก จำนวน 20 ข้อ มีเนื้อหาเกี่ยวกับสภาพปัจจุบันของการประเมินภาคปฏิบัติในสถาบันการศึกษาพยาบาล ในตอนท้ายของแบบสอบถามเป็นคำถามปลายเปิด (Open-ended) เพื่อให้ตัวอย่างได้ให้ความคิดเห็นเพิ่มเติมเกี่ยวกับแนวทางการเตรียมความพร้อมของผู้ประเมิน และความต้องการในการเสริมสร้างสมรรถนะการประเมิน

การตรวจสอบคุณภาพเครื่องมือ ตรวจสอบความตรงเชิงเนื้อหา (content validity) โดยอาจารย์ที่ปรึกษา อาจารย์ที่ปรึกษาร่วม และผู้ทรงคุณวุฒิ 3 ท่าน (รายชื่อในภาคผนวก ก หน้า 256) คำนวณความตรงเชิงเนื้อหาโดยใช้ค่า Item Objective Congruence Index (IOC) มีค่า IOC เท่ากับ 0.60 - 1.00

การรวบรวมข้อมูล ผู้วิจัยทำหนังสือขอความอนุเคราะห์จากบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย และผู้วิจัยส่งแบบสอบถามให้ตัวอย่างด้วยตนเองตามวัน เวลา ที่นัดหมายและรวบรวมคืนโดยไปรษณีย์ตอบกลับ จำนวนแบบสอบถามที่ได้รับคืน 96 ฉบับ (ร้อยละ 86.48)

การวิเคราะห์ข้อมูล ข้อมูลเชิงปริมาณใช้การวิเคราะห์ความถี่ และร้อยละ ข้อมูลจากคำถามปลายเปิดใช้การวิเคราะห์เนื้อหา

กลุ่มผู้ให้ข้อมูลกลุ่มที่ 2 ผู้บริหารฝ่ายวิชาการจำนวน 5 ท่านจาก 5 สถาบัน

ให้ข้อมูลโดยการสัมภาษณ์เชิงลึกเกี่ยวกับสภาพปัจจุบันของการประเมินภาคปฏิบัติใน
นัศึกษพยาบาลระดับปริญญาตรี

เครื่องมือการวิจัย ได้แก่

1. ตัวผู้วิจัย

2. เทปบันทึกเสียง และแบบบันทึกข้อมูลส่วนบุคคล

3. แบบสัมภาษณ์แบบกึ่งโครงสร้าง สำหรับการสัมภาษณ์เชิงลึกโดยได้แนวคำถามจากแบบสอบถามสำรวจสภาพปัจจุบันของการประเมินภาคปฏิบัติ ใช้แนวคำถามหลัก คือ แนวคิดหลักที่นำมาวางแผนในการจัดการเรียนการสอนภาคปฏิบัติของสถาบันท่านเป็นอย่างไร มีเกณฑ์ในการสร้างเครื่องมือ การเตรียมความพร้อมของอาจารย์ในการฝึกภาคปฏิบัติอย่างไร ปัญหา หรืออุปสรรคในการจัดการเรียนการสอนภาคปฏิบัติ และ แนวทางแก้ไขเป็นอย่างไร

การตรวจสอบคุณภาพเครื่องมือ แนวคำถามจากแบบสอบถามสำรวจสภาพปัจจุบันของการประเมินภาคปฏิบัติที่ผ่านการตรวจสอบคุณภาพมาแล้ว นำไปทดลองสัมภาษณ์เพื่อตรวจสอบความเข้าใจ และความเหมาะสมของแนวคำถามกับผู้บริหารด้านที่เกี่ยวข้องกับงานวิชาการจำนวน 3 ท่าน หลังจากนั้นนำไปปรับปรุงแนวคำถามให้เหมาะสมก่อนนำไปสัมภาษณ์จริง

การรวบรวมข้อมูล ผู้วิจัยทำหนังสือขอความอนุเคราะห์จากบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย และผู้วิจัยไปเก็บข้อมูลด้วยตนเองตามวัน เวลา ที่นัดหมาย

การวิเคราะห์ข้อมูล

1. ข้อมูลเชิงปริมาณวิเคราะห์ด้วยสถิติเชิงบรรยาย ความถี่ ร้อยละ

2. ข้อมูลเชิงคุณภาพจากคำถามปลายเปิด และการสัมภาษณ์เชิงลึกวิเคราะห์ด้วยวิธีการวิเคราะห์เนื้อหา

2. สำนวจความต้องการในการใช้รุกรบริการประเมินภาคปฏิบัติ และการจัดระดับความยากง่ายของกิจกรรมการพยาบาล

ผู้ให้ข้อมูลเป็นผู้ทรงคุณวุฒิภายในคณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต จำนวน 7 ท่าน (ภาคผนวก ก หน้า 254)

เครื่องมือ

เครื่องมือในการสำวจความต้องการในการใช้รุกรบริการประเมินภาคปฏิบัติของผู้ทรงคุณวุฒิภายในคณะพยาบาลศาสตร์ฯ ประกอบด้วยแบบสอบถามสำวจความต้องการในการใช้รุกรบริการประเมิน สอบถามเกี่ยวกับจำนวนของระดับคุณภาพที่ควรเลือกใช้ และนิยามที่ใช้ และการจัดระดับความยากง่ายของกิจกรรมการพยาบาลพื้นฐานทั้งหมด

การตรวจสอบคุณภาพเครื่องมือ

แบบสอบถามความต้องการในการใช้รุกรบริการประเมินภาคปฏิบัติและตรวจสอบความตรงเชิงเนื้อหา (content validity) โดยอาจารย์ที่ปรึกษา อาจารย์ที่ปรึกษาร่วม และผู้ทรงคุณวุฒิ 3 ท่าน คำนวณความตรงเชิงเนื้อหาโดยใช้ค่า Item Objective Congruence Index (IOC) มีค่า IOC เท่ากับ 0.80 - 1.00

การรวบรวมข้อมูล

ผู้วิจัยทำหนังสือขอความอนุเคราะห์จากบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย และส่งแบบสอบถามให้กลุ่มตัวอย่างด้วยตนเองตามวัน เวลา ที่นัดหมาย

การวิเคราะห์ข้อมูล วิเคราะห์ความถี่ และร้อยละ

ขั้นตอนที่ 2 การพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล

การพัฒนาโมดูลการฝึกอบรมมีวัตถุประสงค์เพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลในการประเมินภาคปฏิบัตินักศึกษาพยาบาลในระดับปริญญาตรี ประกอบด้วยขั้นตอนย่อย 3 ขั้นตอนดังต่อไปนี้ 1) การสร้างชุดประเมินการฝึกภาคปฏิบัติด้วยรุกร 2) สร้างสื่อประกอบการฝึกอบรม 3) การออกแบบวิธีการ และกระบวนการฝึกอบรม

1. การสร้างชุดประเมินการฝึกภาคปฏิบัติ

ผู้วิจัยนำผลการวิจัยจากขั้นตอนที่ 1 ขั้นตอนย่อย 1.2 ความต้องการใช้รูบรีคในการประเมินภาคปฏิบัติ มาเป็นขอบเขตของเนื้อหาในการสร้าง ชุดประเมินการฝึกภาคปฏิบัติด้วยรูบรีค โดยบูรณาการแนวคิดที่เกี่ยวข้องได้แก่ 1) ผลลัพธ์การเรียนรู้ตามกรอบมาตรฐานคุณวุฒิระดับอุดมศึกษา แห่งชาติ (Thai Qualifications Framework for Higher Education, TQF: HEd) 2) เอกลักษณ์ของมหาวิทยาลัยสวนดุสิต 3) การประเมินในคลินิกสำหรับพยาบาล (Lasater, 2011; Tanner, 2006) 4) ทฤษฎีการเรียนรู้ของบลูม (Bloom's taxonomy) 5) การสร้างรูบรีคประเภท analytic rubric (Reddy, 2011) และคำอธิบายเกณฑ์ในการประเมินที่จำแนกคุณภาพการปฏิบัติเป็นระดับ (performance level descriptors) มีขั้นตอนการสร้างชุดประเมิน ดังนี้

1.1 การสร้างรูบรีคการประเมินการฝึกภาคปฏิบัติ เพื่อให้ได้รูบรีคฉบับสมบูรณ์ มีขั้นตอนในการสร้าง และตรวจสอบคุณภาพเครื่องมือ ดังต่อไปนี้

1.1.1 ร่างคำอธิบายนิยาม และเกณฑ์ในการประเมินที่จำแนกคุณภาพการปฏิบัติเป็นระดับจำนวน 3 ชุดคือ (1) คำอธิบาย และเกณฑ์การประเมิน 5 ระดับ (2) คำอธิบาย และเกณฑ์การประเมิน 4 ระดับที่มีการตัดระดับคุณภาพสูงสุดทิ้ง (3) คำอธิบาย และเกณฑ์การประเมิน 4 ระดับที่มีการตัดระดับต่ำสุดทิ้ง ตามคำแนะนำของผู้ทรงคุณวุฒิ 7 ท่าน (รายละเอียดในภาคผนวก ก หน้า 254)

1.1.2 นำร่างทั้ง 3 ไปขอความคิดเห็นเพื่อหาข้อสรุปในการสร้างรูบรีคฉบับสมบูรณ์จากผู้ทรงคุณวุฒิ 7 ท่านข้างต้น (รายละเอียดในภาคผนวก ก หน้า 254)

1.1.3 ผลการสำรวจความต้องการในการใช้รูบรีคการประเมินคือ ต้องการใช้รูบรีค 4 ระดับคุณภาพที่ตัดระดับคุณภาพสูงสุดทิ้ง เพื่อให้สอดคล้องกับสมรรถนะของนักศึกษาพยาบาลชั้นปีที่ 2 นำไปสร้างสมรรถนะย่อย และคำอธิบายระดับคุณภาพของแต่ละสมรรถนะจนได้รูบรีคฉบับสมบูรณ์ที่ประกอบด้วย 7 สมรรถนะ 26 สมรรถนะย่อย (รายละเอียดในภาคผนวก ข หน้า 268)

1.1.4 ตรวจสอบความตรงเชิงเนื้อหาโดยอาจารย์ที่ปรึกษา และอาจารย์ที่ปรึกษาร่วม และให้ผู้ทรงคุณวุฒิจำนวน 9 ท่าน (รายชื่อในภาคผนวก ก หน้า 255)

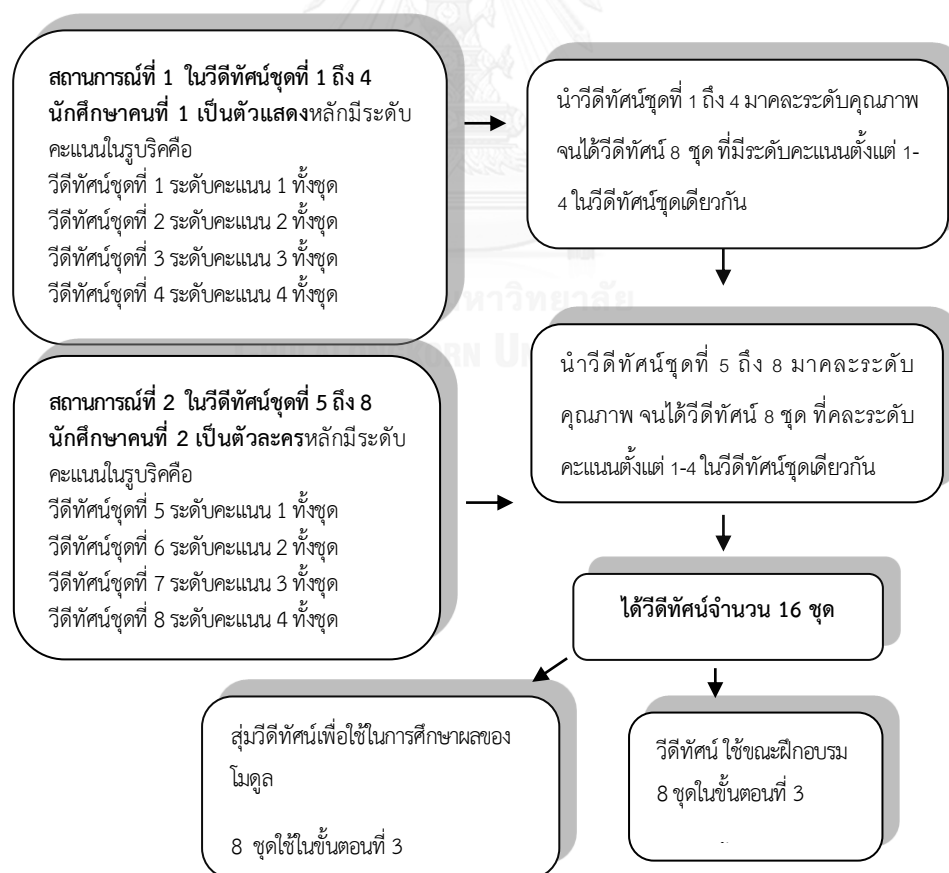
1.1.5 ตรวจสอบความเที่ยงระหว่างผู้ประเมินร่วมกับวิธีทัศนสถานการณ์จำลองจำนวน 8 ชุด ด้วยค่า Intra-class Correlation Coefficient (ICC) และค่า Rater agreement index (RAI)

1.2 สื่อวีดิทัศน์ประกอบการประเมินด้วยรูบริควีดิทัศน์ มีขั้นตอนในการสร้าง และตรวจสอบคุณภาพ คือ

1.2.1 ร่างสถานการณ์จำลองการขึ้นฝึกปฏิบัติงานของนักศึกษาพยาบาลชั้นปีที่ 2 รายวิชาปฏิบัติการพยาบาลพื้นฐาน โดยผ่านการตรวจสอบความตรงเชิงเนื้อหาโดยผู้ทรงคุณวุฒิ 3 ใน 7 ท่าน (รายชื่อในภาคผนวก ก หน้า 254)

1.2.2 ถ่ายทำสถานการณ์จำลอง 2 สถานการณ์ แต่ละสถานการณ์มีตัวแสดงหลักสถานการณ์ละ 1 คน จำนวนสถานการณ์ละ 4 ชุด รวม 8 ชุด

1.2.3 สถานการณ์ทั้ง 2 มีระดับคะแนนสอดคล้องกับรูบริกการประเมินภาคปฏิบัติที่ระดับ 1 ถึง ระดับ 4 นำมาคละระดับคุณภาพจนได้วีดิทัศน์จำนวน 16 ชุด จากนั้นทำการสุ่มเลือกวีดิทัศน์ 8 ชุดสำหรับใช้ในชั้นทดลองใช้โมดูลการฝึกอบรวมฯ และ 10 ชุดสำหรับการใช้โมดูลการฝึกอบรวมฯจริง ดังแสดงในภาพที่ 3



ภาพที่ 3 ขั้นตอนการสุ่มเลือกชุดวีดิทัศน์

การตรวจสอบคุณภาพเครื่องมือ

1. รูปรีการประเมินภาคปฏิบัติ ตรวจสอบความตรงเชิงเนื้อหาโดยผู้ทรงคุณวุฒิ 9 ท่าน (รายชื่อในภาคผนวก ก หน้า 255) ได้ค่า IOC เท่ากับ 0.67 – 1.00

2. ชุดประเมินการฝึกภาคปฏิบัติด้วยรูปรีประกอบด้วย 2 ส่วนคือ รูปรีการประเมินภาคปฏิบัติ และวิธีทัศนสถานการณ์จำลอง ตรวจสอบความเที่ยงระหว่าง ผู้ประเมิน (Inter rater reliability) โดยผู้ทรงคุณวุฒิ 5 ท่าน (รายชื่อในภาคผนวก ก หน้า 257) ได้ค่า Intra-class Correlation Coefficient (ICC) เท่ากับ 0.45 – 0.80 และค่า Rater agreement index (RAI) เท่ากับ 0.84 - 0.92

2. **สร้างสื่อประกอบการฝึกอบรม** ในรูปคู่มือ และเนื้อหาในหน่วยการเรียนรู้ โดยนำข้อมูลจากการสำรวจสภาพปัญหาการประเมินภาคปฏิบัติในข้อคำถามย่อย ความต้องการในการฝึกอบรม และวิธีที่ต้องการใช้ในการฝึกอบรม และการบูรณาการองค์ความรู้จากการทบทวนวรรณกรรม เกี่ยวกับโมดูลมาตรฐานที่ใช้ในการฝึกอบรม (Bernadin, 1978; Borman, 1979; Holmboe & Hawkins, 2008 ; Kobak et al., 2003; Latham, 1975; Elaine D Pulakos, 1984; Thornton III & Zorich, 1980; Warmke, 1979) มีขั้นตอนในการสร้าง และตรวจสอบคุณภาพเครื่องมือ ดังนี้

2.1 ร่างลำดับของเนื้อหาตามลักษณะที่ดีของคู่มือดัดแปลงจากการพัฒนาคู่มือของ นันทนัช อ่อนพวน (2553)

2.2 ร่างเนื้อหาตามหน่วยการเรียนรู้ 5 หน่วยคือ 1) ความรู้พื้นฐานด้านการประเมินภาคปฏิบัติ 2) ความรู้พื้นฐานเกี่ยวกับความคลาดเคลื่อนจากผู้ประเมิน 3) กรอบของการอ้างอิง 4) การฝึกการสังเกตพฤติกรรม และ 5) การฝึกการตัดสินใจ

2.3 ตรวจสอบความตรงเชิงเนื้อหาโดย อาจารย์ที่ปรึกษา อาจารย์ที่ปรึกษาร่วม และผู้ทรงคุณวุฒิ 9 ท่าน (รายชื่อในภาคผนวก ก หน้า 255)

3. การออกแบบวิธีการ และกระบวนการฝึกอบรม

การออกแบบวิธีการ และกระบวนการฝึกอบรม โดยนำข้อมูลที่ได้จากขั้นตอนที่ 1 การสำรวจสภาพปัญหาการประเมินภาคปฏิบัติที่เกี่ยวกับวิธีที่เลือกใช้ในการฝึกอบรม มาออกแบบ การเรียนรู้ ร่วมกับการบูรณาการแนวคิดของโมดูลมาตรฐานที่ใช้ในการเตรียมผู้ประเมิน (Bernadin, 1978; Borman, 1979; Holmboe & Hawkins, 2008 ; Kobak et al., 2003; Latham, 1975; Elaine D Pulakos, 1984; Thornton III & Zorich, 1980; Warmke, 1979) ได้เนื้อหา 5 หน่วยการเรียนรู้ดังนี้

3.1 ความรู้พื้นฐานด้านการประเมินภาคปฏิบัติ ประกอบด้วยความหมายของการประเมินภาคปฏิบัติ กระบวนการประเมินภาคปฏิบัติ ที่เริ่มต้นตั้งแต่การออกแบบการประเมิน เลือกรงานการประเมินภาคปฏิบัติให้ชัดเจน ออกแบบกิจกรรมสำหรับใช้ประเมิน กำหนดวิธีวัดทักษะภาคปฏิบัติ กำหนดเครื่องมือที่ใช้ในการวัดทักษะภาคปฏิบัติ และกำหนดวิธีการประเมินผล และรายงานผล โดยมีกระบวนการในการฝึกอบรมคือ การบรรยายเนื้อหา และทำแบบวัดความรู้ท้ายบท พร้อมการเฉลยคำตอบ

3.2 ความรู้พื้นฐานเกี่ยวกับความคลาดเคลื่อนจากผู้ประเมิน (Rater error training) มีเนื้อหาประกอบด้วยความหมาย สาเหตุ และการแก้ไขปัญหา ความคลาดเคลื่อนจากผู้ประเมิน ที่ประกอบด้วยความคลาดเคลื่อนแบบฮาล ความคลาดเคลื่อนจากการกด ปล่อยคะแนน และความคลาดเคลื่อนจากการให้คะแนน เกาะกลุ่มตรงกลาง และความคลาดเคลื่อนอื่น ๆ กระบวนการในการฝึกอบรมคือ การบรรยายเนื้อหา ที่มีตัวอย่างของข้อมูลที่แสดงถึง ความคลาดเคลื่อนประเภทต่าง ๆ และการทำกิจกรรมท้ายบท เพื่อฝึกการวิเคราะห์หาข้อมูลเบื้องต้นว่ามีความคลาดเคลื่อนแบบใด จากนั้นจึงอภิปรายภายในกลุ่ม

3.3 กรอบของการอ้างอิง (Frame of reference training) เป็นการกล่าวถึงระดับคุณภาพของงานภาคปฏิบัติ และตัวอย่างของงานภาคปฏิบัติ ซึ่งจะช่วยให้ผู้ประเมินมีความเข้าใจที่ตรงกันในระดับคุณภาพของงานประเมินเดียวกัน มีกระบวนการฝึกอบรมคือ การบรรยาย และนำเสนอรูปกริดที่ใช้ในการประเมินภาคปฏิบัติโดยอภิปรายร่วมกันถึงนิยาม และระดับคุณภาพของรูปกริดในแต่ละสมรรถนะ จากนั้นนำเสนอวิธีทัศนตัวอย่างสถานการณ์ที่มีระดับคะแนนการประเมินตั้งแต่ระดับ 1 ถึง 4 และอภิปรายร่วมกันเพื่อสร้างความเข้าใจให้ตรงกันระหว่างผู้ประเมินก่อนการประเมินงานภาคปฏิบัติจริง

3.4 การฝึกการสังเกตพฤติกรรม (Behavioral observation training) ประกอบด้วยการฝึกทักษะการสังเกตพฤติกรรม ซึ่งจะช่วยให้เข้าใจในเทคนิคการสังเกต และนำมาใช้ประกอบการตัดสินใจคุณภาพของงานภาคปฏิบัติ มีกระบวนการฝึกอบรมคือ การบรรยาย และนำเสนอวิธีทัศนเพื่อให้ผู้เข้ารับการฝึกอบรมได้ฝึกการสังเกต และอภิปรายร่วมกันในกลุ่ม

3.5 การฝึกการตัดสินใจ (Decision making training) เป็นเนื้อหาที่ว่าด้วยกฎของการตัดสินใจเพื่อใช้ในการกำกับตนเอง มีกระบวนการฝึกอบรมคือ การบรรยาย และนำเสนอวิธีทัศนเพื่อให้ผู้เข้ารับการฝึกอบรมได้ฝึกการสังเกต และอภิปรายร่วมกันในกลุ่ม

ขั้นตอนที่ 3 ทดลองใช้โมเดลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล

การทดลองใช้โมเดลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินที่ออกแบบไว้ในขั้นตอนที่ 2 มีระเบียบวิธีวิจัยดังต่อไปนี้

ตัวอย่าง ได้แก่อาจารย์คณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต กำหนดขนาดตัวอย่างโดยใช้หลักการ 2 ประการ คือ 1) หลักสถิติทดสอบที และ 2) หลักการฝึกอบรมของAthey และ McIntyre (1987) ที่กล่าวถึงขนาดของกลุ่มที่มีอิทธิพลต่อการเรียนรู้ (Bridge & Sawilowsky, 1999; Campbell, Julious, & Altman, 1995; Fritz, Morris, & Richler, 2012) ดังนั้น ขนาดตัวอย่างในขั้นตอนนี้ จำนวน 6 คน คัดเลือกโดยการเจาะจง และแบ่งกลุ่มควบคุม และกลุ่มทดลองด้วยการเลือกแบบจับคู่ (match) กำหนดเกณฑ์การคัดเลือก (Inclusive criteria) คือ เป็นอาจารย์ที่มีประสบการณ์การสอนในสาขาวิชาเดียวกันหรือใกล้เคียงกัน และมีความพร้อมในการทดลองใช้โมเดลการฝึกอบรมจนเสร็จสิ้นกระบวนการ สุ่มอย่างง่ายที่ละคู่ แบ่งเข้ากลุ่มทดลองใช้โมเดล 1 คน และเข้ากลุ่มใช้โมเดลจริง จำนวน 1 คน จนครบ 6 คู่ ได้ตัวอย่างกลุ่มละ 6 คน

เครื่องมือการวิจัย ได้แก่ เครื่องมือประเมินผลการใช้โมเดลการฝึกอบรม จำนวน 5 ชุด ประกอบด้วย 1) แบบวัดความรู้เกี่ยวกับการประเมินภาคปฏิบัติ 2) แบบวัดทัศนคติต่อการประเมินภาคปฏิบัติ 3) ชุดประเมินการฝึกภาคปฏิบัติ ประกอบด้วยรูบริค ร่วมกับวีดิทัศน์สถานการณ์จำลอง 4) ชุดแบบประเมินความพึงพอใจในการฝึกอบรมประกอบด้วย แบบวัดความพึงพอใจ และ แบบสัมภาษณ์แบบมีโครงสร้างในการสัมภาษณ์เฉพาะกลุ่ม และ 5) คู่มือประกอบการฝึกอบรม มีรายละเอียดดังนี้

1. แบบวัดความรู้เกี่ยวกับการประเมินภาคปฏิบัติ ก่อน และหลังการเข้ารับการฝึกอบรม ผู้วิจัยสร้างแบบวัดความรู้ โดยสร้างข้อคำถามโดยให้ระดับการวัด และจำนวนข้อของข้อสอบสอดคล้องกับมวลเนื้อหา และวิธีการให้ความรู้ ดังแสดงในตารางที่ 7

ตารางที่ 7 ผังข้อสอบในการสร้างแบบวัดความรู้เรื่องการประเมินภาคปฏิบัติ

หน่วยการเรียนรู้	ระดับการวัด		จำนวนข้อ
	รู้จำ	เข้าใจ	
1. ความรู้พื้นฐานด้านการประเมินภาคปฏิบัติ	4	5	9
2. Rater error training*	8	1	9
3. Frame of reference training*	2	1	3
4. Behavioral observation training*	2		2
5. Decision making training*	2		2
รวม	18	7	25

* เป็นหน่วยที่มีการฝึกปฏิบัติ

2. แบบวัดทัศนคติต่อการประเมินภาคปฏิบัติ ก่อน และหลังการเข้ารับการฝึกอบรม ผู้วิจัยสร้างข้อคำถามจำนวน 10 ข้อ ลักษณะคำตอบเป็นมาตรวัดแบบ Likert scale มีเกณฑ์การแปลความหมายค่าเฉลี่ยดังนี้

ค่าเฉลี่ย	ความหมาย
4.50 - 5.00	มีทัศนคติที่ดีอยู่ในระดับมากที่สุด
3.50 - 4.49	มีทัศนคติที่ดีอยู่ในระดับมาก
2.50 - 3.49	มีทัศนคติที่ดีอยู่ในระดับปานกลาง
1.50 - 2.49	มีทัศนคติที่ดีอยู่ในระดับน้อย
1.00 - 1.49	มีทัศนคติที่ดีอยู่ในระดับน้อยที่สุด

3. ชุดประเมินการฝึกภาคปฏิบัติ ประกอบด้วยรูปประกอบการประเมินภาคปฏิบัติจำนวน 7 สมรรถนะหลัก รวมทั้งสิ้น 26 สมรรถนะย่อย และวิดีโอทัศนจำนวน 8 ชุด เฉลี่ยระยะเวลาชุดละ 30 นาที ที่ผ่านการสุ่มอย่างง่ายจากขั้นตอนที่ 2 (รายละเอียดภาพที่ 3 หน้า 110)

4. ชุดประเมินความพึงพอใจในการฝึกอบรมประกอบด้วย 1) แบบวัดความพึงพอใจต่อการฝึกอบรม และ 2) แบบสัมภาษณ์เฉพาะกลุ่ม แบบมีโครงสร้าง มีรายละเอียดดังนี้

4.1 ผู้วิจัยสร้างข้อคำถามจำนวน 10 ข้อ ลักษณะคำตอบเป็นมาตรวัดแบบ Likert scale มีเกณฑ์ในการแปลความหมายดังนี้

ค่าเฉลี่ย	ความหมาย
4.50 - 5.00	มีความพึงพอใจต่อเรื่องนั้นอยู่ในระดับมากที่สุด
3.50 - 4.49	มีความพึงพอใจต่อเรื่องนั้นอยู่ในระดับมาก
2.50 - 3.49	มีความพึงพอใจต่อเรื่องนั้นอยู่ในระดับปานกลาง
1.50 - 2.49	มีความพึงพอใจต่อเรื่องนั้นอยู่ในระดับน้อย
1.00 - 1.49	มีความพึงพอใจต่อเรื่องนั้นอยู่ในระดับน้อยที่สุด

4.2 แบบสัมภาษณ์แบบมีโครงสร้าง (structured interview) ใช้ในการสัมภาษณ์เฉพาะกลุ่มเพื่อประเมินความพึงพอใจในการฝึกอบรม ประกอบด้วยข้อคำถามดังนี้ 1) ความรู้ที่ได้รับมีประโยชน์กับท่านอย่างไร 2) รู้สึกอย่างไรต่อการประเมินภาคปฏิบัติเมื่อเปรียบเทียบกับระหว่างก่อน และหลังการฝึกอบรม 3) การฝึกอบรมมีผลต่อทักษะการประเมินหรือไม่อย่างไร และ 4) จุดเด่นในการฝึกอบรม และจุดที่ควรปรับปรุง

5. คู่มือประกอบการฝึกอบรม ที่ผ่านการสร้าง และตรวจสอบคุณภาพเสร็จสิ้นแล้วในขั้นตอนที่ 2

การตรวจสอบคุณภาพเครื่องมือ

1. แบบวัดความรู้เกี่ยวกับการประเมินภาคปฏิบัติ ตรวจสอบโดยผู้ทรงคุณวุฒิ 9 ท่าน (รายชื่อภาคผนวก ก หน้า 255) ได้ค่าความตรงเชิงเนื้อหา (IOC) เท่ากับ 0.54 – 1.00 ระดับความยากง่ายของข้อสอบ (p) (Level of difficulty of the items) เท่ากับ 0.31 -0.85 อำนาจจำแนกของข้อสอบ (r) (Discrimination power of the items) เท่ากับ 0.20 – 0.54 และค่าสัมประสิทธิ์ความเที่ยงของ ครอนบาค แอลฟา (Cronbach's Alpha Coefficient: α) เท่ากับ 0.72

2. แบบวัดทัศนคติต่อการประเมินภาคปฏิบัติโดยผู้ทรงคุณวุฒิจำนวน 9 ท่าน (รายชื่อในภาคผนวก ก หน้า 255) ได้ค่าความตรงเชิงเนื้อหา (IOC) เท่ากับ 0.67 – 1.00 ตรวจสอบรายข้อด้วยค่า item total correlation ค่าเท่ากับ 0.30 – 0.82 ตรวจสอบความเที่ยงของแบบวัดทัศนคติด้วยสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Alpha Coefficient: α) ได้เท่ากับ 0.83

3. ชุดฝึกประเมินทักษะภาคปฏิบัติผ่านการตรวจสอบคุณภาพแล้วในขั้นตอนที่ 2

4. ชุดประเมินความพึงพอใจในการฝึกอบรม แบ่งเป็น 2 ส่วนคือ

4.1 แบบวัดความพึงพอใจ ผ่านการตรวจสอบความตรงเชิงเนื้อหา โดยอาจารย์ที่ปรึกษา อาจารย์ที่ปรึกษาร่วม และผู้ทรงคุณวุฒิจำนวน 9 ท่าน (รายชื่อในภาคผนวก ก หน้า 255) ด้วยค่า Item Objective Congruence Index (IOC) IOC เท่ากับ 0.89 – 1.00

4.2 แบบสัมภาษณ์กลุ่มเฉพาะ ใช้แบบมีโครงสร้าง (structured interview) ใช้คำถามหลักจากแบบวัดความพึงพอใจในการฝึกอบรมที่ผ่านการตรวจสอบคุณภาพแล้ว

5. คู่มือประกอบการฝึกอบรม ที่ผ่านการสร้าง และตรวจสอบคุณภาพเสร็จสิ้นแล้วในขั้นตอนที่ 2

การรวบรวมข้อมูล ดำเนินการดังนี้

1. นัด วัน เวลา สถานที่ และแจ้งกำหนดการล่วงหน้าแก่ตัวอย่าง
2. ดำเนินกิจกรรมการฝึกอบรม ตามร่างกำหนดการในการเก็บรวบรวมข้อมูล ดังแสดงในตารางที่ 8

ตารางที่ 8 กำหนดการจัดการฝึกอบรมการเสริมสร้างสมรรถนะการประเมิน

วันที่ 20 กรกฎาคม 2568	กิจกรรม
8:30-9:00 น.	ลงทะเบียน
9:00-09:30 น.	- ผู้วิจัยชี้แจงวัตถุประสงค์ของกรฝึกอบรม และชี้แจงบทบาทของผู้เข้ารับกรฝึกอบรมในแต่ละกิจกรรม
09:30-10:00 น.	- ประเมินความรู้, ความคิดเห็น และทักษะการประเมินภาคปฏิบัติ
10:00-10:30 น.	- ประเมินทักษะการประเมินภาคปฏิบัติกับงาน 8 ชุด
10:30-10:40 น.	- พักรับประทานอาหารว่าง
10:40-12:00 น.	- ประเมินทักษะการประเมินภาคปฏิบัติกับงาน 8 ชุด (ต่อ)
12:00-13:00 น.	- พักรับประทานอาหารกลางวัน
13:00-14:30 น.	- ประเมินทักษะการประเมินภาคปฏิบัติกับงาน 8 ชุด (ต่อ)
วันที่ 20 กรกฎาคม 2568	กิจกรรม
14:30-14:40 น.	- พักรับประทานอาหารว่าง
14:40-16:30 น.	- ความรู้พื้นฐานด้านการประเมินภาคปฏิบัติ - การสร้างความเข้าใจเรื่องความคลาดเคลื่อนจากผู้ประเมิน - การฝึกอบรมเรื่องกรอบของการอ้างอิง (Frame of Reference training) - การฝึกการสังเกตพฤติกรรม (Behavioral observation training: BOT) - การฝึกการตัดสินใจ (Decision making training)

ตารางที่ 8 กำหนดการจัดการฝึกอบรมการเสริมสร้างสมรรถนะการประเมิน (ต่อ)

วันที่ 21 กรกฎาคม 2568	
9.00-10.30 น.	- ฝึกการประเมินภาคปฏิบัติกับงานภาคปฏิบัติในรูปวีดิทัศน์จำนวนครั้งละ 1 ชุดจนกว่าจะได้ผลการประเมินที่สอดคล้องกับผู้เชี่ยวชาญร้อยละ 70 ขึ้นไป
10.30-10.40 น.	พักรับประทานอาหารว่าง
10.40-12.00 น.	- ฝึกการประเมินภาคปฏิบัติกับงานภาคปฏิบัติในรูปวีดิทัศน์จำนวนครั้งละ 1 ชุดจนกว่าจะได้ผลการประเมินที่สอดคล้องกับผู้เชี่ยวชาญร้อยละ 70 ขึ้นไป หรือ จนหมดเวลาในการฝึกอบรมในช่วงนี้
12.00-13.00 น.	พักรับประทานอาหาร
13.00-16.00 น.	- ประเมินความรู้ ทักษะ และความคิดเห็นหลังการฝึกอบรม และประเมินความพึงพอใจต่อการเข้าร่วมอบรม
16.00-16.30 น.	- Focus group interview

3. ก่อนการฝึกอบรม ตัวอย่างทำแบบวัดที่ออกแบบไว้ในขั้นตอนที่ 2 และ ขั้นตอนที่ 3 เครื่องมือประกอบด้วย แบบวัดความรู้ และแบบวัดทัศนคติเกี่ยวกับการประเมินภาคปฏิบัติ และชุดแบบวัดทักษะการประเมินภาคปฏิบัติ ที่ประกอบด้วยรូบริกการประเมินภาคปฏิบัติ และวีดิทัศน์สถานการณ์จำลอง จำนวน 8 สถานการณ์

4. เมื่อสิ้นสุดการฝึกอบรม ตัวอย่างทำแบบวัดชุดเดียวกัน และตอบแบบวัดความพึงพอใจ และการสัมภาษณ์เฉพาะกลุ่มตามความสมัครใจ (Focus group interview)

5. สรุปความพึงพอใจในการฝึกอบรมเพื่อนำความคิดเห็นไปใช้ในการปรับปรุงโมดูลเพื่อการใช้งานจริงในขั้นตอนที่ 4

การวิเคราะห์ข้อมูล

1. ข้อมูลพื้นฐานของตัวอย่าง วิเคราะห์ด้วยสถิติเชิงบรรยาย ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน
2. ผลของการใช้โมดูลฯ ต่อความรู้ และทัศนคติ ของการประเมินภาคปฏิบัติก่อน และหลังการใช้โมดูล วิเคราะห์เปรียบเทียบค่าเฉลี่ยโดยใช้สถิติทดสอบที

3. ผลของการใช้โมดูลฯ ต่อทักษะการประเมินภาคปฏิบัติ มีฟาเซทที่ต้องวิเคราะห์ 4 ฟาเซทคือ
1) ผู้ประเมิน 11 คน 2) ผู้ถูกประเมิน ในที่นี้ให้ถือวีดิทัศน์ 1 ชุดที่มีนักศึกษาแสดงหลัก 1 คน เป็นเสมือนผู้ถูกประเมิน 1 คน มีวีดิทัศน์ 8 ชุด จึงมีผู้ถูกประเมินทั้งสิ้น 8 คน 3) ฟาเซทสมรรถนะที่ประเมิน 7 สมรรถนะหลัก และ 4) ฟาเซทของเวลาในการวัด 2 ครั้ง วิเคราะห์ด้วย โมเดลหลายองค์ประกอบของราล์ซ ด้วยโปรแกรม FACETS version 3.71.4 (Linacre, 2014)

4. ข้อมูลจากการสัมภาษณ์เฉพาะกลุ่ม วิเคราะห์ข้อมูลโดยการวิเคราะห์เนื้อหา

ขั้นตอนที่ 4 การประเมินผลการใช้โมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับ อาจารย์พยาบาล

การใช้โมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินที่ผ่านการปรับปรุงหลังการทดลองใช้ในขั้นตอนที่ 3 มีระเบียบวิธีวิจัยดังนี้

ตัวอย่าง ได้แก่อาจารย์คณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต ที่ผ่านการจับคู่ และสุ่มอย่างง่ายเข้ากลุ่มใช้โมดูลจำนวน 6 คน และอาจารย์ในทุกสาขาวิชาในคณะพยาบาลศาสตร์สวนดุสิต ที่สมัครใจเข้ารับการฝึกอบรมอีก 5 คน รวมเป็น 11 คน

เครื่องมือ

แบ่งเป็น 2 ส่วนคือ

1. เครื่องมือที่ใช้ในการศึกษาผลของการฝึกอบรมการเสริมสร้างสมรรถนะการประเมินต่อ ความรู้ทัศนคติ การประเมินภาคปฏิบัติของอาจารย์พยาบาล ใช้เครื่องมือที่ผ่านการปรับปรุงจากขั้นตอนที่ 3
2. ชุดประเมินทักษะการประเมินภาคปฏิบัติ ที่ใช้ในขั้นตอนที่ 3 นำผลการประเมินความพึงพอใจมาปรับแก้วิธีทัศนให้มียุทธศาสตร์สั้นลง เฉลี่ยระยะเวลาของวิธีทัศนชุดละ 20 นาที จำนวน 10 ชุด ส่วนที่เหลือ 6 ชุดนำไปใช้ในขั้นตอนการฝึกอบรม

การรวบรวมข้อมูล

1. นัดหมายวันเวลา และแจ้งกำหนดการ ฝึกอบรมไว้ล่วงหน้า
2. ดำเนินการเก็บรวบรวมข้อมูล ก่อน และหลังการฝึกอบรม และหลังการฝึกอบรม 3 เดือน โดยใช้โมดูลการฝึกอบรมการเสริมสร้างสมรรถนะการประเมินกับตัวอย่างจำนวน 11 คน เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลเป็นเครื่องมือที่ผ่านการพัฒนามาจากขั้นตอนที่ 2 และผ่านการทดลองใช้และปรับปรุงมาจากขั้นตอนที่ 3 โดยดำเนินกิจกรรมเช่นเดียวกับขั้นตอนที่ 3 ตามกำหนดการเดียวกับกำหนดการในตารางที่ 18 โดยจัดกิจกรรมการฝึกอบรมในวันที่ 28 – 29 กรกฎาคม พ.ศ. 2568

การวิเคราะห์ข้อมูล

1. ข้อมูลพื้นฐานของตัวอย่าง วิเคราะห์ด้วยสถิติเชิงบรรยาย ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน
2. ผลของการใช้โมดูล ต่อความรู้ และทัศนคติ ของการประเมินภาคปฏิบัติก่อน และหลังการใช้โมดูลทันที และหลังการใช้โมดูล 3 เดือน เพื่อประเมินความยั่งยืน ด้วยการวิเคราะห์ความแปรปรวน แบบวัดซ้ำมิติเดียว (One- way repeated measures ANOVA)

โดยดำเนินการตามขั้นตอนคือ (Field, 2012)

2.1 ทดสอบข้อตกลงเบื้องต้นของการทดสอบแบบวัดซ้ำมิติเดียว ที่กล่าวว่าค่าความแปรปรวนของกลุ่มประชากรในการวัดทุกครั้งมีค่าเท่ากัน (Sphericity) ด้วยวิธีทดสอบของ Mauchly ตัวแปรที่ต้องการวัด คือคะแนนรวมของความรู้พื้นฐานด้านการประเมินภาคปฏิบัติ และมี 3 ตัวแปรคือ คะแนนความรู้ และทัศนคติ ในระยะก่อน และหลังการใช้โมดูลการฝึกอบรมทันที และหลังการใช้โมดูลการฝึกอบรม 3 เดือน ผลการทดสอบจะต้องไม่มีนัยสำคัญทางสถิติ ($p > .05$) แสดงว่าค่าความแปรปรวนของตัวแปรทั้งสามไม่แตกต่างกัน หากการทดสอบไม่เป็นไปตามข้อตกลงเบื้องต้น สามารถปรับ degree of freedom ด้วยการคูณด้วยค่า sphericity ที่ประมาณค่าไว้ล่วงหน้า การลด degree of freedom จะทำให้ค่า F-ratio มีความหมายมากขึ้น คือทำให้ค่า degree of freedom มากขึ้น และทำให้ไม่มีนัยสำคัญทางสถิติ หากค่าที่ประมาณได้ (ϵ) มีค่ามากกว่า .75 ให้ใช้ค่า Huynh-Feldt แทนค่าเดิม และถ้ามีค่าน้อยกว่า .75 ให้ใช้ค่า Greenhouse-Geisser แทนค่าเดิม

2.2 วิเคราะห์ความแปรปรวนแบบวัดซ้ำมิติเดียวของตัวแปรทั้งสามด้วยค่า Mean square ซึ่งคำนวณจากค่า ผลหารของ Sum of square หารด้วย degree of freedom ในที่นี้ degree of freedom เท่ากับตัวอย่างที่ศึกษา ลบด้วย 1 คูณด้วยจำนวนลักษณะงาน คือ 3 ลบด้วย 1 หากมีค่า $p > .05$ สรุปได้ว่าคะแนนการทดสอบความรู้ และทัศนคติฯ ในระยะก่อน หลังการใช้โมดูลการฝึกอบรมทันที และหลังการใช้โมดูล 3 เดือนไม่มีความแตกต่าง ยุติการวิเคราะห์ แต่หากพบค่า $p < .05$ แสดงว่ามีนัยสำคัญทางสถิติ สรุปได้ว่าคะแนนการทดสอบ ความรู้ ทักษะ และทัศนคติ ในระยะก่อน หลังการใช้โมดูลการฝึกอบรมทันที และหลังการใช้โมดูลการฝึกอบรม 3 เดือนมีความแตกต่างกันอย่างมีนัยสำคัญ

2.3 ทำการวิเคราะห์ต่อด้วย post hoc tests ด้วยวิธีของ Bonferroni เพื่อเปรียบเทียบอิทธิพลหลัก (main effects) ของตัวแปรทั้งสามที่แตกต่างกัน พิจารณา within- subjects effects จากค่า significant และค่า F-ratio หากค่า significant $< .05$ แสดงว่าการวัดความรู้ และทัศนคติ ในเวลาต่างกัันมีอิทธิพลต่อ

คะแนนการประเมินทั้งสามด้านอย่างมีนัยสำคัญ และพิจารณาปฏิสัมพันธ์ระหว่างตัวแปรแต่ละตัวว่ามีอิทธิพลต่อกันอย่างไร โดยใช้หลักการเดียวกัน

3. ผลของการใช้โมดูล ต่อทักษะการประเมินภาคปฏิบัติ มีฟาเซทที่ต้องวิเคราะห์ 4 ฟาเซทคือ 1) ผู้ประเมิน 11 คน 2) ผู้ถูกประเมิน ในที่นี้ให้ถือวีดีทัศน์ 1 ชุดที่มีนักศึกษาแสดงหลัก 1 คน เป็นเสมือนผู้ถูกประเมิน 1 คน มีวีดีทัศน์ 10 ชุด จึงมีผู้ถูกประเมินทั้งสิ้น 10 คน 3) ฟาเซทสมรรถนะที่ประเมิน 7 สมรรถนะหลัก และ 4) ฟาเซทของเวลาในการวัด 3 ครั้ง คือ ก่อน หลังการใช้โมดูล และหลังการใช้โมดูล 3 เดือน เพื่อประเมินความยั่งยืน วิเคราะห์ด้วย โมเดลหลายองค์ประกอบของราล์ซ ด้วยโปรแกรม FACETS version 3.71.4 (Linacre, 2014) ใช้ฟาเซทผู้ประเมินในการศึกษาใช้ค่าพารามิเตอร์ และการแปลผลเหมือนฟาเซทผู้ประเมินในหัวข้อที่ 4

4. ผลของการใช้โมดูล ต่อความคลาดเคลื่อนจากการประเมิน ศึกษาในการวัด 2 ครั้งคือ ก่อน และหลังการใช้โมดูลทันที วิเคราะห์ด้วยโมเดลหลายองค์ประกอบของราล์ซ ด้วยโปรแกรม FACETS version 3.71.4 (Linacre, 2014) ทำการศึกษาความคลาดเคลื่อนจากผู้ประเมิน คือ การกด ปล่อยคะแนน การให้คะแนนเกาะกลุ่มตรงกลาง และความคลาดเคลื่อนแบบฮาลโล ใน 4 ฟาเซทคือ

4.1 ผู้ประเมิน 11 คน ฟาเซทของผู้ประเมินใช้ศึกษาความคลาดเคลื่อนจากการ กด ปล่อยคะแนน มีพารามิเตอร์ที่ใช้ และการแปลผลดังนี้

4.1.1 Fixed chi-square test ของฟาเซทผู้ประเมิน ทดสอบสมมติฐานว่างที่ว่า การวัดการกด ปล่อยคะแนนของผู้ประเมิน ไม่มีความแตกต่างกัน ดังนั้นผู้ประเมิน ให้ผลการประเมินในระดับการกด ปล่อยคะแนนในระดับเดียวกัน หรือ ในทางตรงข้าม หากทดสอบสมมติฐานแล้วพบว่า มีนัยสำคัญทางสถิติ แสดงว่า อย่างน้อย 2 คนในกลุ่มผู้ประเมินมีความแตกต่างในระดับการกด ปล่อยคะแนนอย่างมีนัยสำคัญทางสถิติ

4.1.2 Rater separation ratio เป็นสัดส่วนของการวัดการกระจายของการกด ปล่อยคะแนนสัมพันธ์กับความแม่นยำของการวัด หรือความคลาดเคลื่อนจากการวัด แปลค่าจากสัดส่วนที่ได้ว่ามีระดับการกด ปล่อยคะแนนเป็นกี่เท่าของความคลาดเคลื่อนจากการวัด (Myford & Wolfe, 2003) ค่าบวก แสดงถึงผู้ประเมินปล่อยคะแนน (leniency error) ถ้าค่าต่ำหรือติดลบ แสดงถึงการกดคะแนน (severity error) (Linacre, 2014)

4.1.3 Reliability of rater separation index เป็นค่าพารามิเตอร์ที่บอกให้ทราบว่า สามารถแยกลักษณะของผู้ประเมินว่ามีความเที่ยงในการประเมินมากน้อยเพียงใด มีหลักการ

เหมือนการหาความเที่ยงด้วยวิธีของ Cronbach หรือ KR 20 แต่แปลความหมายตรงกันข้ามกัน ค่าที่ดีคือ ค่าที่ต่ำเข้าใกล้ศูนย์แสดงว่าผู้ประเมินแต่ละคนมีระดับการคาดคะเนน หรือปล่อยคะเนน ใกล้เคียงกัน (Myford & Wolfe, 2003)

4.2 ผู้ถูกประเมิน ใช้ในการศึกษาความคลาดเคลื่อนแบบ ให้คะเนนเกาะกลุ่มตรงกลาง เท่ากับวิธีทัศน์ 1 ชุดเสมือนผู้ถูกประเมิน 1 คน มีวิธีทัศน์ 10 ชุด เท่ากับมีผู้ถูกประเมิน 10 คน ใช้พารามิเตอร์ และเกณฑ์ในการแปลผลดังนี้

4.2.1 Fixed chi-square test เป็นการทดสอบสมมติฐานที่ว่าผู้ถูกประเมิน (ratee) ได้รับการประเมินในระดับคะเนนบนมาตรฐานการประเมินที่ไม่แตกต่างกัน หากผลการทดสอบสมมติฐานว่างพบว่าไม่มีนัยสำคัญทางสถิติ แสดงว่าผู้ถูกประเมินได้รับการประเมินที่มีความคลาดเคลื่อนแบบ เกาะกลุ่มตรงกลาง

4.2.2 Ratee separation ratio เป็นสัดส่วนของการวัดการกระจายของการผลการปฏิบัติงานของผู้ถูกประเมินสัมพันธ์กับความแม่นยำในการวัด การแปลค่า คือค่าที่น้อยแสดงถึงการมีความคลาดเคลื่อนแบบเกาะกลุ่มตรงกลาง (Myford & Wolfe, 2003)

4.2.3 Reliability of ratee separation index เป็นการวัดการกระจายของการวัดการปฏิบัติงานของผู้ถูกประเมินสัมพันธ์กับความแม่นยำในการวัด ที่สะท้อนว่าผู้ประเมินสามารถแยกแยะระดับการปฏิบัติงานของผู้ถูกประเมินได้ ค่าที่เข้าใกล้ศูนย์แสดงว่าผู้ถูกประเมินทั้งกลุ่มถูกประเมินให้มีระดับคะเนนบนมาตรฐานการประเมินเกาะกลุ่มตรงกลาง ค่ามากแสดงว่ามีการกระจายของระดับความสามารถ จึงไม่เกิดความคลาดเคลื่อนจากการให้คะเนนเกาะกลุ่มตรงกลาง (Myford & Wolfe, 2004)

4.3 ฟาเชทสมรรถนะที่ประเมิน 7 สมรรถนะหลัก ใช้พารามิเตอร์ และเกณฑ์ในการแปลผลดังนี้

4.3.1 Fixed chi-square test เป็นการทดสอบสมมติฐานที่ว่าสมรรถนะการวัดทุกสมรรถนะได้รับการปรับเทียบให้มีค่าความยากให้อยู่ในระดับเดียวกันแล้ว หากไม่มีนัยสำคัญทางสถิติของการทดสอบ chi-square แสดงว่าไม่พบความแตกต่างกันของระดับการประเมินในแต่ละสมรรถนะที่ผู้ถูกประเมินได้รับการประเมิน แสดงว่ามีความคลาดเคลื่อนแบบฮาโล แต่หากมีนัยสำคัญทางสถิติ แสดงว่ามีสมรรถนะอย่างน้อย 2 สมรรถนะที่มีระดับความยากแตกต่างกัน แสดงว่าไม่พบความคลาดเคลื่อนแบบฮาโลในระดับกลุ่ม

4.3.2 Trait separation ratio เป็นสัดส่วนของการวัดการกระจายของความยากของสมรรถนะที่วัดสัมพันธ์กับความแม่นยำจากการวัด หากค่าต่ำแสดงว่าการประเมินที่ได้มีความ

คลาดเคลื่อนแบบฮาลอ เนื่องจากสัดส่วนของความยากของสมรรถนะที่วัดมีค่าใกล้เคียงกับความแม่นยำในการวัดนั้น (Myford & Wolfe, 2004)

4.3.3 Trait separation index เป็นตัวบ่งชี้ระดับความแตกต่างระหว่างสมรรถนะว่ามีช่วงชั้นของความยากแตกต่างกันเป็นกี่เท่าของความคลาดเคลื่อนมาตรฐานอย่างมีนัยสำคัญทางสถิติหรือไม่ (statistically distinct levels of trait difficulty that are apart from standard error) (Myford & Wolfe, 2003) การแปลความหมาย ใช้หลักเกณฑ์เดียวกับ separation ratio

4.3.4 Reliability of trait separation index เป็นการวัดการกระจายของสมรรถนะที่วัดว่าสามารถแยกความยากของแต่ละสมรรถนะได้ดีหรือไม่ หรือ ผู้ประเมินสามารถประเมินโดยแยกความแตกต่างของแต่ละสมรรถนะได้ดีหรือไม่ ค่าที่เข้าใกล้ศูนย์ หรือน้อยกว่า 1 แสดงว่ามีความไม่แตกต่างกัน หรือแตกต่างกันน้อยมากในการประเมินความยากของสมรรถนะ เกิดความคลาดเคลื่อนแบบฮาลอ ค่าเข้าใกล้ 1 แสดงว่ามีความหลากหลายในการประเมินความยากของสมรรถนะแสดงว่าไม่มีความคลาดเคลื่อนแบบฮาลอ (Myford & Wolfe, 2003)

4.3.4 ฟาเซทของเวลาในการวัด 2 ครั้ง เป็นตัวแปรหุ่นใช้ในการเปรียบเทียบโลจิทการวัดในแต่ละครั้ง

4.4 ค่า Infit และ Outfit mean-square มีค่าอยู่ระหว่าง 0 ถึง infinity ค่ามีคาดหวังคือค่าที่เข้าใกล้ 1 ในการศึกษาครั้งนี้กำหนดให้มีขอบเขตระหว่าง 0.6 ถึง 1.4 หากค่าน้อยกว่า 0.6 แสดงว่ามีค่าเข้าใกล้ค่าที่คาดหวังมากกว่าที่โมเดลคาดการณ์เรียกภาวะนี้ว่า over fit ถ้าค่ามากกว่า 1.4 แสดงว่าคะแนนการประเมินห่างไกลจากค่าที่คาดหวังกว่าที่โมเดลคาดการณ์ เรียกภาวะนี้ว่า misfit โดยจะให้ความสำคัญกับ misfit เนื่องจากแสดงถึงปัญหาในการประเมินที่ต้องศึกษาถึงสาเหตุต่อไป อย่างไรก็ตาม Linacre ได้แนะนำให้พิจารณาค่า Fit mean-square ที่มากกว่า 2.00 ขึ้นไปว่าเป็นค่าที่สะท้อนปัญหาการประเมินอย่างชัดเจน (Linacre, 1994)

4.5 Variable map เป็นการนำเสนอภาพแผนผังในทุกฟาเซทที่ศึกษาในระดับบุคคลรายคน หรือรายสมรรถนะ ที่แสดงให้เห็นภาพของการกระจายตัวของข้อมูลในการกด ปุ่มย่อยคะแนนความสามารถของผู้ถูกประเมินเป็นรายบุคคล และความยากของสมรรถนะรายด้านมีหน่วยการให้เทียบ 2 หน่วยคือ หน่วยโลจิท และหน่วยการวัดบนมาตรการวัดที่ใช้ในรูปบริค คือ 1 ถึง 4 สำหรับการศึกษานี้ผู้วิจัยได้เขียนคำสั่งการวิเคราะห์ให้ฟาเซทผู้ประเมินมีค่าพารามิเตอร์เป็นบวก ส่วนค่าพารามิเตอร์ของฟาเซทอื่น ๆ เป็นลบ แปลความหมายได้ดังนี้

4.4.1 พารามิเตอร์ของผู้ประเมินค่าบวกแสดงว่ามีการปล่อยคะแนน ค่าลบแสดงว่ามีการกดคะแนนมากกว่าที่โมเดลคาดการณ์

4.4.2 พารามิเตอร์ของผู้ถูกประเมินค่าบวกแสดงว่ามีความสามารถน้อย ค่าลบแสดงว่ามีความสามารถมากกว่าที่โมเดลคาดการณ์

4.4.3 พารามิเตอร์ของสมรรถนะ ค่าบวกแสดงว่ามีความยากมาก ค่าลบแสดงว่ามีความง่ายมากกว่าที่โมเดลคาดการณ์

5. ผลการใช้โมเดล ด้านทักษะโดยนำข้อมูลที่ได้จากการประเมินภาคปฏิบัติมาวิเคราะห์ความสอดคล้อง และความเที่ยงในการให้คะแนนระหว่าง ผู้ประเมิน กับผู้เชี่ยวชาญ (inter-rater reliability) ด้วย 2 วิธีการคือ

5.1 Rater agreement index (RAI) ใช้เงื่อนไขของผู้ประเมิน 2 คน ทำการประเมินงาน การประเมินชุดเดียวกัน เกณฑ์การแปลผลคือ ค่าอยู่ระหว่าง 0 ถึง 1 หากมีค่าเป็น 1 แสดงว่าผลการประเมินมีความสอดคล้องกันของผู้ประเมินอย่างสมบูรณ์ 100% หรือ หากมีค่าเข้าใกล้ 1 มากเท่าใด แสดงถึงความสอดคล้องของการให้คะแนนของผู้ประเมินมากเท่านั้น (Judith A Burry-Stock, Dale G Shaw, Cecelia Laurie, & Brad S Chissom, 1996)

5.2 Intra-class correlation coefficient (ICC) ใช้ในเงื่อนไขผู้ประเมิน 2 คนทำการประเมินงาน การประเมินชุดเดียวกัน เกณฑ์ในการแปลผลคือ (Hallgren, 2012)

ต่ำกว่า 0.40 ระดับต่ำ

ค่าอยู่ระหว่าง 0.40 ถึง 0.59 ระดับปานกลาง

ค่าอยู่ระหว่าง 0.60 ถึง 0.74 ระดับดี

ค่าตั้งแต่ 0.75 ขึ้นไปถึง 1 ระดับดีเยี่ยม

6. ข้อมูลจากการสัมภาษณ์เฉพาะกลุ่ม วิเคราะห์ข้อมูลโดยการวิเคราะห์เนื้อหา การเตรียมข้อมูลสำหรับการวิเคราะห์ด้วยโปรแกรม FACETS มีขั้นตอนดังนี้

7. การแปลงคะแนนที่มีระดับการวัดเป็นมาตราช่วง (interval scale) ให้เป็นระดับการวัดที่เป็นมาตราบัญญัติ (nominal scale) ตามลักษณะธรรมชาติของการวิเคราะห์ด้วยโมเดลหลายองค์ประกอบของราล์ซ (Iramaneerat & Myford, 2006) โดยระบุบริคมีระดับคะแนน 1 ถึง 4 มีสมรรถนะที่ต้องการวัดทั้งหมด 7 ด้าน คือ

1) คุณธรรมจริยธรรม มีสมรรถนะย่อย 5 ข้อ คือ (1) ความซื่อสัตย์ รับผิดชอบมีวินัย (2) ความสามารถในการควบคุมตนเอง (3) ยอมรับความแตกต่างของความเชื่อของผู้ใช้บริการ หรือญาติ (4) จิตอาสา (5) รักษาจรรยาบรรณวิชาชีพ พินิจพิเคราะห์ของผู้ใช้บริการ คะแนนรวม 20 คะแนน

2) ความรู้ มีสมรรถนะย่อย 2 ข้อ คือ (1) บอกเหตุผล และหลักการสำคัญของการปฏิบัติกิจกรรมการพยาบาลต่าง ๆ (2) มีส่วนร่วมในการอภิปรายแสดงถึงการมีความรู้ความเข้าใจในสาระสำคัญของศาสตร์ที่เป็นพื้นฐานชีวิต และวิทยาศาสตร์สุขภาพ และศาสตร์ทางวิชาชีพพยาบาล คะแนนรวม 8 คะแนน

3) ทักษะทางปัญญามีสมรรถนะย่อย 3 ข้อ คือ (1) ระบุจุดแข็ง จุดอ่อน และสิ่งที่ต้องพัฒนาตนเองเกี่ยวกับความรู้ และ ทักษะการปฏิบัติงาน (2) ค้นคว้าได้ตรงประเด็นจากหลากหลายแหล่งข้อมูล มาใช้ในการปฏิบัติ (3) วางแผนการพยาบาลตามกระบวนการพยาบาลสอดคล้องกับสถานการณ์ของผู้รับบริการ และแก้ไขปัญหาได้เหมาะสมกับสถานการณ์บนหลักการที่ถูกต้อง คะแนนรวม 12 คะแนน

4) ความสัมพันธ์ระหว่างบุคคลและความรับผิดชอบ มีสมรรถนะย่อย 4 ข้อ คือ (1) ปฏิสัมพันธ์ต่อบุคลากรในทีมสุขภาพ เพื่อน อาจารย์ หรือผู้ให้บริการ และการปรับตัวเชิงวิชาชีพ (2) แสดงออกถึงการคิดบวกต่อวิชาชีพ และต่อผู้อื่น (3) แสดงตนเป็นสมาชิกที่ดีของกลุ่ม ตามบทบาทของตน (4) พัฒนาตนเองอย่างต่อเนื่อง คะแนนรวม 16 คะแนน

5) ทักษะการวิเคราะห์เชิงตัวเลข การสื่อสาร และการใช้เทคโนโลยีสารสนเทศ มีสมรรถนะย่อย 3 ข้อ คือ (1) การเข้าใจความหมายของตัวเลข และนำข้อมูลตัวเลขมาใช้ได้ถูกต้อง (2) ทักษะการสื่อสาร ด้านการฟัง พูด อ่านเขียน ภาษาไทย และภาษาต่างประเทศที่ถูกต้อง ครบถ้วน (3) การสืบค้นข้อมูลที่เป็นปัจจุบันจากแหล่งต่าง ๆ ทั้งจำนวนเอกสารอ้างอิง แหล่งอ้างอิง ความน่าเชื่อถือของแหล่งอ้างอิง คะแนนรวม 12 คะแนน

6) ทักษะการปฏิบัติทางวิชาชีพ สาขาพยาบาลศาสตร์ มีสมรรถนะย่อย 6 ข้อ

คือ (1) ปฏิบัติการพยาบาลโดยใช้กระบวนการพยาบาลอย่างเป็นองค์รวมที่สามารถสะท้อนถึงการสร้างเสริมสุขภาพ การป้องกัน การรักษา และการฟื้นฟูสุขภาพ (2) ปฏิบัติการพยาบาลถูกหลักการพยาบาล ในกิจกรรมพยาบาลนั้นๆ (3) ปฏิบัติการพยาบาลถูกหลักการป้องกันการกระจายเชื้อ (4) ปฏิบัติการพยาบาล ด้วยความเมตตา กรุณา และเอื้ออาทร และทันเหตุการณ์สอดคล้องกับความต้องการของผู้ใช้บริการ (5) ปฏิบัติการพยาบาลโดยคำนึงถึงความเป็นปัจเจกบุคคล และความหลากหลายทางวัฒนธรรม (6) ปฏิบัติการพยาบาลโดยยึดหลักความปลอดภัย คะแนนรวม 24 คะแนน

7) สมรรถนะของนักศึกษาตามเอกลักษณ์ของมหาวิทยาลัยมีสมรรถนะย่อย 3 ข้อ คือ (1) แต่งกายถูกระเบียบ สะอาดเรียบร้อย ผมเรียบร้อยยุ่งเหยิง รองเท้าหุ้มส้นถูกระเบียบสะอาด ทำทางสง่าในแต่ละอิริยาบถ ทั้งขณะนั่ง ยืนหรือเดิน (2) หน้าตายิ้มแย้ม ทำทางเป็นมิตร ขณะพูดคุยเสียงดัง ฟังชัด ฉะฉาน ด้วยทำทางเชื่อมั่นในตัวเอง (3) ความประณีต พิถีพิถัน คະแนนรวม 12 คະแนน

นำคะแนนการประเมินในแต่ละสมรรถนะย่อยมาหาค่าเฉลี่ยเพื่อเป็นข้อมูลของสมรรถนะหลัก และมีเกณฑ์ในการปรับคะแนนสมรรถนะย่อย 26 ข้อซึ่งเป็นข้อมูลระดับ interval scale ให้เป็นสมรรถนะหลัก 7 ข้อ เพื่อให้ลักษณะของข้อมูลมีลักษณะเป็น category สอดคล้องกับลักษณะของข้อมูลที่จะใช้ในการวิเคราะห์ด้วยโปรแกรม FACETS (Linacre, 2014) ดังแสดงในตารางที่ 9

ตารางที่ 9 เกณฑ์ในการแปลความหมายระดับคุณภาพของสมรรถนะแต่ละสมรรถนะ

ระดับคะแนน	การแปลความหมาย
คะแนนเฉลี่ย 3.50 - 4.00	สูงกว่าระดับมาตรฐาน ระดับคะแนน 4
คะแนนเฉลี่ย 2.50 - 3.49	ระดับมาตรฐาน ระดับคะแนน 3
คะแนนเฉลี่ย 1.50 - 2.49	ระดับพื้นฐาน ระดับคะแนน 2
คะแนนเฉลี่ย 0.00 - 1.49	ต่ำกว่าระดับพื้นฐาน ระดับคะแนน 1

ขั้นตอนที่ 5 ปรับปรุงโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์
พยาบาล.

ภายหลังการใช้โมดูลการฝึกอบรม นำผลการประเมินการใช้โมดูลมาปรับปรุง และสรุปเป็น
โมดูลมาตรฐาน ในการเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้ผู้วิจัยใช้ระเบียบการวิจัย และพัฒนา (Research and Development) มีวัตถุประสงค์เพื่อ 1) พัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล 2) ศึกษาผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล ต่อความรู้ ทักษะการประเมินภาคปฏิบัติ 3) ศึกษาผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล ต่อความคลาดเคลื่อนจากผู้ประเมินผลการวิเคราะห์ข้อมูลตามวัตถุประสงค์การวิจัยนำเสนอตามวัตถุประสงค์ของการวิจัย 3 ตอนดังต่อไปนี้

ตอนที่ 1 ผลการพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล

ผลการพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล นำเสนอตามขั้นตอนการพัฒนา คือ (1) การสำรวจสภาพปัจจุบันของการประเมินภาคปฏิบัติ (2) การออกแบบโมดูลการฝึกอบรม (3) การสร้างเครื่องมือ และ ตรวจสอบคุณภาพเครื่องมือที่ใช้ในโมดูล และเครื่องมือที่ใช้ในการศึกษาผลของโมดูลการฝึกอบรม ต่อความรู้ ทักษะการประเมินภาคปฏิบัติ และต่อความคลาดเคลื่อนจากผู้ประเมิน

1. ผลการสำรวจสภาพปัจจุบันของการประเมินภาคปฏิบัติ

การสำรวจสภาพปัจจุบันของการประเมินภาคปฏิบัติในอาจารย์พยาบาลมีการสำรวจ 2 ส่วน คือ ส่วนที่ 1 สำรวจสภาพปัญหาการประเมินภาคปฏิบัติในอาจารย์พยาบาลทำหน้าที่สอนภาคปฏิบัตินักศึกษาพยาบาลในระดับปริญญาตรี และผู้บริหารฝ่ายวิชาการ และส่วนที่ 2 เป็นการสำรวจความต้องการในการใช้ระบบการประเมินภาคปฏิบัติ ได้ผลการวิเคราะห์แจกแจงตามกลุ่มผู้ให้ข้อมูลดังนี้

1.1 สำรวจสภาพปัญหาจากอาจารย์พยาบาลที่ทำหน้าที่เป็นอาจารย์ผู้สอนภาคปฏิบัติ ของแต่ละสถาบัน 5 สถาบัน ดำเนินงานตามขั้นตอนคือ

1.1.1 สํารวจข้อมูลพื้นฐานของสถาบันการศึกษาพยาบาลจากสถาบันของแต่ละภาค จาก 4 ภาค จำนวนภาคละ 1 สถาบัน เลือกแบบเจาะจง 1 สถาบันคือ คณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต จำนวนตัวอย่าง 96 คน ดังแสดงในตารางที่ 10

ตารางที่ 10 จำนวนผู้ให้ข้อมูลที่ได้ผลตอบกลับจำแนกตามคณะพยาบาลศาสตร์ และสาขาวิชาที่รับผิดชอบ

คณะพยาบาล ศาสตร์ มหาวิทยาลัย	จำนวน	จำนวนผู้ให้ข้อมูลจำแนกตามสาขาวิชา					
		การพยาบาลมารดาทารก และผดุงครรภ์	การพยาบาลเด็ก และวัยรุ่น	การพยาบาลผู้ใหญ่ และผู้สูงอายุ	การพยาบาลสุขภาพจิต และจิตเวช	การพยาบาลชุมชน และ การรณรงค์โรคเบื้องต้น	
เชียงใหม่	43	7	4	21	4	7	
ขอนแก่น	31	7	5	13	3	3	
ราชภัฏสวนสุนันทา	12	2	2	6	1	1	
ราชภัฏสุราษฎร์ธานี	4	1	1	2	0	0	
สวนดุสิต	11	2	2	5	1	1	
รวม	96	18	13	41	9	15	

1.1.2 ข้อมูลทั่วไปของตัวอย่าง ส่วนใหญ่เป็นเพศหญิง (ร้อยละ 89.58) เพศชาย (ร้อยละ 10.42) ระดับการศึกษา ส่วนใหญ่ระดับการศึกษาปริญญาโท (ร้อยละ 48.89) รองลงมาระดับปริญญาเอก (ร้อยละ 47.78) และ ปริญญาตรี (ร้อยละ 3) ตามลำดับ ตำแหน่งทางวิชาการ ตัวอย่างส่วนใหญ่ ตำแหน่งอาจารย์ (ร้อยละ 50) รองลงมาคือ ตำแหน่งผู้ช่วยศาสตราจารย์ (ร้อยละ 30.21) และ รองศาสตราจารย์ (ร้อยละ 19.79) ตามลำดับ สำหรับสาขาวิชาที่อาจารย์สังกัด ส่วนใหญ่สังกัดสาขาวิชาการพยาบาลผู้ใหญ่ (ร้อยละ 42.71) รองลงมา สาขาการพยาบาลมารดาทารก และผดุงครรภ์ (ร้อยละ 18.75) การพยาบาลชุมชน (ร้อยละ 15.63) การพยาบาลเด็ก และวัยรุ่น (ร้อยละ 13.54) และ สาขาการพยาบาลจิตเวช (ร้อยละ 9.37) ตามลำดับ ข้อมูลด้านประสบการณ์การสอนภาคปฏิบัติ ส่วนใหญ่มีประสบการณ์ฯ มากกว่า 20 ปี (ร้อยละ 35.42) ประสบการณ์ น้อยกว่า 3 ปี (ร้อยละ 22.92) ประสบการณ์ 11 ถึง 19 ปี (ร้อยละ 19.79) ประสบการณ์ 3 ถึง 5 ปี (ร้อยละ 15.63) และ 6 ถึง 10 ปี (ร้อยละ 6.25) ตามลำดับ ด้านการฝึกอบรม เคยเข้าฝึกอบรมการประเมินภาคปฏิบัติเป็นส่วนใหญ่ (ร้อยละ 80.21) และไม่เคยเข้าฝึกอบรม (ร้อยละ 19.79) ดังแสดงในตาราง 11

ตารางที่ 11 ข้อมูลทั่วไป

ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม	จำนวน	ร้อยละ
เพศ		
หญิง	86	89.58
ชาย	10	10.42
ระดับการศึกษา		
ปริญญาตรี	3	3.13
ปริญญาโท	50	48.89
ปริญญาเอก	43	47.78
ตำแหน่งทางวิชาการ		
อาจารย์	48	50
ผู้ช่วยศาสตราจารย์	29	30.21
รองศาสตราจารย์	19	19.79
สาขาวิชา		
การพยาบาลมารดาทารก และผดุงครรภ์	18	18.75
การพยาบาลเด็ก และวัยรุ่น	13	13.54
การพยาบาลผู้ใหญ่	41	42.71
การพยาบาลชุมชน	15	15.63
การพยาบาลจิตเวช	9	9.38
ประสบการณ์การสอนภาคปฏิบัติ		
น้อยกว่า 3 ปี	22	22.92
3 - 5 ปี	15	15.63
6 - 10 ปี	6	6.25
11 - 19 ปี	19	19.79
ตั้งแต่ 20 ปีขึ้นไป	34	35.42
การเข้าฝึกอบรมการประเมินภาคปฏิบัติ		
เคยรับการฝึกอบรม	77	80.21
ไม่เคยรับการฝึกอบรม	19	19.79

n = 96

1.1.3 ผลการสำรวจสภาพปัญหาการประเมินภาคปฏิบัติ มีรายละเอียดดังนี้ สมรรถนะที่ใช้เป็นกรอบแนวคิดของเครื่องมือที่ใช้ประเมิน ส่วนใหญ่ใช้กรอบมาตรฐานคุณวุฒิระดับอุดมศึกษา (ทักษะการปฏิบัติทางวิชาชีพ) (ร้อยละ 66.67) รองลงมาใช้แนวคิดของสมรรถนะของบัณฑิตตามเอกลักษณ์ของสถาบัน (ร้อยละ 32.29) ส่วนใหญ่มีคู่มือสำหรับอาจารย์ (ร้อยละ 92.71) เนื้อหาในคู่มือประกอบด้วยความรู้พื้นฐานเกี่ยวกับการประเมินภาคปฏิบัติ เป็นส่วนใหญ่ (ร้อยละ 69.79) เครื่องมือที่ใช้ในการประเมินภาคปฏิบัติมีใช้หลายประเภทในการประเมินครั้งเดียวกัน ส่วนใหญ่ใช้แบบตรวจสอบรายการ (ร้อยละ 80.21) รองลงมาใช้มาตราประมาณค่า (ร้อยละ 79.17) รูบริค (ร้อยละ 52.08)

และใช้ระเบียบพหุติการณ์ (ร้อยละ 51.04) ตามลำดับ ส่วนใหญ่ไม่มีโมดูล หรือหลักสูตรระยะสั้นในการเตรียมความพร้อมอาจารย์ในการทำหน้าที่ประเมินภาคปฏิบัติของนักศึกษาพยาบาล (ร้อยละ 56.25) ส่วนใหญ่เห็นว่าหน่วยงานของตนจำเป็นต้องมีโปรแกรม หรือหลักสูตรระยะสั้นเพื่อเสริมสร้างสมรรถนะการประเมิน (ร้อยละ 80.21) ส่วนใหญ่ประเมินตนเองว่าจำเป็นต้องได้รับการเตรียมความพร้อมในการเป็นผู้ประเมินภาคปฏิบัติของนักศึกษาพยาบาล (ร้อยละ 78.13)

สำหรับเนื้อหาในการฝึกอบรม ส่วนใหญ่เห็นว่าควรฝึกอบรมในเรื่อง ความรู้พื้นฐานเกี่ยวกับการวัด และประเมินผล (ร้อยละ 72.92) รองลงมาเรื่อง การฝึกการสังเกตพฤติกรรม (ร้อยละ 71.88) การฝึกอบรมเพื่อทำความเข้าใจกับความคลาดเคลื่อนจากผู้ประเมิน (ร้อยละ 64.58) กรอบของการอ้างอิง (ร้อยละ 56.25) การฝึกอบรมการตัดสินใจ (ร้อยละ 47.92) ตามลำดับ รูปแบบการเตรียมความพร้อมที่ต้องการ ส่วนใหญ่ต้องการระบบการปฐมนิเทศโดยใช้อาจารย์พี่เลี้ยง (ร้อยละ 58.33) รองลงมาต้องการ ระบบการสอนงาน และให้คำปรึกษา (ร้อยละ 57.29) การฝึกอบรม (ร้อยละ 54.17) ตามลำดับ

สภาพปัญหา และการแก้ไขปัญหาด้านเครื่องมือประเมิน ส่วนใหญ่ไม่พบปัญหาในการสร้างเครื่องมือ (ร้อยละ 68.75) ส่วนใหญ่ไม่พบปัญหาในการเลือกใช้เกณฑ์ หรือสมรรถนะการประเมินเพื่อเป็นแนวคิดพื้นฐาน ในการสร้างเครื่องมือ (ร้อยละ 85.42) ส่วนใหญ่ไม่พบปัญหาในการสร้างเครื่องมือตามเกณฑ์ (ร้อยละ 67.71) ส่วนใหญ่ไม่พบปัญหาตัวบ่งชี้คุณภาพการปฏิบัติไม่ครอบคลุมวัตถุประสงค์ของการเรียนรู้ (ร้อยละ 78.13) ส่วนใหญ่ไม่พบปัญหาตัวบ่งชี้คุณภาพการปฏิบัติที่ไม่สะท้อนความแตกต่างของคุณภาพงานที่ชัดเจน (ร้อยละ 66.67) และ ส่วนใหญ่ ไม่พบปัญหาระดับคุณภาพของเกณฑ์สูงเกินระดับสมรรถนะของนักศึกษา (ร้อยละ 79.17)

ปัญหาจากผู้ประเมิน ส่วนใหญ่ไม่พบปัญหาจากผู้ประเมินในด้านความรู้พื้นฐานในการประเมิน (ร้อยละ 65.63) ส่วนใหญ่พบปัญหาความเข้าใจในวัตถุประสงค์การประเมินไม่ตรงกัน (ร้อยละ 55.21) ส่วนใหญ่พบปัญหาการรับรู้ระดับคุณภาพของเกณฑ์ ไม่ตรงกันระหว่างอาจารย์ผู้สอนภาคปฏิบัติ (ร้อยละ 55.21) การเตรียมอาจารย์ผู้สอนภาคปฏิบัติเพื่อประเมินภาคปฏิบัตินักศึกษาส่วนใหญ่ไม่พบปัญหา (ร้อยละ 66.67) รายละเอียดดังแสดงในตาราง 12

ตารางที่ 12 ผลการสำรวจปัญหาของการประเมินภาคปฏิบัติ ปัญหาอุปสรรคและแนวทางแก้ไข

สภาพปัญหา และแนวทางแก้ไข ของการประเมินภาคปฏิบัติ	จำนวน	ร้อยละ
1. สมรรถนะที่ใช้เป็นกรอบแนวคิดในการสร้างเครื่องมือในการประเมิน		
- กรอบมาตรฐานคุณวุฒิระดับอุดมศึกษา(ทักษะการปฏิบัติทางวิชาชีพ)	64	66.67
- สมรรถนะของบัณฑิตตามเอกลักษณ์ของสถาบัน	31	32.29
- อื่นๆ	2	2.08
2. มีคู่มือสำหรับอาจารย์ที่ทำหน้าที่สอนภาคปฏิบัติเพื่อใช้ในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล		
- มี	89	92.71
- ไม่มี	7	7.29
3. กรณีมีคู่มือสำหรับอาจารย์ที่ทำหน้าที่สอนภาคปฏิบัติเพื่อใช้ในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล สารสำคัญประกอบด้วย *		
- ความรู้พื้นฐานเกี่ยวกับการประเมินภาคปฏิบัติ	67	69.79
- การใช้รูบรีคที่บ่งชี้ระดับคุณภาพได้ชัดเจน	44	45.83
- อื่นๆ		
4. เครื่องมือที่ใช้ในการประเมินภาคปฏิบัติ		
- แบบตรวจสอบรายการ (Checklist)	77	80.21
- ระเบียบพฤติกรรม (Anecdotal Record)	49	51.04
- มาตรฐานค่า (Rating scale)	76	79.17
- รูบรีค (Scoring rubric)	50	52.08
- อื่นๆ	1	1.04
5. มีโมดูล หรือหลักสูตรระยะสั้นในการเตรียมความพร้อมด้านสมรรถนะการประเมินของอาจารย์ผู้สอนภาคปฏิบัติที่ทำหน้าที่ประเมินภาคปฏิบัติของนักศึกษาพยาบาล		
- มี	42	43.75
- ไม่มี	54	56.25
6. หน่วยงานของท่านจำเป็นต้องมีโมดูล หรือหลักสูตรระยะสั้นเพื่อเสริมสร้างสมรรถนะในการประเมินหรือไม่		
- จำเป็น	77	80.21
- ไม่จำเป็น	19	19.79
7. ท่านคิดว่าตนเองจำเป็นต้องได้รับการเตรียมความพร้อมในการเป็นผู้ประเมินภาคปฏิบัติของนักศึกษาพยาบาลเพื่อเสริมสร้างสมรรถนะในการประเมินหรือไม่		
- จำเป็น	75	78.13
- ไม่จำเป็น	21	21.88
8. เครื่องมือที่ใช้ในการประเมินภาคปฏิบัติ		
- แบบตรวจสอบรายการ (Checklist)	77	80.21
- ระเบียบพฤติกรรม (Anecdotal Record)	49	51.04
- การฝึกอบรมกรอบของการอ้างอิง (Frame of reference: FOR)	54	56.25
- การฝึกอบรมการสังเกตพฤติกรรม (Behavioral observation training: BOT)	69	71.88
- การฝึกอบรมการตัดสินใจ (Decision making training)	46	47.92

ตารางที่ 12 ผลการสำรวจสภาพการณ์ของการประเมินภาคปฏิบัติ ปัญหาอุปสรรคและแนวทางแก้ไข (ต่อ)

สภาพการณ์ ปัญหาอุปสรรคและแนวทางแก้ไข ของการประเมินภาคปฏิบัติ	จำนวน	ร้อยละ
9. หากมีโครงการ หรือหลักสูตรระยะสั้นเพื่อเสริมสร้างสมรรถนะการประเมินภาคปฏิบัติแก่อาจารย์ผู้สอนภาคปฏิบัติในนักศึกษายาบาล ท่านต้องการให้ใช้รูปแบบการเตรียมความพร้อมแบบใด		
* <ul style="list-style-type: none"> - การใช้ระบบการปฐมนิเทศโดยใช้อาจารย์พี่เลี้ยง (Internship) - การศึกษาจากคู่มือ (Writing materials) - การใช้เทคโนโลยี (Technology) - การประชุม (Meeting) - การตั้งคำถามเชิงบวก (Appreciative inquiry) - ชุมชนแห่งการปฏิบัติ (Communities of practice) - การฝึกอบรม (Training) - การเข้ามามีส่วนร่วมในกระบวนการประเมิน (Involvement in an evaluation process) - การให้ความช่วยเหลือทางเทคนิค (Technical assistance) - การสอนงาน/ระบบพี่เลี้ยงและการให้คำปรึกษา (Mentoring /coaching) 		
10. การใช้เกณฑ์ในการประเมินผลการเรียนรู้ของผู้เรียนในหน่วยงานของท่าน*		
- กำหนดเกรดเป็นระดับ เช่น A, B, C, D, F	96	100
- กำหนดเกรดเป็น ผ่าน/ไม่ผ่าน	0	0
- อื่นๆ		
11. สภาพปัญหาการประเมินภาคปฏิบัติและวิธีการแก้ไขด้านเครื่องมือประเมิน		
11.1 ไม่มีการสร้างเครื่องมือที่มีการอธิบายระดับคุณภาพของงาน		
- พบปัญหา	30	31.25
- ไม่พบปัญหา	66	68.75
11.2 การเลือกใช้เกณฑ์ หรือสมรรถนะในการประเมินเพื่อเป็นแนวคิดพื้นฐานในการสร้างเครื่องมือ		
- พบปัญหา	12	12.5
- ไม่พบปัญหา	82	85.42
11.3 การสร้างเครื่องมือตามเกณฑ์หรือสมรรถนะ		
- พบปัญหา	30	31.25
- ไม่พบปัญหา	65	67.71
- ไม่แน่ใจว่ามีปัญหาหรือไม่	1	1.04
11.4 ตัวบ่งชี้คุณภาพการปฏิบัติไม่ครอบคลุมวัตถุประสงค์ของการเรียนรู้		
- พบปัญหา	21	21.88
- ไม่พบปัญหา	75	78.13
11.5 ตัวบ่งชี้คุณภาพการปฏิบัติงานไม่สะท้อนความแตกต่างของคุณภาพงานที่ชัดเจน ทำให้แยกพฤติกรรมที่ต้องการประเมินในแต่ละระดับคุณภาพได้ยาก		
- พบปัญหา	32	33.33
- ไม่พบปัญหา	64	66.67

ตารางที่ 12 ผลการสำรวจสภาพการณ์ของการประเมินภาคปฏิบัติ ปัญหาอุปสรรคและแนวทางแก้ไข (ต่อ)

สภาพการณ์ ปัญหาอุปสรรคและแนวทางแก้ไข ของการประเมินภาคปฏิบัติ	จำนวน	ร้อยละ
11.6 ระดับคุณภาพของเกณฑ์ และ/หรือ สมรรถนะที่ต้องการประเมินสูงเกินสมรรถนะในระดับที่นักศึกษาจะปฏิบัติได้		
- พบปัญหา	19	19.79
- ไม่พบปัญหา	77	80.21
12. ปัญหาจากผู้ประเมิน		
12.1 ความรู้พื้นฐานในการประเมิน		
- พบปัญหา	32	33.33
- ไม่พบปัญหา	63	65.63
- ไม่แน่ใจ	1	1.04
12.2 ความเข้าใจในวัตถุประสงค์การประเมินไม่ตรงกัน		
- พบปัญหา	53	55.21
- ไม่พบปัญหา	43	44.79
12.3 การรับรู้ระดับคุณภาพของเกณฑ์ไม่ตรงกันระหว่างอาจารย์ผู้สอนภาคปฏิบัติ		
- พบปัญหา	53	55.21
- ไม่พบปัญหา	43	44.79
12.4 การเตรียมอาจารย์ผู้สอนภาคปฏิบัติเพื่อเป็นผู้ประเมินนักศึกษาพยาบาล		
- พบปัญหา	32	33.33
- ไม่พบปัญหา	64	66.67

หมายเหตุ: * แบบสอบถามข้อที่ตอบได้หลายคำตอบ n = 96

1.1.4 ข้อมูลจากการสอบถามความคิดเห็นของอาจารย์ผู้สอนภาคปฏิบัติในนักศึกษาพยาบาล เกี่ยวกับแนวปฏิบัติ หรือปัญหาที่พบในการประเมินภาคปฏิบัติ ด้วยคำถามปลายเปิดสามารถสรุปประเด็นสำคัญได้ดังแสดงในตารางที่ 13

ตารางที่ 13 สรุปความคิดเห็นของอาจารย์ผู้สอนภาคปฏิบัติต่อการประเมินภาคปฏิบัติ

ข้อที่	คำถาม	เหตุผล/ แนวทางแก้ไข
1.	หน่วยงานของท่านจำเป็นต้องมีโมดูล หรือ หลักสูตรระยะสั้นเพื่อเสริมสร้างสมรรถนะในการประเมินหรือไม่	จำเป็นเนื่องจาก 1. มีการเปลี่ยนแปลงสมรรถนะที่ต้องการ/คาดหวังในนักศึกษา/ผู้เรียน หรือ มีการเปลี่ยนแปลงของศาสตร์ใหม่ๆ 2. เพื่อให้มาตรฐานการวัดและประเมินผลเท่ากัน เข้าใจตรงกัน ผลการประเมินมีความถูกต้องน่าเชื่อถือมากที่สุด มีความเที่ยงเบนน้อยที่สุด และมีประสิทธิภาพมากขึ้น

ตารางที่ 13 สรุปความคิดเห็นของอาจารย์ผู้สอนภาคปฏิบัติต่อการประเมินภาคปฏิบัติ (ต่อ)

ข้อที่	คำถาม	เหตุผล/ แนวทางแก้ไข
		<p>3. เพื่อเตรียมอาจารย์ใหม่ และทบทวนอาจารย์ปัจจุบัน โดย แลกเปลี่ยนเรียนรู้ร่วมกัน และเสริมความพร้อม และ สมรรถนะของอาจารย์ผู้สอน</p> <p>4. เพื่อลดปัญหา ความแตกต่างของคะแนนจากอาจารย์ที่ร่วมกัน สอนหลายคนในรายวิชาเดียวกัน</p> <p>5. บุคลากรส่วนใหญ่ของหน่วยงานมาจากผู้ปฏิบัติ ยังไม่ผ่านการ เตรียมความพร้อมด้านสมรรถนะการเป็นอาจารย์พยาบาล</p> <p>6. อาจารย์ที่ไม่ได้รับการฝึกอบรมมักจะประเมินผลไม่สอดคล้อง กับความเป็นจริง และ ไม่รู้หลักการที่ถูกต้องด้านการวัด และ ประเมินผลภาคปฏิบัติ</p>
		<p><u>ไม่จำเป็นเนื่องจาก</u></p> <p>1. เพราะมีระบบบริหารจัดการด้านการประเมินภาคปฏิบัติอยู่แล้ว เช่น มีการประชุมปรึกษาหารือทุก ๆ 2-4 สัปดาห์ และทุก ๆ กลุ่ม</p> <p>2. ที่หน่วยงานมีหลักสูตรให้อบรมอยู่แล้ว และมีระบบที่เลี้ยงที่ ติดตามอาจารย์ที่หมุนเวียนมาใหม่อยู่แล้ว</p>
2	ท่านคิดว่าตนเองจำเป็นต้องได้รับการเตรียมความพร้อมในการเป็น ผู้ประเมินภาคปฏิบัติของ นักศึกษาพยาบาล เพื่อเสริมสร้างสมรรถนะในการ ประเมินหรือไม่	<p><u>จำเป็นเนื่องจาก</u></p> <p>1. เป็นการพัฒนาตนเองอย่างต่อเนื่อง และตามทันการ เปลี่ยนแปลงขององค์ความรู้ที่มีหลากหลาย และการปฏิบัติ เกี่ยวกับการวัดและประเมินนักศึกษา ในประเด็น ใหม่ ๆ</p> <p>2. เป็นการเสริมศักยภาพด้านการประเมินภาคปฏิบัติของตนเอง เพื่อความโปร่งใส ยุติธรรม ไม่มีอคติ มีหลักการที่ถูกต้อง</p> <p>3. มีประสบการณ์น้อย และขาดทักษะ ควรได้รับความรู้ที่ถูกต้อง ตามหลักการวัด และประเมินผล เพื่อจะได้ประเมินผล ภาคปฏิบัติได้อย่างมีคุณภาพ</p> <p>4. มีการเปลี่ยนแปลงหลักสูตร และปรับปรุงแบบประเมินอย่าง ต่อเนื่อง จำเป็นต้องนำความเข้าใจในเรื่องการประเมินให้ตรงกัน</p> <p>5. เพื่อเป็นการเตรียมความพร้อมก่อนลงมือปฏิบัติงาน โดยกำหนด ข้อตกลงเบื้องต้น และเป็นการวางแผนร่วมกันของผู้สอน ร่วม ทั้งชี้แจงข้อสงสัยต่างๆ ในแบบประเมิน ขจัดปัญหาการประเมิน ที่ผู้ประเมินเข้าใจไม่ตรงกันที่ทำให้คะแนนการวัดต่างกันมาก มี แนวทางปฏิบัติเดียวกันทั้งองค์กร</p>

ตารางที่ 13 สรุปความคิดเห็นของอาจารย์ผู้สอนภาคปฏิบัติต่อการประเมินภาคปฏิบัติ (ต่อ)

ข้อที่	คำถาม	เหตุผล/ แนวทางแก้ไข
		6. ทำให้เกิดการแลกเปลี่ยนเรียนรู้ระหว่างผู้ประเมิน ไม่จำเป็นเนื่องจาก
		1. มีประสบการณ์การสอนภาคปฏิบัติมานานแล้ว
		2. ได้เรียนรู้จากครูพี่เลี้ยงแล้ว
		3. เคยได้รับการฝึกอบรมมาแล้ว
		4. ใกล้เกษียณแล้ว
3	ปัญหาเกี่ยวกับเครื่องมือประเมิน	
	ปัญหาที่พบ	<u>แนวทางแก้ไข</u>
	1. เครื่องมือในการประเมินไม่ครอบคลุมเนื้อหาในการประเมิน	1. ใช้ scoring rubric
	2. มีเพียงบางรายวิชาที่มีการสร้างเครื่องมือที่มีการอธิบายด้านคุณภาพ (รูบรีค) บางวิชาไม่มี	2. ใช้ anecdote record ประกอบ
		3. กำลังพัฒนาแบบประเมินสมรรถนะของนักศึกษาโดยเป็น rubric
4	ปัญหาเกี่ยวกับการเลือกใช้เกณฑ์การประเมิน	
	ปัญหาที่พบ	<u>แนวทางแก้ไข</u>
	1. ผู้สอนแต่ละคนมีแนวคิดที่แตกต่างกันในการตีความ ให้คะแนน	1. ใช้ผู้ประเมินหลายคน และนำผลมาอภิปรายร่วมกัน
	2. อาจารย์ใช้เกณฑ์ประเมินไม่เหมือนกัน	2. ใช้กรรมการทวนสอบ
		1. ประชุมทำความเข้าใจตกลงกันในกลุ่มผู้สอน
5	ปัญหาเกี่ยวกับความรู้พื้นฐานด้านการประเมินของผู้ประเมิน	
	ปัญหาที่พบ	<u>แนวทางแก้ไข</u>
	1. ผู้ประเมินมีมาตรฐานไม่ตรงกันเข้าใจไม่ตรงกัน ประเมินผลแตกต่างกัน บางคนให้คะแนนสูงเกินไป	1. อบรมอาจารย์
		2. อบรมฝึกประสบการณ์ผ่าน Mentors
		3. ประชุมผู้ประเมินเพื่อทำความเข้าใจ อธิบายเกณฑ์ และวิธีการประเมินผลก่อน และหลังการประเมิน
	2. ขาดความรู้ในเรื่อง	1. ปรึกษาอาจารย์ที่มีประสบการณ์ ผู้เชี่ยวชาญ ใช้ระบบอาจารย์พี่เลี้ยง
	- พื้นฐานของผู้ประเมิน ขาดประสบการณ์ ที่ถูกต้องชัดเจน ทำให้ขาดความมั่นใจในการประเมิน	2. สัมมนาเตรียมอาจารย์ และอภิปรายร่วมกันเพื่อการจัดการเรียนการสอน และก่อนจะจัดการเรียนการสอนต่อไป
	- การพัฒนา และการใช้เครื่องมือเพื่อให้ความน่าเชื่อถือ	
	3. ไม่ยอมเข้าใจ หรือปรับตัวเอง	1. ทำใจ ปล่อยวาง

ตารางที่ 13 สรุปความคิดเห็นของอาจารย์ผู้สอนภาคปฏิบัติต่อการประเมินภาคปฏิบัติ (ต่อ)

ข้อที่	คำถาม	เหตุผล/ แนวทางแก้ไข
6.	การเตรียมอาจารย์ผู้สอนภาคปฏิบัติเพื่อเป็น ผู้ประเมินนักศึกษาพยาบาล	
	ปัญหาที่พบ	<u>แนวทางแก้ไข</u>
1.	ภาระงานมาก ทำให้ไม่มีเวลาความเข้าใจให้ ชัดเจน และประเมินไม่ครอบคลุม	1. ไม่ได้ระบุแนวทางแก้ไข
2.	ไม่มีการเตรียมในเรื่องการประเมินภาคปฏิบัติ โดยเฉพาะ	2. การสัมมนาร่วมกันถึงปัญหาที่พบ
3.	อาจารย์ใหม่มีประสบการณ์ภาคปฏิบัติน้อย ได้รับการเตรียมน้อยเกินไป ทำให้ไม่ เข้าใจวิธีการ	3. ให้คำแนะนำ อธิบายรายบุคคล หรือปรึกษาอาจารย์ที่มี ประสบการณ์
4.	จำนวนอาจารย์นิเทศไม่เพียงพอต่อสัดส่วนของ นักศึกษา	4. ให้บุคลากรทางคลินิกภายนอกร่วมในการประเมินนักศึกษา
5.	อาจารย์เก่ามีความเป็นบุคคล เป็นอึดน้อยสูง	5. มีการประชุมเพื่อหาข้อสรุปร่วมกันในการประเมิน
6.	อาจารย์บางท่านไม่สามารถเข้าร่วมกิจกรรม การเตรียมความพร้อมได้(ติดภารกิจ)	6. ไม่ได้ระบุแนวทางแก้ไข
7.	ขาดคู่มือ หรือพี่เลี้ยงของอาจารย์ผู้สอน	7. ศึกษาเรียนรู้ด้วยตนเอง

n = 96

1.2 ผลการสัมภาษณ์ เพื่อสำรวจสภาพปัจจุบันของสถาบันการศึกษาพยาบาล ใน 4 ภาค และ
คณะพยาบาลศาสตร์สวนดุสิต จากผู้บริหารฝ่ายวิชาการของสถาบัน 5 ท่าน ประกอบด้วย สถาบันที่มี
การเปิดการเรียนการสอนมากกว่า 10 ปี 2 แห่ง และสถาบันที่มีการเปิดการเรียนการสอนมาน้อย
กว่า 10 ปี 3 แห่ง ใช้แนวคำถามกึ่งโครงสร้างสำหรับการสัมภาษณ์เชิงลึก โดยใช้แนวคำถามหลัก คือ
แนวคิดหลักที่นำมาวางแผนในการจัดการเรียนการสอนภาคปฏิบัติของสถาบันท่านเป็นอย่างไร มี
เกณฑ์ในการสร้างเครื่องมือ การเตรียมความพร้อมของอาจารย์ในการฝึกภาคปฏิบัติอย่างไร ปัญหา
หรืออุปสรรคในการจัดการเรียนการสอนภาคปฏิบัติ และ แนวทางแก้ไขเป็นอย่างไร ในระหว่าง
สัมภาษณ์เชิงลึกผู้วิจัยขออนุญาตบันทึกเสียง ใช้เวลาในการสัมภาษณ์เฉลี่ย 60 นาที ผลการวิเคราะห์
เนื้อหา มีประเด็นสำคัญ 4 ประเด็น คือ 1) วัตถุประสงค์หลักของการจัดการเรียนการสอนภาคปฏิบัติ
คือ ผลลัพธ์การเรียนรู้ (learning outcome) 2) แนวคิดพื้นฐานในการเลือกเกณฑ์ที่ใช้ในการสร้าง
เครื่องมือประเมิน และ การเลือกใช้เครื่องมือ คือ TQF 3) การเตรียมความพร้อมด้านการฝึกเพื่อทำ

การประเมินภาคปฏิบัติที่สำคัญคือการเตรียมอาจารย์ 4) ปัญหา และ แนวทางแก้ไข ในการกำหนดผลการเรียนภาคปฏิบัติ ดังรายละเอียดต่อไปนี้

ประเด็นที่ 1 ผลลัพธ์การเรียนรู้ ของการเรียนการสอนภาคปฏิบัติ ผู้บริหารแต่ละสถาบันการศึกษาพยาบาล จะกล่าวถึงเป้าหมายหลักของการจัดการเรียนการสอนภาคปฏิบัติว่าต้องการให้ผู้เรียนบรรลุตามเป้าหมายของหลักสูตร โดยจัดการเรียนการสอนที่ส่งเสริมให้เกิดผลลัพธ์การเรียนรู้ที่วางแผนไว้ ผู้บริหารทุกคนใช้หลักการเดียวกันในการจัดการเรียนการสอนภาคปฏิบัติคือ กรอบมาตรฐานคุณวุฒิระดับอุดมศึกษา (TQF) และอาจมีแนวคิดเกี่ยวกับสมรรถนะบัณฑิตที่พึงประสงค์ ในทศวรรษ 21 หรือตามเอกลักษณ์ของสถาบัน ดังรายละเอียดต่อไปนี้

“ตอนนี้ กรอบที่เราใช้... ใช้ TQF เป็นหลัก ก่อนหน้านี้ใช้เกณฑ์ของสมรรถนะของการพยาบาลที่กำหนดโดยสภาพยาบาล ส่วนของสมรรถนะหลักของสภาการพยาบาล ของสภาพยาบาลเป็นพื้นฐานอยู่แล้ว และ TQF นี้ต้องสอดคล้องกับ มคอ. ที่ให้ความสำคัญกับการประเมินความองงามของเด็ก ... learning outcome ... ความองงามอาจใช้สมุดบันทึกของเด็กให้เด็กจดบันทึก ว่าทำอะไรไปแล้วพัฒนาตัวเองถึงไหนแล้ว ครูก็เอามาคุยกับเด็กวางแผนกันว่าพัฒนาอย่างไร นโยบายของมหาวิทยาลัยว่าเราจำเป็นที่จะต้องเพิ่มทักษะในศตวรรษที่ 21 เราจะเริ่มปีหน้า”

ผู้ให้ข้อมูลคนที่ 1

(สัมภาษณ์ ณ วันที่ 20 ธันวาคม 2566)

“การพัฒนาหลักสูตรเราก็จะเอาหลักมาจาก TQF กับสภาการพยาบาล ในเชิงวิชาชีพกับสมรรถนะทั่วไป ถ้าส่วนของวิชาชีพกับสมรรถนะที่จะวัดเราก็เน้นเรื่อง nursing process และเรื่องการสร้างเสริมสุขภาพ กระบวนการวิธีการเรียนการสอนกำลังคิดว่าจะทำอย่างไรให้เด็กทำอย่างนี้ได้ ซึ่งเราจะตั้งตามกรอบของมคอ. ... มหาวิทยาลัยเค้าก็จะมีวิสัยทัศน์ซึ่งสะท้อนเอกลักษณ์ของมหาวิทยาลัย นักศึกษา ก็คือต้อง ready to work มีทักษะชีวิตในการทำงาน มีทักษะในเชิงวิชาชีพ ... จะให้เน้นเรื่องของ self - directed learning และ critical thinking”

ผู้ให้ข้อมูลคนที่ 2

(สัมภาษณ์ ณ วันที่ 24 ตุลาคม 2566)

“ก็วางหลักสูตรปรับให้สอดคล้องกับนโยบายการศึกษาระดับชาติ อย่างตอนนี้ก็ต้องใช้เกณฑ์ของ TQF และก็เกณฑ์ของสภาวิชาชีพ ทำอย่างไรให้ได้ผลลัพธ์การเรียนรู้อย่างที่เขาวางไว้”

ผู้ให้ข้อมูลคนที่ 3

(สัมภาษณ์ ณ วันที่ 11 ธันวาคม 2566)

“วันนี้ที่ไปประชุม สมศ.นะ คำก็พูดถึง TQF แบบประเมิน 5-6 ตัวเนี่ย คำก็ยังไม่ถึงไหน คือสาขาอื่น แต่พยาบาลไปเยอะแล้ว ถ้าภาคปฏิบัติของพยาบาลมันก็มีความรู้ จริยธรรม ปัญญา ทักษะสื่อสาร เทคโนโลยี สารสนเทศ การวิชาชีพ 6 ตัว ที่กำลังปรับหลักสูตรให้สอดคล้องตรงนั้นนะ ตอนนั้นที่ใช้จัดหลักสูตรภาคปฏิบัติที่ใช้ก็อิงของสภาการพยาบาล”

ผู้ให้ข้อมูลคนที่ 4

(สัมภาษณ์ ณ วันที่ 22 พฤศจิกายน 2566)

“หลักสูตรของเราตอนนี้ก็ใช้หลักเกณฑ์ในการประเมินผลการเรียนรู้โดยมีฐานคิดมาจาก TQF ตอนนี้พยาบาลของเราก็นำร่อง ในหลายที่แล้ว ของเราก็ดลองทำมาก่อนที่เขาจะเริ่มบังคับใช้”

ผู้ให้ข้อมูลคนที่ 5

(สัมภาษณ์ ณ วันที่ 3 มกราคม 2567)

ประเด็นที่ 2 แนวคิดพื้นฐานในการเลือกเกณฑ์ที่ใช้ในการสร้างเครื่องมือประเมิน และการเลือกใช้เครื่องมือ ในการจัดการเรียนการสอนภาคปฏิบัติ ทุกสถาบันการศึกษาพยาบาล มีการจัดเตรียมความพร้อมเพื่อการเรียนการสอนภาคปฏิบัติคือ การเตรียมคน และการเตรียมเครื่องมือที่จะใช้ในการประเมิน ไปในทิศทางเดียวกันคือ ใช้กรอบมาตรฐานคุณวุฒิระดับอุดมศึกษาทั้ง 6 ด้าน ด้านที่ 6 มีเนื้อหาเกี่ยวกับทักษะการปฏิบัติทางวิชาชีพ สาขาพยาบาลศาสตร์ ร่วมกับสมรรถนะหลักด้านการพยาบาลของสภาการพยาบาล เครื่องมือที่ใช้การประเมินส่วนใหญ่ใช้รูบริก ระดับคะแนน 5 ระดับ และมี 1 สถาบันใช้ 4 ระดับ ดังข้อมูลจากบทสัมภาษณ์ ต่อไปนี้

“การเตรียมเครื่องมือประเมินเราใช้เกณฑ์ของสมรรถนะหลักของการพยาบาลที่กำหนดโดยสภาพยาบาล เป็นพื้นฐานอยู่แล้ว และร่วมกับ TQF ด้วย...ในเกณฑ์การให้คะแนนภาคปฏิบัติเนี่ยนะคะ เราจะweightน้ำหนัก 5 สเกล ก็ถ้าจะมีน้ำหนัก 5 เนี่ยก็คือจะปฏิบัติได้เองโดยไม่ต้องการคำแนะนำ เป็นรูบริกทั้งหมดเลย”

ผู้ให้ข้อมูลคนที่ 1

(สัมภาษณ์ ณ วันที่ 20 ธันวาคม 2566)

“อย่างที่บอกไปนะว่าเน้นหลักสูตร ก็ TQF กับวิสัยทัศน์ของมหาวิทยาลัย นี่ที่ก็เลยเปลี่ยนวิธีการเขียนการพัฒนาตัว item ของการประเมิน และitem ของตัวrubricให้สะท้อนให้เห็นตัว nursing process ของการสร้างเสริมสุขภาพ แต่ outcome ที่จะสร้างเด็กสมรรถนะอื่นก็คือว่าต้องให้เค้า critical thinking เพราะฉะนั้นแบบประเมินใหม่ที่ทำอยู่ได้ 2 – 3 วิชานี้ก็ถือว่าแบบประเมินrubricจะสะท้อนว่า ถ้าได้อย่างนี้แล้วเด็กต้อง critical thinking... ปัจจุบัน ใช้ 5 สเกล”

ผู้ให้ข้อมูลคนที่ 2

(สัมภาษณ์ ณ วันที่ 24 ตุลาคม 2566)

“เวลาเตรียมความพร้อม ก็เตรียมด้านคนที่ไหนก็ไม่ค่อยพอนะ แล้วก็กับเกณฑ์ที่ใช้ที่นี่ใช้ของสภาเอามาเลยแล้วเอามาทำที่เป็นรายชื่อย่อยของแต่ละวิชาแล้วก็สร้างrubricวิชาขึ้นมา ใครจะได้ 4 3 2 1 ก็คิดเป็นสเกลเหมือนกัน ทุกสาขาวิชาต้องมีrubricเป็นของตัวเองด้วย เด็กทำอย่างนี้ ทำอย่างไรจะได้คะแนน 4 3 2 1 จะใช้ 4 สเกล”

ผู้ให้ข้อมูลคนที่ 3

(สัมภาษณ์ ณ วันที่ 11 ธันวาคม 2566)

“การประเมินภาคปฏิบัติของที่นี่จะใช้ของทางศิริราชที่อิงของสภา แต่รู้สึกตอนนี้เขากเปลี่ยนตาม TQF หรือเปล่า ก็อย่างที่ว่ามาที่เพิ่งไปประชุม...ถามว่าเรามาพัฒนาอะไรไหม การพัฒนาอย่างที่ว่าเนี่ยมันกว้างไปมันแคบไป แต่เราก็ยังไม่ได้พัฒนาถึงขั้นดีเพราะเรามีปัญหาในLab ขาดอาจารย์ อาจารย์ไม่ค่อยพอ ขาดนักวัดประเมินผล... คือยังขาดคนที่ช่วยสร้างเกณฑ์ หรือพื้นฐานด้านการวัดประเมิน”

ผู้ให้ข้อมูลคนที่ 4

(สัมภาษณ์ ณ วันที่ 22 พฤศจิกายน 2566)

“ในการสอนภาคปฏิบัติเราใช้ข้อมูล 2 ชุดขึ้นไป rubricก็อันหนึ่งใช้ anecdotal ประกอบ การใช้ทั้ง 2 ตัวเพื่อให้ข้อมูลครอบคลุมรอบด้าน การทำฟอร์มเมทนั่นก็จะช่วยไกด์ให้อาจารย์ให้มีแนวทางว่าจะบันทึกอะไรบ้าง อาจจะทำเป็นตุ๊กตาแล้วให้แต่ละสาขาไปปรับเอง เนื่องจากมันมีความหลากหลายของรายวิชา ส่วนประกอบก็คือrubricที่ตอนนี้เราใช้ 5 สเกล เป็นแบบประเมินกลาง แต่คาดหวังให้มีใช้เอาไปปรับใช้ให้มีความเฉพาะของแต่ละสาขา”

ผู้ให้ข้อมูลคนที่ 5

(สัมภาษณ์ ณ วันที่ 3 มกราคม 2567)

ประเด็นที่ 3 การเตรียมความพร้อมอาจารย์เพื่อทำการประเมินภาคปฏิบัติ ส่วนใหญ่ผู้ให้ข้อมูลมีความคิดเห็นสอดคล้องกันว่า หลักสำคัญของการจัดการเรียนการสอนภาคปฏิบัติคือ ปริมาณ และคุณภาพของอาจารย์ผู้สอนภาคปฏิบัติ ควรมีการเตรียมความพร้อมของอาจารย์ก่อนการขึ้นทำการสอนในสถานการณ์จริง บนหอผู้ป่วย ดังบทสัมภาษณ์ต่อไปนี้

“เวลาเตรียมความพร้อมก่อนขึ้นฝึกจริง กระบวนการมันต้องเริ่มจากเตรียมครู ของเราต้องผ่านการฝึกอบรบศาสตร์ และศิลป์ซึ่งเราจัดอยู่แล้ว ...ครูคลินิกเรากำหนดได้ว่าเค้าผ่านการฝึกอบรบ 4 เดือน...เราอบรมเชิงทฤษฎี และตัวของอาจารย์ใหม่ทุกคนถึงแม้จะจบปริญญาเอก เรายังมี mentor และในช่วงแรก ต้องขึ้นคู่กับmentor จะได้ว่า record พฤติกรรมเด็กยังไง จะ feedbackยังไง จะดู progressของเด็กยังไง ต้องขึ้นคู่กับ mentor ประมาณ 1 ปีคะ นักศึกษา 1 คนใน 1 setting ส่วนใหญ่จะไม่มีครูคนเดียว เราจะมีครูหลายคน แล้วก็ก็เป็นลักษณะการที่จะให้ครูหลายๆ คนมีโอกาสได้มาดูนักศึกษาหนึ่งคน ก็เพื่อให้ความเป็นธรรมต่อเด็กด้วย...อย่างเรื่องที่อาจารย์กำลังจะทำที่คิดว่าสำหรับที่นี่เราจะมีระบบmentorอยู่แล้วก็เลยมองว่าไม่จำเป็น”

ผู้ให้ข้อมูลคนที่ 1

(สัมภาษณ์ ณ วันที่ 20 ธันวาคม 2566)

“การเตรียมการสอนก็เน้นด้านเตรียมครูเป็นสำคัญ..มองเป็นเชิงกระบวนการการพัฒนา มันก็ใช้กระบวนการปรับเปลี่ยน knowledge skill คนที่เข้ามาเป็นอาจารย์มันต้องมี step มหาวิทยาลัย ก็จะมีหลักสูตรอบรมประจำทุกปีเค้าก็จะประชาสัมพันธ์มา มหาวิทยาลัยจะมีสำนักนวัตกรรมการสมัครออนไลน์... พอมีอาจารย์ใหม่เข้ามาเราจะปฐมนิเทศอาจารย์ใหม่... เรื่องทั่วไป... การจัดอบรมการเรียนการสอน การประเมิน อบรมหลักสูตรครูที่เลี้ยง เนื้อหาเหมือนกันเป็นไปตามหลักสูตรของสภา ศาสตร์และศิลป์การสอนอบรม 4 เดือน ถ้าจบปริญญาโทผ่านหลักสูตร 4 เดือน ต้องนิเทศได้ กับอีกกลุ่มหนึ่งที่ต้องนิเทศส่งไปอบรมสั้นๆ 1 ระบบ ไปอยู่กับอาจารย์ TA คณะจะจัดอบรมอาจารย์ที่เลี้ยง 1 อาทิตย์ หลักสูตรนี้อาจารย์จะเข้าไปได้ อบรมครูที่เลี้ยง 5 วัน การจัดการเรียนการสอน การประเมิน และmicro teaching แต่การประเมินนี้ยังไม่มีการขึ้นไปที่เวิร์คนะ การฝึกอบรบนี้ผู้เข้าร่วมอบรมมีความรู้ทักษะเพิ่มขึ้น...คือว่าเราจะรับปริญญาตรีเข้ามาแทนไปเป็นอาจารย์เพราะฉะนั้นกลุ่มอาจารย์ปีนี้ พอเค้าเข้ามาต้องมาเรียนรู้ฝึกตั้งแต่ประเมิน แล้วพื้ก็เอามาเช็คกับเค้า นี้คือระบบmentor ...”

ผู้ให้ข้อมูลคนที่ 2

(สัมภาษณ์ ณ วันที่ 24 ตุลาคม 2566)

“หลักสำคัญของการฝึกภาคปฏิบัติเเหอ อยู่ที่ครู ครูเราไม่ค่อยพอ...พอจะเตรียมครูเราก็ก๊ยกให้กับอาจารย์หัวหน้ากลุ่ม แล้วเวลาขึ้นไปนิเทศเราก็จะให้รุ่นพี่ประกบไปก่อน debatenกันไปก่อนที่นี้ก็ใช้การ KM การตั้งเป็นชมรมเป็น

คณะกรรมการ คล้าย ๆ สภาภาแพ มั่นอยู่ที่บริหารการทำงานว่าจะมาเจอกันได้ไหม ...แล้วก็ให้อบรมหลักสูตรศาสตร์และศิลป์ 4 เดือนถ้าไม่เคยเป็นครูมาก่อน...อีกอย่างอยากให้อาจารย์ทุกคนต้องมานั่งร่วมกันทำใบประเมิน ทำรูปรีด คิดดูว่าจะวัดอะไร debate กันไปเลย”

ผู้ให้ข้อมูลคนที่ 3

(สัมภาษณ์ ณ วันที่ 11 ธันวาคม 2566)

“เตรียมครูนิเทศสำคัญมาก แต่ถ้าเป็นtrainingก็จะเป็นหลักสูตรแค่ 1 เดือน เพราะว่าเมื่อก่อน มศว.เค้า จัดหลักสูตรอบรม 4 เดือน ไม้รู้ว่าตอนนี้ยังมีอยู่รึเปล่า แต่พอตอนนี้ ความที่เราขาดคนถ้าจะทำ 4 เดือนไม่ไหว ต้องจัดcourse สั้นๆเฉพาะหน้า น่าจะดีต่อพยาบาล แต่ไม่รู้ว่าจะแก้ปัญหาได้ไหม ยังมีปัญหาเรื่องเตรียมครูอยู่เยอะ...ถ้ามีหลักสูตรที่อาจารย์คนไหนว่างก็เข้ามาทำเป็นระบบคล้ายๆ ออนไลน์ได้เลยจะดีมาก เพราะถ้ารอให้หลายๆคนมาอบรมอย่างนี้ไม่ได้ เดี่ยวไม่มีคนทำงาน ถ้าที่น้องจะทำเป็นอย่างนี้ก็ดีนะ ถ้าดีจริงจะขอเอามาใช้บ้าง”

ผู้ให้ข้อมูลคนที่ 4

(สัมภาษณ์ ณ วันที่ 22 พฤศจิกายน 2566)

“หัวใจสำคัญของการจัดการเรียนการสอนภาคปฏิบัติก็คงอยู่ที่อาจารย์นิเทศ ...เรามีอาจารย์หมุนเวียนตลอดเวลา มีอาจารย์น้องใหม่เข้ามายังไม่คุ้นเคย ระบบของครูพี่เลี้ยง ระบบการมาประชุมปรึกษาหารือกัน และระบบการเทรนนิ่งที่ที่อ่านจากเอกสารของน้อง พี่คิดว่าถ้าทำได้จริงๆจะดีมากมีความจำเป็น...ของเราขณะนี้มีการส่งอาจารย์ไปเข้ารับการฝึกอบรม หลักสูตรการเรียนการสอนที่จัด 4 เดือน”

ผู้ให้ข้อมูลคนที่ 5

(สัมภาษณ์ ณ วันที่ 3 มกราคม 2567)

ประเด็นที่ 4 ปัญหา และ แนวทางแก้ไขในการกำหนดผลการเรียนภาคปฏิบัติ ผลการสัมภาษณ์สามารถแยกประเด็นตามบริบทใหญ่ของสถานศึกษาได้ว่า ปัญหาที่ทั้งสถาบันเก่า และใหม่ เผชิญเหมือนกันคือ ความไม่สอดคล้องกันของผลการประเมิน ความคลาดเคลื่อนจากผู้ประเมิน ส่วนปัญหาที่ต่างกันคือ ปริมาณของอาจารย์ผู้สอน และคุณภาพของกระบวนการในการเตรียมอาจารย์ผู้สอนภาคปฏิบัติสถาบันเก่ามีปริมาณอาจารย์เพียงพอ และมีระบบการเตรียมอาจารย์ที่มีคุณภาพ และพบปัญหาด้านนี้น้อย ในขณะที่สถาบันใหม่ประสบปัญหามีอาจารย์ไม่เพียงพอ อาจารย์มีอัตราการลาออกบ่อยกว่า มีระบบการเตรียมอาจารย์ที่ไม่ครบถ้วนสมบูรณ์ เช่น ไม่มีระบบ mentor ที่ชัดเจน ดังข้อมูลจากบทสัมภาษณ์

“ปัญหาที่พบมากที่สุด เรื่องการตัดเกรด พอทำกระบวนการแล้วที่นี้เกรดจะต้องไป approveตามลำดับ ลำดับนั้นก็ถือว่า 1. ในกลุ่มคุยกัน 2. ไปapproveในกลุ่มวิชา 3. มาapproveในกลุ่มกรรมการบริหารวิชาการ ประเมิน 4. ไปapproveในกรรมการบริหารคณะ แล้วเกรดจะส่งมาได้ต้องมีคะแนนดี และต้องมีการเซ็น approveบ้าง ได้มีการตรวจสอบ ใครบ้างที่มีการตรวจสอบมาตามลำดับ ในการประชุมครั้งไหนอย่างไรตามลำดับ มีหมด จนกระทั่งมาถึงกรรมการใหญ่สุดของคณะคือกรรมการบริหารประจำคณะ บางทีเราก็จะเจอคำถามว่าทำไม เด็กบางคน เรียน lecture ได้ความรู้ประมาณ C ประมาณ B แต่practice ทำไม่ได้ A อย่างนั้นจะคะ สัดส่วนของ ความรู้ตรงนั้นเนี่ยมันก็ค่อนข้างเยอะอยู่พอสมควร ตรงนั้นเราก็ให้ความรู้ประมาณ 30 ที่นั้นจะคะ แล้วก็practice 50 attitude 20 แต่ถ้าattitude ไม่ผ่านนี่จบเลิกเลยนะคะ อีกด้านหนึ่งนะคะเตรียมเอาreference เตรียมเอาตัวรูปrik มาดูว่า ... มีปัญหาอะไร เอา recordขึ้นมาดูสิ และเราจะสามารถดูได้ ทั้งชิ้นงาน แต่บางทีก็มีปัญหาเรื่องความเป็น ตัวตน หรือ คนนี้ advisee มัน เราจะรู้จักกันตอนคุยกันในกลุ่ม ถึงจะมีระบบชัดเจนอย่างนี้ก็มีปัญหา กว่าที่จะผ่านแต่ละ ขั้นตอน”

ผู้ให้ข้อมูลคนที่ 1

(สัมภาษณ์ ณ วันที่ 20 ธันวาคม 2566)

“ปัญหาการตัดสินผลการฝึก ...ด้วยความที่เราไม่ได้มีทักษะด้านนี้พอ ที่นี้จะมีปัญหาคือ เราก็จะมีคนอายุ มาก คนอายุน้อย เด็กก็จะให้เยอะ คนแก่จะให้หน่อย ก็จะมีปัญหาถ้าเป็นรุ่นเก่ามันก็จะเป๊ะกว่านี้ เด็กรุ่นใหม่ก็จะเข้ายาก ไม่ค่อยเป๊ะเท่าไร มันก็จะมีปัญหา ควรจะให้ข้อมูลตรงนี้ใหม่ วิธีการแก้ สมมติว่าเรามีวิชา ทุกวิชาที่มีคนนิเทศเยอะ เราก็ต้องมานั่งประชุม debateกัน เราก็คุยคะแนน 3 วอร์ดเธอนี้มันโด้ง คะแนนวอร์ดเธอนี้ต่ำมาก มันไม่ตรงกัน เอาpaper มาดู หลัง ๆ นี้เรียกว่า total record เดิมนี้เราจะบอกเลยว่า การขึ้นนิเทศเราต้องมีสมุดน้อย จด พฤติกรรมรายอาทิตย์เลยว่าเป็นอย่างไร บางคนอาจารย์ก็ไม่ได้ซีเรียสกับการขึ้นเวร บางคนซีเรียสมาก วันนี้ขึ้นซ้ำ วันนี้เป็นอย่างนี้ เค้าก็จะโน้ดเลยว่าคุณทำไม่itemนี้ในมุมมองฉันฉันว่ามันดี ...เค้าจะเอาตรงนี้มาdebateกัน ”

ผู้ให้ข้อมูลคนที่ 2

(สัมภาษณ์ ณ วันที่ 24 ตุลาคม 2566)

“ส่วนใหญ่เราจะรับอาจารย์เป็นอาจารย์นิเทศเยอะเพราะอาจารย์นิเทศไม่พอ ที่จะมาเป็นอาจารย์ประจำ เลขเซอร์ก็มีบ้าง ว่ามีปัญหาเวลาตัดเกรดส่งไป 8 คนได้หมดเลย คืออาจารย์ใหม่ เราไม่เข้าใจ ไม่กล้าให้คะแนน ปล่อยไปตามเรื่องตามราว 20% ไม่กล้าหักเค้าได้ ไม่สามารถหักเด็กได้ ข้อนี้เด็กได้ฟรี 20% ครูไม่สามารถแยกได้ เพราะเด็กดีหมดเลย หรือเพราะครูไม่สามารถแยกแยะความต่างได้ ถ้าแยกไม่ได้คุณไปดูรูปrik เช็ครูปrik ”

ผู้ให้ข้อมูลคนที่ 3

(สัมภาษณ์ ณ วันที่ 22 พฤศจิกายน 2566)

“ตอนนี้ทุกคนมีปัญหา พี่เลี้ยงมีปัญหาว่าไม่แฟร์ คนนี้มาอยู่กับน้องคะแนนต่ำเตี้ย คนนี้มีอยู่กับพี่คะแนน ฟุงโสด อาจารย์ที่ไปนิเทศก็ไม่ค่อยรู้ พี่ก็ไม่รู้ว่าศาสตร์ และศิลป์เค้าสอนอะไรบ้าง เพราะว่าดูแล้วน้องก็ยังไม่รู้ เค้าก็ไปฝึกสอนกันนะ ไปฝึกสอนภาคปฏิบัติ เน้นไปทางทฤษฎีหรือเปล่าคะ เค้าก็ฝึกทั้งสองแบบ แล้ววิชาที่ตัดเกรดมันจะไปเน้นในวิชาที่เกี่ยวข้องวิชาภาคทฤษฎีอย่างเช่น พี่ก็ไปดูที่วอร์ดนะ แต่เค้าก็ไม่ได้ลงลึก น่าจะเป็นการดูทั่วไป อาจารย์ที่ไปอบรมมาแล้วก็ยังไม่ดีขึ้น เวลาตัดเกรดกันก็ยังไม่แม่นในหลักการว่าจะนำความรู้พวกนี้มาใช้อย่างไร ให้ลดปัญหาความคิดเห็นที่ไม่ตรงกัน... ก็พยายามแก้ด้วยการประชุมปรึกษาหารือกัน แนะนำกัน จดบันทึกให้มากขึ้น”

ผู้ให้ข้อมูลคนที่ 4

(สัมภาษณ์ ณ วันที่ 11 ธันวาคม 2566)

“ ปัญหาของเราคือ มาตรฐานในการตัดสินคุณภาพผู้เรียนเรามีระบบการทวนสอบ การประชุมปรึกษาหารือกันในสาขาก่อนการตัดสินผลการฝึก แต่ยังเป็นที่น่าสงสัยว่ามันไม่work ตัวอย่างก็ของเรานี้เด็กปี 4 คนหนึ่งที่ชื่อ... เพิ่งฝึกอันนี้ผ่านไปทางพี่ที่วอร์ดซึ่งได้ประเมินมา ในข้อที่มีดาว (*) เค้าได้แค่ 2 ซึ่งตามเกณฑ์เด็กคนนี้ต้องได้ F นะ ...ในที่สุดเราก็ส่งเด็กคนนี้ไปปรึกษาตอนนี้กำลังไปอยู่ที่... พอเด็กมาอยู่ในสถานการณ์ที่กดดันเค้าไม่สามารถทำงานโฟกัสกับงานได้เป็นเหมือนว่าไม่รู้จะทำอะไรเลย set priority ของงานไม่ได้ เราก็อยกกลับมาดูว่าเด็กคนนี้ผ่านวอร์ดอื่น ๆ เนี่ยเค้าผ่านมาได้ในระดับที่ดีนะ เวลาที่defendเกรดนี้พี่จะจดไว้เสมอว่าเกรด 5 คนสุดท้ายไม่มีชื่อเค้าเลยแสดงว่าเค้าก็เป็นที่มีสติปัญญาพอสมควรนะ ทั้งภาคทฤษฎีภาคปฏิบัติไม่เคย topless แต่พอมายปี 4 มาเจอพฤติกรรมแบบนี้ก็คิดว่ามันเกิดจากอะไร นี่ถ้าเราดูจากที่มีลักษณะตัวดาวกำหนดเนี่ยเราจะstrict กับมันแค่ไหน มันก็เป็นเกณฑ์อย่างหนึ่งนะที่เราให้น้ำหนักให้ความสำคัญกับเรื่องนั้นมากกว่าเพราะว่าแต่ละข้อ 5 4 3 2 1 คะแนนก็เท่ากัน น้ำหนักแต่ละข้อเท่ากันหมดเลย แต่ถ้าเป็นด้านสอบข้อเขียนมันชัดเจนสมมติว่าออกข้อสอบกาหวักระดาษ อันนั้นคือจริยธรรมแม้ว่ามีความรู้เต็มที่แต่ไม่ได้สอบ แต่ในภาคปฏิบัติอาจจะไม่ชัดเท่าไหร่นัก คงต้องกลับมาทบทวน”

ผู้ให้ข้อมูลคนที่ 5

(สัมภาษณ์ ณ วันที่ 3 มกราคม 2567)

1.3 ผลการสำรวจความต้องการในการใช้รูบริค จากผู้ทรงคุณวุฒิ ภายในคณะพยาบาลศาสตร์ จำนวน 7 ท่าน (รายชื่อในภาคผนวก ก หน้า 243) และการขอความคิดเห็นจากรองคมนตรีฝ่ายวิชาการของคณะพยาบาลศาสตร์สวนดุสิต ในการสร้างรูบริคการประเมินภาคปฏิบัติ และบทวิทัศน์ที่ใช้เป็นงานสำหรับการประเมินร่วมกับรูบริค ได้ผลสรุปของการสำรวจความต้องการในการใช้รูบริคการประเมินภาคปฏิบัติได้ประเด็นสำคัญคือ ต้องการใช้อูบริคการประเมินแบบ 4 ระดับ และได้

กรอบแนวคิดที่จะนำมาใช้ในการสร้างรูปก็คือ กรอบคุณวุฒิระดับอุดมศึกษาแห่งชาติ ร่วมกับเอกลักษณ์ของมหาวิทยาลัย และได้สถานการณ์ที่จะใช้ในการทำวิทัศน์

ตัวอย่างสถานการณ์ที่ 1 ผู้ป่วยหญิงอายุ 56 ปี มาด้วยอาการ 1 วันก่อนมารพ. หายใจเหนื่อย หน้าบวม ปัสสาวะไม่ออก ประเมินสัญญาณชีพพบว่า BP = 130/90 mmHg PR 88 /min RR = 22 /min T = 38. °C แพทย์ให้นอนโรงพยาบาล และมีแผนการรักษาให้ O₂ cannula 3 L/min On 5% D/N/2 1000 ml vein drip 80 ml/hr. ใส่คาสายสวนปัสสาวะ เพื่อประเมินการขับถ่ายปัสสาวะ และให้สังเกตอาการเปลี่ยนแปลง และให้ยาทางปาก

สรุปประเด็นสำคัญจากการสำรวจสภาพปัญหาการประเมินภาคปฏิบัติ

สรุปประเด็นสำคัญจากการสำรวจสภาพปัญหา ในผู้ให้ข้อมูล 3 กลุ่ม มีรายละเอียดดังนี้ ผลการสำรวจปัญหาจากข้อมูลเชิงปริมาณพบว่า ส่วนใหญ่ไม่พบปัญหา แต่เมื่อพิจารณาความต้องการในการเตรียมความพร้อมด้านสมรรถนะการประเมินเพื่อการประเมินการฝึกปฏิบัติงานของนักศึกษาพยาบาลส่วนใหญ่คิดเห็นว่าตนเองมีความจำเป็นต้องได้รับการเตรียมความพร้อม และต้องการให้หน่วยงานจัดหลักสูตรระยะสั้นเพื่อเสริมสร้างสมรรถนะการประเมินด้วย ซึ่งสอดคล้องกับข้อมูลเชิงคุณภาพจากคำถามปลายเปิด ซึ่งสะท้อนให้เห็นปัญหา ซึ่งขัดแย้งกับคำตอบของคำถามปลายปิดว่าส่วนใหญ่ ไม่มีปัญหา ร่วมกับการพิจารณาข้อมูลเชิงคุณภาพที่ได้รับจากผู้บริหารฝ่ายวิชาการที่พบปัญหาเกี่ยวกับการตัดสินผลการประเมิน การใช้ความเป็นอัตโนมัติของผู้ประเมิน หรือความลำเอียงต่างๆ จึงสรุปได้ว่าจำเป็นต้องพัฒนาโมดูลเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลในการประเมินภาคปฏิบัตินักศึกษาพยาบาล และ ควรอยู่ในรูปแบบระบบปฐมนิเทศด้วยการใช้ครูพี่เลี้ยง รองลงมาคือ การฝึกอบรม แต่เนื่องจากระบบครูพี่เลี้ยงมีข้อจำกัดในหลายสถาบันที่มีจำนวนอาจารย์ไม่เพียงพอ จึงควรใช้การเตรียมความพร้อมด้วยวิธีการฝึกอบรมโดยเลือกใช้กลยุทธ์ในการเสริมสร้างสมรรถนะคือ การฝึกอบรม การสอนงาน ระบบพี่เลี้ยงและการให้คำปรึกษา เนื่องจากมีความเป็นไปได้ และเป็นวิธีที่นิยมใช้ และวิธีที่ผู้ให้ข้อมูลเลือกใช้หากมีการฝึกอบรมคือ ความรู้พื้นฐานเกี่ยวกับการวัดและประเมินผลภาคปฏิบัติ การฝึกอบรมเกี่ยวกับความคลาดเคลื่อนจากผู้ประเมิน การฝึกอบรมการสังเกต การฝึกอบรมกรอบของการอ้างอิง และการฝึกอบรมการตัดสินใจตามลำดับ

2. ผลการออกแบบโมดูลการฝึกอบรม

การออกแบบโมดูลการฝึกอบรม ได้นำข้อมูลจากการสำรวจสภาพปัจจุบัน มาบูรณาการกับ มโนทัศน์เกี่ยวกับการพัฒนาโมดูลทำให้ได้หน่วยการเรียนรู้ 5 หน่วยดังนี้

2.1 หน่วยการเรียนรู้ที่ 1 ความรู้พื้นฐานด้านการประเมินภาคปฏิบัติ

2.1.1 เนื้อหาประกอบด้วย ความหมาย หลักการประเมินผลภาคปฏิบัติ ขั้นตอนการ ดำเนินการประเมินภาคปฏิบัติ เครื่องมือ และเกณฑ์ในการแปลความหมายของเครื่องมือที่ใช้ในการ ประเมินภาคปฏิบัติ การตรวจสอบคุณภาพการประเมินภาคปฏิบัติ การตัดสินผลการประเมิน ภาคปฏิบัติ โดยมีน้ำหนักของเนื้อหาเน้นที่การตัดสินผลการประเมิน ตามที่ระบุในสภาพปัญหาที่พบใน ขั้นตอนการสำรวจสภาพปัญหา

2.2.2 กิจกรรมการฝึกอบรมประกอบด้วย การบรรยาย และการทดสอบความรู้จากแบบ วัดความรู้ท้ายบท เผลยคำตอบถูก และอภิปรายร่วมกัน

2.2 หน่วยการเรียนรู้ที่ 2 ความรู้เกี่ยวกับความคลาดเคลื่อนจากผู้ประเมิน มีเนื้อหาของ หน่วยการเรียนรู้ และกิจกรรมดังนี้

2.2.1 เนื้อหาประกอบด้วย ความหมาย คุณลักษณะ สาเหตุ และวิธีแก้ไข ความคลาด เคลื่อนที่เกิดจากผู้ประเมินแต่ละประเภท และการหลีกเลี่ยงการเกิดความคลาดเคลื่อนจากผู้ประเมิน

2.2.2 กิจกรรมการฝึกอบรมประกอบด้วย

- (1) การบรรยายเนื้อหาโดยมีข้อมูลในรูปแบบกราฟของความคลาดเคลื่อนแต่ละประเภท
- (2) แบ่งกลุ่ม แจกใบงานที่มีข้อมูลของการให้คะแนนการประเมินที่มีความคลาดเคลื่อนแต่ ละประเภทอยู่โดยให้อภิปรายในกลุ่มว่าข้อมูลที่ได้เป็นความคลาดเคลื่อนประเภทใด และเฉลยคำตอบที่ ถูกต้อง และอภิปรายผลร่วมกัน

ตัวอย่างใบงานใน กิจกรรมท้ายหน่วยการเรียนรู้ที่ 2

จงนำข้อมูลต่อไปนี้ไปวิเคราะห์เพื่อตอบคำถามต่อไปนี้

1. จากข้อมูล เป็นการประเมินรูบริคในมิติด้านทักษะการปฏิบัติกิจกรรมพยาบาลมีระดับคะแนนตั้งแต่ 1-5 ด้วยอาจารย์ 3 คน ต่อนักศึกษา 8 คน

อาจารย์	นักศึกษาคนที่							
	1	2	3	4	5	6	7	8
คนที่1	2	3	3	2	2	3	3	2
คนที่ 2	4	4	4	4	4	4	5	5
คนที่ 3	1	4	2	3	4	5	4	2

ผู้วิจัยตั้งคำถามให้ผู้เข้ารับการฝึกอบรมร่วมกันตอบ

คำถามที่ 1 อาจารย์ท่านใดมีแนวโน้มที่จะเกิดความคลาดเคลื่อนจากการให้คะแนนแบบปล่อยคะแนน

คำถามที่ 2 อาจารย์ท่านใดมีแนวโน้มที่จะเกิดความคลาดเคลื่อนจากการให้คะแนนแบบกดคะแนน

2. จากข้อมูล เป็นการประเมินรูบริคในมิติต่าง ๆ 6 มิติ มีระดับคะแนนตั้งแต่ 1-5 ด้วยอาจารย์ 3 คน ในการประเมินนักศึกษา 1 คน

อาจารย์	มิติกรวัด 6 มิติ					
	ทักษะปฏิบัติ	ประสิทธิภาพปฏิบัติ	การป้องกันอันตราย	ความปลอดภัย	การตัดสินใจ	การระดมทรัพยากร
คนที่1	2	2	2	2	1	1
คนที่ 2	4	5	4	3	4	4
คนที่ 3	3	4	2	3	4	5

ผู้วิจัยตั้งคำถามให้ผู้เข้ารับการฝึกอบรมร่วมกันตอบ

ข้อคำถาม อาจารย์ท่านใดมีแนวโน้มที่จะเกิดความคลาดเคลื่อนแบบฮาโล

2.3 หน่วยการเรียนรู้ที่ 3 กรอบของการอ้างอิง เน้นให้เห็นความสำคัญว่าวิธีการนี้จะช่วยสร้างความเข้าใจให้ตรงกันในการใช้เกณฑ์การประเมิน เนื่องจากพบปัญหานี้ในขั้นตอนการสำรวจสภาพปัญหา มีเนื้อหา และกิจกรรมดังนี้

2.3.1 เนื้อหาประกอบด้วย ความหมาย ความสำคัญ และการอธิบายความหมายของระดับคุณภาพของเครื่องมือที่ใช้ในการประเมิน

2.3.2 กิจกรรมการฝึกอบรมประกอบด้วย

- (1) บรรยายเนื้อหา และการใช้รูปรีดประกอบการบรรยาย
- (2) เข้ากลุ่ม กลุ่มละไม่เกิน 7 ให้ศึกษานิยามของแต่ละสมรรถนะจากใบงานที่มอบหมาย จากนั้นอภิปรายกลุ่มถึงนิยามของแต่ละสมรรถนะ ควรมีความลักษณะอย่างไร
- (3) ศึกษาคำอธิบายระดับคุณภาพของงานการปฏิบัติจากรูปรีด คุณลักษณะของงานในแต่ละระดับคุณภาพ และอภิปรายว่าแต่ละคุณลักษณะควรมีลักษณะเป็นอย่างไร
- (4) ศึกษาสื่อวีดิทัศน์ที่มีงานการปฏิบัติที่มีระดับคุณภาพของงาน ไล่ระดับจากต่ำไประดับสูงจำนวน 4 ระดับตามลำดับ
- (5) ทดลองให้คะแนนการประเมินจากสื่อที่ได้จัดไว้ให้ในระดับคุณภาพสูงสุด ปานกลาง และต่ำสุด และนำผลการประเมินมาพิจารณาเปรียบเทียบกับผลการให้คะแนนจากผู้เชี่ยวชาญตามระดับคุณภาพงานที่ควรจะเป็น
- (6) วิทยากรให้ข้อมูลย้อนกลับกับกลุ่มผู้เข้ารับการฝึกอบรมเกี่ยวกับผลการประเมินของผู้เข้ารับการฝึกอบรมกับผลการประเมินที่ควรจะเป็นเทียบกับผลของผู้เชี่ยวชาญ เพื่อให้เกิดการพัฒนาการประเมินที่มีคุณภาพใกล้เคียงกับงานการประเมินที่แท้จริง

2.4 หน่วยการเรียนรู้ที่ 4 การฝึกการสังเกต มีเนื้อหา และกิจกรรมดังนี้

2.4.1 เนื้อหาประกอบด้วย ความหมาย ความสำคัญ กระบวนการของการสังเกต กฎของการสังเกต กลยุทธ์ที่ใช้สร้างทักษะการสังเกต

2.4.2 กิจกรรมการฝึกอบรมประกอบด้วย

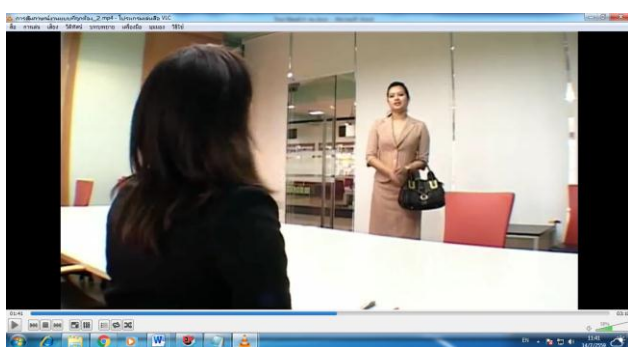
- (1) การบรรยายเนื้อหา และการนำเสนอตัวอย่างการจดบันทึกพฤติกรรม
- (2) หลังการบรรยาย ทำกิจกรรมกลุ่มโดยแบ่งกลุ่มย่อย กลุ่มละไม่เกิน 7 คน ให้ชมวีดิทัศน์ ประมาณ 5 นาที

(3) เขียนคำตอบในกระดาษว่าง (ห้ามพูดคุยกัน) ตอบคำถามว่าเกิดอะไรขึ้นในวิดีโอ และเขียนอธิบายพฤติกรรมของบุคคลในวิดีโอ

(4) นำผลการตอบคำถามมาเปรียบเทียบกันภายในกลุ่ม อภิปรายกลุ่มถึง เหตุผลที่ทำให้คำตอบของแต่ละคนแตกต่างกัน

(5) วิทยากรประจำกลุ่มเฉลยคำตอบ และอภิปรายภายในกลุ่ม

ตัวอย่างวิดีโอการสัมภาษณ์งานใช้ในกิจกรรมท้ายหน่วยการเรียนรู้ที่ 4



2.5 หน่วยการเรียนรู้ที่ 5 การฝึกการตัดสินใจ มีเนื้อหา และกิจกรรมดังนี้

2.5.1 เนื้อหาประกอบด้วย ความสำคัญ ทฤษฎีพื้นฐาน กระบวนการตัดสินใจตามสัญชาตญาณ และกระบวนการตัดสินใจตามหลักการที่ถูกต้อง

2.5.2 กิจกรรมการฝึกอบรมประกอบด้วย

(1) การบรรยายเนื้อหา

(2) ให้ชมวิดีโอสั้น ๆ ประมาณ 5 นาที ให้เกณฑ์ในการประเมิน และให้ผู้เข้ารับการฝึกอบรมเขียนกระบวนการตัดสินใจของตนเอง จากนั้นสุ่มถามกระบวนการตัดสินใจของผู้เข้ารับการอบรม และอภิปรายร่วมกันภายในกลุ่ม

ตัวอย่างวิดีโอการอาบน้ำเด็กแบบบ้าน ๆ ใช้ในกิจกรรมท้ายหน่วยการเรียนรู้ที่ 5



เมื่อสิ้นสุดเนื้อหาของการฝึกอบรม ฝึกการประเมินต่อไปเป็นระยะๆ โดยให้ผู้เข้ารับการฝึกอบรมทุกคนทำการประเมินงาน 1 ชุด และหยุดทำการตรวจสอบ และอภิปรายผลการให้คะแนนร่วมกับกลุ่ม และให้ข้อมูลย้อนกลับกลับไปจนเมื่อใกล้สิ้นสุดระยะเวลาที่กำหนด จึงมีการตรวจสอบความแม่นยำของการประเมินอีกครั้งโดยผู้ประเมินจะผ่านเกณฑ์การเป็นผู้ประเมินที่ดี เมื่อมีคะแนนการประเมินที่สอดคล้องกับผลการประเมินจากผู้ทรงคุณวุฒิ ร้อยละ 70 ขึ้นไป โดยพิจารณาจากค่า Rater agreement index (RAI)

3. การสร้างเครื่องมือ และตรวจสอบคุณภาพเครื่องมือ

การสร้างเครื่องมือ และตรวจสอบคุณภาพของเครื่องมือที่ใช้ในโมดูลการฝึกอบรม และใช้ในการศึกษาผลของโมดูลการความรู้ ทักษะ และทัศนคติ รวมถึงความคลาดเคลื่อนจากผู้ประเมิน ประกอบด้วย (1) ชุดฝึกประเมินทักษะภาคปฏิบัติ ประกอบด้วย ระบุวิธีการประเมินภาคปฏิบัติ และวิธีทัศนสถานการณ์จำลองการฝึกปฏิบัติงานของนักศึกษาพยาบาลชั้นปีที่ 2 ในรายวิชาปฏิบัติการพยาบาลพื้นฐาน (2) ชุดประเมินความพึงพอใจ ประกอบด้วย แบบวัดความพึงพอใจต่อการใช้โมดูลการฝึกอบรม และแบบสัมภาษณ์เฉพาะกลุ่ม (3) แบบวัดความรู้เกี่ยวกับการประเมินภาคปฏิบัติ (4) แบบวัดทัศนคติเกี่ยวกับการประเมินภาคปฏิบัติ และ (5) คู่มือที่ใช้ประกอบการฝึกอบรม ดังรายละเอียดต่อไปนี้

3.1 การสร้างเครื่องมือที่ใช้ในการฝึกอบรม

3.1.1 ชุดฝึกทักษะการประเมินภาคปฏิบัติ ประกอบด้วยระบุโรค และวิธีทัศน

3.1.1.1 การสร้างระบุวิธีการประเมินภาคปฏิบัติได้ผลการสำรวจความต้องการในการใช้ระบุโรคเพื่อการประเมินภาคปฏิบัติ ของผู้ทรงคุณวุฒิภายในคณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต มีมติให้ใช้ระบุโรคที่มีระดับคะแนน 4 และใช้กรอบมาตรฐานคุณวุฒิระดับอุดมศึกษาแห่งชาติ 6 สมรรถนะ คือ 1) คุณธรรม จริยธรรม 2) ความรู้ 3) ทักษะทางปัญญา 4) ความสัมพันธ์ระหว่างบุคคลและความรับผิดชอบ 5) ทักษะการวิเคราะห์เชิงตัวเลข การสื่อสาร และการใช้เทคโนโลยีสารสนเทศ และสำหรับการพยาบาลศึกษามีการเพิ่มมาตรฐานคุณวุฒิอีกด้านหนึ่งคือ 6) ทักษะการปฏิบัติทางวิชาชีพ สาขาพยาบาลศาสตร์ และ 7) เอกลักษณ์ของ คือ มีบุคลิกภาพที่ดี มีความประณีต และพิถีพิถันในการทำงาน การสำรวจความต้องการในการใช้ระบุโรคการประเมินภาคปฏิบัติของผู้ทรงคุณวุฒิภายในคณะพยาบาลศาสตร์สวนดุสิต (ภาคผนวก ก หน้า 243) จากนั้นนำผลที่ได้ไปสร้างระบุวิธีการประเมินภาคปฏิบัติดังแสดงในตาราง 14

ตารางที่ 14 การสำรวจความต้องการในการใช้รูปบริการประเมินภาคปฏิบัติ

ข้อเสนอแนะจากผู้ทรงคุณวุฒิ ภายในคณะพยาบาลศาสตร์	IOC	ข้อคิดเห็น หรือข้อเสนอแนะ
1. นิยามของแต่ละสมรรถนะ	0.71	ควรทบทวนนิยามของแต่ละสมรรถนะใหม่ และดึงคำสำคัญของแต่ละสมรรถนะมา หรือการตีความแต่ละสมรรถนะให้ชัดเจน ควรตัดระดับที่ 5 advance ที่เนื่องจากไม่สอดคล้องกับสมรรถนะของ นค. ปี 2 และขยับระดับ 4 ขึ้นมาเป็นระดับ 5 ระดับ 1 ขึ้นมาเป็นระดับ 2 เพิ่มระดับ 1 ที่มีระดับ below standard มากขึ้น (คะแนนค่อนข้างต่ำด้าน ลบมากกว่า)
2. สมรรถนะที่ 1 คุณธรรม จริยธรรม	0.86	เกณฑ์ที่ 1 มีความละเอียดมากเกินไป ทำให้กระชับกว่านี้ โดยควรปรับ เกณฑ์เน้นที่จุดสำคัญของแต่ละสมรรถนะ เช่น สมรรถนะทั้งหมดนิยามให้ กระชับ เกณฑ์ข้อ 1 แสดงพฤติกรรมสะท้อนคุณธรรมจริยธรรม เกณฑ์ข้อ 2 การควบคุมตนเอง เกณฑ์ข้อ 3 ยอมรับความแตกต่าง ความเชื่อและให้เกียรติ เกณฑ์ข้อ 4 วินัย ตรงต่อเวลา เกณฑ์ข้อ 5 จิตอาสา เกณฑ์ข้อ 6 รักษาจรรยาบรรณวิชาชีพ เกณฑ์ข้อ 7 เป็นแบบอย่างที่ดีด้านสุขภาพ เกณฑ์ข้อ 8 พึงกษัตริย์ของผู้ป่วย
3. สมรรถนะที่ 2 ความรู้	0.57	เกณฑ์ที่ 2-7 เกณฑ์ไม่ค่อยตรงประเด็นของสมรรถนะ ควรตัดออก
4. สมรรถนะที่ 3 ทักษะทาง ปัญญา	1.00	ทำให้ข้อความมีความกระชับมากกว่านี้
5. สมรรถนะที่ 4 ความสัมพันธ์ ระหว่างบุคคลและความ รับผิดชอบ	1.00	ทำให้ข้อความมีความกระชับมากกว่านี้
6. สมรรถนะที่ 5 ทักษะการ วิเคราะห์เชิงตัวเลขฯ	1.00	ทำให้ข้อความมีความกระชับมากกว่านี้
7. สมรรถนะที่ 6 ทักษะการ ปฏิบัติทางวิชาชีพ	1.00	ทำให้ข้อความมีความกระชับกว่านี้
8. สมรรถนะที่ 7 เอกลักษณ์ มหาวิทยาลัย	1.00	ควรทำให้ข้อความกระชับกว่านี้ และ เพิ่มเอกลักษณ์ด้านการทำงานด้วย ความพิถีพิถัน

n = 7

หลังการตรวจสอบคุณภาพของรูปบริการประเมินภาคปฏิบัติ โดยผู้ทรงคุณวุฒิภายในคณะ
พยาบาลศาสตร์แล้วนำไปปรับปรุง และให้ผู้ทรงคุณวุฒิ 9 ท่าน (ภาคผนวก ก หน้า 257) ตรวจสอบ
ความตรงเชิงเนื้อหาจนได้นิยามที่ใช้ในแต่ละระดับคุณภาพดังนี้

การให้นิยามการแปลค่าคะแนน 4 ระดับ

คะแนน 1 คือ ต่ำกว่ามาตรฐาน ต้องปรับปรุง (Below basic) นักศึกษาแสดงให้เห็นถึงการมีความรู้
ทักษะและมีความพร้อมในการปฏิบัติ ในเรื่องนั้น อย่างจำกัด หรือไม่สามารถปฏิบัติ
ด้วยตนเองได้ต้องอาศัยการชี้แนะ และกำกับอย่างใกล้ชิด บ่อยครั้ง หรือ อยู่เสมอ
ความถี่ในการแสดงสมรรถนะ พบพฤติกรรมตั้งแต่ร้อยละ 60 ของเหตุการณ์ขึ้นไป

คะแนน 2 คือ ระดับพื้นฐาน (พอใช้) (Basic) นักศึกษาแสดงให้เห็นถึงการมีความรู้ ทักษะ และความ
พร้อมในการปฏิบัติ ในเรื่องนั้นบางส่วนไม่ครอบคลุมสาระสำคัญทั้งหมด สามารถ
ปฏิบัติได้ด้วยตนเองบางส่วน ต้องอาศัยการชี้แนะบางส่วนต้องอาศัยการกำกับอย่าง
ใกล้ชิดบางครั้ง ความถี่ของการแสดงสมรรถนะนั้นตั้งแต่ประมาณร้อยละ 60 ขึ้นไป
ของเหตุการณ์

คะแนน 3 คือ ระดับมาตรฐาน (ดี) (Meet standard) นักศึกษาแสดงให้เห็นว่ามีความรู้ ทักษะ และ
ความพร้อมในการปฏิบัติ ครอบคลุมสาระสำคัญทั้งหมด สามารถปฏิบัติด้วยตนเองได้
เป็นส่วนใหญ่ ต้องการการชี้แนะเป็นส่วนน้อยความถี่ของการแสดงสมรรถนะนั้น
บ่อยครั้ง หรือเป็นส่วนใหญ่ คือประมาณตั้งแต่ร้อยละ 80 ของเหตุการณ์ขึ้นไป
(ประมาณร้อยละ 80 แต่ไม่ถึงร้อยละ 90)

คะแนน 4 คือ ระดับเหนือมาตรฐาน (ดีมาก) (Above standard) นักศึกษาแสดงให้เห็นว่ามีความรู้
และทักษะในเรื่องนั้น ครอบคลุมสาระสำคัญทั้งหมด อาจครอบคลุมรายละเอียดเกือบ
ทั้งหมด หรือ ทั้งหมด สามารถปฏิบัติด้วยตนเองได้อาจไม่ต้องการข้อชี้แนะ หรือ
ต้องการเพียงเล็กน้อย และมีความถี่ในการแสดงสมรรถนะนั้น อย่างสม่ำเสมอ คือ
ประมาณร้อยละ 90 ของเหตุการณ์ขึ้นไป (ประมาณ 90-100%)

เมื่อได้นิยามของแต่ละระดับคุณภาพของงานภาคปฏิบัติแล้วนำไปเขียนระดับคุณภาพของแต่ละ
สมรรถนะจนครบ 26 สมรรถนะย่อยใน 7 สมรรถนะหลัก และได้ระบุบริสำหรับประเมินภาคปฏิบัติ
(ภาคผนวก ข หน้า 268) ดังตัวอย่างต่อไปนี้

ตารางที่ 15 ตัวอย่างรูปบริการประเมินภาคปฏิบัติ สมรรถนะที่ 1

สมรรถนะ/เกณฑ์ ที่ต้องประเมิน	ระดับคุณภาพ			
	ต่ำกว่าพื้นฐาน (คะแนน 1)	พื้นฐาน (คะแนน 2)	มาตรฐาน (คะแนน 3)	สูงกว่ามาตรฐาน (คะแนน 4)
1. คุณ ธรรม จริยธรรม 1.1 ความซื่อสัตย์ รับผิดชอบ มี วินัย	1. - แสดงพฤติกรรม สะท้อน ความซื่อสัตย์ น้อยกว่าร้อยละ 100 ของเหตุการณ์ - มีความรับผิดชอบ มี วินัย น้อยกว่าร้อยละ 80 ของ เหตุ การณ์ ทั้งหมดของเวลาขึ้นฝึก	2. - แสดงพฤติกรรม สะท้อน ความซื่อสัตย์ ร้อยละ 100 ของ เหตุการณ์ - มีความรับผิดชอบ มี วินัย ประมาณร้อยละ 80 ของเหตุการณ์ ทั้งหมดของเวลาขึ้นฝึก	3. - แสดงพฤติกรรม สะท้อน ความซื่อสัตย์ ร้อยละ 100 ของ เหตุการณ์ - มีความรับผิดชอบ มี วินัยประมาณร้อยละ 90 ของเหตุการณ์ ทั้งหมดของเวลาขึ้นฝึก	4. - แสดงพฤติกรรม สะท้อน ความ ซื่อสัตย์ร้อยละ100 ของเหตุการณ์ - มีความรับผิดชอบ มีวินัยร้อยละ 100 ของ เหตุ การณ์ ทั้งหมดของเวลาขึ้น ฝึก
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<p>หมายเหตุ</p> <p>เกณฑ์ข้อ 1 หากพบว่าอยู่ในระดับต่ำกว่า 2 ถือว่าไม่ผ่านการพิจารณาการประเมินภาคปฏิบัติ</p> <p>ตัวอย่างพฤติกรรมสะท้อน ความซื่อสัตย์ รับผิดชอบ มีวินัย</p> <ul style="list-style-type: none"> - ความซื่อสัตย์ เช่น ไม่คัดลอกผลงานของผู้อื่น ไม่แจ้ง หรือไม่ใช้ข้อมูลอันเป็นเท็จ หรือไม่บิดเบือนข้อมูลอันจะทำให้เกิดอันตรายต่อผู้ป่วย - ความมีวินัย เช่น ขึ้นปฏิบัติงานตรงเวลา ส่งงานที่ได้รับมอบหมายตรงเวลา - ความรับผิดชอบ เช่น ปฏิบัติงานที่ได้รับมอบหมายเสร็จสิ้นครบถ้วน หรือ ไม่ละทิ้งหน้าที่ 				

ผลจากการสำรวจความต้องการในการใช้รูปบริการประเมินภาคปฏิบัติของผู้ทรงคุณวุฒิ
ภายในคณะฯ แยกระดับกิจกรรมการพยาบาลเป็นระดับ ง่าย ปานกลาง ยากดังแสดงในตารางที่ 16
เพื่อนำไปจัดทำสถานการณ์จำลอง และบทวีดีทัศน์ที่มีกิจกรรมการพยาบาลในทุกระดับความยาก
ที่สอดคล้องกับสมรรถนะของนักศึกษาชั้นปีที่ 2 ที่จะต้องใช้ร่วมกับรูปบริการประเมินภาคปฏิบัติ

ตารางที่ 16 ความคิดเห็นของผู้ทรงคุณวุฒิภายในฯ ในการเลือกระดับความยากง่ายของกิจกรรมการพยาบาล

หัวข้อกิจกรรมการพยาบาล	ความถี่ของความคิดเห็นของผู้ทรงคุณวุฒิต่อระดับความยากง่ายของกิจกรรมการพยาบาล			
	ง่าย	ปานกลาง	ยาก	สรุป
1. หลักการพยาบาลและเทคนิคการป้องกันและควบคุมการแพร่กระจายเชื้อ				
1.1 การล้างมือในห้องพักตัด	6	1		ง่าย
1.2 การใส่เสื้อห้องพักตัด	5	2		ง่าย
1.3 การใส่ถุงมือ	5	2		ง่าย
1.4 การหยิบจับของปราศจากเชื้อ	5	2		ง่าย
1.5 การทำแผล		4	3	ปานกลาง
1.6 การพันผ้า	1	4	1	ปานกลาง
2. การดูแลความสะอาดและความปลอดภัย				
2.1 การออกกำลังกาย	3	4		ปานกลาง
2.2 การจัดท่านอน	4	1	2	ง่าย
2.3 ทำเตียง	5	1		ง่าย
3. การอาบน้ำบนเตียง การสวนล้างช่องคลอด				
3.1 การทำความสะอาดผิวหนัง ผม ปาก หู ตา จมูก เท้าเล็บ	1	5		ปานกลาง
3.2 การทำความสะอาดอวัยวะสืบพันธุ์		5	2	ปานกลาง
3.3 การสวนล้างช่องคลอด	1	1	5	ยาก
4. การใช้ความร้อนเย็นในการบำบัด				
4.1 ประคบร้อน	5	2		ง่าย
4.2 ประคบเย็น	5	2		ง่าย
4.3 เช็ดตัวลดไข้		5	2	ปานกลาง
4.4 นิ่งแช่กัน	3	4		ปานกลาง
5. หลักและวิธีการวัดและบันทึกสัญญาณชีพ		5	1	ปานกลาง
6. การรับใหม่และจำหน่ายผู้ป่วย	1	4	2	ปานกลาง
7. หลักการและเทคนิคการพยาบาลเพื่อตอบสนองความต้องการด้านอาหารและน้ำ				
7.1 การประเมินภาวะขาดน้ำและโภชนาการ	1	2	4	ปานกลาง
7.2 การทำ/O	3	4		ปานกลาง
7.3 การใส่NG		1	6	ยาก

ตารางที่ 16 ความคิดเห็นของผู้ทรงคุณวุฒิภายใน ในการเลือกระดับความยากง่ายของกิจกรรมการพยาบาล (ต่อ)

หัวข้อกิจกรรมการพยาบาล	ความถี่ของความคิดเห็นของผู้ทรงคุณวุฒิต่อระดับความยากง่ายของกิจกรรมการพยาบาล			
	ง่าย	ปานกลาง	ยาก	สรุป
8. หลักการและเทคนิคการพยาบาลเพื่อตอบสนองความต้องการการ ซึบถ่าย				
8.1 สอนปัสสาวะ			7	ยาก
8.2 ใส่ถุงยางเพื่อระบายปัสสาวะ	1	4	2	ปานกลาง
8.3 การสวนอุจจาระ	2	4	1	ปานกลาง
8.4 การควักอุจจาระ	4	3		ง่าย
8.5 การใส่ท่อระบายลม	1	5	1	ปานกลาง
8.6 การสวนอุจจาระทางทวารเทียม		5	2	ปานกลาง
9. หลักการและเทคนิคการพยาบาลเพื่อตอบสนองความต้องการ ออกซิเจน				
9.1 การให้ออกซิเจน	1	3	3	ปานกลาง
9.2 การดูดเสมหะ			7	ยาก
10. หลักการและเทคนิคการพยาบาลในการให้ยาและการให้สารน้ำทาง หลอดเลือดดำ				
10.1 การให้ยาทางปาก	2	5		ปานกลาง
10.2 การให้ยาทางintra dermal IM IV			7	ยาก
10.3 การให้IV			7	ยาก
10.4 การให้เลือด		1	6	ยาก
11. การเก็บสิ่งส่งตรวจเพื่อประกอบการวินิจฉัย				
11.1 การเก็บสิ่งส่งตรวจปัสสาวะ	4	3		ง่าย
11.2 การเก็บสิ่งส่งตรวจอุจจาระ	6	1		ยาก
11.3 การเก็บสิ่งส่งตรวจ เสมหะ	2	5		ปานกลาง
11.4 การเก็บสิ่งส่งตรวจเลือด		1	6	ยาก
11.5 การเก็บสิ่งส่งตรวจ น้ำไขสันหลัง	3	2	2	ง่าย
11.6 เก็บสิ่งส่งตรวจสิ่งคัดหลั่งต่างๆ	2	5		ปานกลาง
11.7 การจัดทำในการเก็บสิ่งส่งตรวจ	1	6		ปานกลาง
11.8 การแปลผลการตรวจทางห้องปฏิบัติการการ		2	5	ยาก

n = 7

3.1.1.2 การสร้างวิดิทัศน์สถานการณ์จำลองการฝึกปฏิบัติงานของนักศึกษาพยาบาล โดยนำผลจากการแสดงความคิดเห็นของผู้ทรงคุณวุฒิที่เป็นเอกฉันท์ ในการระบุระดับความยากมาประกอบการเขียนสถานการณ์จำลองที่จะต้องมิจิจกรรมการพยาบาลที่มีระดับความง่าย ปาน

กลาง และยาก จากนั้นนำสถานการณ์จำลองไปให้ผู้ทรงคุณวุฒิภายในคณะ 3 ท่านตรวจสอบ แล้วจึงนำมาเขียนบทวีดีทัศน์

ตัวอย่างสถานการณ์ที่ใช้ในการสร้างบทวีดีทัศน์

ผู้ป่วยหญิงอายุ 56 ปี มาด้วยอาการ 1 วันก่อนมารพ. หายใจเหนื่อย หน้าบวม ปัสสาวะไม่ออก ประเมินสัญญาณชีพพบว่า BP 130/90 mmHg PR 88 /min RR 22 /min T 38. °C แพทย์ให้นอนโรงพยาบาล และมีแผนการรักษาให้ O2 cannula 3 U/min On 5% D/N/2 1000 ml vein drip 80 ml/hr. ใส่คาสายสวนปัสสาวะ เพื่อประเมินการขับถ่ายปัสสาวะ และให้สังเกตอาการเปลี่ยนแปลง และให้ยาทางปาก โดยการหยิบจับของปราศจากเชื้อ เป็นกิจกรรมการพยาบาลอย่างง่าย การวัดสัญญาณชีพ และการให้ยาทางปากเป็นกิจกรรมการพยาบาลระดับปานกลาง และการสวนปัสสาวะเป็นกิจกรรมการพยาบาลอย่างยาก

บทวีดีทัศน์สร้างให้สอดคล้องกับสถานการณ์จำลองจำนวน 8 ตอน แต่ละตอนประกอบด้วยตอนย่อย (Shot) จำนวน 8 ตอนย่อย โดยผ่านการตรวจสอบจากผู้ทรงคุณวุฒิภายในคณะฯ จำนวน 3 ท่าน นำเสนอในตารางที่ 17

ตารางที่ 17 การตรวจสอบคุณภาพบทวีดีทัศน์ให้สอดคล้องกับรูปบริการประเมิน และสถานการณ์

บทวีดีทัศน์	IOC	ข้อเสนอแนะ	การแก้ไข
ตอนที่ 1			
Shot ที่ 1	1.00	จะทราบได้อย่างไรว่าเด็กขึ้นเวรสาย อาจต้องเขียนบรรยายเหนือภาพ	เขียนบรรยายเหนือภาพ “ภาพนาฬิกาบอก เวลา 8.20 น. เวลารับเวร 7.30 น.”
Shot ที่ 2	0.67	Shot ที่ 2 ปรับคำพูดในบทให้สื่อถึงการมาสายเช่นระบุจำนวนครั้งของการมาสาย	แก้ไขบทพูดของอาจารย์ “นี่เป็นครั้งที่ 3 ที่หนูมาสายนะ”
Shot ที่ 3	1.00		ไม่มีการแก้ไข
Shot ที่ 4	1.00	ควรเปลี่ยนเรื่องของความเชื่อ เป็นเรื่องอื่นๆ ที่เป็นเรื่องกว้างๆ เช่น ใช้พวงมาลัยไหว้พระ	แก้ไขบทพูด ให้เป็นเรื่องเอาพวงมาลัยมาไหว้พระแล้วเอามาวางไว้ใกล้หมอน

ตารางที่ 17 การตรวจสอบคุณภาพบทวิดีโอที่ค้นหีสอดคล้องกับรูปบริการประเมิน และสถานการณ์ (ต่อ)

บทวิดีโอที่ค้น	IOC	ข้อเสนอแนะ	การแก้ไข
Shotที่ 5	1.00	ตรวจสอบความถูกต้องกับขั้นตอนการปฏิบัติของกิจกรรมนั้นอีกครั้ง	ตรวจสอบกับรายการแบบตรวจสอบรายการของกิจกรรมการพยาบาลการวัดสัญญาณชีพในรายวิชาการพยาบาลพื้นฐาน แล้วสอดคล้องกันดี
Shotที่ 6	1.00	ตรวจสอบความถูกต้องกับขั้นตอนการปฏิบัติของกิจกรรมนั้นอีกครั้ง	ตรวจสอบกับรายการแบบตรวจสอบรายการของกิจกรรมการพยาบาลการวัดใส่คาสายสวนปัสสาวะในรายวิชาการพยาบาลพื้นฐาน แล้วสอดคล้องกันดี
Shotที่ 7	1.00	ตรวจสอบความถูกต้องกับขั้นตอนการปฏิบัติของกิจกรรมนั้นอีกครั้ง	ตรวจสอบกับรายการแบบตรวจสอบรายการของกิจกรรมการพยาบาลการให้ยาทางปากในรายวิชาการพยาบาลพื้นฐาน แล้วสอดคล้องกันดี
Shotที่ 8	0.67	ตรวจสอบความครบถ้วนของบทพูดว่าครอบคลุมสมรรถนะทั้ง 7 ด้าน 26 ข้อย่อย	ปรับเพิ่มบทพูดให้สามารถประเมินสมรรถนะในรูปรีกให้ได้ครอบคลุม
ตอนที่ 2			
Shotที่ 1	1.00		ไม่มีการแก้ไข
Shotที่ 2	0.67	Shot ที่ 2 ปรับคำพูดในบทให้สื่อถึงพฤติกรรมที่ไม่รับผิดชอบ	แก้ไขบทพูดของอาจารย์ “นี่เป็นครั้งที่ 2 ที่หนูมาสายนะ”
Shotที่ 3	1.00		ไม่มีการแก้ไข
Shotที่ 4	1.00		ไม่มีการแก้ไข
Shotที่ 5	1.00	ควรเปลี่ยนเรื่องของความเชื่อ ตามตอนที่ 5 และเปลี่ยนบทพูดให้สอดคล้องกับระดับคะแนนด้วย	แก้ไขบทพูด ให้มีระดับคะแนน 2
Shotที่ 6	1.00	ตรวจสอบความถูกต้องกับขั้นตอนการปฏิบัติของกิจกรรมนั้นอีกครั้ง	ตรวจสอบกับรายการแบบตรวจสอบรายการของกิจกรรมการพยาบาลการวัดสัญญาณชีพในรายวิชาการพยาบาลพื้นฐาน แล้วสอดคล้องกันดี
Shotที่ 7	1.00	ตรวจสอบความถูกต้องกับขั้นตอนการปฏิบัติของกิจกรรมนั้นอีกครั้ง	ตรวจสอบกับรายการแบบตรวจสอบรายการของกิจกรรมการพยาบาลการวัดใส่คาสายสวนปัสสาวะในรายวิชาการพยาบาลพื้นฐาน แล้วสอดคล้องกันดี

ตารางที่ 17 การตรวจสอบคุณภาพบทวีดิทัศน์ให้สอดคล้องกับรูบริกการประเมิน และสถานการณ์ (ต่อ)

บทวีดิทัศน์	IOC	ข้อเสนอแนะ	การแก้ไข
Shotที่ 8	1.00	ตรวจสอบความครบถ้วนของบทพูดว่าครอบคลุมสมรรถนะทั้ง 7 ด้าน 26 ข้อ	ปรับเพิ่มบทพูดให้สามารถประเมินสมรรถนะในรูบริกให้ได้ครอบคลุมในระดับคะแนน 2 ย่อย
ตอนที่ 3			
Shotที่ 1	1.00		ไม่มีการแก้ไข
Shotที่ 2	1.00	Shot ที่ 2 ปรับคำพูดในบทให้สื่อถึงพฤติกรรมความรับผิดชอบในระดับดี	แก้ไขบทพูดของอาจารย์
Shotที่ 3	1.00		ไม่มีการแก้ไข
Shotที่ 4	1.00		ไม่มีการแก้ไข
Shotที่ 5	1.00	ควรเปลี่ยนเรื่องของความเชื่อ ตามตอนที่ 5 และเปลี่ยนบทพูดให้สอดคล้องกับระดับคะแนนด้วย	แก้ไขบทพูด ให้มีระดับคะแนน 3
Shotที่ 6	1.00	ตรวจสอบความถูกต้องกับขั้นตอนการปฏิบัติของกิจกรรมนั้นอีกครั้ง	ตรวจสอบกับรายการแบบตรวจสอบรายการของกิจกรรมการพยาบาลการวัดสัญญาณชีพในรายวิชาการพยาบาลพื้นฐาน แล้วสอดคล้องกันดี
Shotที่ 7	1.00	ตรวจสอบความถูกต้องกับขั้นตอนการปฏิบัติของกิจกรรมนั้นอีกครั้ง	ตรวจสอบกับรายการแบบตรวจสอบรายการของกิจกรรมการพยาบาลการวัดใส่คาสายสวนปัสสาวะในรายวิชาการพยาบาลพื้นฐาน แล้วสอดคล้องกันดี
Shotที่ 8	0.67	ตรวจสอบความครบถ้วนของบทพูดว่าครอบคลุมสมรรถนะทั้ง 7 ด้าน 26 ข้อ	ปรับเพิ่มบทพูดให้สามารถประเมินสมรรถนะในรูบริกให้ได้ครอบคลุมในระดับคะแนน 3 ย่อย
ตอนที่ 4			
Shotที่ 1	1.00		ไม่มีการแก้ไข
Shotที่ 2	1.00		ไม่มีการแก้ไข
Shotที่ 3	1.00	Shot ที่ 2 ปรับคำพูดในบทให้สื่อถึงพฤติกรรมความรับผิดชอบในระดับดี	แก้ไขบทพูดของอาจารย์
Shotที่ 4	1.00		ไม่มีการแก้ไข

ตารางที่ 17 การตรวจสอบคุณภาพบทวีดิทัศน์ให้สอดคล้องกับรูปบริการประเมิน และสถานการณ์ (ต่อ)

บทวีดิทัศน์	IOC	ข้อเสนอแนะ	การแก้ไข
hotที่ 5	1.00	ควรเปลี่ยนเรื่องของความเชื่อ ตามตอนที่ 5 และเปลี่ยนบทพูดให้สอดคล้องกับระดับคะแนนด้วย	แก้ไขบทพูด ให้มีระดับคะแนน 4
Shotที่ 6	1.00	ตรวจสอบความถูกต้องกับขั้นตอนการปฏิบัติของกิจกรรมนั้นอีกครั้ง	ตรวจสอบกับรายการแบบตรวจสอบรายการของกิจกรรมการพยาบาลการวัดสัญญาณชีพในรายวิชาการพยาบาลพื้นฐาน แล้วสอดคล้องกันดี
Shotที่ 7	1.00	ตรวจสอบความถูกต้องกับขั้นตอนการปฏิบัติของกิจกรรมนั้นอีกครั้ง	ตรวจสอบกับรายการแบบตรวจสอบรายการของกิจกรรมการพยาบาลการวัดใส่คาสายสวนปัสสาวะในรายวิชาการพยาบาลพื้นฐาน แล้วสอดคล้องกันดี
Shotที่ 8	1.00	ตรวจสอบความครบถ้วนของบทพูดว่าครอบคลุมสมรรถนะทั้ง 7 ด้าน 26 ข้อย่อย	ปรับเพิ่มบทพูดให้สามารถประเมินสมรรถนะในรูปรีกให้ได้ครอบคลุมในระดับคะแนน 4
ตอนที่ 5			
Shotที่ 1	1.00	ไม่มีการแก้ไข	
Shotที่ 2	1.00	ไม่มีการแก้ไข	
Shotที่ 3	1.00	ความถี่ของพฤติกรรมไม่ส่งงานไม่ชัดเจน	ปรับบทพูดให้ระบุจำนวนครั้งของการส่งงานสาย เพื่อให้ง่ายต่อการประเมินพฤติกรรมความรับผิดชอบ
Shotที่ 4	1.00	ควรเปลี่ยนเรื่องของความเชื่อ เป็นเรื่องอื่นๆ ที่ไม่ซ้ำซ้อนกับตอนที่ 1-4	แก้ไขบทพูด ให้เป็นความเชื่อเกี่ยวกับการรักษา ด้วยยาสมุนไพร
Shotที่ 5	1.00	ตรวจสอบความถูกต้องกับขั้นตอนการปฏิบัติของกิจกรรมนั้นอีกครั้ง	ตรวจสอบกับรายการแบบตรวจสอบรายการของกิจกรรมการพยาบาลการวัดสัญญาณชีพในรายวิชาการพยาบาลพื้นฐาน แล้วสอดคล้องกันดี
Shotที่ 6	1.00	ตรวจสอบความถูกต้องกับขั้นตอนการปฏิบัติของกิจกรรมนั้นอีกครั้ง	ตรวจสอบกับรายการแบบตรวจสอบรายการของกิจกรรมการพยาบาลการวัดใส่คาสายสวนปัสสาวะในรายวิชาการพยาบาลพื้นฐาน แล้วสอดคล้องกันดี
Shotที่ 8	0.67	ตรวจสอบความครบถ้วนของบทพูดว่าครอบคลุมสมรรถนะทั้ง 7 ด้าน 26 ข้อย่อยเหมือนตอนที่ 1-4	ปรับเพิ่มบทพูดให้สามารถประเมินสมรรถนะในรูปรีกให้ได้ครอบคลุม

ตารางที่ 17 การตรวจสอบคุณภาพบทวีดิทัศน์ให้สอดคล้องกับรูปบริการประเมิน และสถานการณ์ (ต่อ)

บทวีดิทัศน์	IOC	ข้อเสนอแนะ	การแก้ไข
Shotที่ 7	1.00	ตรวจสอบความถูกต้องกับขั้นตอนการปฏิบัติของกิจกรรมนั้นอีกครั้ง	ตรวจสอบกับรายการแบบตรวจสอบรายการของกิจกรรมการพยาบาลการให้ยาทางปากในรายวิชาการพยาบาลพื้นฐาน แล้วสอดคล้องกันดี
ตอนที่ 6			
Shotที่ 1	1.00		ไม่มีการแก้ไข
Shotที่ 2	1.00		ไม่มีการแก้ไข
Shotที่ 3	0.67	ควรระบุความถี่ของพฤติกรรมที่ไม่ส่งงาน	ปรับบทพูดให้ระบุความถี่ของพฤติกรรมที่ไม่ส่งงานให้ชัดเจน
Shotที่ 4	1.00		ไม่มีการแก้ไข
Shotที่ 5	0.67	ควรเปลี่ยนเรื่องของความเชื่อ ตามตอนที่ 5 และเปลี่ยนบทพูดให้สอดคล้องกับระดับคะแนนด้วย	แก้ไขบทพูด ให้มีระดับคะแนน 2
Shotที่ 6	1.00	ตรวจสอบความถูกต้องกับขั้นตอนการปฏิบัติของกิจกรรมนั้นอีกครั้ง	ตรวจสอบกับรายการแบบตรวจสอบรายการของกิจกรรมการพยาบาลการวัดสัญญาณชีพในรายวิชาการพยาบาลพื้นฐาน แล้วสอดคล้องกันดี
Shotที่ 7	1.00	ตรวจสอบความถูกต้องกับขั้นตอนการปฏิบัติของกิจกรรมนั้นอีกครั้ง	ตรวจสอบกับรายการแบบตรวจสอบรายการของ
Shotที่ 8	1.00	ตรวจสอบความครบถ้วนของบทพูดว่าครอบคลุมสมรรถนะทั้ง 7 ด้าน 26 ข้อย่อย	กิจกรรมการพยาบาลการวัดใส่คาสายสวนปัสสาวะในรายวิชาการพยาบาลพื้นฐาน แล้วสอดคล้องกันดี
ตอนที่ 7			
Shotที่ 1	1.00		ไม่มีการแก้ไข
Shotที่ 2	1.00		ไม่มีการแก้ไข
Shotที่ 3	1.00		ไม่มีการแก้ไข
Shotที่ 4	1.00		ไม่มีการแก้ไข
Shotที่ 5	1.00	ควรเปลี่ยนเรื่องของความเชื่อ ตามตอนที่ 5 และเปลี่ยนบทพูดให้สอดคล้องกับระดับคะแนนด้วย	แก้ไขบทพูด ให้มีระดับคะแนน 3

ตารางที่ 17 การตรวจสอบคุณภาพบทวีดิทัศน์ให้สอดคล้องกับรูปบริการประเมิน และสถานการณ์ (ต่อ)

บทวีดิทัศน์	IOC	ข้อเสนอแนะ	การแก้ไข
Shotที่ 6	1.00	ตรวจสอบความถูกต้องกับขั้นตอนการปฏิบัติของกิจกรรมนั้นอีกครั้ง	ตรวจสอบกับรายการแบบตรวจสอบรายการของกิจกรรมการพยาบาลการวัดสัญญาณชีพในรายวิชาการพยาบาลพื้นฐาน แล้วสอดคล้องกันดี
Shotที่ 7	1.00	ตรวจสอบความถูกต้องกับขั้นตอนการปฏิบัติของกิจกรรมนั้นอีกครั้ง	ตรวจสอบกับรายการแบบตรวจสอบรายการของกิจกรรมการพยาบาลการวัดใส่คาสายสวนปัสสาวะในรายวิชาการพยาบาลพื้นฐาน แล้วสอดคล้องกันดี
Shotที่ 8	0.67	ตรวจสอบความครบถ้วนของบทพูดว่าครอบคลุมสมรรถนะทั้ง 7 ด้าน 26 ข้อย่อย	ปรับเพิ่มบทพูดให้สามารถประเมินสมรรถนะในรูปรีกให้ได้ครอบคลุมในระดับคะแนน 3
ตอนที่ 8			
Shotที่ 1	1.00		ไม่มีการแก้ไข
Shotที่ 2	1.00		ไม่มีการแก้ไข
Shotที่ 3	1.00	บทพูดของนักศึกษาไม่ค่อยสอดคล้องกับระดับคุณภาพที่ได้เพราะอยู่ระดับคะแนน 3	แก้ไขบทพูดของนักศึกษาว่ามีพฤติกรรมการส่งงานในระดับดีอย่างไร
Shotที่ 4	1.00		ไม่มีการแก้ไข
Shotที่ 5	1.00	ควรเปลี่ยนเรื่องของความเชื่อ ตามตอนที่ 5 และเปลี่ยนบทพูดให้สอดคล้องกับระดับคะแนนด้วย	แก้ไขบทพูด ให้มีระดับคะแนน 3
Shotที่ 6	1.00	ตรวจสอบความถูกต้องกับขั้นตอนการปฏิบัติของกิจกรรมนั้นอีกครั้ง	ตรวจสอบกับรายการแบบตรวจสอบรายการของกิจกรรมการพยาบาลการวัดสัญญาณชีพในรายวิชาการพยาบาลพื้นฐาน แล้วสอดคล้องกันดี
Shotที่ 7	1.00	ตรวจสอบความถูกต้องกับขั้นตอนการปฏิบัติของกิจกรรมนั้นอีกครั้ง	ตรวจสอบกับรายการแบบตรวจสอบรายการของกิจกรรมการพยาบาลการวัดใส่คาสายสวนปัสสาวะในรายวิชาการพยาบาลพื้นฐาน แล้วสอดคล้องกันดี
Shotที่ 8	1.00	ตรวจสอบความครบถ้วนของบทพูดว่าครอบคลุมสมรรถนะทั้ง 7 ด้าน 26 ข้อย่อย	ปรับเพิ่มบทพูดให้สามารถประเมินสมรรถนะในรูปรีกให้ได้ครอบคลุมในระดับคะแนน 4

n = 3

3.1.2 สร้างแบบวัดความพึงพอใจต่อการใช้โมดูลการฝึกอบรมฯ จำนวน 10 ข้อ โดยดัดแปลงจากแบบวัดความพึงพอใจในการประเมินโครงการฝึกอบรมตามแนวคิดของ Kirkpatrick (2006)

3.1.3 สร้างแบบวัดความรู้เกี่ยวกับการประเมินภาคปฏิบัติ จำนวน 25 ข้อ จากเนื้อหาของการหน่วยการเรียนรู้ 5 หน่วยการเรียนรู้ในโมดูลการฝึกอบรมฯ

3.1.4 สร้างแบบวัดทัศนคติเกี่ยวกับการประเมินภาคปฏิบัติ จำนวน 10 ข้อ โดยดัดแปลงจากแบบวัดทัศนคติในการประเมินโครงการฝึกอบรมตามแนวคิดของ Kirkpatrick (2006)

3.1.5 สร้างคู่มือที่ใช้ประกอบการฝึกอบรม โดยใช้แนวคิดในการสร้างคู่มือดัดแปลงจากแนวทางการสร้างคู่มือของนนทนซ์ อ่อนพวน (2553)

3.2 การตรวจสอบคุณภาพของเครื่องมือแต่ละประเภทตามลำดับต่อไปนี้

3.2.1 ชุดฝึกทักษะการประเมินภาคปฏิบัติ ประกอบด้วย รุบริกการประเมินภาคปฏิบัติ ร่วมกับวีดิทัศน์สถานการณ์จำลอง

3.2.1.1 รุบริกที่ใช้ในการประเมินทักษะ ในระยะก่อนการใช้โมดูลการฝึกอบรม หลังการใช้โมดูลการฝึกอบรมทันที และหลังการใช้โมดูลการฝึกอบรม 3 เดือน ร่วมกับงานที่ใช้ในการประเมินภาคปฏิบัติอยู่ในรูปสื่อวีดิทัศน์การปฏิบัติการพยาบาลในสถานการณ์จำลองการขึ้นฝึกปฏิบัติงานบนหอผู้ป่วยของนักศึกษาพยาบาลชั้นปีที่ 2 ในรายวิชาปฏิบัติการพยาบาลพื้นฐาน จำนวน 8 ชุด ตรวจสอบความเที่ยงระหว่างผู้ประเมินด้วยค่า Intra-class correlation coefficient (ICC) มีค่าระหว่าง .45 ถึง 0.80 และค่า Rater agreement index (RAI) ในรูป Mean RAI มีค่าระหว่าง 0.85 ถึง 0.96 ดังแสดงในตารางที่ 18

ตารางที่ 18 การตรวจสอบคุณภาพของรุบริกที่ใช้ในการประเมินภาคปฏิบัติ

วีดิทัศน์	Mean RAI	ICC
ชุดที่ 1	0.90	0.58
ชุดที่ 2	0.90	0.49
ชุดที่ 3	0.92	0.80
ชุดที่ 4	0.85	0.63
ชุดที่ 5	0.95	0.55
ชุดที่ 6	0.84	0.45
ชุดที่ 7	0.85	0.45
ชุดที่ 8	0.91	0.75

n = 5

3.2.1.2 วิธีทัศนสถานการณ์จำลองการฝึกปฏิบัติงานของนักศึกษาพยาบาล จำนวน 8 ชุด ผ่านการตัดต่อจนได้วิธีทัศน 16 ชุด ดังแสดงในภาพที่ 3 หน้า 110 สุ่มอย่างง่ายนำมาใช้ในการศึกษา ผลของการฝึกอบรมฯ 10 ชุด ส่วนที่เหลือนำมาใช้ในการทำกิจกรรมระหว่างการทำฝึกอบรมฯ



ภาพที่ 4 ตัวอย่างตอนหนึ่งในวิธีทัศน สถานการณ์จำลองเพื่อเสริมสร้างสมรรถนะการประเมิน สำหรับอาจารย์พยาบาล

3.2.2 แบบวัดความพึงพอใจต่อโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมิน ด้วยการฝึกอบรม

การตรวจสอบคุณภาพของแบบวัดความพึงพอใจต่อโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมิน โดยผู้ทรงคุณวุฒิกลุ่มที่ 2 จำนวน 9 ท่าน พิจารณาค่าความตรงเชิงเนื้อหา ใช้ค่า Item Objective Congruence Index (IOC) มีค่าเท่ากับ 0.89 - 1.00 แสดงว่าแบบวัดมีความสอดคล้องระหว่างข้อคำถามกับวัตถุประสงค์การวัดเป็นอย่างดีทุกข้อ แต่มีข้อเสนอแนะให้ปรับแก้เพิ่มเติมจำนวนข้อของแบบวัด ดึงนำเสนอในตารางที่ 19

ตารางที่ 19 การตรวจสอบคุณภาพของแบบวัดความพึงพอใจต่อโมดูลฯ

รายการประเมิน	IOC	ข้อเสนอแนะ	การแก้ไขหลังได้รับข้อเสนอแนะ
1. สถานที่ มีความสะดวก สบายและเหมาะสมสำหรับการฝึกอบรมครั้งนี้	1.00	ไม่มีการแก้ไข	1. สถานที่ มีความสะดวก สบาย และเหมาะสมสำหรับการฝึกอบรมครั้งนี้
2. ระยะเวลาการฝึกอบรมมีความเหมาะสม	1.00	ไม่มีการแก้ไข	2. ระยะเวลาการฝึกอบรมมีความเหมาะสม
3. ปริมาณ และคุณภาพของอาหารว่าง	0.89	ควรรวมหัวข้อที่ 3 และ 4 เข้าไว้ด้วยกัน เพราะเป็นข้อมูลประเภทเดียวกันไม่ใช่สาระสำคัญของ การฝึกอบรม	3. ปริมาณ และคุณภาพของอาหารว่าง และอาหารกลางวัน
4. ปริมาณ และคุณภาพของอาหารกลางวัน	0.89	รวมหัวข้อกับข้อ 3	4. ความรู้ความสามารถของ ผู้ดำเนินการฝึกอบรม และ วิทยากร
5. ความรู้ ความสามารถของ ผู้ดำเนินการฝึกอบรม และ วิทยากร	1.00	ไม่มีการแก้ไข (เลื่อนไปไว้ ข้อถัดไป)	5. ประโยชน์ของการฝึกอบรมต่อตัว ท่าน
6. เนื้อหาของการฝึกอบรมเป็น ประโยชน์กับท่านมาก	1.00	เพิ่มหัวข้อประเมินโดยแยก เป็นประโยชน์ต่อตนเอง 1 ข้อ และ ประโยชน์ต่อองค์กรอีก 1 ข้อ	6. ประโยชน์ของการฝึกอบรมต่อ องค์กรของท่าน
7. ความน่าสนใจของกิจกรรมการ ฝึกอบรม	1.00	ไม่มีการแก้ไข	7. ความน่าสนใจของกิจกรรมการ ฝึกอบรม
8. ความพร้อมของสื่อ โสตทัศนูปกรณ์	1.00	ไม่มีการแก้ไข	8. ความพร้อมของสื่อ โสตทัศนูปกรณ์
9. ความชัดเจนของเนื้อหาและ ความน่าสนใจของเอกสาร ประกอบการฝึกอบรม	1.00	ปรับแก้เป็นความน่าสนใจ ของกิจกรรมเรียงลำดับไว้ ข้อต้นๆ	9. ความชัดเจนของเนื้อหาในการ ฝึกอบรม
10. ความคุ้มค่ากับเวลาที่เสียไปกับการ เข้ารับการฝึกอบรม	1.00	ปรับแก้คำให้มีความกระชับ ชัดเจน เข้าใจง่าย	10. ความคุ้มค่ากับการใช้เวลาใน การฝึกอบรม

n = 9

3.2.3 แบบวัดความรู้ด้านการประเมินภาคปฏิบัติ

การตรวจสอบคุณภาพของแบบวัดความรู้ด้านการประเมินภาคปฏิบัติ ดำเนินการตรวจสอบ 2 วิธีคือ การตรวจสอบความตรงเชิงเนื้อหา และการวิเคราะห์ข้อสอบรายข้อ และชุดข้อสอบ

3.2.3.1 การตรวจสอบความตรงเชิงเนื้อหา ใช้ค่า Item Objective Congruence Index (IOC) มีค่าเท่ากับ 0.45 ถึง 1.00 ทำการปรับปรุงตามคำแนะนำของผู้ทรงคุณวุฒิ ดังแสดงในตารางที่ 20

ตารางที่ 20 การตรวจสอบคุณภาพด้านความสอดคล้องระหว่างข้อคำถามกับวัตถุประสงค์ของการวัดของแบบวัดความรู้ด้านการประเมินภาคปฏิบัติ

ข้อคำถาม	IOC	ข้อเสนอแนะ	การแก้ไข
1. การประเมินภาคปฏิบัติเป็นการประเมินที่สมรรถนะส่วนตัวของผู้เรียนโดยอาศัยความเป็นอันดับของผู้สอนภาคปฏิบัติเป็นสำคัญ	1.00	ปรับคำถามให้เข้าใจง่ายแปลความหมายของอันดับให้ชัดเจน	1. หลักสำคัญในการประเมินภาคปฏิบัติ คือการใช้ประสบการณ์ส่วนตัวของผู้สอนภาคปฏิบัติในการตัดสินผลการเรียนรู้ภาคปฏิบัติของผู้เรียน
2. งานที่ใช้ในการประเมินภาคปฏิบัติ เช่น การทดสอบความรู้ด้วยแบบสอบหลายตัวเลือก	0.78	เปลี่ยนแบบสอบหลายตัวเลือกเป็นงานภาคปฏิบัติจริงๆ น่าจะดีกว่าเพราะผิดชัดเจนไป	2. การขึ้นฝึกปฏิบัติงานบนหอผู้ป่วยของนักศึกษาพยาบาล จัดเป็นงานภาคปฏิบัติประเภท งานที่เกิดขึ้นตามธรรมชาติ
3. มาตรฐานค่า (Rating scale) มีลักษณะที่แตกต่างจากแบบตรวจรายการ คือบอกระดับคุณภาพได้ แต่ไม่สามารถบอกรายการพฤติกรรมได้	0.89	ปรับแก้คำถามเหลือประเด็นเดียว	3. มาตรฐานค่า (Rating scale) มีลักษณะที่แตกต่างจากแบบตรวจรายการ คือบอกรายการพฤติกรรมไม่ได้
4. การหาความเที่ยงระหว่างผู้ประเมินเป็นการหาความสอดคล้องในผลการประเมินระหว่างผู้ประเมิน	1.00	ไม่มีการแก้ไข	
5. Rubric เป็นวิธีหนึ่งในการแก้ปัญหาการประเมินภาคปฏิบัติ เนื่องจากทำให้การสร้าง ความเข้าใจให้ตรงกันระหว่าง ผู้ประเมิน	1.00	ปรับการจัดเรียงคำถามให้เข้าใจง่ายขึ้น	5. การสร้างความเข้าใจให้ตรงกันระหว่างผู้ประเมิน ในการใช้เครื่องมือในการประเมินภาคปฏิบัติ เช่น Rubric เป็นวิธีหนึ่งในการแก้ปัญหาการประเมินภาคปฏิบัติ
6. Rater error training เป็นการฝึกอบรมโดยการคัดเลือกผู้ประเมินที่มีปัญหาเกี่ยวกับการประเมินภาคปฏิบัติอันเป็นผลจากความคลาดเคลื่อนจากการประเมินมาก ๆ เข้ารับการฝึกอบรมให้มีคุณภาพการประเมินดีขึ้น	0.89	ปรับ คำถาม ให้มีความกระชับมากขึ้น	6. Rater error training เป็นการฝึกอบรมโดยการคัดเลือกผู้ประเมินที่มีความคลาดเคลื่อนจากการประเมินมาก ๆ เข้ารับการฝึกอบรมให้มีคุณภาพการประเมินดีขึ้น

ตารางที่ 20 การตรวจสอบคุณภาพด้านความสอดคล้องระหว่างข้อคำถามกับวัตถุประสงค์ของการวัดของแบบวัดความรู้ด้านการประเมินภาคปฏิบัติ (ต่อ)

ข้อคำถาม	IOC	ข้อเสนอแนะ	การแก้ไข
7. ความคลาดเคลื่อนจากการกดคะแนนคือ การที่ผู้ประเมินมือคติดกับผู้ถูกประเมินแล้วให้คะแนนต่ำกว่าความสามารถที่แท้จริง	0.89	ควรระบุว่าการกดคะแนนเกิดกับคนคนเดียวหรือ ทั้งกลุ่ม เพื่อจะได้เกิดความชัดเจน	7. ความคลาดเคลื่อนจากการกดคะแนน เป็น การที่ผู้ประเมินมือคติดกับผู้ถูกประเมินคนหนึ่งแล้วให้คะแนนต่ำกว่าความสามารถที่แท้จริง
8. การฝึกอบรมเรื่องกรอบของการอ้างอิง เป็นการสร้างความเข้าใจให้ตรงกันระหว่างผู้ประเมินในเรื่องมาตรฐานการประเมินว่า รุบริคหรือมาตรฐานที่จะถูกนำมาใช้เป็น เกณฑ์ในการประเมินเดียวกัน ในการประเมินงานภาคปฏิบัติขึ้นเดียวกัน	1.00	ปรับคำเชื่อมประโยค และ ประโยค ให้มีความกระชับมากขึ้น	8. การฝึกอบรมเรื่องกรอบของการอ้างอิง เป็นการสร้างความเข้าใจให้ตรงกันระหว่างผู้ประเมินในการใช้เครื่องมือในการประเมินงานภาคปฏิบัติ ในชั้นงานเดียวกัน
9. การฝึกการสังเกตพฤติกรรมเป็นการฝึกด้าน 1) การพัฒนาทักษะด้านการค้นหา 2) การรับรู้ และ 3) การทบทวนความจำต่อพฤติกรรมการปฏิบัติงาน	1.00	ไม่มีการแก้ไข	
10. การฝึกการตัดสินใจในการประเมินภาคปฏิบัติ มีข้อดีคือ จะทำให้งานภาคปฏิบัติมีคุณภาพมากขึ้น มีข้อเสียคือทำให้ระยะเวลาในการตัดสินผลการประเมินนานขึ้น	0.55	ควรเลือกถามเพียงประเด็นเดียว	10. การฝึกอบรมเรื่อง การตัดสินใจในการประเมินภาคปฏิบัติ มีข้อเสียคือทำให้ระยะเวลาในการตัดสินผลการประเมินนานขึ้น
11. ข้อใดคือตัวอย่างของงานที่ผู้สอนมอบหมายให้ผู้เรียนในการเรียนการสอนภาคปฏิบัติ	1.00	ไม่มีการแก้ไข	11. ข้อใดคือตัวอย่างของงานที่ผู้สอนมอบหมายให้ผู้เรียนในการเรียนการสอนภาคปฏิบัติ
12. จากกราฟแสดงผลการประเมินภาคปฏิบัติ อาจารย์คนที่ 2 ความคลาดเคลื่อนจากผู้ประเมินแบบใด	0.89	แก้ไขตัวโจทย์จาก เฉลยแบบฮาโลเป็นแบบกด ปล่อยจะเข้าใจได้ง่ายกว่า	
13. ข้อใดแสดงถึงสาเหตุของความคลาดเคลื่อนที่เกิดขึ้นในการประเมินงานภาคปฏิบัติ	1.00	ไม่มีการแก้ไข	
14. เครื่องมือชนิดใดที่เป็นที่ยอมรับว่าช่วยเพิ่มความแม่นยำในการประเมิน	0.89	ปรับแก้โจทย์แปลความหมายของปรนัยเป็นภาษาที่เข้าใจง่าย	14. เครื่องมือชนิดใดที่เป็นที่ยอมรับว่าช่วยเพิ่มความสอดคล้องในการให้คะแนนระหว่างอาจารย์ในการให้คะแนนผู้เรียนคนเดียวกัน (ความเป็นปรนัย)ในการประเมิน
15. ข้อใดคือการตรวจสอบคุณภาพการประเมินภาคปฏิบัติ	0.89	ไม่มีการแก้ไข	

ตารางที่ 20 การตรวจสอบคุณภาพด้านความสอดคล้องระหว่างข้อความกับวัตถุประสงค์ของการวัดของแบบวัดความรู้ด้านการประเมินภาคปฏิบัติ (ต่อ)

ข้อความ	IOC	ข้อเสนอแนะ	การแก้ไข
16. การนำคะแนนสอบของผู้เรียนทั้งหมดมาหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานแล้วนำมาใช้แปลงเป็นคะแนนมาตรฐานเพื่อกำหนดตำแหน่งของผู้เรียนเป็นการกำหนดเกรดแบบใด	1.00	ไม่มีการแก้ไข	
17. ความคลาดเคลื่อนจากผู้ประเมิน (Rater error or rater effect) หมายถึงความคลาดเคลื่อนที่เกิดจากสาเหตุใด	0.89	แก้ไขตัวเลือก	
18. วิธีการแก้ปัญหาความคลาดเคลื่อนจากผู้ประเมินแบบใดที่จะช่วยแก้ปัญหาได้ตรงประเด็นปัญหามากที่สุด	0.78	ไม่มีการแก้ไข	18. วิธีการแก้ปัญหาความคลาดเคลื่อนจากผู้ประเมินแบบใดเป็นการแก้ปัญหาได้ตรงประเด็นปัญหามากที่สุด
19. หลักการสังเกตพฤติกรรมที่ดีคือข้อใด	0.89	แก้ไขตัวเลือก	
20. ข้อใดเป็นสถานการณ์ที่อาจารย์ผู้สอนภาคปฏิบัติมีการใช้กฎของการตัดสินใจอย่างเหมาะสม	0.78	แก้ไขตัวเลือก	
21. การภาคปฏิบัติเน้นการประเมิน 2 ประเด็นสำคัญ คือการประเมินกระบวนการปฏิบัติ และการประเมินผลผลิตจากกระบวนการปฏิบัติ	1.00	ไม่มีการแก้ไข	
22. เครื่องมือที่ใช้ในการประเมินภาคปฏิบัติ เช่น มาตรฐานประมาณค่า (rating scale)	1.00	ไม่มีการแก้ไข	
23. งานที่มีความเสี่ยงสูง เช่น การช่วยฟื้นคืนชีพผู้ใหญ่ ไม่ควรประเมินความรู้ก่อนการขึ้นปฏิบัติงานด้วยการทำข้อสอบแบบเขียนตอบ	1.00	ไม่มีการแก้ไข	
24. แบบตรวจสอบรายการ (Checklist) มีช่องว่างให้ระบุว่าได้ทำหรือไม่ได้ทำพฤติกรรมนั้น แต่ไม่สามารถบอกระดับคุณภาพได้	0.90	ปรับ คำพูดให้ชัดเจนลดความคลาดเคลื่อนโดยระบุสมรรถนะของเครื่องมือให้ชัดเจน กระชับ	24. แบบตรวจสอบรายการ (Checklist) ประกอบด้วยรายการของพฤติกรรมการเข้า ำ ใจ ที่แสดงออกตามลำดับขั้นตอนการปฏิบัติ มีช่องว่างให้ระบุว่าได้ทำหรือไม่ได้ทำพฤติกรรมนั้น
25. วิธีการทางสถิติเป็นวิธีที่ดีที่สุดที่ใช้ในการแก้ไขความคลาดเคลื่อนจากผู้ประเมิน	1.00	ไม่มีการแก้ไข	

n = 9

3.2.3.2 การวิเคราะห์แบบสอบ รายข้อด้วยค่าความยาก (item difficulty: p) ค่าอำนาจจำแนก (r) และวิเคราะห์ความเที่ยงของชุดข้อสอบด้วยค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's alpha coefficient) ค่าความยากอยู่ระหว่าง 0.31 ถึง 0.79 โดยมีค่าความยากอยู่ในที่ยอมรับได้ อำนาจจำแนก (Discrimination power of the items: r) ระหว่าง 0.20 ถึง 0.54 และวิเคราะห์คุณภาพชุดข้อสอบด้วยค่า Cronbach's Alpha Coefficient: α เท่ากับ .72 ดังแสดงในตารางที่ 21

ตารางที่ 21 การตรวจสอบความตรงเชิงเนื้อหาของแบบทดสอบความรู้พื้นฐานเกี่ยวกับการประเมินภาคปฏิบัติ

ข้อคำถาม	p	r
1. หลักสำคัญในการประเมินภาคปฏิบัติ คือการใช้ประสบการณ์ส่วนตัวของผู้สอนภาคปฏิบัติในการตัดสินผลการเรียนรู้ภาคปฏิบัติของผู้เรียน	0.62	0.20
2. การขึ้นฝึกปฏิบัติงานบนหอผู้ป่วยของนักศึกษาพยาบาล จัดเป็นงานภาคปฏิบัติประเภท งานที่เกิดขึ้นตามธรรมชาติ	0.69	0.28
3. มาตรฐานค่า (Rating scale) มีลักษณะที่แตกต่างจากแบบตรวจรายการ คือบอกรายการพฤติกรรมไม่ได้	0.49	0.41
4. การหาความเที่ยงระหว่างผู้ประเมินเป็นการหาความสอดคล้องในผลการประเมินระหว่างผู้ประเมิน	0.69	0.28
5. การสร้างความเข้าใจให้ตรงกันระหว่างผู้ประเมิน ในการใช้เครื่องมือในการประเมินภาคปฏิบัติ เช่น Rubric เป็นวิธีหนึ่งในการแก้ปัญหาการประเมินภาคปฏิบัติ	0.31	0.29
6. Rater error training เป็นการฝึกอบรมโดยการคัดเลือกเฉพาะผู้ประเมินที่มีความคลาดเคลื่อนจากการประเมินมาก ๆ เข้ารับการฝึกอบรมให้มีคุณภาพการประเมินดีขึ้น	0.62	0.24
7. ความคลาดเคลื่อนจากการคาดคะเน เป็นการที่ผู้ประเมินมีอคติกับผู้ถูกประเมินคนหนึ่งแล้วให้คะแนนต่ำกว่าความสามารถที่แท้จริง	0.79	0.30
8. การฝึกอบรมเรื่องกรอบของการอ้างอิง เป็นการสร้างความเข้าใจให้ตรงกันระหว่างผู้ประเมินในการใช้เครื่องมือในการประเมินงานภาคปฏิบัติ ในชิ้นงานเดียวกัน	0.79	0.41
9. การฝึกการสังเกตพฤติกรรมเป็นการฝึกด้าน 1) การพัฒนาทักษะด้านการค้นหา 2) การรับรู้ และ 3) การทบทวนความจำต่อพฤติกรรมการปฏิบัติงาน	0.41	0.39
10. การฝึกอบรมเรื่อง การตัดสินใจในการประเมินภาคปฏิบัติ มีข้อเสียคือทำให้ระยะเวลาในการตัดสินผลการประเมินนานขึ้น	0.67	0.39
11.. ข้อใดคือตัวอย่างของงานที่ผู้สอนมอบหมายให้ผู้เรียนในการเรียนการสอนภาคปฏิบัติ	0.62	0.54
12. จากกราฟแสดงผลการประเมินภาคปฏิบัติ อาจารย์คนที่ 2 ความคลาดเคลื่อนจากผู้ประเมินแบบใด	0.69	0.31
13. ข้อใดแสดงถึงสาเหตุของความคลาดเคลื่อนที่เกิดขึ้นในการประเมินงานภาคปฏิบัติ	0.56	0.22
14. เครื่องมือชนิดใดที่เป็นที่ยอมรับว่าช่วยเพิ่มความสอดคล้องในการให้คะแนนระหว่างอาจารย์ในการให้คะแนนผู้เรียนคนเดียวกัน (ความเป็นปรนัยในการประเมิน)	0.33	0.28
15. ข้อใดคือการตรวจสอบคุณภาพการประเมินภาคปฏิบัติ	0.74	0.26

ตารางที่ 21 การตรวจสอบความตรงเชิงเนื้อหาของแบบทดสอบความรู้พื้นฐาน เกี่ยวกับการ
ประเมินภาคปฏิบัติ (ต่อ)

ข้อคำถาม	p	r
16. การนำคะแนนสอบของผู้เรียนทั้งหมดมาหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน แล้วนำมาใช้แปลงเป็นคะแนนมาตรฐานเพื่อกำหนดตำแหน่งของผู้เรียนเป็นการกำหนดเกรดแบบใด	0.64	0.39
17. ความคลาดเคลื่อนจากผู้ประเมิน (Rater error or rater effect) หมายถึงความคลาดเคลื่อนที่เกิดจากสาเหตุใด	0.59	0.32
18. วิธีการแก้ปัญหาความคลาดเคลื่อนจากผู้ประเมินแบบใดเป็นการแก้ปัญหาได้ตรงประเด็นปัญหามากที่สุด	0.85	0.22
19. หลักการสังเกตพฤติกรรมที่ดีคือข้อใด	0.64	0.24
20. ข้อใดเป็นสถานการณ์ที่อาจารย์ผู้สอนภาคปฏิบัติมีการใช้กฎของการตัดสินใจอย่างเหมาะสม	0.72	0.47
21. การภาคปฏิบัติเน้นการประเมิน 2 ประเด็นสำคัญ คือการประเมินกระบวนการปฏิบัติ และ การประเมินผลผลิตจากกระบวนการปฏิบัติ	0.69	0.54
22. เครื่องมือที่ใช้ในการประเมินภาคปฏิบัติ เช่น มาตรฐานประมาณค่า (rating scale) 3) การทบทวนความจำต่อพฤติกรรมการปฏิบัติงาน	0.51	0.39
23. งานที่มีความเสี่ยงสูง เช่น การช่วยฟื้นคืนชีพผู้ใหญ่ ไม่ควรประเมินความรู้ก่อนการขึ้นปฏิบัติงานด้วยการทำข้อสอบแบบเขียนตอบ	0.69	0.43
24. แบบตรวจสอบรายการ (Checklist) มีช่องว่างให้ระบุว่าทำได้หรือไม่ทำพฤติกรรมนั้น แต่ไม่สามารถบอกระดับคุณภาพได้	0.41	0.22
25. วิธีการทางสถิติเป็นวิธีที่ดีที่สุดที่ใช้ในการแก้ไขความคลาดเคลื่อนจากผู้ประเมิน	0.69	0.50
n = 39	Cronbach's alpha = 0.72	

3.2.4 แบบวัดทัศนคติต่อการประเมินภาคปฏิบัติ

3.2.4.1 การตรวจสอบความตรงเชิงเนื้อหา ของแบบวัดทัศนคติต่อการประเมินภาคปฏิบัติ

ผ่านการตรวจสอบจากผู้ทรงคุณวุฒิ 9 ท่าน โดยใช้ค่า Item Objective Congruence Index (IOC) มีค่าเท่ากับ 0.67 ถึง 1.00 ดังแสดงในตารางที่ 22

ตารางที่ 22 การตรวจสอบความตรงเชิงเนื้อหาของแบบประเมินทัศนคติต่อการประเมินภาคปฏิบัติ

รายการประเมิน	IOC	ข้อเสนอแนะ	การแก้ไข
1. ท่านรู้สึกดีที่ได้เตรียมการสอนภาคปฏิบัติ	1.00	ไม่มีการแก้ไข	1. ท่านรู้สึกดีที่ได้เตรียมการสอนภาคปฏิบัติ
2. ท่านคิดว่าไม่ใช่เรื่องยากในการเตรียมงานเพื่อใช้ในการประเมินภาคปฏิบัติ	0.67	ปรับคำไม่ให้ความหมายปฏิเสธซ้อนปฏิเสธ	2. ท่านคิดว่าเป็นเรื่องง่ายมากในการเตรียมงานเพื่อใช้ในการประเมินภาคปฏิบัติ
3. ท่านคิดว่าการใช้รูปрикเพื่อการประเมินงานภาคปฏิบัติเป็นสิ่งจำเป็น	1.00	ไม่มีการแก้ไข	3. ท่านคิดว่าการใช้รูปริกเพื่อการประเมินงานภาคปฏิบัติเป็นสิ่งจำเป็น
4. ท่านรู้สึกดีที่ได้มีส่วนร่วมในการประเมินภาคปฏิบัติ	1.00	ไม่มีการแก้ไข	4. ท่านรู้สึกดีที่ได้มีส่วนร่วมในการประเมินภาคปฏิบัติ
5. ท่านรู้สึกว่าการประเมินภาคปฏิบัติเป็นเรื่องน่าเบื่อ	1.00	ไม่มีการแก้ไข	5. ท่านรู้สึกว่าการประเมินภาคปฏิบัติเป็นเรื่องน่าเบื่อ
6. ถ้าไม่จำเป็นท่านจะไม่เข้ามามีส่วนร่วมในการประเมินภาคปฏิบัติ	0.89	ปรับคำไม่ใช่ปฏิเสธซ้อนปฏิเสธ	6. ถ้าเป็นไปได้ ท่านอยากให้บุคคลอื่นเข้ามามีส่วนร่วมในการประเมินภาคปฏิบัติแทนท่าน
7. ท่านคิดว่าการประเมินภาคปฏิบัติต้องสิ้นเปลืองทรัพยากรทั้งด้านเวลาและกำลังคน	1.00	ไม่มีการแก้ไข	7. ท่านคิดว่าการประเมินภาคปฏิบัติต้องสิ้นเปลืองทรัพยากรทั้งด้านเวลาและกำลังคน
8. การประเมินภาคปฏิบัติของนักศึกษาพยาบาล มีกระบวนการที่ยุ่งยาก	1.00	ไม่มีการแก้ไข	8. การประเมินภาคปฏิบัติของนักศึกษาพยาบาล มีกระบวนการที่ยุ่งยาก
9. การประเมินภาคปฏิบัติของนักศึกษาพยาบาล เป็นเรื่องยากมากที่จะทำให้เกิดความเป็นปรนัย	1.00	ขยายความคำว่าปรนัยให้ชัดเจน	9. การประเมินภาคปฏิบัติของนักศึกษาพยาบาล เป็นเรื่องยากมากที่จะทำให้เกิดความเข้าใจที่ตรงกัน และให้คะแนนตรงกันระหว่างอาจารย์ผู้ประเมิน
10. การแก้ปัญหาที่เกิดกับการประเมินภาคปฏิบัติฯ เป็นเรื่องที่ไม่มีทางเป็นไปได้จริง	0.78	ปรับให้คำมีความสละสลวยขึ้น	10. การแก้ปัญหาที่เกิดกับการประเมินภาคปฏิบัติฯ เป็นเรื่องในอุดมคติ

n = 26

3.2.4.2 การตรวจสอบความสัมพันธ์รายข้อของแบบวัดทัศนคติด้วยค่าสหสัมพันธ์ (item total correlation) มีค่าเท่ากับ 0.30 ถึง 0.81 ความเที่ยงของแบบวัดทั้งฉบับ ใช้วิธีสัมประสิทธิ์แอลฟาของ ครอนบาค (Cronbach's Alpha Coefficient: α เท่ากับ .83 ดังตารางที่ 23

ตารางที่ 23 การตรวจสอบความสัมพันธ์รายข้อของแบบวัดทัศนคติเกี่ยวกับการประเมินภาคปฏิบัติ

รายการประเมิน	r
1. ท่านรู้สึกดีที่ได้เตรียมการสอนภาคปฏิบัติ	0.81
2. ท่านคิดว่าไม่ใช่เรื่องยากในการเตรียมงานเพื่อใช้ในการประเมินภาคปฏิบัติ	0.62
3. ท่านคิดว่าการใช้รูปรีดเพื่อการประเมินงานภาคปฏิบัติเป็นสิ่งจำเป็น	0.37
4. ท่านรู้สึกดีที่ได้มีส่วนร่วมในการประเมินภาคปฏิบัติ	0.30
5. ท่านรู้สึกว่า การประเมินภาคปฏิบัติเป็นเรื่องน่าเบื่อ	0.55
6. ถ้าไม่จำเป็นท่านจะไม่เข้ามามีส่วนร่วมในการประเมินภาคปฏิบัติ	0.66
7. ท่านคิดว่าการประเมินภาคปฏิบัติต้องสิ้นเปลืองทรัพยากรทั้งด้านเวลาและกำลังคน	0.78
8. การประเมินภาคปฏิบัติของนักศึกษาพยาบาล มีกระบวนการที่ยุ่งยาก	0.72
9. การประเมินภาคปฏิบัติของนักศึกษาพยาบาล เป็นเรื่องยากมากที่จะทำให้เกิดความเป็นปรนัย	0.57
10. การแก้ปัญหาที่เกิดกับการประเมินภาคปฏิบัติ เป็นเรื่องที่ไม่มีทางเป็นไปได้จริง	0.35
n = 26	$\alpha = .83$

3.2.5 คู่มือสำหรับอบรมการเสริมสร้างสมรรถนะการประเมินมีการตรวจสอบความตรงเชิงเนื้อหาจากผู้ทรงคุณวุฒิจำนวน 9 ท่าน ใช้ค่า Item Objective Congruence Index (IOC) ค่าเท่ากับ 0.78 ถึง 0.89. มีค่า IOC ปรับปรุงตามข้อเสนอแนะ ดังตารางที่ 24

ตารางที่ 24 การตรวจสอบความตรงเชิงเนื้อหาของคู่มือสำหรับอบรมการเสริมสร้างสมรรถนะการประเมิน ด้วยค่า Item Objective Congruence Index (IOC)

รายการประเมิน	IOC	ข้อเสนอแนะ	การแก้ไข
ปกหน้าคู่มือ	.89	ควรนำรูปภาพมาตกแต่ง ปก หรือให้ความรู้สีที่น่าจับขึ้นมาอ่าน	นำรูปภาพบางตอนของการถ่ายวีดิทัศน์มาใช้เป็นปก
ขนาดและรูปแบบของคู่มือ	1	ขนาดกำลังดี	ไม่มีการแก้ไข
แบบและขนาดอักษร	.56	ตัวอักษรแน่นมาก อ่านไม่สบายตา	ปรับแก้เนื้อหาให้กระชับมากขึ้น และจัดรูปแบบไม่ให้เนื้อหาแน่นไป
คำนำ	1		ไม่มีการแก้ไข
สารบัญ	1		ไม่มีการแก้ไข
จุดประสงค์	1		ไม่มีการแก้ไข
คำแนะนำ	1		ไม่มีการแก้ไข

ตารางที่ 24 การตรวจสอบความตรงเชิงเนื้อหาของคู่มือสำหรับอบรมการเสริมสร้างสมรรถนะการประเมิน ด้วยค่า

Item Objective Congruence Index (IOC) (ต่อ)

รายการประเมิน	IOC	ข้อเสนอแนะ	การแก้ไข
การเรียงเรียงลำดับเนื้อเรื่อง	1		ไม่มีการแก้ไข
รายการอ้างอิง	0.89	ควรปรับมาใช้เป็นบรรณานุกรม	แก้ไขตามคำแนะนำ
เนื้อหาบทที่ 1 บทนำ	1		ไม่มีการแก้ไข
เนื้อหาบทที่ 2 ความรู้พื้นฐานด้านการประเมินภาคปฏิบัติ	0.89	เนื้อหาเกินไป ปรับให้กระชับกว่านี้ ตรวจสอบเกี่ยวกับการกำหนดเกณฑ์ที่ใช้ในวิชาภาคปฏิบัติใหม่	ปรับให้เนื้อหามีความกระชับ ตรวจสอบตามคำแนะนำแล้วยังคงเนื้อหาเดิมไว้ เนื่องจากอ้างอิงจากแหล่งอ้างอิงที่น่าเชื่อถือ
เนื้อหาบทที่ 3 การสร้างความเข้าใจเรื่องความคลาดเคลื่อนจากผู้ประเมิน	1.00		ไม่มีการแก้ไข
เนื้อหาบทที่ 4 กรอบของกรอบอ้างอิง (Frame of Reference training)	0.78	เนื้อหาแน่นมาก ควรสรุปสาระสำคัญ	ปรับเนื้อหาให้กระชับขึ้น
เนื้อหาบทที่ 5 การฝึกการสังเกตพฤติกรรม (Behavioral observation training: BOT)	0.89	เนื้อหาแน่นมาก ควรสรุปสาระสำคัญ	ปรับเนื้อหาให้กระชับขึ้น
เนื้อหาบทที่ 6 การฝึกการตัดสินใจ	.81	เนื้อหาทวนซ้ำมากเกินไป	ปรับเนื้อหาให้กระชับมากขึ้น

n = 9

ตอนที่ 2 ผลการวิเคราะห์ผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับ อาจารย์พยาบาล ต่อความรู้ ทักษะคติ และทักษะการประเมินภาคปฏิบัติ

การวิเคราะห์ผลของโมดูลการฝึกอบรมฯ ต่อความรู้ ทักษะคติ และทักษะการประเมินภาคปฏิบัติ แบ่งเป็น 2 ส่วนคือ ส่วนที่ 1 การวิเคราะห์ผลการทดลองใช้โมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินในตัวอย่าง 6 คนที่ผ่านการจับคู่เข้ากลุ่มทดลองใช้โมดูลการฝึกอบรม ส่วนที่ 2 การวิเคราะห์ผลการใช้โมดูลการฝึกอบรม ในตัวอย่าง 11 คนในกลุ่มใช้โมดูลจริง ตัวอย่างเป็นอาจารย์ที่รับผิดชอบการสอนภาคปฏิบัติของคณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต

1. ผลการทดลองใช้โมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์ พยาบาล

ผลการทดลองใช้โมดูลการฝึกอบรม นำเสนอผลการศึกษาใน 2 ประเด็น คือ 1) ข้อมูลทั่วไปของตัวอย่าง 2) ผลการวัดความรู้ ทักษะคติ และทักษะ ต่อการประเมินภาคปฏิบัติ ก่อน และหลังการฝึกอบรม 3) ผลการวัดความพึงพอใจของผู้เข้ารับการฝึกอบรมหลังสิ้นสุดการฝึกอบรมทันที และ ผลการสัมภาษณ์เฉพาะกลุ่มหลังสิ้นสุดการฝึกอบรม

1.1 ข้อมูลทั่วไปของตัวอย่าง ตัวอย่างที่ใช้ในการทดลองการใช้อินเตอร์เน็ตฝึกอบรม เลือกขนาดตัวอย่างจำนวน 6 คน ตัวอย่างมีข้อมูลทั่วไปดังนี้ ส่วนใหญ่เป็นเพศหญิง จำนวน 5 คน เป็นเพศชายจำนวน 1 คน ส่วนใหญ่มีระดับการศึกษาปริญญาโท จำนวน 5 คน ปริญญาตรีจำนวน 1 คน ส่วนใหญ่สังกัดสาขาวิชา การพยาบาลผู้ใหญ่ จำนวน 3 คน สาขาวิชาเด็ก และวัยรุ่น 2 คน และสาขาวิชาการพยาบาลชุมชน 1 คน ดังแสดงในตาราง 25

ตารางที่ 25 ข้อมูลทั่วไป

รายการประเมิน	จำนวน
เพศ	
หญิง	5
ชาย	1
ระดับการศึกษา	
ปริญญาตรี	1
ปริญญาโท	5
สาขาวิชา	
การพยาบาลผู้ใหญ่	3
การพยาบาลชุมชน	1
การพยาบาลเด็ก และวัยรุ่น	2
ประสบการณ์ในการสอนภาคปฏิบัติ	
น้อยกว่า 3 ปี	3
3 ถึง 5 ปี	2
6 ถึง 10 ปี	1
ประสบการณ์ในการเข้ารับการฝึกอบรมเป็นผู้ประเมิน	
เคย	0
ไม่เคย	6

n = 6

1.2 ผลการวัดความรู้ ทักษะ และทักษะการประเมินภาคปฏิบัติ นำเสนอผลได้ดังนี้

1.2.1 ผลการวัดความรู้ด้านการประเมินภาคปฏิบัติของอาจารย์พยาบาล ก่อน และ หลังการเข้ารับการฝึกอบรม พบว่าค่าเฉลี่ยของคะแนนความรู้ด้านการประเมินภาคปฏิบัติหลังรับการฝึกอบรม สูงกว่าค่าเฉลี่ยของคะแนนความรู้ฯ ก่อนการเข้ารับการฝึกอบรมอย่างมีนัยสำคัญทางสถิติ ที่ระดับ .05 ดังแสดงในตาราง 26

ตารางที่ 26 การเปรียบเทียบค่าเฉลี่ยคะแนนความรู้เกี่ยวกับการประเมินภาคปฏิบัติของอาจารย์พยาบาล ก่อนและหลังการใช้โมดูลฯ

การวัด	Mean	SD	t	p
ก่อน	11.50	2.59	3.25	.02 *
หลัง	16.00	1.90		

*p < .05

1.2.2 ผลการวัดทัศนคติด้านการประเมินภาคปฏิบัติของอาจารย์พยาบาล ก่อน และ หลังการเข้ารับการฝึกอบรมโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล พบว่าค่าเฉลี่ยของคะแนนทัศนคติด้านการประเมินภาคปฏิบัติ ก่อน และหลังการฝึกอบรม ไม่แตกต่างกันที่ระดับนัยสำคัญ .05 ดังแสดงในตารางที่ 27

ตารางที่ 27 การเปรียบเทียบค่าเฉลี่ยของคะแนนทัศนคติต่อการประเมินภาคปฏิบัติของอาจารย์พยาบาล ก่อนและหลังการใช้โมดูลการฝึกอบรมการเสริมสร้างสมรรถนะการประเมิน

การวัด	Mean	SD	t	p
ก่อน	34.83	5.53	-.35	.74
หลัง	34.33	8.04		

p < .05

1.2.3 ผลการวัดทักษะการประเมินภาคปฏิบัติ ก่อน และหลังการฝึกอบรมทันที ในการประเมินของผู้ประเมิน 6 คน ทำการประเมินงานภาคปฏิบัติจำนวน 8 ชุดเดียวกัน เมื่อทดสอบด้วยสถิติระดับกลุ่ม Fixed Chi-Square พบว่ามีผู้ประเมินอย่างน้อย 2 คน ใน 6 คนมีระดับการกด ปล่อยคะแนนแตกต่างกัน $\chi^2 = 98.1, d.f. = 5, p = .00, Mean = 0.86, SD = 0.04$) เมื่อพิจารณาค่า Infit mean square ผลการประเมินของผู้ประเมินทุกคนประเมินสอดคล้องกับที่ โมเดลคาดการณ์ หรืออยู่ในค่าที่คาดการณ์โดยโมเดล (ค่า Fit mean-square 0.98 – 1.37)

เมื่อพิจารณารายบุคคลโดยการทดสอบความแตกต่างที่เกิดจากการประเมินในแต่ละ ครั้ง (ทดสอบความแตกต่างของ target contrast/ Joint S.E.) (Linacre, 2014) ผลการวิเคราะห์ นำเสนอผลต่างของค่าที่สังเกตได้กับค่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) มีหน่วยการวัด เป็นโลจิท (target measurement) ค่าบวกแสดงว่ามีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วย โมเดล ค่าลบแสดงว่ามีการกดคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล ค่า t เป็นการทดสอบสมมติฐาน ที่ว่าไม่พบความแตกต่างของ target contrast/Joint S.E. ผลการวิเคราะห์พบความแตกต่างอย่างมี ระดับนัยสำคัญทางสถิติที่ระดับ .05 ในผู้ประเมินคนที่ 1 และคนที่ 2 และพบความแตกต่างอย่างมี นัยสำคัญทางสถิติที่ระดับ .01 ในผู้ประเมินคนที่ 6 แต่ ไม่พบความแตกต่างของผลการประเมินที่ ระดับนัยสำคัญ .05 ในผู้ประเมินคนที่ 3 คนที่ 4 และคนที่ 5 การฝึกอบรมทำให้คุณภาพการประเมิน ของผู้ประเมินคนที่ 1 และคนที่ 6 ดีขึ้น ดังรายละเอียดต่อไปนี้

ผู้ประเมินคนที่ 1

วัดครั้งที่ 1 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.10
และหน่วยการวัดเป็นโลจิท (target measurement) = 2.71

ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ - 0.10 (1.85 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) (0.23 - 0.08) = 0.86 โลจิท

มีความต่างของคะแนนอย่างมีนัยสำคัญทางสถิติที่ระดับ.05 ($t = 2.16, p = .03$) และหลังการฝึกอบรมมีการปล่อยคะแนนลดลงแสดงว่ามีคุณภาพการประเมินดีขึ้น

ผู้ประเมินคนที่ 2

วัดครั้งที่ 1 มีการกดคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = - 0.10
และหน่วยการวัดเป็นโลจิท (target measurement) = - 0.02 โลจิท

ครั้งที่ 2 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ฯ 0.10 (0.70 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast)(- 0.02 - 0.70) = -.72 โลจิท

มีความแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ.05 ($t = -1.99, p = .04$) ก่อนการฝึกอบรมมีการกดคะแนน หลังการฝึกอบรมมีการปล่อยคะแนนเมื่อพิจารณาค่า Infit mean-square มีค่าใกล้เคียงกัน ทำให้ไม่สามารถสรุปได้ชัดเจนว่าการฝึกอบรมทำให้คุณภาพการประเมินดีขึ้น

ผู้ประเมินคนที่ 3

วัดครั้งที่ 1 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.01
และหน่วยการวัดเป็นโลจิท (target measurement) = 1.85 โลจิท

ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ -0.01 (1.78 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) (- 0.22- -0.41) = 0.07 โลจิท แต่ไม่พบความต่างกัน ที่ระดับนัยสำคัญ .05 ($t = 0.19, p = .84$)

ผู้ประเมินคนที่ 4

วัดครั้งที่ 1 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.02
และหน่วยการวัดเป็นโลจิท (target measurement) = 2.23 โลจิท

วัดครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ - 0.02 (2.08 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) (2.23 - 2.08) = 0.15 โลจิท แต่ไม่พบความต่างกัน ที่ระดับนัยสำคัญ .05 ($t = 0.39, p = .69$)

ผู้ประเมินคนที่ 5

วัดครั้งที่ 1 มีการกดคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = - 0.07 และหน่วยการวัดเป็นโลจิท (target measurement) = 0.63 โลจิท

ครั้งที่ 2 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ฯ 0.07 (1.16 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) (0.63-1.16) = - 0.53 โลจิท แต่ไม่พบความต่างของคะแนนที่ระดับนัยสำคัญ .05 ($t = -1.45, p = .14$)

ผู้ประเมินคนที่ 6

วัดครั้งที่ 1 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.12 และหน่วยการวัดเป็นโลจิท (target measurement) = 2.89 โลจิท

ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ - 0.11 (1.85 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) (2.89 -1.85) = 1.04 โลจิท มีความต่างของคะแนน อย่างมีนัยสำคัญทางสถิติที่ระดับ.01 ($t = 2.57, p = .01$) การฝึกอบรมทำให้คุณภาพการให้คะแนนดีขึ้น สังเกตได้จากค่าผลต่างระหว่างค่าที่สังเกตได้กับค่าที่คาดการณ์ด้วยโมเดลลดลงภายหลังการฝึกอบรม ดังแสดงในตารางที่ 28

ตารางที่ 28 ผลการวิเคราะห์ความคลาดเคลื่อนของปฏิสัมพันธ์ระหว่างผู้ประเมิน กับเวลาในการวัด 2 ครั้ง (Rater X Time interactions) จากการวัดด้วยวิธีทัศนชุดเดียวกัน

ผู้ประเมิน	ครั้งที่	Target Measr	Infit MnSq	Outfit MnSq	S.E.	Obs.-Exp. Average	Target contrast	Joint S.E.	t	p
คนที่ 1	1	2.71	0.98	0.90	.27	0.10	.86	.40	2.16	.03
	2	1.85	1.06	0.97		-0.10				
คนที่ 2	1	-.02	0.63	0.59	.26	-0.10	-.72	.36	-1.99	.04
	2	.70	0.62	0.57		0.10				
คนที่ 3	1	1.85	1.37	1.44	.27	0.01	.07	.38	.19	.84
	2	1.78	1.35	1.42		-0.01				
คนที่ 4	1	2.23	1.26	1.23	.27	0.02	.15	.39	.39	.69
	2	2.08	1.23	1.13		-0.02				
คนที่ 5	1	.63	0.81	1.21	.26	-.07	-.53	.36	-1.45	.14
	2	1.16	0.81	1.36		.07				
คนที่ 6	1	2.89	0.69	0.59	.27	.12	1.04	.40	2.57	.01
	2	1.85	0.68	0.59		-.11				

Summary statistics: Fixed (all=0), Chi-square: 98.1, d.f.= 5, $p = .00$, Mean = 0.86, SD = 0.04

2. ผลการวัดความพึงพอใจของผู้เข้ารับการฝึกอบรมหลังสิ้นสุดการฝึกอบรมทันที

ผลการวัดความพึงพอใจของผู้เข้ารับการฝึกอบรมหลังสิ้นสุดการฝึกอบรมทันที ผู้เข้ารับการฝึกอบรมมีความพึงพอใจเรียงลำดับจากพึงพอใจในระดับมากที่สุด ถึงปานกลางได้ดังนี้ ประโยชน์ของการฝึกอบรมต่อตัวท่าน (M = 4.67, SD = 0.52) ประโยชน์ของการฝึกอบรมต่อองค์กรของท่าน (M = 4.67, SD = 0.52) ปริมาณ และคุณภาพของอาหาร และอาหารว่าง (M = 4.67, SD = 0.52) ความคุ้มค่ากับการเข้ารับการฝึกอบรม (M = 4.50, SD = 0.55) พึงพอใจในระดับมาก คือ ความรู้ความสามารถของผู้ดำเนินการฝึกอบรม และวิทยากร (M = 4.33, SD = 0.52) และ ความน่าสนใจของกิจกรรมการฝึกอบรม (M = 4.33, SD = 0.82) ความชัดเจนเข้าใจง่ายของเนื้อหาในการฝึกอบรม (M = 4.00, SD = 0.63) สถานที่ มีความสะดวก สบาย และเหมาะสำหรับใช้ในการฝึกอบรมครั้งนี้ (M = 4.00, SD = 0.63) ความพร้อมของสื่อโสตทัศนูปกรณ์ (M = 3.83, SD = 0.75) และมีความพึงพอใจในระดับปานกลางต่อประเด็น ระยะเวลาการฝึกอบรมมีความเหมาะสม (M = 2.50, SD = 0.84) ดังแสดงในตารางที่ 29

ตารางที่ 29 ผลการวัดความพึงพอใจของผู้เข้ารับการฝึกอบรมกลุ่มทดลองใช้โมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล

หัวข้อประเมิน	ความพึงพอใจ		
	Mean	SD	ระดับความพึงพอใจ
1. ประโยชน์ของการฝึกอบรมต่อตัวท่าน	4.67	0.52	มากที่สุด
2. ประโยชน์ของการฝึกอบรมต่อองค์กรของท่าน	4.67	0.52	มากที่สุด
3. ปริมาณ และคุณภาพของอาหาร และอาหารว่าง	4.67	0.52	มากที่สุด
4. ความคุ้มค่ากับการเข้ารับการฝึกอบรม	4.50	0.55	มากที่สุด
5. ความรู้ ความสามารถของผู้ดำเนินการฝึกอบรม และวิทยากร	4.33	0.52	มาก
6. ความน่าสนใจของกิจกรรมการฝึกอบรม	4.33	0.82	มาก
7. ความชัดเจนเข้าใจง่ายของเนื้อหาในการฝึกอบรม	4.00	0.63	มาก
8. สถานที่ มีความสะดวก สบายและเหมาะสำหรับใช้ในการฝึกอบรมครั้งนี้	4.00	0.63	มาก
9. ความพร้อมของสื่อโสตทัศนูปกรณ์	3.83	0.75	ปานกลาง
10. ระยะเวลาการฝึกอบรมมีความเหมาะสม	2.50	0.84	ปานกลาง

n = 6

3. ผลการสัมภาษณ์เฉพาะกลุ่มหลังสิ้นสุดการฝึกอบรมทันที

หลังสิ้นสุดการฝึกอบรมทันที มีการสัมภาษณ์เฉพาะกลุ่มกับตัวอย่างที่สะดวกที่จะสนทนาตามความสมัครใจ ในการสนทนาครั้งนี้ตัวอย่างอยู่ร่วมสนทนากลุ่ม 5 คนจากจำนวน 6 คน ผลการสัมภาษณ์ ใช้แนวคำถามมีโครงสร้างสำหรับการสัมภาษณ์เฉพาะกลุ่ม เพื่อให้ได้ข้อมูลเพิ่มเติมจากข้อมูลเชิงปริมาณจากแบบสอบถามในด้านความรู้ ทักษะคติ และทักษะที่ได้รับหลังการฝึกอบรม และความพึงพอใจต่อการฝึกอบรม ใช้เวลาในการสนทนาประมาณ 20 นาที สนทนา ณ วันที่ 21 กรกฎาคม 2568 โดยใช้แนวคำถามหลัก คือ 1) การฝึกอบรมทำให้เกิดความรู้เกี่ยวกับการประเมินภาคปฏิบัติอย่างไร 2) การฝึกอบรมทำให้ท่านมีทักษะในการประเมินภาคปฏิบัติมากขึ้นหรือไม่อย่างไร 3) เมื่อเทียบระหว่างก่อน และหลังการฝึกอบรม ท่านรู้สึกอย่างไรกับการเข้าไปเกี่ยวข้องกับการประเมินภาคปฏิบัติในนักศึกษาพยาบาล และ 4) จุดเด่นของการฝึกอบรม และจุดที่ควรปรับปรุงในการฝึกอบรมคืออะไร ควรดำเนินการอย่างไรต่อไป ผลการวิเคราะห์เนื้อหาประเด็นสำคัญ 4 ประเด็น คือ 1) ความรู้ใหม่ที่ได้รับหลังการฝึกอบรม 2) การรับรู้ถึงความยุ่งยากในการเป็นผู้ประเมินที่มีคุณภาพ 3) ความเชื่อมั่นในทักษะการประเมิน 4) สิ่งที่ต้องปรับปรุงในการฝึกอบรม ดังรายละเอียดต่อไปนี้

3.1 ความรู้ใหม่ที่ได้รับหลังการฝึกอบรม ผู้เข้ารับการอบรมส่วนใหญ่ 4 ใน 5 คน กล่าวถึงเนื้อหาการฝึกอบรมว่า เป็นเรื่องใหม่ บางประเด็นเคยได้ยิน เช่นการกดปล่อยคะแนน บางประเด็นไม่เคยได้ยิน เช่น ความคลาดเคลื่อนแบบฮาโล จึงเป็นเรื่องที่ต้องอาศัยการทำความเข้าใจ ดังบทสนทนาต่อไปนี้

“ก็ทำงานมานานก็จริง แต่เพิ่งมาเป็นครู อายุมากขึ้นการเรียนรู้เรื่องใหม่ๆ ก็อาจจะเข้าไปหน่อย ... แต่เนื้อหาเป็นประโยชน์มาก”

ผู้ให้ข้อมูลคนที่ 1

“เนื้อหาประโยชน์จริง เหมือนเป็นเรื่องใหม่ แต่พอดูเนื้อหาจริงๆ เนี่ยเป็นความเป็นจริงที่เราพบกันอยู่ แต่เอามาพูดเป็นศัพท์ วิชาการหน่อย อย่างอันไหนนะ...similar to me เออ เราก็เคยให้คะแนนเด็กอย่างนี้ ก็คิดถึงตัวเราตอนเป็นนักเรียน... ไม่คิดว่าจะผิดหลักการ ”

ผู้ให้ข้อมูลคนที่ 2

“ฟังตอนแรกงง ไม่เคยได้ยิน แต่พอได้ฟังคำอธิบายในความหมายก็พอเข้าใจ แต่เนื้อหาเยอะนะ ตอนทำข้อสอบก่อนอบรมนี่มัน พออบรมเสร็จก็ยังมีอยู่ คะแนนไม่น่าจะเพิ่มขึ้น”

ผู้ให้ข้อมูลคนที่ 4

“พอมีความรู้ด้านวัด และประเพณีผลอยู่เพราะจบมาโดยตรง แต่เนื้อหาที่ได้รับเจาะลึกลงไปมาก... เนื้อหาเยอะ แต่ดี ”

ผู้ให้ข้อมูลคนที่ 5

“ไม่เคยเป็นอาจารย์มาก่อน ศาสตร์ และศิลป์ก็ยังไม่ได้เรียน...เนื้อหาค่อนข้างมาก แต่ก็มีประโยชน์ สำหรับอาจารย์ใหม่ ”

ผู้ให้ข้อมูลคนที่ 6

3.2 การรับรู้ถึงความยุ่งยากในการเป็นผู้ประเมินที่มีคุณภาพ

“รู้สึกที่เราต้องระวังตัวไม่ให้เกิดความลำเอียง สาระพียงตลอดเวลา มันก็เป็นธรรมชาติของคนนะ ”

ผู้ให้ข้อมูลคนที่ 1

“ อย่างนี้ ตอนขึ้นวอร์ดก็ต้องเตือนตัวอย่างอย่าเผลอ ลำเอียงโดยไม่รู้ตัว ลำบากนะ ”

ผู้ให้ข้อมูลคนที่ 2

“ต้องมีสติ คิดถึงความลำเอียงที่จะเกิดขึ้นอยู่เสมอ ถ้าต้องให้คะแนนจริง”

ผู้ให้ข้อมูลคนที่ 3

“ก็ยากอยู่ บางทีคนเราก็จะมีความรู้สึก bias อยู่บ้างใช่ไหม ”

ผู้ให้ข้อมูลคนที่ 4

“ก็พอรู้หลักการแหละ แต่ก็นะ ความเป็นคน พอได้รู้พื้นความรู้ ความจำที่ดี จะได้จำใส่ใจให้มากขึ้น”

ผู้ให้ข้อมูลคนที่ 5

3.2 ความเชื่อมั่นในทักษะการประเมิน ตัวอย่างส่วนใหญ่เชื่อว่าความรู้ และทักษะที่ได้รับจากการฝึกอบรมจะสามารถนำไปใช้ประโยชน์ในสถานการณ์จริงของการประเมินภาคปฏิบัติของนักศึกษาพยาบาล โดยมีข้อมูลจากการสัมภาษณ์เฉพาะกลุ่มดังนี้

“พอได้อบรมแล้ว รู้สึกว่าเรื่องบางเรื่องที่เรามองข้ามจะมองข้ามต่อไปไม่ได้แล้ว คงต้องเอาไปใช้ประโยชน์ แต่จะทำได้มากได้น้อยก็ไม่น่าใจต้องเอาไปลองทำก่อน ถึงจะบอกได้ว่าความรู้ที่ได้มัน work”

ผู้ให้ข้อมูลคนที่ 1

“ สำหรับพี่นะ พี่ว่าน่าจะเอาไปใช้ได้จริง ถ้าเอาไปใช้จริงมันก็ต้องเห็นผลสิ ว่าผลการให้คะแนนเด็กมันดูดีขึ้น มีหลักการมากขึ้น ”

ผู้ให้ข้อมูลคนที่ 2

“เป็นประโยชน์กับอาจารย์ใหม่อย่างผมมาก ครับ ทำให้มั่นใจขึ้นเวลาจะต้องนิเทศเด็กๆ ตอนนี่ก็มีมาช่วยงานนิเทศบ้าง จะได้เอาไปใช้จริงเลย ”

ผู้ให้ข้อมูลคนที่ 3

“ น่าสนใจนะ... คิดว่าถ้าอาจารย์ทุกคนเข้าใจตรงกันได้ แล้วเอาหลักการนี้ไปใช้ อย่างในคู่มือที่บอกวิธีหลีกเลี่ยง bias แต่ละประเภท คงจะดี ”

ผู้ให้ข้อมูลคนที่ 4

“ถ้าเอาหลักการพวกนี้ไปใช้ลดความ bias เวลาประเมินนักศึกษา หรือเวลาจะตัดเกรด มันน่าจะดีขึ้น แต่ก็มันแหละ พอความเป็นจริง แต่ละคนก็มีตัวตน ”

ผู้ให้ข้อมูลคนที่ 5

3.4 สิ่งที่ควรปรับปรุงในการฝึกอบรม เป็นเรื่องเกี่ยวกับประโยชน์ของเนื้อหาที่ได้รับในการฝึกอบรม ความแปลกใหม่ ของสื่อในรูปวีดิทัศน์ สิ่งที่ควรปรับปรุง ในประเด็นสำคัญที่สามารถนำไปปรับปรุงได้คือ ระยะเวลาในการฝึกอบรมน้อยเกินไปเมื่อเทียบกับเนื้อหา และระยะเวลาในการฝึกประเมินก่อน และหลังการฝึกอบรมมากเกินไป วีดิทัศน์มีระยะเวลานานเกินไป ดังตัวอย่างบทสนทนาต่อไปนี้

“วิดีโอยาวมากๆ พอดูตอนแรกๆ ก็ตื่นเต้นดี พอดูไปๆ มันล้า ทำให้ระชับกว่านี้ได้จะดี”

ผู้ให้ข้อมูลคนที่ 1

“ คลิปวิดีโอยาวไป กินเวลามาก ทำให้เบียดเวลาการบรรยายที่เนื้อหาที่แน่นเหมือนกัน ”

ผู้ให้ข้อมูลคนที่ 2

“เนื้อหาเยอะครับ น่าจะจัดสัก 3 วัน ใช้เวลากับวิดีโอเยอะด้วยครับ ”

ผู้ให้ข้อมูลคนที่ 3

“ฝึกอบรม 2 วันไม่พอสอน เนื้อหาเยอะ โดยเฉพาะคลิปวิดีโอ ยาวไปค่ะ ”

ผู้ให้ข้อมูลคนที่ 4

“เช่นเดียวกับที่ อาจารย์...ค่ะว่าวิดีโอยาวไป น่าจะย่อเวลาให้สั้นกว่านี้ จับประเด็นสำคัญจริงๆ มาเป็นบทพูด ”

ผู้ให้ข้อมูลคนที่ 5

2. ผลการใช้โมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล

ผลการวิเคราะห์ข้อมูลการใช้โมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล ในอาจารย์พยาบาลที่ทำการสอนภาคปฏิบัติของคณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต จำนวน 11 คน ผลการศึกษาประกอบด้วยประเด็นต่อไปนี้ 1) ข้อมูลทั่วไปของตัวอย่าง 2) ผลการวัดความรู้ ทักษะ และทักษะ ต่อการประเมินภาคปฏิบัติ ก่อน และ หลังการฝึกอบรมทันที และหลักการฝึกอบรม 3 เดือน 3) ผลของโมดูลต่อความคลาดเคลื่อนจากผู้ประเมิน คือ ความคลาดเคลื่อนแบบ กต ปล่อยคะแนน ให้คะแนนเกาะกลุ่มตรงกลาง และ ความคลาดเคลื่อนแบบ ฮาโล 4) ความสอดคล้อง และความเที่ยงของผลการประเมินภาคปฏิบัติระหว่างผู้ประเมินที่เข้าอบรม และระหว่างผู้เข้ารับการฝึกอบรมกับผู้ทรงวุฒิ 5) ผลการวัดความพึงพอใจของผู้เข้ารับการฝึกอบรมหลังสิ้นสุดการฝึกอบรมทันที ร่วมกับการสัมภาษณ์กลุ่มเฉพาะและผลการสัมภาษณ์เฉพาะกลุ่มหลังสิ้นสุดการฝึกอบรมทันที และหลังการฝึกอบรม 3 เดือน ดังรายละเอียดต่อไปนี้

2.1 ข้อมูลทั่วไปของตัวอย่าง ส่วนใหญ่เป็นเพศหญิง (ร้อยละ 90.9) ระดับการศึกษาส่วนใหญ่ ระดับปริญญาโท (ร้อยละ 81.8) รองลงมาปริญญาเอก (ร้อยละ 18.2) ตัวอย่างส่วนใหญ่สังกัดสาขาวิชาการพยาบาลผู้ใหญ่ และการพยาบาลชุมชน (ร้อยละ 27.3) รองลงมาสังกัดการพยาบาลเด็ก

และวัยรุ่น และการพยาบาลจิตเวช (ร้อยละ 18.2) และการพยาบาลมารดาทารก และผดุงครรภ์ (ร้อยละ 9.1) ส่วนใหญ่มีประสบการณ์การสอน 6 ถึง 10 ปี (ร้อยละ 45.5) รองลงมา มีประสบการณ์ฯ ไม่ถึง 3 ปี (ร้อยละ 27.3) มีประสบการณ์ฯ 11 ถึง 19 ปี (ร้อยละ 18.2) และมีประสบการณ์ฯ 3 ถึง 5 ปี และส่วนใหญ่ไม่เคยเข้ารับการฝึกอบรมเป็นผู้ประเมินภาคปฏิบัติ (ร้อยละ 90.9) ดังแสดงในตารางที่ 30

ตารางที่ 30 ข้อมูลทั่วไป

รายการประเมิน	จำนวน	ร้อยละ
เพศ		
หญิง	10	90.9
ชาย	1	9.1
ระดับการศึกษา		
ปริญญาโท	9	81.8
ปริญญาเอก	2	18.2
สาขาวิชา		
การพยาบาลผู้ใหญ่	3	27.3
การพยาบาลชุมชน	3	27.3
การพยาบาลจิตเวช	2	18.2
การพยาบาลเด็ก และวัยรุ่น	2	18.2
การพยาบาลมารดา ทารกและผดุงครรภ์	1	9.1
ประสบการณ์ในการสอนภาคปฏิบัติ		
น้อยกว่า 3 ปี	3	27.3
3 ถึง 5 ปี	1	9.1
6 ถึง 10 ปี	5	45.5
11 ถึง 19 ปี	2	18.2
ประสบการณ์ในการเข้ารับการฝึกอบรมเป็นผู้ประเมิน		
ไม่เคย	10	90.9
เคย	1	9.1

n = 11

2.2 ผลการวัดความรู้ต่อการประเมินภาคปฏิบัติ ก่อน และ หลังการใช้โมดูลการฝึกอบรมทันที และ หลังการใช้ 3 เดือน

2.2.1 ผลการทดสอบเงื่อนไข sphericity ของค่าเฉลี่ยคะแนนของคะแนนความรู้เกี่ยวกับการประเมินภาคปฏิบัติที่ได้จากการวัดซ้ำนั้น ทำการตรวจสอบความเป็น sphericity ของเมทริกซ์ ความแปรปรวน ความแปรปรวนร่วมของคะแนนความรู้เกี่ยวกับการประเมินภาคปฏิบัติที่ได้จากการวัด 3 ครั้ง ด้วยวิธีทดสอบ

Mauchly's test of sphericity ซึ่งตัวแปรที่ต้องการวัดมีมิติเดียว คือคะแนนรวมของความรู้พื้นฐานด้านการประเมินภาคปฏิบัติ และมี 3 ตัวแปรคือ คะแนนความรู้ ในระยะก่อนการฝึกอบรม หลังการฝึกอบรมทันที และหลังการฝึกอบรม 3 เดือน ผลการทดสอบ พบว่าค่าสถิติของการวัดทั้งสามครั้งมีค่าเท่ากับ .55 ค่า Approx. ChiSquare มีค่าเท่ากับ 5.31 ค่า df มีค่าเท่ากับ 2 ค่า p เท่ากับ .07 แสดงว่าตัวแปรการวัดซ้ำ มีเมทริกซ์ ความแปรปรวน และความแปรปรวนร่วมเป็นเมทริกซ์เอกลักษณะ (compound symmetry) ที่ระดับนัยสำคัญ .05 ซึ่งเป็นไปตามข้อตกลงเบื้องต้นของการใช้สถิติ Repeated measurement ANOVA จึงใช้ค่า Sphericity Assumed ในการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยคะแนนความรู้เกี่ยวกับการประเมินภาคปฏิบัติ และผลของการวิเคราะห์ความแปรปรวน ด้วยค่า Sphericity Assumed ค่า Greenhouse-Geisser, Huynh-Feldt และค่า Lower-bound ดังตารางที่ 31 ผลการวิเคราะห์ Sphericity Assumed พบว่ามีความแตกต่างของค่าเฉลี่ยอย่างน้อย 1 คู่ จึงทำการศึกษเปรียบเทียบค่าเฉลี่ยรายคู่ต่อ

ตารางที่ 31 ผลการทดสอบ sphericity และผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยของคะแนนความรู้พื้นฐานด้านการประเมินภาคปฏิบัติ จากการวัดความรู้ 3 ครั้งด้วยแบบทดสอบชุดเดียวกัน

แหล่งข้อมูล	Mauchly's W	Approx. Chi-Square	Type III Sum of Squares	df	Mean Square	F	Sig.	Epsilon
Within Subjects Effect (เวลาในการวัดซ้ำ 3 ครั้ง)	.554	5.312		2			.070	
Greenhouse-Geisser								.692
Huynh-Feldt								.767
Lower-bound								.500
เวลาในการ วัดซ้ำ 3 ครั้ง								
Sphericity Assumed			80.182	2	40.091	15.474	.000	
Greenhouse-Geisser			80.182	1.383	57.963	15.474	.001	
Huynh-Feldt			80.182	1.534	52.275	15.474	.000	
Lower-bound			80.182	1.000	80.182	15.474	.003	
Error (เวลาในการวัดซ้ำ 3 ครั้ง)								
Sphericity Assumed			51.818	20	2.591			
Greenhouse-Geisser			51.818	1.383	3.746			
Huynh-Feldt			51.818	15.338	3.378			
Lower-bound			51.818	10.000	5.182			

2.2.2 ผลการเปรียบเทียบค่าเฉลี่ยคะแนนความรู้ เกี่ยวกับการประเมินภาคปฏิบัติ ก่อน และ หลังการฝึกอบรมทันที และ หลังการฝึกอบรม 3 เดือน การศึกษาเพื่อเปรียบเทียบค่าเฉลี่ยรายคู่ด้วย แบบทดสอบชูดเดียวกัน ผลการศึกษาพบว่า ค่าเฉลี่ยของคะแนนความรู้ฯ จากการวัดครั้งที่ 1 และครั้งที่ 2 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนค่าเฉลี่ยของคะแนนความรู้ฯ จากการวัดครั้งที่ 2 กับ ครั้งที่ 3 และ ครั้งที่ 1 กับครั้งที่ 3 ไม่แตกต่างกันที่ระดับนัยสำคัญ .05 ดังแสดงในตารางที่ 32

ตารางที่ 32 ผลการวิเคราะห์ความแปรปรวน และผลการเปรียบเทียบค่าเฉลี่ยรายคู่ของค่าเฉลี่ยของ คะแนนความรู้เกี่ยวกับการประเมินภาคปฏิบัติ ระหว่างเวลาของข้อมูลชุดวัดซ้ำด้วย แบบสอบฉบับเดิม

แหล่งข้อมูล	df	Type III SS	Mean Difference	MS	SE	F	Sig.
เวลาในการวัดซ้ำ 3 ครั้ง	2.00	80.18		40.09		15.47	.000**
ความคลาดเคลื่อนจากการวัด 3 ครั้ง	20.00	51.82		2.59			
ครั้งที่ 2 กับครั้งที่ 1			3.82		0.42		.000**
ครั้งที่ 3 กับครั้งที่ 2			-1.91		0.72		.071
ครั้งที่ 3 กับครั้งที่ 1			1.91		0.85		.137

**P < .001

2.3 ผลการวัดทัศนคติ ก่อนและ หลังการใช้โมดูลการฝึกอบรมฯ ทันที และหลังการใช้ 3 เดือน

2.3.1 ผลการทดสอบเงื่อนไข sphericity ของค่าเฉลี่ยคะแนนของคะแนนทัศนคติเกี่ยวกับการ ประเมินภาคปฏิบัติที่ได้จากการวัดซ้ำนั้น ผู้วิจัยได้ทำการตรวจสอบความเป็น sphericity ของเมทริกซ์ ความแปรปรวน ความแปรปรวนร่วมของคะแนนทัศนคติเกี่ยวกับการประเมินภาคปฏิบัติที่ได้จากการ วัด 3 ครั้ง ด้วยวิธีทดสอบของ Mauchly ผลการทดสอบ Mauchly's test of sphericity พบว่าค่าสถิติ Mauchly ของการวัดทั้งสามครั้งมีค่าเท่ากับ .827 ค่า Approx. Chi-Square มีค่าเท่ากับ 1.707 ค่า df มีค่า เท่ากับ 2 ค่า p เท่ากับ .426 แสดงว่าตัวแปรการวัดซ้ำ มีเมทริกซ์ ความแปรปรวน และความแปรปรวน ร่วมเป็นเมทริกซ์เอกลักษณะ (compound symmetry) ที่ระดับนัยสำคัญ .05 ซึ่งเป็นไปตามข้อตกลง

เบื้องต้นของการใช้สถิติ Repeated measurement ANOVA จึงใช้ค่า Sphericity Assumed ในการวิเคราะห์ความแปรปรวนของคะแนนเฉลี่ยของทัศนคติในการวัดซ้ำ และผลของการวิเคราะห์ความแปรปรวน ด้วยค่า Sphericity Assumed และการวิเคราะห์ความแปรปรวนด้วยค่า Greenhouse-Geisser, Huynh-Feldt และค่า Lower-bound นำเสนอดังตารางที่ 33

ตารางที่ 33 ผลการทดสอบ sphericity และผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยของคะแนนทัศนคติเกี่ยวกับการประเมินภาคปฏิบัติ จากการวัดความรู้ 3 ครั้งด้วยวัดฉบับเดียวกัน

แหล่งข้อมูล	Mauchl y's W	Approx. Chi- Square	Type III Sum of Squares	df	Mean Square	F	Sig	Epsilon
Within Subjects Effect (เวลา ในการวัดซ้ำ 3 ครั้ง)	.827	1.707		2			.426	
Greenhouse-Geisser								.853
Huynh-Feldt								1.000
Lower-bound								.500
เวลาในการ วัดซ้ำ 3 ครั้ง								
Sphericity Assumed			39.818	2	19.909	2.455	.111	
Greenhouse-Geisser			39.818	1.705	23.350	2.455	.122	
Huynh-Feldt			39.818	2.000	19.909	2.455	.111	
Lower-bound			39.818	1.000	39.818	2.455	.148	
Error (เวลาในการวัดซ้ำ 3 ครั้ง)								
Sphericity Assumed			162.182	20	8.109			
Greenhouse-Geisser			162.182	17.053	9.510			
Huynh-Feldt			162.182	20.000	8.109			
Lower-bound			162.182	10.000	16.218			

2.3.2 ผลการวัดทัศนคติเกี่ยวกับการประเมินภาคปฏิบัติ ก่อน และ หลังการใช้โมดูลการฝึกอบรมทันที และ หลังการใช้ 3 เดือน ได้ผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยของคะแนนความรู้ฯ จากการวัดความรู้ทั้ง 3 ครั้ง พบว่าการวัดทั้ง 3 ครั้งไม่แตกต่างกัน ที่ระดับนัยสำคัญ .05 จึงไม่มีการเปรียบเทียบค่าเฉลี่ยรายคู่ ดังแสดงในตารางที่ 34

ตารางที่ 34 ผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยของคะแนนทัศนคติต่อการประเมิน

ภาคปฏิบัติ จากการวัดซ้ำ 3 ครั้งด้วยแบบวัดฉบับเดียวกัน

แหล่งข้อมูล		SS	MS	F	Sig.
เวลาในการวัดซ้ำ 3 ครั้ง	2	39.82	19.91	2.46	.11
ความคลาดเคลื่อนจากการวัด 3 ครั้ง	20	162.18	8.11		

$p < .05$

24 ผลการวัดทักษะซ้ำ 3 ครั้ง คือ ก่อน และหลังการใช้โมดูลการฝึกอบรมทันที และหลังการใช้ 3 เดือน พิจารณาในภาพรวม ของพฤติกรรมกรรมการประเมินของผู้ประเมิน โดยการทดสอบด้วยสถิติ Fixed Chi-Square พบว่ามีผู้ประเมินอย่างน้อย 2 คน ใน 11 คน ให้คะแนนการประเมินแตกต่างกันระหว่างครั้งที่ 1, 2 และครั้งที่ 3 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ($\chi^2 = 98.1$, $df = 5$, $p = .00$, $Mean = 0.86$, $SD = 0.04$)

เมื่อเปรียบเทียบค่าเฉลี่ยรายคู่ของคะแนนทักษะจากการวัดก่อน และหลังการใช้โมดูลการฝึกอบรมทันที และหลังการใช้ 3 เดือน จากการประเมินงานภาคปฏิบัติทั้งสิ้น 10 ชุด ผลต่างของค่าที่สังเกตได้กับค่าที่คาดการณ์ด้วยโมเดล ($Obs - Exp. Average$) มีหน่วยการวัดเป็นโลจิท (target measurement = observed measurement + Bias size), หากค่า $Obs - Exp. Average$ เป็นบวกแสดงว่ามีการปล่อยคะแนน มากกว่าที่โมเดลคาดการณ์ ค่าลบแสดงว่ามีการกดคะแนนมากกว่าที่โมเดลคาดการณ์ ค่าเป็นศูนย์แสดงว่าค่าที่ให้ตรงกับที่โมเดลคาดการณ์ ค่า t เป็นการทดสอบสมมติฐานของ target contrast/Joint S.E. ที่ว่าไม่พบความแตกต่างของผลการประเมินของผู้ประเมิน ดังรายละเอียดต่อไปนี้ พบว่าการทดสอบความแตกต่างที่เกิดจากการวัดในแต่ละครั้ง (ทดสอบความแตกต่างของ target contrast/ Joint S.E.) (Linacre, 2014) พบความแตกต่างอย่างมีนัยสำคัญที่ระดับ .05 ในการเปรียบเทียบการวัดครั้งที่ 2 และครั้งที่ 3 ในผู้ประเมินคนที่ 3 และคนที่ 4 และพบความแตกต่างอย่างมีนัยสำคัญที่ระดับ .01 ใน ผู้ประเมินคนที่ 11 โดย มีโลจิทการปล่อยคะแนนมากขึ้นในการวัดครั้งที่ 3 ผลของ Infit mean-square ในการวัดทั้ง 3 ครั้ง พบว่าผู้ประเมินทุกคน ให้ผลการประเมินเป็นไปตามการคาดการณ์ของโมเดล (ค่าที่คาดหวัง 0.6 – 1.4) ดังแสดงในตารางที่ 35

ผู้ประเมินคนที่ 1

วัดครั้งที่ 1 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.04
(target measurement) = 0.23 โลจิต

ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ -0.14 (target measurement = 0.08 โลจิต)

ครั้งที่ 3 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ฯ 0.10 (target measurement = 0.28 โลจิต)

ครั้งที่ 1 มีโลจิตการให้คะแนนมากกว่าครั้งที่ 2 (target contrast) (0.23 - 0.08) = 0.15 โลจิต
แต่พบว่าไม่มีความแตกต่างกันของโลจิตที่ระดับนัยสำคัญ .05 ($t = 0.99, p = .32$)

ครั้งที่ 2 แตกต่างจากครั้งที่ 3 (0.08 - 0.28) = - 0.20 โลจิต แต่ไม่พบความแตกต่าง
กันของการให้คะแนนที่ระดับนัยสำคัญ .05 ($t = - 1.30, p = .20$)

ผู้ประเมินคนที่ 2

วัดครั้งที่ 1 มีการกดคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = -0.05
(target measurement) = 0.22 โลจิต

วัดครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล - 0.10 (0.18 โลจิต)

ครั้งที่ 3 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ฯ 0.15 (0.39 โลจิต)

ครั้งที่ 1 มีการให้คะแนนการประเมินแตกต่างจากครั้งที่ 2 (target contrast)
(0.22-0.18) = 0.04 โลจิต แต่ไม่มีความแตกต่างของโลจิตที่ระดับนัยสำคัญ .05 ($t = 0.23, p = .82$)

ครั้งที่ 2 มีโลจิตการให้คะแนนน้อยกว่าครั้งที่ 3 (0.18-0.39) = - 0.21 โลจิต แต่ไม่พบ
ความแตกต่างของคะแนนที่ระดับนัยสำคัญ .05 ($t = - 1.32, p = .19$)

ผู้ประเมินคนที่ 3

วัดครั้งที่ 1 มีการกดคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = - 0.07
(target measurement) = - 0.22 โลจิต

ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ -0.29 (target measurement = 0.41 โลจิต)

ครั้งที่ 3 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ฯ 0.36 (target measurement = 0.13 โลจิต)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) (- 0.22 - - 0.41) = 0.19 โลจิต แต่
ไม่พบความแตกต่างกันของการให้คะแนนที่ระดับนัยสำคัญ .05 ($t = 0.23, p = 0.82$)

ครั้งที่ 2 แตกต่างจากครั้งที่ 3 (- 0.41 - 0.13) = - 0.54 โลจิต พบว่ามีความแตกต่าง
กันของการให้คะแนนอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($t = - 3.41, p = .001$)

ผู้ประเมินคนที่ 4

วัดครั้งที่ 1 มีการกดคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = - 0.10
และหน่วยการวัดเป็นโลจิท (target measurement) = 0.20 โลจิท

ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ - 0.17 (0.14 โลจิท)

ครั้งที่ 3 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ฯ 0.27 (0.52 โลจิท)

ครั้งที่ 1 มีการให้คะแนนแตกต่างจากครั้งที่ 2 (target contrast) (0.20 - 0.14) = 0.06 โลจิท
แต่ไม่พบความแตกต่างกันที่ระดับนัยสำคัญ .05 ($t = 0.38, p = .70$)

ครั้งที่ 2 มีการให้คะแนนน้อยกว่าครั้งที่ 3 (0.14 - 0.52) = - 0.39 โลจิท และพบว่ามี
ความแตกต่างกันของการให้คะแนนอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($t = - 2.42, p = .01$)

ผู้ประเมินคนที่ 5

การวัดครั้งที่ 1 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.02
(target measurement) = 0.23 โลจิท

ครั้งที่ 2 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ฯ 0.01 (0.22 โลจิท)

ครั้งที่ 3 มีการกดคะแนนน้อยกว่าที่คาดการณ์ฯ - 0.03 (0.18 โลจิท)

ครั้งที่ 1 แตกต่างจาก ครั้งที่ 2 (target contrast) (0.23 - 0.22) = 0.01 โลจิท
แต่ไม่พบความแตกต่างกันของการให้คะแนนที่ระดับนัยสำคัญ .05 ($t = 0.08, p = .94$)

ครั้งที่ 2 แตกต่างจากครั้งที่ 3 (0.22-0.18) = 0.04 โลจิท แต่ไม่พบความแตกต่างกัน
ของการให้คะแนนที่ระดับนัยสำคัญ .05 ($t = 0.23, p = .82$)

ผู้ประเมินคนที่ 6

วัดครั้งที่ 1 มีการกดคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs- Exp. Average) = - 0.11
(target measurement) = - 0.04 โลจิท

ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ - 0.07 (0.00 โลจิท)

ครั้งที่ 3 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ฯ 0.19 (0.21 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast)(-0.04-0.00) = 0.03 โลจิท แต่ไม่พบ
ความแตกต่างของการให้คะแนนที่ระดับนัยสำคัญ .05 ($t = - 0.23, p = .82$)

ครั้งที่ 2 แตกต่างจากครั้งที่ 3 (0.00-0.21) = - 0.21 โลจิท แต่ไม่พบความแตกต่างกัน
ของการให้คะแนนที่ระดับนัยสำคัญ .05 ($t = -1.37, p = .17$)

ผู้ประเมินคนที่ 7

วัดครั้งที่ 1 มีการกดคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs- Exp. Average) = - 0.10
และหน่วยการวัดเป็นโลจิท (target measurement) = 0.02 โลจิท

ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ - 0.11 (0.01 โลจิท)

ครั้งที่ 3 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ฯ 0.21 (0.28 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast)(0.02-0.01) = 0.01 โลจิท แต่ไม่พบ
ความแตกต่างที่ระดับนัยสำคัญ .05 ($t = 0.08, p = .94$)

ครั้งที่ 2 แตกต่างจากครั้งที่ 3 (0.01-0.28) = - 0.27 โลจิท แต่ไม่พบความแตกต่างที่
ระดับนัยสำคัญ .05 ($t = 1.75, p = .08$)

ผู้ประเมินคนที่ 8

การวัดครั้งที่ 1 มีการกดคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs- Exp. Average) = - 0.10
target measurement = 0.13 โลจิท

ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ - 0.13 (0.10 โลจิท)

ครั้งที่ 3 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ฯ 0.23 (0.40 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast)(0.13-0.10) = 0.02 โลจิท แต่ไม่พบ
ความแตกต่างที่ระดับนัยสำคัญ .05 ($t = 0.15, p = .88$)

ครั้งที่ 2 แตกต่างจากครั้งที่ 3 (0.10-0.40) = - 0.30 โลจิท แต่ไม่พบความแตกต่างที่
ระดับนัยสำคัญ .05 ($t = -1.93, p = .06$)

ผู้ประเมินคนที่ 9

การวัดครั้งที่ 1 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.15
target measurement = 0.26 โลจิท

ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ - 0.19 (- 0.02 โลจิท)

ครั้งที่ 3 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ฯ 0.04 (0.16 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) (0.26- -0.02) = 0.28 โลจิท
แต่ไม่พบความแตกต่างที่ระดับนัยสำคัญ .05 ($t = 1.82, p = .07$)

ครั้งที่ 2 แตกต่างจากครั้งที่ 3 (- 0.02 - 0.16) = -0.18 โลจิท แต่ไม่พบความแตกต่าง
ที่ระดับนัยสำคัญ .05 ($t = -1.21, p = .23$)

ผู้ประเมินคนที่ 10

วัดครั้งที่ 1 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs- Exp. Average) = 0.16
target measurement = 0.47 โลจิท

ครั้งที่ 2 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ฯ 0.11 (0.28 โลจิท)

ครั้งที่ 3 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ - 0.11 (0.23 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) (0.47-0.28) = 0.19 โลจิท แต่ไม่พบ
ความแตกต่างที่ระดับนัยสำคัญ.05 (t = 1.19, p = .24)

ครั้งที่ 2 แตกต่างจากครั้งที่ 3 (0.28 - 0.23) = 0.05 โลจิท (t = 0.31, p = .76) แต่ไม่พบ
ความแตกต่างที่ระดับนัยสำคัญ .05

ผู้ประเมินคนที่ 11

การวัดครั้งที่ 1 มีการกดคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs- Exp. Average) = - 0.01
และหน่วยการวัดเป็นโลจิท (target measurement) = 0.15 โลจิท

ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ - 0.21 (- 0.01 โลจิท)

ครั้งที่ 3 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ฯ 0.22 (0.34 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) (0.15 - - 0.01) = 0.16 โลจิท
แต่ไม่พบความแตกต่างที่ระดับนัยสำคัญ .05 (t = 1.06, p = .29)

ครั้งที่ 2 แตกต่างจากครั้งที่ 3 (- 0.01 - 0.34) = - 0.35 โลจิท พบความแตกต่างอย่าง
มีนัยสำคัญที่ระดับ .05 (t = - 2.28, p = .02)

ตารางที่ 35 ผลการเปรียบเทียบรายคู่ของคะแนนทักษะจากการวัดก่อน และหลังการ

ฝึกอบรมทันที และหลังอบรม 3 เดือน ด้วยรูบรีค และวิธีทัศนชุดเดียวกัน

ผู้ประเมิน	ครั้งที่วัด	Target measure	Obs- Exp. Average	Infit MnSq	Outfit MnSq	Target contrast	Joint S.E.	t	p
คนที่1	ครั้งที่ 1	0.23	0.04	1.31	1.28	0.15	0.15	0.99	.32
	ครั้งที่ 2	0.08	-0.14	1.06	1.03				
	ครั้งที่ 3	0.28	0.10	1.12	1.06	-0.20	0.15	-1.30	.20

ตารางที่ 35 ผลการเปรียบเทียบรายคู่ของคะแนนทักษะจากการวัดก่อน และหลังการฝึกอบรมทันที และหลังอบรม 3 เดือน ด้วยรูบริค และวิธีทัศนชุดเดียวกัน (ต่อ)

ผู้ประเมิน	ครั้งที่วัด	Target measure	Obs- Exp. Average	Infit MnSq	Outfit MnSq	Target contrast	Joint S.E.	t	p
คนที่2	ครั้งที่1	0.22	-0.05	1.19	1.17	0.04	0.15	0.23	.82
	ครั้งที่ 2	0.18	-0.10	1.28	1.29				
	ครั้งที่ 3	0.39	0.15	1.04	1.03	-0.21	0.16	-1.32	.19
คนที่3	ครั้งที่ 1	-0.22	-0.07	0.64	0.64	0.19	0.16	1.18	.24
	และ 2	-0.41	-0.29	0.64	0.64				
	ครั้งที่ 3	0.13	0.36	0.62	0.71	-0.54	0.16	-3.41	.00**
คนที่4	ครั้งที่1	0.20	-0.10	1.11	1.08	0.06	0.15	0.38	.70
	ครั้งที่ 2	0.14	-0.17	0.94	0.93				
	ครั้งที่ 3	0.52	0.27	1.08	1.12	-0.39	0.16	-2.42	.01
คนที่5	ครั้งที่ 1	0.23	0.02	0.77	0.75	0.01	0.15	0.08	.94
	ครั้งที่ 2	0.22	0.01	0.82	0.82				
	ครั้งที่ 3	0.18	-0.03	0.97	0.94	0.04	0.15	0.23	.82
คนที่6	ครั้งที่ 1	-0.04	-0.11	0.83	0.83	-0.03	0.15	-0.23	.82
	ครั้งที่ 2	0.00	-0.07	0.92	0.91				
	ครั้งที่ 3	0.21	0.19	1.00	0.94	-0.21	0.15	-1.37	.17
คนที่7	ครั้งที่ 1	0.02	-0.10	1.07	1.06	0.01	0.15	0.08	.94
	ครั้งที่ 2	0.01	-0.11	1.28	1.27				
	ครั้งที่ 3	0.28	0.21	1.30	1.29	-0.27	0.15	-1.75	.082
คนที่8	ครั้งที่ 1	0.13	-0.10	1.13	1.09	0.02	0.15	0.15	.88
	ครั้งที่ 2	0.10	-0.13	1.16	1.15				
	ครั้งที่ 3	0.40	0.23	1.31	1.24	-0.30	0.16	-1.93	.056
คนที่9	ครั้งที่ 1	0.26	0.15	0.89	0.88	0.28	0.15	1.82	.07
	ครั้งที่ 2	-0.02	-0.19	0.79	0.79				
	ครั้งที่ 3	0.16	0.04	0.82	0.74	-0.18	0.15	-1.21	.23
คนที่10	ครั้งที่ 1	0.47	0.16	1.10	1.08	0.19	0.16	1.19	.24
	ครั้งที่ 2	0.28	0.11	1.33	1.27				
	ครั้งที่ 3	0.23	-0.11	0.88	0.86	0.05	0.15	0.31	.75
คนที่11	ครั้งที่ 1	0.15	-0.01	0.99	0.98	0.16	0.15	1.06	.29
	ครั้งที่ 2	-0.01	-0.21	0.82	0.80				
	ครั้งที่ 3	0.34	0.22	0.89	0.81	-0.35	0.15	-2.28	.23

Fixed (all same) Chi-Square = 49.1, d.f. = 10, p = .00,

2.6 ผลของความสอดคล้อง และความเที่ยงระหว่างผู้ประเมิน และผู้ทรงคุณวุฒิ ทั้งในระยะก่อนและ หลังการใช้โมดูลการฝึกอบรมทันที และหลังการใช้โมดูล 3 เดือน พบว่า ผู้ประเมินทุกคนมีความสอดคล้องของการให้คะแนนกับผู้ทรงคุณวุฒิ (RAI) = 0.77 ถึง 0.99 ในระดับสูง และมีความเที่ยงระหว่างผู้ประเมิน (ICC) ระหว่างผู้ประเมินกับผู้ทรงคุณวุฒิ อยู่ระหว่าง 0.69 ถึง 0.89 ซึ่งมีความเที่ยงในระดับมาก ถึงมากที่สุด ดังตารางที่ 36

ตารางที่ 36 ความสอดคล้องระหว่างผู้ประเมิน และความเที่ยงภายในผู้ประเมิน

ผู้ประเมิน	RAI1	RAI2	RAI3	ICC1	ICC2	ICC3
คนที่ 1	0.96	0.98	0.95	0.80	0.81	0.84
คนที่ 2	0.98	0.85	0.94	0.80	0.80	0.82
คนที่ 3	0.84	0.89	0.89	0.82	0.82	0.77
คนที่ 4	0.99	0.93	0.96	0.83	0.81	0.85
คนที่ 5	0.98	0.93	0.94	0.86	0.81	0.85
คนที่ 6	0.90	0.93	0.94	0.83	0.83	0.85
คนที่ 7	0.92	0.97	0.96	0.82	0.82	0.84
คนที่ 8	0.96	0.97	0.99	0.85	0.85	0.86
คนที่ 9	0.98	0.77	0.93	0.80	0.80	0.69
คนที่ 10	0.91	0.95	0.86	0.83	0.83	0.88
คนที่ 11	0.99	0.97	0.97	0.86	0.87	0.89

หมายเหตุ RAI1 = RAI ก่อนการฝึกอบรม, RAI2 = RAI หลังการฝึกอบรมทันที, RAI3 = RAI หลังการฝึกอบรม 3 เดือน ICC1 = ICC ก่อนการฝึกอบรม, ICC2 = ICC หลังการฝึกอบรมทันที, ICC3 = ICC หลังการฝึกอบรม 3 เดือน

2.7 ผลการวัดความพึงพอใจของผู้เข้ารับการฝึกอบรมหลังสิ้นสุดการฝึกอบรมทันที ร่วมกับการสัมภาษณ์เฉพาะกลุ่ม

2.7.1 ผลการวัดความพึงพอใจของผู้เข้ารับการฝึกอบรมหลังสิ้นสุดการฝึกอบรมทันที

ผลการวัดความพึงพอใจของผู้เข้ารับการฝึกอบรมหลังสิ้นสุดการฝึกอบรมทันที ผู้เข้ารับการฝึกอบรมมีความพึงพอใจเรียงลำดับจากมากที่สุด ถึงปานกลางได้ดังนี้ ประโยชน์ของการฝึกอบรมต่อองค์กรของท่าน (M = 4.82, SD = 0.40) ปริมาณ และคุณภาพของอาหาร และอาหารว่าง (M = 4.82, SD = 0.40)

ประโยชน์ของการฝึกอบรมต่อตัวท่าน (M = 4.73, SD = 0.47) ความคุ้มค่ากับการเข้ารับการฝึกอบรม (M = 4.55, SD = 0.52) และความรู้ ความสามารถของผู้ดำเนินการฝึกอบรม และวิทยากร (M = 4.55, SD = 0.52) มีความพึงพอใจในระดับมากในประเด็นของ ความน่าสนใจของกิจกรรมการฝึกอบรม (M = 4.45,) ความชัดเจนเข้าใจง่ายของเนื้อหาในการฝึกอบรม (M = 4.18, SD = 0.60) สถานที่ มีความสะดวก สบาย และเหมาะสมสำหรับการใช้ในการฝึกอบรมครั้งนี้ (M = 4.09, SD = 0.54) ความพร้อมของสื่อโสตทัศนูปกรณ์ (M = 3.91, SD = 0.83) และมีความพึงพอใจในระดับปานกลางต่อประเด็นระยะเวลาการฝึกอบรมมีความเหมาะสม (M= 3.18, SD = 1.25) ดังแสดงในตารางที่ 37

ตารางที่ 37 ผลการวัดความพึงพอใจของผู้เข้ารับการฝึกอบรมกลุ่มใช้โมดูลการฝึกอบรมเพื่อเสริมสร้าง สมรรถนะการประเมินสำหรับอาจารย์พยาบาล

หัวข้อประเมิน	ความพึงพอใจ		
	Mean	SD	ระดับความพึงพอใจ
1. ประโยชน์ของการฝึกอบรมต่อองค์กรของท่าน	4.82	0.40	มากที่สุด
2. ปริมาณ และคุณภาพของอาหาร และอาหารว่าง	4.82	0.40	มากที่สุด
3. ประโยชน์ของการฝึกอบรมต่อตัวท่าน	4.73	0.47	มากที่สุด
4. ความคุ้มค่ากับการเข้ารับการฝึกอบรม	4.55	0.52	มากที่สุด
5. ความรู้ ความสามารถของผู้ดำเนินการฝึกอบรม และวิทยากร	4.55	0.52	มากที่สุด
6. ความน่าสนใจของกิจกรรมการฝึกอบรม	4.45	0.82	มาก
7. ความชัดเจนเข้าใจง่ายของเนื้อหาในการฝึกอบรม	4.18	0.60	มาก
8. สถานที่ มีความสะดวก สบายและเหมาะสมสำหรับการฝึกอบรมครั้งนี้	4.09	0.54	มาก
9. ความพร้อมของสื่อโสตทัศนูปกรณ์	3.91	0.83	มาก
10. ระยะเวลาการฝึกอบรมมีความเหมาะสม	3.18	1.25	ปานกลาง

n = 11

2.7.2 การสัมภาษณ์เฉพาะกลุ่มหลังสิ้นสุดการฝึกอบรมทันที

หลังสิ้นสุดการฝึกอบรมทันที มีการสัมภาษณ์เฉพาะกลุ่มกับตัวอย่างที่สะดวกที่จะสนทนาต่อตามความสมัครใจ ในการสนทนาครั้งนี้ตัวอย่างอยู่ร่วมสนทนากลุ่ม 5 คนจากจำนวน 11 คน ผลการสัมภาษณ์ ใช้แนวคำถามมีโครงสร้างสำหรับการสัมภาษณ์เฉพาะกลุ่ม เพื่อให้ได้ข้อมูลเพิ่มเติมจากข้อมูลเชิงปริมาณจากแบบสอบถามในด้านความรู้ ทักษะคิด และทักษะที่ได้รับหลังการฝึกอบรม

และความพึงพอใจต่อการฝึกอบรม ใช้เวลาในการสนทนาประมาณ 30 นาที สนทนา ณ วันที่ 29 กรกฎาคม 2568 โดยใช้แนวคำถามหลัก คือ 1) การฝึกอบรมทำให้เกิดความรู้เกี่ยวกับการประเมินภาคปฏิบัติอย่างไรบ้าง 2) การฝึกอบรมทำให้ท่านมีทักษะในการประเมินภาคปฏิบัติมากขึ้นหรือไม่อย่างไร 3) ท่านรู้สึกอย่างไรกับการเข้าไปเกี่ยวข้องกับการประเมินภาคปฏิบัติในนักศึกษาพยาบาล และ 4) จุดเด่นของการฝึกอบรม และจุดที่ควรปรับปรุงในการฝึกอบรมคืออะไร ควรดำเนินการอย่างไรต่อไป ผลการวิเคราะห์เนื้อหา มีประเด็นสำคัญ 4 ประเด็นคือ 1) ความรู้ใหม่ที่ได้รับหลังการฝึกอบรม 2) ประโยชน์ของการฝึกทักษะการประเมินที่อาจต้องนำไปปรับใช้กับสถานการณ์จริง 3) ทักษะที่ดีต่อการประเมินภาคปฏิบัติ 4) สิ่งที่ควรปรับปรุงในการฝึกอบรม ดังรายละเอียดต่อไปนี้

2.7.2.1 ความรู้ใหม่ที่ได้รับหลังการฝึกอบรม ผู้เข้ารับการอบรม กล่าวถึงเนื้อหาการฝึกอบรม ว่าเป็นเรื่องใหม่ ซึ่งเป็นเรื่องเกี่ยวกับการวัดและประเมินผลการเรียนรู้เดิมซึ่ง อาจารย์แต่ละท่านมีพื้นฐานความรู้อยู่แล้ว แต่มีเนื้อหาเฉพาะด้านที่ต้องทำความเข้าใจมากขึ้น ดังตัวอย่างของบทสนทนาต่อไปนี้

“เป็นงานที่ดีมีคุณค่ามาก เนื้อหาที่ความลึกกว่าที่เคยรู้มา”

ผู้ให้ข้อมูลคนที่ 1

“ได้ความรู้ใหม่ เยอะขึ้นมากเลยคะ”

ผู้ให้ข้อมูลคนที่ 2

“เหมือนเป็นเรื่องพื้นๆ ที่เรามองข้ามเพราะไม่คิดว่าเรื่องที่เราทำตอนให้คะแนนเด็กจะเป็นความลำเอียงโดยไม่รู้ตัว ”

ผู้ให้ข้อมูลคนที่ 3

“เป็นอาจารย์มาหลายปีมีบางเรื่องที่เราที่ยังไม่รู้ไม่ลึกพอ พอเข้าอบรมทำให้รู้ขึ้นมาก”

ผู้ให้ข้อมูลคนที่ 4

“พอเคยได้ยินมาบ้าง การเข้าอบรมก็ทำให้รู้ได้ละเอียดขึ้น ”

ผู้ให้ข้อมูลคนที่ 5

2.7.2.2 ประโยชน์ของการฝึกทักษะการประเมินที่สามารถนำไปปรับใช้กับสถานการณ์จริง ผู้ให้ข้อมูลส่วนใหญ่กล่าวถึงการฝึกทักษะที่ใช้ในการฝึกอบรมว่าเป็นทักษะที่คุ้นเคยอยู่แล้ว เพราะผู้เข้ารับการฝึกอบรมทุกคนมีประสบการณ์การเป็นพยาบาลประจำการบนหอผู้ป่วยมาก่อน แม้บางคนมีประสบการณ์การเป็นอาจารย์มาได้ไม่นาน จึงทำให้เมื่อฝึกทักษะการประเมินกับสื่อวีดิทัศน์สถานการณ์จำลอง สามารถประเมินผลได้สอดคล้องกับคะแนนมาตรฐานของวีดิทัศน์ชุดนั้น ๆ ประกอบกับรูบริกมีความชัดเจนสอดคล้องกับพฤติกรรมที่ต้องการประเมิน จึงทำให้บางคนไม่แน่ใจว่าจะนำไปใช้ในสถานการณ์จริงจะทำได้มากน้อยเพียงใด เนื่องจากอาจไม่เจอสถานการณ์ที่ต้องการประเมิน บางส่วนมีความคิดที่จะลองนำไปใช้จริง ตัวอย่างบทสนทนาบางส่วนของแต่ละประเด็น

“รูบริคค่อนข้างชัดเจน สอดรับกับคลิปวีดิโอทำให้ให้คะแนนได้ไม่ยาก ”

ผู้ให้ข้อมูลคนที่ 1

“ ให้คะแนนไม่ยากเพราะรูบริคชัด พฤติกรรมเด็กในวีดิโอก็ชัด เดี่ยวจะลองเอาไปใช้จริงดู ”

ผู้ให้ข้อมูลคนที่ 2

“ ตอนดูคลิปวีดิโอ ก็ดูง่าย เห็นพฤติกรรมชัด ครบทุกพฤติกรรมที่ต้องการ รูบริคก็ชัด แต่อย่าลืมว่าเวลาขึ้นฝึกจริงเราต้องดูนักเรียนตั้งหลายคน แล้วถ้าเราบังเอิญไม่เจอพฤติกรรมที่เราต้องการประเมินละ เราจะให้คะแนนเด็กอย่างไร ”

ผู้ให้ข้อมูลคนที่ 5

2.7.2.3 ทักษะที่ดีต่อการประเมินภาคปฏิบัติ ผู้ให้ข้อมูลกล่าวเป็นเอกฉันท์ว่าการเป็นอาจารย์พยาบาลจะต้องมีทัศนคติที่ดีต่อการฝึกภาคปฏิบัติเป็นจุดเริ่มต้น หากขาดคุณสมบัตินี้ จะขาดคุณสมบัติสำคัญของการเป็นอาจารย์พยาบาลที่ดี เนื่องจากการฝึกภาคปฏิบัติเป็นหัวใจสำคัญของการผลิตบุคลากรด้านการพยาบาล โดยมีข้อมูลจากการสัมภาษณ์เฉพาะกลุ่มดังนี้

“การนิเทศเด็กฝึกงาน เป็นงานที่เลี้ยงไม่ได้สำหรับอาจารย์พยาบาลอยู่แล้ว ถ้าไม่ชอบก็ไม่ควรมาเป็นอาจารย์พยาบาล”

ผู้ให้ข้อมูลคนที่ 1

“ เหมือนที่ที่...พูด มันเสี่ยงไม่ได้เพราะฉะนั้นถ้าไม่รักการประเมินภาคปฏิบัติก็เป็นครูพยาบาลที่ดีไม่ได้ ”

ผู้ให้ข้อมูลคนที่ 2

2.7.2.4 สิ่งที่ต้องปรับปรุงในการฝึกอบรม

ผู้ให้ข้อมูลทุกคนเห็นตรงกันว่าเนื้อหาการฝึกอบรม และกิจกรรมในการฝึกมีประโยชน์มากกับการเสริมสร้างทักษะการประเมิน แต่มีสิ่งที่จะต้องปรับปรุง คือ ระยะเวลาในการฝึกประเมินด้วยวิธีทัศนิกนานเกินไป ดังตัวอย่างบทสนทนาต่อไปนี้

“วิธีโอยยาวไปนะคะ ถ้าสั้นกว่านี้จะดีมาก”

ผู้ให้ข้อมูลคนที่ 1

“วิธีโอยยาวไป ต้องบอกเลยดูจนล้า”

ผู้ให้ข้อมูลคนที่ 2

2.7.3 ผลการสัมภาษณ์หลังสิ้นสุดการใช้โมดูลการฝึกอบรมอบรม 3 เดือน

ผลการสัมภาษณ์เชิงลึกโดยใช้แบบสัมภาษณ์แบบมีโครงสร้าง (Structured interview) เพื่อประเมินความยั่งยืนในประเด็นเกี่ยวกับความรู้ ทักษะ ทัศนคติฯ และการสร้างความยั่งยืนให้สมรรถนะการประเมินภาคปฏิบัติในคณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต โดยสุ่มสัมภาษณ์ตัวอย่างที่สะดวกที่จะเข้ารับการสัมภาษณ์ ได้จำนวน 6 คน จาก 11 คน และการสัมภาษณ์รองคณบดีฝ่ายวิชาการ โดยใช้แนวคำถาม ดังนี้ หลังการฝึกอบรม 3 เดือน ท่านคิดว่าตนเองได้อะไรจากการฝึกอบรม และได้นำผลการฝึกอบรมไปใช้อย่างไร และใช้คำถามว่าจากผลการวิจัยที่สรุปได้ท่านมีแนวคิดจะนำผลที่ได้ไปใช้ประโยชน์อย่างไร ผลการสัมภาษณ์ตัวอย่าง สรุปประเด็นสำคัญได้ 2 ประเด็นคือ 1) ประโยชน์ในระดับบุคคล มีการนำความรู้ไปทดลองใช้จริง และ 3) ประโยชน์ในระดับองค์กร และผลการสัมภาษณ์รองคณบดีฝ่ายวิชาการ สรุปประเด็นสำคัญได้ว่าควรนำผลการวิจัยมาเผยแพร่ จึงค่อยตัดสินใจที่จะวางแผนดำเนินการระดับนโยบายต่อไป

ผลการสนทนาสัมภาษณ์ตัวอย่างจำนวน 6 คน

2.7.3.1 ด้านประโยชน์ในระดับบุคคล ผู้ให้ข้อมูลทุกคนกล่าวถึงประโยชน์ระดับบุคคล ว่าทำให้มีความรู้ใหม่ๆ เพิ่มขึ้น ทำให้เกิดความมั่นใจในการประเมินผลการฝึกของนักศึกษา อย่างมีหลักการที่ถูกต้อง หรือทำให้เกิดความตื่นตัวที่จะทราบสมรรถนะการประเมินของตนเอง ตัวอย่างบางคนแจ้งว่าทดลองนำรูบริค และหลักการของการป้องกันการเกิดความคลาดเคลื่อนจากผู้ประเมินไปใช้ในการประเมินนักศึกษาที่ขึ้นฝึกปฏิบัติงานจริง และรู้สึกมั่นใจในหลักการที่ถูกต้องในการให้คะแนนนักศึกษา ดังตัวอย่างบทสนทนาต่อไปนี้

“เป็นความรู้ที่ค่อนข้างใหม่สำหรับพี่ เป็นความรู้ที่ดีที่ครูควรรู้เวลาขึ้นไปฝึก พี่ก็รู้สึก ว่าความรู้ที่ได้รับมันลึกซึ้งมากกว่าที่เคยรู้”

ผู้ให้ข้อมูลคนที่ 1

“พี่เอารูบริคไปลองใช้งานจริง คู่กับรูบริคเดิมที่ใช้อยู่ พี่ว่าwork นะ โอเคเลย”

ผู้ให้ข้อมูลคนที่ 2

“พี่เอาหลักการ ที่ได้กับรูบริคไปลองใช้งานจริงก็ใช้ได้ระดับหนึ่งนะ แต่อาจต้องปรับรายละเอียด ของแต่ละระดับคะแนนให้ชัดเจนตามบริบทของแต่ละเวอร์ต...รูบริคทำให้ตัดสินใจได้ง่ายขึ้นเวลาให้คะแนน”

ผู้ให้ข้อมูลคนที่ 3

“พี่เคยอบรมเรื่องนี้กับ มศว. นานมาแล้ว แต่มันก็ยังอยู่ในหัวเรื่อง *rater error* นี้แหละ แต่พี่ อยากรู้ว่าวิเคราะห์ลักษณะของผู้ประเมินแต่ละคนว่าเป็นอย่างไรได้ไหม พี่อยากรู้สมรรถนะตัวเองว่าประเมินคงที่ หรือเปล่า”

ผู้ให้ข้อมูลคนที่ 4

“ยังรู้สึกดีมาก ๆ กับความรู้ที่ได้ อย่างที่บอกว่าเป็นครูมาหลายปีก็จริง แต่บางเรื่องก็ยังไม่รู้สึก พอรู้ ก็มั่นใจขึ้นในการประเมินเด็ก”

ผู้ให้ข้อมูลคนที่ 5

“เป็นความรู้ที่ดีค่ะ ถ้าเอาไปใช้ได้จริงคงมีประโยชน์กว่านี้มาก แต่ยังไม่ไดลองใช้ค่ะ”

ผู้ให้ข้อมูลคนที่ 6

2.7.3.2 ด้านประโยชน์ระดับองค์กร ผู้ให้ข้อมูลส่วนใหญ่กล่าวถึงการขยายผลการใช้ไปในระดับองค์กร เช่น ควรมีการกำหนดเชิงนโยบายของผู้บริหาร ควรมีการนำเสนอรูปคดีต่อผู้บริหารเพื่อนำมาทดลองใช้ หรือมีการสำรวจความต้องการในการใช้รูปคดีการประเมินภาคปฏิบัติเพื่อปรับแก้ให้เหมาะสมกับการใช้งานมากที่สุด บางท่านเสนอให้มีการบรรจุเนื้อหาของการฝึกอบรมนี้แก่อาจารย์ที่เข้ามาปฏิบัติงานใหม่ทุกครั้ง และรองคณบดีฝ่ายวิชาการ ผู้บริหารให้ความคิดเห็นว่าควรนำผลการวิจัยนำเสนอต่อผู้บริหาร เพื่อจะได้นำมาพิจารณาบรรจุลงในวาระของแผนการจัดการความรู้ของคณะพยาบาลศาสตร์ต่อไป ดังตัวอย่างบทสนทนาต่อไปนี้

“น่าจะลองนำผลการวิจัยไปเสนอผู้บริหาร จะได้ทำอะไรออกมาให้เห็นเป็นรูปเป็นร่าง”

ผู้ให้ข้อมูลคนที่ 1

“ถ้าจะให้ได้ประโยชน์จากการฝึกอบรมจริงๆ นะ ต้องมีนโยบายให้ปฏิบัติเป็นเรื่องเป็นราว มันจะได้เกิดความต่อเนื่อง ”

ผู้ให้ข้อมูลคนที่ 2

“อาจารย์ใหม่ หรือระบบนิเทศอาจารย์ใหม่น่าจะนำการฝึกนี้ไปใช้ดู น่าจะมีประโยชน์มากขึ้น”

ผู้ให้ข้อมูลคนที่ 3

“ถ้าจะเอามาใช้จริงๆ พี่ว่าต้องเอามาปรับใช้แต่ละสาขา เอามาคุยกันปรับจากแนวปฏิบัติเดิม”

ผู้ให้ข้อมูลคนที่ 4

“ถ้าจะให้เห็นประโยชน์จริงๆ อาจารย์ใหม่ๆ น่าจะได้เตรียมความพร้อมแบบนี้”

ผู้ให้ข้อมูลคนที่ 5

“สมมติหนูเป็นอาจารย์ใหม่ถ้าใช้การฝึกอบรมแบบนี้ จะช่วยได้มาก...เอาเข้าระบบการปฐมนิเทศอาจารย์ใหม่ไปได้เลย”

ผู้ให้ข้อมูลคนที่ 6

“ควรนำผลการวิจัยมานำเสนอ เพื่อจะได้พิจารณาบรรจุเข้าในกิจกรรม KM ของคณะ แล้วค่อยมาดูกันว่า จะนำผลการวิจัยมาใช้อย่างไร แต่ที่ที่คาดหวังคือการนำรูปคดีที่เราช่วยกันคิดมาใช้ ”

รองคณบดีฝ่ายวิชาการ

2.8 ผลการศึกษาความยั่งยืนของการใช้โมดูลการฝึกอบรมต่อ ความรู้ ทักษะคิต และทักษะ การประเมินภาคปฏิบัติ พิจารณาจากผลการเปรียบเทียบการเปลี่ยนแปลงของความรู้ ทักษะคิต และ ทักษะหลังการใช้โมดูลการฝึกอบรมทันที และหลังการใช้โมดูล 3 เดือน ได้ผลการศึกษาดังนี้

2.8.1 ผลการศึกษาความยั่งยืนของการใช้โมดูลการฝึกอบรมต่อ ความรู้เกี่ยวกับการ ประเมินภาคปฏิบัติ พบว่าเมื่อเปรียบเทียบผลของความรู้ ไม่พบความแตกต่างที่ระดับนัยสำคัญ .05 ในการวัด หลังการใช้โมดูลการฝึกอบรมทันที และหลังการใช้โมดูล 3 เดือน แสดงว่ามีความยั่งยืนของ ความรู้

2.8.2 ผลการศึกษาความยั่งยืนของการใช้โมดูลการฝึกอบรมต่อ ทักษะคิตเกี่ยวกับการ ประเมินภาคปฏิบัติ ไม่พบการเปลี่ยนแปลงของทักษะคิตในการวัดทั้ง 3 ครั้ง แสดงว่าหลังการใช้โมดูล การฝึกอบรมไม่มีผลต่อการเปลี่ยนแปลงทักษะคิต

2.8.3 ผลการศึกษาความยั่งยืนของการใช้โมดูลการฝึกอบรมต่อ ทักษะการประเมิน ภาคปฏิบัติ ไม่พบการเปลี่ยนแปลงของทักษะการประเมินในการวัดทั้ง 3 ครั้ง ในผู้ประเมินส่วนใหญ่ แสดงว่าการใช้โมดูลไม่มีผลต่อการเปลี่ยนแปลงทักษะที่ระดับนัยสำคัญ .05 ($t = 0.08 - 1.82$) ยกเว้น ผู้ประเมินคนที่ 3 คนที่ 4 และ คนที่ 11 โดยพบว่าทักษะการประเมินลดลง ($t = - 2.28 - -3.41$) รายละเอียดดังนำเสนอในตอนที่ 2 หัวข้อ 2.4

2.8.4 ผลการสัมภาษณ์เชิงลึกในตัวอย่างที่ใช้โมดูล และรองคณบดีฝ่ายวิชาการของคณะ พยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต ในประเด็นของความยั่งยืนตัวอย่างมีความคิดเห็นว่า หาก ต้องการให้เกิดความยั่งยืนของการประเมิน ควรมีการนำเสนอผู้บริหารของคณะฯ เพื่อพิจารณาบรรจุ โมดูลการฝึกอบรมเข้าเป็นหลักสูตรเพื่อการเตรียมความพร้อมอาจารย์ผู้สอนภาคปฏิบัติโดยเฉพาะการ เตรียมความพร้อมอาจารย์ใหม่ และรองคณบดีฝ่ายวิชาการได้เสนอให้นำเสนอผลการวิจัยเพื่อ พิจารณาใช้ผลการวิจัย เป็นกิจกรรมหนึ่งในโครงการบริหารความรู้ของคณะฯ เพื่อพัฒนาคุณภาพการ เรียนการสอนภาคปฏิบัติ และคาดหวังที่จะนำรูปคดีที่ได้จากการวิจัยไปปรับใช้กับการประเมิน ภาคปฏิบัตินักศึกษาพยาบาลต่อไป

ผลจากการพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล โดยผ่านการทดลองใช้ และการใช้งานจริง นำผลการใช้มาปรับปรุงเพื่อให้ได้โมดูลมาตรฐานที่ ประกอบด้วย การเสริมสร้างสมรรถนะการประเมินด้านความรู้ โดยการฝึกอบรม ร่วมกับการให้คำปรึกษา โดยการบรรยาย พร้อมกับการศึกษาจากคู่มือการเสริมสร้างสมรรถนะการประเมิน และให้ผู้เข้ารับการอบรมทำกิจกรรมท้ายบทของแต่ละบท ซึ่งมีสาระสำคัญประกอบด้วย (1) ความรู้พื้นฐานเกี่ยวกับการประเมินภาคปฏิบัติ (2) การฝึกอบรมเกี่ยวกับความคลาดเคลื่อนจากผู้ประเมิน (3) การฝึกอบรมกรอบของการอ้างอิง (4) การฝึกทักษะการสังเกต และ (5) การฝึกทักษะการตัดสินใจ

ตอนที่ 3 ผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล

ต่อความคลาดเคลื่อนจากผู้ประเมิน

ผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลต่อความคลาดเคลื่อนจากผู้ประเมิน คือ การกด ปล่อยคะแนน (Severity/ Leniency effect) การให้คะแนนเกาะกลุ่มตรงกลาง (Central tendency effect) และความคลาดเคลื่อนแบบฮาโล (Halo effect) จากการประเมินงานภาคปฏิบัติทั้งสิ้น 10 ชุด ในผู้ประเมิน 11 คน โดยตรวจสอบจากค่าพารามิเตอร์ของ 3 ฟาเซทคือ ฟาเซทผู้ประเมิน (rater) ใช้ตรวจสอบความคลาดเคลื่อนแบบกด ปล่อยคะแนน ฟาเซทผู้ถูกประเมิน (ratee) ใช้ตรวจสอบความคลาดเคลื่อนแบบให้คะแนนเกาะกลุ่มตรงกลาง และฟาเซทสมรรถนะ (trait) ใช้ตรวจสอบความคลาดเคลื่อนแบบฮาโล ผลการศึกษานำเสนอเป็น 4 ประเด็น คือ (1) ผลของโมดูลการฝึกอบรม ต่อความคลาดเคลื่อนแบบกด ปล่อยคะแนน (2) ผลของโมดูลการฝึกอบรม ต่อความคลาดเคลื่อนแบบให้คะแนนเกาะกลุ่มตรงกลาง (3) ผลของโมดูลการฝึกอบรม ต่อความคลาดเคลื่อนแบบฮาโล และ (4) ภาพรวมของผลของโมดูลการฝึกอบรมต่อความคลาดเคลื่อนทั้งหมด โดยประเด็นที่ 1 ถึง 3 เป็นการใช่วิธีระดับกลุ่ม และประเด็นที่ 4 เป็นการใช่วิธีระดับบุคคล ดังรายละเอียดต่อไปนี้

1. ผลของโมดูลการฝึกอบรบฯ ต่อความคลาดเคลื่อนแบบกต ปล่อยคะแนน

แบ่งการศึกษาเป็น 2 ขั้นตอนคือ (1) ศึกษาความแตกต่างของระดับการกต ปล่อยคะแนนในการวัดก่อนการฝึกอบรบฯ (ครั้งที่ 1) และหลังการฝึกอบรบฯ ทันที (ครั้งที่ 2) และ (2) ศึกษาสถิติในระดับกลุ่มที่ช่วยในการระบุความคลาดเคลื่อนแบบกต ปล่อยคะแนนในการวัดแต่ละครั้ง

1.1 การศึกษาความแตกต่างของระดับการกต ปล่อยคะแนนในการวัด 2 ครั้ง ศึกษาได้จากผลต่างของค่าที่สังเกตได้กับค่าที่คาดหวังด้วยโมเดล ($Obs - Exp. Average$) มีหน่วยการวัดเป็น โลจิท ($target\ measurement = observed\ measurement + Bias\ size$) และ $Obs - Exp. Average$ หากค่าเป็นบวกแสดงว่ามีการปล่อยคะแนนมากกว่าที่โมเดลคาดการณ์ ค่าลบแสดงว่ามีการกตคะแนนมากกว่าที่โมเดลคาดการณ์ ค่าเป็นศูนย์แสดงว่าค่าที่ได้ตรงกับที่โมเดลคาดการณ์ และค่าสถิติ $fit\ mean\ square$ ค่าที่คาดหวังคือค่าที่เข้าใกล้ 1 ใช้เพื่อทดสอบความสอดคล้องของผลการประเมิน และผลการประเมินที่โมเดลคาดการณ์ ค่า t เป็นการทดสอบสมมติฐานของ $target\ contrast/Joint\ S.E.$ ที่ว่าไม่พบความแตกต่างของระดับการกตปล่อยคะแนนในการวัดทั้ง 2 ครั้ง ผลการศึกษาไม่พบว่า ผู้ประเมินรายใดมีระดับการกต ปล่อยคะแนนแตกต่างกันระหว่างการวัดครั้งที่ 1 และการวัดครั้งที่ 2 ที่ระดับนัยสำคัญ .05 ดังแสดงในตารางต่อไปนี้

ผู้ประเมินคนที่ 1

วัดครั้งที่ 1 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล ($Obs - Exp. Average$) = 0.09 มีหน่วยการวัดเป็นโลจิท ($Target\ measure$) = 0.22

ครั้งที่ 2 มีการกตคะแนนมากกว่าที่คาดการณ์ฯ - 0.09 (0.07 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (0.22- 0.07) 0.15 โลจิท แต่ไม่พบความแตกต่างที่ระดับนัยสำคัญ .05 ($t = 0.99, p = .32$)

ผู้ประเมินคนที่ 2

วัดครั้งที่ 1 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล ($Obs - Exp. Average$) = 0.02 มีหน่วยการวัดเป็นโลจิท ($Target\ measure$) = 0.21 โลจิท

ครั้งที่ 2 มีการกตคะแนนมากกว่าที่คาดการณ์ฯ - 0.02 (0.17 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (0.21- 0.17) 0.04 โลจิท ไม่พบความแตกต่างที่ระดับนัยสำคัญ .05 ($t = 0.23, p = .82$)

ผู้ประเมินคนที่ 3

วัดครั้งที่ 1 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.11
 มีหน่วยการวัดเป็นโลจิท (Target measure) = - 0.23 โลจิท
 ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ - 0.11 (-0.42 โลจิท)
 ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (- 0.23 - - 0.42) 0.18 โลจิท แต่ไม่พบความแตกต่างที่ระดับนัยสำคัญ .05 ($t = 1.17, p = .24$)

ผู้ประเมินคนที่ 4

วัดครั้งที่ 1 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.04
 มีหน่วยการวัดเป็นโลจิท (Target measure) = 0.18 โลจิท
 ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ - 0.04 (0.12 โลจิท)
 ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (0.18 - 0.12) 0.06 โลจิท แต่ไม่พบความแตกต่างที่ระดับนัยสำคัญ .05 ($t = 0.38, p = .70$)

ผู้ประเมินคนที่ 5

วัดครั้งที่ 1 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.01
 มีหน่วยการวัดเป็นโลจิท (Target measure) = 0.22 โลจิท
 ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ - 0.01 (0.21 โลจิท)
 ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (0.22 - 0.21) 0.01 โลจิท ไม่พบความแตกต่างที่ระดับนัยสำคัญ .05 ($t = .08, p = .94$)

ผู้ประเมินคนที่ 6

วัดครั้งที่ 1 มีการกดคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.02
 มีหน่วยการวัดเป็นโลจิท (Target measure) = - 0.05 โลจิท
 ครั้งที่ 2 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ฯ 0.02 (- 0.01 โลจิท)
 ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (- 0.05 - - 0.01) - 0.03 โลจิท แต่ไม่พบความแตกต่างกันที่ระดับนัยสำคัญ .05 ($t = - 0.23, p = .82$)

ผู้ประเมินคนที่ 7

วัดครั้งที่ 1 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.01
 มีหน่วยการวัดเป็นโลจิท (Target measure) = 0.01 โลจิท

ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ - 0.01 (0.00 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (0.01 - 0.00) 0.01 โลจิท ไม่พบความแตกต่างที่ระดับนัยสำคัญ .05 ($t = 0.08, p = .94$)

ผู้ประเมินคนที่ 8

วัดครั้งที่ 1 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.01 มีหน่วยการวัดเป็นโลจิท (Target measure) = 0.11 โลจิท

ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ - 0.01 (0.09 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (0.11 - 0.09) 0.02 โลจิท ไม่พบความแตกต่างที่ระดับนัยสำคัญ .05 ($t = 0.15, p = .88$)

ผู้ประเมินคนที่ 9

วัดครั้งที่ 1 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.17 มีหน่วยการวัดเป็นโลจิท (Target measure) = 0.24 โลจิท

ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ - 0.17 (- 0.04 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (0.24 - - 0.04) 0.28 โลจิท ไม่พบความแตกต่างที่ระดับนัยสำคัญ .05 ($t = 1.82, p = .07$)

ผู้ประเมินคนที่ 10

วัดครั้งที่ 1 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.11 มีหน่วยการวัดเป็นโลจิท (Target measure) = 0.46 โลจิท

ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ - 0.11 (0.27 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (0.46 - 0.27) 0.19 โลจิท ไม่พบความแตกต่างที่ระดับนัยสำคัญ .05 ($t = 1.20, p = .23$)

ผู้ประเมินคนที่ 11

วัดครั้งที่ 1 มีการปล่อยคะแนนมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.10 มีหน่วยการวัดเป็นโลจิท (Target measure) = 0.14 โลจิท

ครั้งที่ 2 มีการกดคะแนนมากกว่าที่คาดการณ์ฯ - 0.17 (- 0.03 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (0.14 - - 0.03) 0.16 โลจิท ไม่พบความแตกต่างที่ระดับนัยสำคัญ .05 ($t = 1.06, p = .29$)

ในภาพรวมไม่พบความแตกต่างของผลการวัดในครั้งที่ 1 และ ครั้งที่ 2 ที่ระดับนัยสำคัญ .05 จากค่า Infit mean square ผลการประเมินของผู้ประเมินทุกค่าอยู่ในค่าที่คาดหวังโดยโมเดล ดังแสดงในตารางต่อไปนี้

ตารางที่ 38 เปรียบเทียบผลการประเมินก่อน และหลังการฝึกอบรม ของฟาเซทผู้ประเมิน

ผู้ประเมิน	ครั้งที่วัด	Target measure	Obs- Exp. Average	Target contrast	Joint S.E.	Infit MnSq	Outfit MnSq	t	p
คนที่1	ครั้งที่ 1	0.22	0.09	0.15	0.15	1.13	1.28	0.99	.32
	ครั้งที่ 2	0.07	-0.09			1.06	1.03		
คนที่2	ครั้งที่ 1	0.21	0.02	0.04	0.15	1.19	1.17	0.23	.82
	ครั้งที่ 2	0.17	-0.02			1.28	1.29		
คนที่3	ครั้งที่ 1	-0.23	0.11	0.18	0.16	0.64	0.64	1.17	.24
	และ 2	-0.42	-0.11			0.64	0.64		
คนที่4	ครั้งที่ 1	0.18	0.04	0.06	0.15	1.11	1.08	0.38	.70
	ครั้งที่ 2	0.12	-0.04			0.94	0.93		
คนที่5	ครั้งที่ 1	0.22	0.01	0.01	0.15	0.77	0.75	.08	.94
	ครั้งที่ 2	0.21	-0.01			0.82	0.82		
คนที่6	ครั้งที่ 1	-0.05	-0.02	-0.03	0.15	0.83	0.83	-0.23	.82
	ครั้งที่ 2	-0.01	0.02			0.92	0.91		
คนที่7	ครั้งที่ 1	0.01	0.01	0.89	0.15	1.07	1.06	.08	.94
	ครั้งที่ 2	0.00	-0.01			1.28	1.27		
คนที่8	ครั้งที่ 1	0.11	0.01	0.02	0.15	1.13	1.09	0.15	.88
	ครั้งที่ 2	0.09	-0.01			1.16	1.15		
คนที่9	ครั้งที่ 1	0.24	0.17	0.28	0.15	0.89	0.88	1.82	.07
	ครั้งที่ 2	-0.04	-0.17			0.79	0.79		
คนที่10	ครั้งที่ 1	0.46	0.11	0.19	0.16	1.10	1.08	1.20	.23
	ครั้งที่ 2	0.27	-0.11			1.33	1.27		
คนที่11	ครั้งที่ 1	0.14	0.10	0.16	0.15	0.99	0.98	1.06	.29
	ครั้งที่ 2	-0.03	-0.17			0.82	0.80		

$p < .05$

1.2 ศึกษาสถิติในระดับกลุ่มที่ช่วยในการระบุความคลาดเคลื่อนแบบกต ปล่อยคะแนนในการวัดแต่ละครั้ง

โดยศึกษาจากพารามิเตอร์ของฟาเซทผู้ประเมิน (rater) พิจารณาจากการทดสอบด้วย Fixed Chi-Square มีสมมติฐานว่างว่า ผู้ประเมินทุกคนมีระดับการกต ปล่อยคะแนนเหมือนกัน (The null hypothesis that all raters were equally severe (or equally lenient)) หากปฏิเสธสมมติฐานว่างแสดงว่ามีผู้ประเมินอย่างน้อย 2 คนให้คะแนนการประเมินต่างกัน อย่างไรก็ตามสถิตินี้มีความไวต่อขนาดของกลุ่มตัวอย่างมาก และเป็นการประมาณการคร่าวๆ ดังนั้นจะต้องพิจารณาร่วมกับสถิติอื่นร่วมด้วย เช่น separation ratio, separation of rater reliability, หรือ separation index

พิจารณาสถิติในระดับกลุ่มของการวัดครั้งที่ 1 กลุ่มผู้ประเมินให้คะแนนการประเมินในระดับการกต ปล่อยคะแนนแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ($\chi^2 = 26.83$, d.f. = 10) แสดงว่ามีผู้ประเมินอย่างน้อย 2 คนใน 11 คนให้คะแนนการกต ปล่อยคะแนนแตกต่างกันที่ระดับนัยสำคัญ.001 ค่าเฉลี่ย และช่วงของการกตปล่อยคะแนนคือ 0.14 และ 0.71 เป็นช่วงที่แคบ ความแปรปรวนของการประมาณค่าการกต ปล่อยคะแนนค่อนข้างน้อย (SD = 0.17) ค่า reliability of rater separation = 0.60 มีค่าค่อนข้างต่ำ (เกณฑ์การพิจารณาใช้เกณฑ์เดียวกับค่าความเที่ยง เช่น Alpha's Cronbach หรือ KR 20) (Linacre, 2014) แสดงว่าผู้ประเมินอยู่ในระดับการกต ปล่อยคะแนนไม่แตกต่างกันมาก ค่าที่พึงประสงค์คือค่าที่เข้าใกล้ศูนย์ separation ratio = 1.23 แสดงว่าการกต ปล่อยคะแนนของผู้ประเมิน เป็น 1 เท่าของความคลาดเคลื่อนจากการวัด ซึ่งมีการกระจายตัวของการให้คะแนนกต ปล่อยไม่แตกต่างกันมาก ผลของ Mean Infit และ Mean Outfit คือ 1.00 และ 0.98 มีผลการประเมินเป็นไปตามที่คาดการณ์ด้วยโมเดล แสดงว่าในระดับกลุ่มไม่พบความคลาดเคลื่อนแบบกต ปล่อยคะแนน

การวัดครั้งที่ 2 ผู้ประเมินอย่างน้อย 2 คนให้คะแนนการประเมินในระดับการกต ปล่อยคะแนนแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ($\chi^2 = 27.1$, d.f. = 10) แสดงว่ามีผู้ประเมินอย่างน้อย 2 คน ใน 11 คน มีระดับการกต ปล่อยคะแนนแตกต่างกันอย่างมีนัยสำคัญทางสถิติ เมื่อพิจารณา ค่าเฉลี่ย = 0.14 และช่วงของการกตปล่อยคะแนน = 0.69 ซึ่งมีช่วงของการกต ปล่อยคะแนนแคบ ความแปรปรวนของการประมาณค่าการกต ปล่อยคะแนนค่อนข้างน้อย และใกล้เคียงกับการวัดครั้งที่ 1 เมื่อพิจารณาสถิติ Separation ratio ซึ่งเป็นสัดส่วนการวัดการกระจาย

ของการกต ปล่อยคะแนนสัมพันธ์กับความแม่นยำของการวัด = 1.28 แสดงว่าการกต ปล่อยคะแนนของผู้ประเมิน เป็น 1 เท่าของความคลาดเคลื่อนจากการวัด ซึ่งมีการกระจายตัวของการให้คะแนนกต ปล่อยไม่แตกต่างกันมาก ค่า reliability of separation = 0.62 ซึ่งเป็นค่าค่อนข้างต่ำ แสดงถึงความแตกต่างของการกต ปล่อยคะแนนของผู้ประเมินไม่มาก ค่า Mean Infit square และ Mean Outfit square อยู่ในช่วง 1.00 และ 0.99 แสดงว่าพฤติกรรมกรรมการประเมินของผู้ประเมินมีค่าใกล้เคียงกับพฤติกรรมกรรมการประเมินที่โมเดลคาดการณ์ ในระดับกลุ่มจึงไม่พบความคลาดเคลื่อนแบบกต ปล่อยคะแนน รายละเอียดดังแสดงในตารางที่ 39

ตารางที่ 39 สถิติสรุปรวมในระดับกลุ่มสำหรับฟาเซทของผู้ประเมิน

ค่าสถิติ	ฟาเซทผู้ประเมิน	
	วัดครั้งที่ 1	วัดครั้งที่ 2
Mean	0.14	0.04
SD	0.17	0.18
N	11	11
Range	0.71	0.69
Infit		
Mean	1.00	1.00
SD	0.19	0.22
Outfit		
Mean	0.98	0.99
SD	0.18	0.22
Reliability of separation	.60	.62
Separation ratio	1.23	1.28
Chi-square	26.8	27.1
<i>p</i>	.00**	.00**

2. ผลของโมดูลการฝึกอบรบฯ ต่อความคลาดเคลื่อนแบบให้คะแนนเกาะกลุ่มตรงกลาง

แบ่งการศึกษาเป็น 2 ขั้นตอนคือ (1) ศึกษาความแตกต่างของระดับความสามารถของผู้ถูกประเมิน ในการวัดก่อนการฝึกอบรบฯ (ครั้งที่ 1) และหลังการฝึกอบรบฯ (ครั้งที่ 2) และ (2) ศึกษาสถิติในระดับกลุ่มที่ช่วยในการระบุความสามารถของผู้ถูกประเมินในการวัดแต่ละครั้ง

2.1 ผลการศึกษาโมดูลการฝึกอบรบฯ ต่อความคลาดเคลื่อนแบบให้คะแนนเกาะกลุ่มตรงกลาง จาก ฟาเซทผู้ถูกประเมิน (ratee) โดยการเปรียบเทียบรายคู่ของโลจิทความสามารถของผู้ถูกประเมิน จากการวัดครั้งที่ 1 และครั้งที่ 2 นำเสนอผลต่างโลจิทของค่าที่สังเกตได้กับค่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average), (target measurement = observed measurement + Bias size) หากค่า Obs- Exp. Average เป็นบวกแสดงว่ามีความสามารถน้อยกว่าที่โมเดลคาดการณ์ ค่าลบแสดงว่ามีความสามารถมากกว่าที่โมเดลคาดการณ์ (เนื่องจากผู้วิจัยเขียนคำสั่งการวิเคราะห์ในโปรแกรม FACETS ให้ฟาเซท ผู้ประเมินเป็นค่า positive ฟาเซทเดียว ดังนั้นฟาเซทอื่นจึงอ่านค่าเป็น negative คือค่ามากหมายถึงมีความสามารถน้อย ค่าน้อยหมายถึงมีความสามารถมาก) ค่าเป็นศูนย์แสดงว่าความสามารถตรงกับที่โมเดลคาดการณ์ ค่า t เป็นการทดสอบสมมติฐานของ target contrast/Joint S.E. ที่ว่าไม่พบความแตกต่างของความสามารถของผู้ถูกประเมินในการวัดครั้งที่ 1 และครั้งที่ 2 ผลการศึกษา ดังรายละเอียดต่อไปนี้

ผู้ถูกประเมินคนที่ 1

วัดครั้งที่ 1 ผู้ถูกประเมินมีความสามารถน้อยกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.07 มีหน่วยการวัดเป็นโลจิท (target measure) = 0.09 โลจิท

ครั้งที่ 2 มีความสามารถมากกว่าที่คาดการณ์ฯ - 0.07 (0.20 โลจิท)

ครั้งที่ 1 มีระดับความสามารถแตกต่างจากครั้งที่ 2 (target contrast) (0.09-0.20) = - 0.11 โลจิท แต่ไม่พบความแตกต่างมีระดับนัยสำคัญ .05 ($t = - 0.77, p = .44$)

ผู้ถูกประเมินคนที่ 2

วัดครั้งที่ 1 มีความสามารถน้อยกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.05 มีหน่วยการวัดเป็นโลจิท (Target measure) = - 0.38 โลจิท

ครั้งที่ 2 มีความสามารถมากกว่าที่คาดการณ์ฯ 0.05 (- 0.29 โลจิท)

ครั้งที่ 1 มีระดับความสามารถแตกต่างจากครั้งที่ 2 (target contrast) $(-0.38 - -0.29) = 0.09$ โลจิต แต่ไม่พบความแตกต่างที่ระดับนัยสำคัญ .05 ($t = -0.61, p = .54$)

ผู้ถูกประเมินคนที่ 3

วัดครั้งที่ 1 มีความสามารถน้อยกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.07 มีหน่วยการวัดเป็นโลจิต (Target measure) = 0.07 โลจิต

ครั้งที่ 2 มีความสามารถมากกว่าที่คาดการณ์ฯ 0.07 (0.18 โลจิต)

ครั้งที่ 1 ระดับความสามารถแตกต่างจากครั้งที่ 2 (target contrast) $(0.07-0.18) = -0.11$ โลจิต แต่ไม่พบความแตกต่างที่ระดับนัยสำคัญ .05 ($t = -0.77, p = .44$)

ผู้ถูกประเมินคนที่ 4

การวัดครั้งที่ 1 มีความสามารถน้อยกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.05 มีหน่วยการวัดเป็นโลจิต (Target measure) = -0.01 โลจิต

ครั้งที่ 2 มีความสามารถมากกว่าที่คาดการณ์ฯ 0.05 (0.07 โลจิต)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 $(-0.01-0.07)$ (target contrast) = -0.08 โลจิต แต่ไม่พบความแตกต่างของระดับความสามารถที่ระดับนัยสำคัญ .05 ($t = -0.56, p = .58$)

ผู้ถูกประเมินคนที่ 5

วัดครั้งที่ 1 มีความสามารถน้อยกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.06 มีหน่วยการวัดเป็นโลจิต (Target measure) = 0.08 โลจิต

ครั้งที่ 2 มีความสามารถมากกว่าที่คาดการณ์ฯ -0.07 (0.18 โลจิต)

ครั้งที่ 1 มีระดับความสามารถแตกต่างจากครั้งที่ 2 (target contrast) $(0.08 - 0.18) = -0.10$ โลจิต แต่ไม่พบความแตกต่างของระดับความสามารถที่ระดับนัยสำคัญ .05 ($t = -0.70, p = .49$)

ผู้ถูกประเมินคนที่ 6

วัดครั้งที่ 1 มีความสามารถน้อยกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.12 หน่วยการวัดเป็นโลจิต (target measure) = 0.52 โลจิต

ครั้งที่ 2 มีความสามารถมากกว่าที่คาดการณ์ฯ -0.12 (0.74 โลจิต)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) $(0.52 - 0.74) = -0.22$ โลจิต แต่ไม่พบความแตกต่างของระดับความสามารถที่ระดับนัยสำคัญ .05 ($t = -1.40, p = .16$)

ผู้ถูกประเมินคนที่ 7

การวัดครั้งที่ 1 มีความสามารถน้อยกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.06
มีหน่วยการวัดเป็นโลจิท (target measure) = - 0.32 โลจิท

ครั้งที่ 2 มีความสามารถมากกว่าที่คาดการณ์ฯ 0.07 (- 0.21 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) (- 0.32 - - 0.21) = - 0.11 โลจิท แต่ไม่พบ
ความแตกต่างของระดับความสามารถที่ระดับนัยสำคัญ .05 ($t = - 0.75, p = .46$)

ผู้ถูกประเมินคนที่ 8

วัดครั้งที่ 1 มีความสามารถมากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = - 0.05
มีหน่วยการวัดเป็นโลจิท (Target measure) = 0.18 โลจิท

ครั้งที่ 2 มีความสามารถน้อยกว่าที่คาดการณ์ฯ = 0.05 (0.11 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) (0.18 - 0.11) = 0.07 โลจิท แต่ไม่พบ
ความแตกต่างของระดับความสามารถที่ระดับนัยสำคัญ .05 ($t = 0.49, p = .63$)

ผู้ถูกประเมินคนที่ 9

วัดครั้งที่ 1 มีความสามารถน้อยกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.12
มีหน่วยการวัดเป็นโลจิท (target measure) = - 0.03 โลจิท

ครั้งที่ 2 มีความสามารถมากกว่าที่คาดการณ์ฯ - 0.12 (0.15 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) (- 0.03 - 0.15) = - 0.18 โลจิท แต่ไม่พบ
ความแตกต่างของระดับความสามารถที่ระดับนัยสำคัญ .05 ($t = - 1.26, p = .21$)

ผู้ถูกประเมินคนที่ 10

วัดครั้งที่ 1 มีความสามารถน้อยกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.02
มีหน่วยการวัดเป็นโลจิท (Target measure) = - 0.69 โลจิท

ครั้งที่ 2 มีความสามารถมากกว่าที่คาดการณ์ฯ = - 0.02 (- 0.64 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) (- 0.69 - - 0.64) = - 0.05 โลจิท แต่ไม่พบ
ความแตกต่างของระดับความสามารถที่ระดับนัยสำคัญ .05 ($t = - 0.27, p = .79$)

จากผลการศึกษาเมื่อพิจารณาค่า Fit mean square ระดับความสามารถของผู้ถูกประเมิน
ส่วนใหญ่อยู่ในค่าที่คาดการณ์โดยโมเดล ยกเว้น ผู้ถูกประเมินคนที่ 8 ที่มีทั้ง Infit และ Outfit
mean-square เกินเกณฑ์ที่ตั้งไว้ เล็กน้อยคือ 1.45 – 1.57 (เกณฑ์ที่ตั้งไว้ 0.6 – 1.4) แสดงถึงภาวะ misfit

แสดงว่าผู้ถูกประเมินคนที่ 8 มีระดับความสามารถห่างไกลจากการคาดการณ์ของโมเดล เมื่อพิจารณาในระดับบุคคลโดยใช้ข้อมูลจาก variable map ในภาพที่ 5 หน้า 117 พาเซทของผู้ประเมินพบว่า ผู้ประเมินคนที่ 3 มีการกตคะแนนมากที่สุด และคนที่ 10 มีระดับการปล่อยคะแนนมากที่สุด และมีระดับการกต ปล่อยคะแนน แตกต่างจากกลุ่มผู้ประเมิน แต่เมื่อพิจารณาในพาเซทของผู้ถูกประเมินพบว่าผู้ถูกประเมินคนที่ 8 อยู่ในกลุ่มที่มีระดับความสามารถใกล้เคียงกับผู้ถูกประเมินส่วนใหญ่ จากข้อมูลนี้จึงยังไม่สามารถอธิบายเหตุผลของการเกิดภาวะ misfit ได้ อย่างไรก็ตามค่า misfit ที่ถือว่ามีความผิดปกติมากจะมีค่า 2.00 ขึ้นไปจึงไม่ได้ทำการวิเคราะห์สาเหตุต่อ จึงสรุปได้ว่าระดับความสามารถของผู้ถูกประเมิน ในการวัด 2 ครั้ง ไม่แตกต่างกันที่ระดับนัยสำคัญ .05 และการใช้โมดูลการฝึกอบรมไม่ได้ทำให้เกิดการเปลี่ยนแปลงต่อการประเมินระดับความสามารถของผู้ถูกประเมิน (Linacre, 1994) รายละเอียดดั่งนำเสนอในตารางที่ 40

ตารางที่ 40 เปรียบเทียบผลการวัดระดับความสามารถของผู้ถูกประเมิน ก่อน และหลังการฝึกอบรม

ผู้ถูกประเมิน	ครั้งที่วัด	Target measure	Obs-Exp. Average	Target contrast	Joint S.E.	Infit MnSq	Outfit MnSq	t	p
คนที่1	ครั้งที่ 1	0.09	0.07	-0.11	0.14	0.90	0.89	-0.77	.44
	ครั้งที่ 2	0.20	-0.07			0.99	0.98		
คนที่2	ครั้งที่ 1	-0.38	0.05	-0.09	0.15	1.02	1.00	-0.61	.54
	ครั้งที่ 2	-0.29	-0.05			1.00	0.97		
คนที่3	ครั้งที่ 1	0.07	0.07	-0.11	0.14	0.90	0.89	-0.77	.44
	ครั้งที่ 2	0.18	-0.07			0.94	0.95		
คนที่4	ครั้งที่ 1	-0.01	0.05	-0.08	0.14	0.97	0.96	-0.56	.58
	ครั้งที่ 2	0.07	-0.05			0.91	0.91		
คนที่5	ครั้งที่ 1	0.08	0.06	-0.10	0.14	1.00	1.02	-0.70	.49
	ครั้งที่ 2	0.18	-0.07			1.04	1.05		
คนที่6	ครั้งที่ 1	0.52	0.12	-0.22	0.16	0.93	0.92	-1.40	.16
	ครั้งที่ 2	0.74	-0.12			0.86	0.83		
คนที่7	ครั้งที่ 1	-0.32	0.06	-0.11	0.15	1.21	0.92	-0.75	.46
	ครั้งที่ 2	-0.21	-0.07			1.18	1.17		
คนที่8	ครั้งที่ 1	0.18	-0.05	0.07	0.14	1.45	1.46	0.49	.63
	ครั้งที่ 2	0.11	0.05			1.55	1.57		
คนที่9	ครั้งที่ 1	-0.03	0.12	-0.18	0.14	0.80	0.79	-1.26	.21
	ครั้งที่ 2	0.15	-0.12			0.62	0.62		
คนที่10	ครั้งที่ 1	-0.69	0.02	-0.05	.18	0.79	0.74	-0.27	.79
	ครั้งที่ 2	-0.64	--0.02			0.91	0.90		

$p < .05$

2.2 ศักยภาพในระดักลุ่มที่ช่วยในการระบุความคลาดเคลื่อนแบบให้คะแนนเกาะกลุ่มตรงกลาง ในการวัดแต่ละครั้ง

พารามิเตอร์ของฟาเซทผู้ถูกประเมิน (ratee) ใช้พิจารณาความคลาดเคลื่อนแบบให้คะแนนเกาะกลุ่มตรงกลาง โดยมีสมมติฐานว่างว่าผู้ถูกประเมินทุกคนได้รับระดับคะแนนการประเมินผลงานในระดับเดียวกัน (the null hypothesis is that all examinees exhibited the same performance measurement) หากยอมรับสมมติฐานว่างแสดงว่ามีความคลาดเคลื่อนแบบให้คะแนนเกาะกลุ่มตรงกลาง หากปฏิเสธสมมติฐานว่างแสดงว่าไม่มีความคลาดเคลื่อนแบบให้คะแนนเกาะกลุ่มตรงกลาง

ในการวัดครั้งที่ 1 เมื่อพิจารณาในรายกลุ่ม ผู้ถูกประเมินอย่างน้อย 2 คน ได้รับคะแนนการประเมินในระดับที่แตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ($\chi^2 = 80.8$, d.f. = 9) เมื่อพิจารณา ค่าเฉลี่ย และช่วงของการได้รับคะแนนคือ 0.00 และ 1.24 มีช่วงของคะแนนไม่กว้างมาก ความแปรปรวนของการประมาณค่าการให้คะแนนค่อนข้างน้อย (SD = 0.33) แสดงว่าแม้จะมีช่วงคะแนนค่อนข้างแคบ อย่างไรก็ตามค่า reliability of separation = 0.89 ค่าสูงแสดงว่าโดยเฉลี่ยแล้ว ผู้ประเมินสามารถแยกความสามารถของผู้ถูกประเมินได้ในระดับดี ค่าที่พึงประสงค์ คือค่าที่เข้าใกล้ 1 และค่า Separation ratio = 2.92 แสดงว่าภายในกลุ่มของผู้ถูกประเมินมีผลการประเมินเป็น 3 เท่าของความคลาดเคลื่อนจากการวัด มีการกระจายตัวกันของระดับความสามารถของผู้ถูกประเมินในระดับดี ผลของ Mean Infit และ Mean Outfit คือ 1.00 และ 0.99 แสดงให้เห็นว่าพฤติกรรมกรรมการประเมินของผู้ประเมินมีค่าใกล้เคียงกับพฤติกรรมกรรมการประเมินที่โมเดลคาดการณ์ ในระดับกลุ่มไม่พบความคลาดเคลื่อนแบบ ให้คะแนนแบบเกาะกลุ่มตรงกลาง

การวัดครั้งที่ 2 พบว่ากลุ่มผู้ถูกประเมินอย่างน้อย 2 คนได้รับการประเมินในระดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.001 ($\chi^2 = 89.6$, d.f. = 9) ค่าเฉลี่ย และช่วงของการให้คะแนนคือ 0.00 และ 0.69 ตามลำดับ ความแปรปรวนของการประมาณค่าการให้คะแนนเกาะกลุ่มตรงกลางสูงกว่าครั้งที่ 1 เล็กน้อย (SD = 0.39) เพื่อพิจารณาสถิติ reliability of separation = 0.91 มีค่าเข้าใกล้ 1 แสดงว่าโดยเฉลี่ยแล้วผู้ประเมินสามารถแยกระดับความสามารถของผู้ถูกประเมินได้ในระดับสูง และค่า Separation ratio = 3.17 ภายในกลุ่มของผู้ถูกประเมินมีผลการประเมินเป็น 3 เท่า ของความคลาดเคลื่อนจากการวัด แสดงถึงผู้ถูกประเมินมีระดับความสามารถที่แตกต่างกัน ผลของ Mean

Infit และ Mean Outfit คือ 1.00 และ 0.99 แสดงให้เห็นว่าระดับความสามารถของผู้ถูกประเมินที่ได้รับ การประเมินมีค่าใกล้เคียงกับที่โมเดลคาดการณ์ สรุปคือไม่พบความคลาดเคลื่อนแบบให้คะแนนแบบ เกาะกลุ่มตรงกลาง ในระดับกลุ่ม รายละเอียดดังแสดงในตารางที่ 41

ตารางที่ 41 สถิติสรุปรวมในระดับกลุ่มสำหรับฟาเซทผู้ถูกประเมิน

ค่าสถิติ	ฟาเซทผู้ถูกประเมิน	
	การวัดครั้งที่ 1	การวัดครั้งที่ 2
Mean	0.00	0.00
SD	0.33	0.35
N	10	10
Range	1.24	1.38
Infit		
Mean	1.00	1.00
SD	0.19	0.23
Outfit		
Mean	0.99	0.99
SD	0.20	0.23
Reliability of separation	.89	.91
Separation ratio	2.92	3.17
Chi-square	80.8	89.6
<i>p</i>	.00**	.00**

3. ผลของโมดูลการฝึกอบรมฯ ต่อความคลาดเคลื่อนแบบฮาโล

แบ่งการศึกษาเป็น 2 ขั้นตอนคือ (1) ศึกษาความแตกต่างของระดับความยากของสมรรถนะ ในการวัดก่อนการฝึกอบรม (ครั้งที่ 1) และหลังการฝึกอบรมทันที (ครั้งที่ 2) และ (2) ศึกษาสถิติใน ระดับกลุ่มที่ช่วยในการระบุความยากของแต่ละสมรรถนะในการวัดแต่ละครั้ง

3.1 ศึกษาความแตกต่างของระดับความยากของสมรรถนะ ในการวัดก่อนการฝึกอบรมฯ (ครั้งที่ 1) และหลังการใช้โมดูลการฝึกอบรมทันที (ครั้งที่ 2) เมื่อเปรียบเทียบค่าเฉลี่ยรายคู่ของ คะแนนทักษะจากการวัดก่อน และหลังการใช้โมดูลการฝึกอบรมทันที จากสมรรถนะที่ประเมิน 7 ด้านคือ 1) คุณธรรม จริยธรรม 2) ความรู้ 3) ทักษะทางปัญญา 4) ความสัมพันธ์ระหว่างบุคคล และ

ความรับผิดชอบ 5) ทักษะการวิเคราะห์เชิงตัวเลข การสื่อสาร และการใช้เทคโนโลยีสารสนเทศ 6) ทักษะการปฏิบัติทางวิชาชีพ สาขาพยาบาลศาสตร์ และ 7) สมรรถนะของนักศึกษาตามเอกลักษณ์ของมหาวิทยาลัย ผลการวิเคราะห์นำเสนอผลต่างของค่าที่สังเกตได้กับค่าที่คาดหวังด้วยโมเดล (Obs - Exp. Average) มีหน่วยการวัดเป็นโลจิท (target measurement = observed measurement + Bias size), ค่า Obs - Exp. Average หากค่าเป็นบวกแสดงว่ามีความยากมากกว่าที่โมเดลคาดการณ์ ค่าลบแสดงว่ามีความง่ายมากกว่าที่โมเดลคาดการณ์ (เนื่องจากผู้วิจัยเขียนคำสั่งให้ฟาเซทผู้ประเมินเป็นค่า positive ฟาเซทเดียว ดังนั้นฟาเซทอื่นจึงอ่านค่าเป็น negative คือค่ามากหมายถึงมีความยากมาก ค่าน้อยหมายถึงมีความง่ายมาก) ค่าเป็นศูนย์แสดงว่าค่าที่ให้ตรงกับที่โมเดลคาดการณ์ ค่า t เป็นการทดสอบสมมติฐานของ target contrast/Joint S.E. ที่ว่าไม่พบความแตกต่างของระดับความง่าย ดังรายละเอียดต่อไปนี้

สมรรถนะที่ 1

วัดครั้งที่ 1 มีความยากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.05 มีหน่วยการวัด (target measure) = -0.05 โลจิท

ครั้งที่ 2 มีความง่ายกว่าที่คาดการณ์ฯ 0.05 (0.04 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) (- 0.05 - 0.04) = - 0.09 โลจิท แต่ไม่พบความแตกต่างของระดับความง่ายของสมรรถนะที่ระดับนัยสำคัญ .05 ($t = - 0.73, p = .46$)

สมรรถนะที่ 2

วัดครั้งที่ 1 มีความยากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.06 มีหน่วยการวัดเป็นโลจิท (target measure) = 0.14 โลจิท

ครั้งที่ 2 มีความง่ายกว่าที่คาดการณ์ฯ 0.06 (0.24 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (0.14 - 0.24) (target contrast) = - 0.10 โลจิท แต่ไม่พบความแตกต่างของระดับความง่ายของสมรรถนะที่ระดับนัยสำคัญ .05 ($t = - 0.85, p = .40$)

สมรรถนะที่ 3

วัดครั้งที่ 1 มีความยากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.10 มีหน่วยการวัดเป็นโลจิท = - 0.17 โลจิท

ครั้งที่ 2 มีความง่ายกว่าที่คาดการณ์ฯ 0.11 (0.01 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) $(- 0.17 - 0.01) = - 0.17$ โลจิท แต่ไม่พบความแตกต่างของระดับความยากง่ายของสมรรถนะที่ระดับนัยสำคัญ .05 ($t = - 1.41, p = .16$)

สมรรถนะที่ 4

วัดครั้งที่ 1 มีความยากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.03 มีหน่วยการวัดเป็นโลจิท (target measure) = 0.03 โลจิท

ครั้งที่ 2 มีความง่ายกว่าที่คาดการณ์ฯ - 0.03 (0.07 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) $(0.03 - 0.07) = - 0.04$ โลจิท แต่ไม่พบความแตกต่างของระดับความยากง่ายของสมรรถนะที่ระดับนัยสำคัญ .05 ($t = - 0.36, p = .72$)

สมรรถนะที่ 5

วัดครั้งที่ 1 มีความยากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.08 มีหน่วยการวัดเป็นโลจิท (target measure) = - 0.08 โลจิท

ครั้งที่ 2 มีความง่ายกว่าที่คาดการณ์ฯ - 0.08 (0.05 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) $(- 0.08 - 0.05) = - 0.13$ โลจิท แต่ไม่พบความแตกต่างของระดับความยากง่ายของสมรรถนะที่ระดับนัยสำคัญ .05 ($t = - 1.04, p = .30$)

สมรรถนะที่ 6

วัดครั้งที่ 1 มีความยากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.04 มีหน่วยการวัดเป็นโลจิท (target measure) = - 0.26 โลจิท

ครั้งที่ 2 มีความง่ายกว่าที่คาดการณ์ฯ - 0.04 (- 0.18 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) $(- 0.26 - - 0.18) = - 0.07$ โลจิท แต่ไม่พบความแตกต่างของระดับความยากง่ายของสมรรถนะที่ระดับนัยสำคัญ .05 ($t = - 0.57, p = .57$)

สมรรถนะที่ 7

วัดครั้งที่ 1 มีความยากกว่าที่คาดการณ์ด้วยโมเดล (Obs - Exp. Average) = 0.04 มีหน่วยการวัดเป็นโลจิท (target measure) = 0.04 โลจิท

ครั้งที่ 2 มีความง่ายกว่าที่คาดการณ์ฯ - 0.04 (0.11 โลจิท)

ครั้งที่ 1 แตกต่างจากครั้งที่ 2 (target contrast) $(- 0.04 - 0.11) = - 0.07$ โลจิท แต่ไม่พบความแตกต่างของระดับความยากง่ายของสมรรถนะที่ระดับนัยสำคัญ .05 ($t = - 0.55, p = .59$)

การทดสอบความแตกต่างที่เกิดจากการประเมินในแต่ละครั้ง ไม่พบความแตกต่าง ที่ระดับนัยสำคัญ.05 แสดงว่าการวัดผลการประเมินก่อน และหลังการฝึกอบรมไม่แตกต่างกัน หรือการฝึกอบรมไม่ได้ส่งผลต่อการรับรู้ของผู้ประเมินต่อระดับความยากง่ายของสมรรถนะ เมื่อพิจารณาสถิติ fit mean square พบว่าความยากของสมรรถนะส่วนใหญ่อยู่ในค่าที่คาดการณ์โดยโมเดล ยกเว้นสมรรถนะที่ 7 ที่พบภาวะ misfit ในการประเมินครั้งที่ 1 (Infit mean-square = 1.47) เมื่อพิจารณา Infit mean-square ของผู้ประเมินร่วมด้วย (ตาราง 38) พบว่าผู้ประเมินมีพฤติกรรมกรรมการประเมินเป็นไปตามที่โมเดลคาดการณ์ อย่างไรก็ตามค่า misfit ที่ถือว่ามีความผิดปกติมากจะมีค่า 2.00 ขึ้นไป จึงไม่ได้ทำการวิเคราะห์สาเหตุต่อ ดังนำเสนอในตารางที่ 42

ตารางที่ 42 เปรียบเทียบผลการวัดก่อน และหลังการใช้โมดูลฝึกอบรมของสมรรถนะที่ใช้วัด

สมรรถนะที่	ครั้งที่	Target measure	Obs- Exp. Average	Target contrast	Infit MnSq	Outfit MnSq	Joint S.E.	t	p
1. คุณธรรม จริยธรรม	ครั้งที่ 1	-0.05	0.05	-0.09	0.77	0.74	0.12	-0.73	.46
	ครั้งที่ 2	0.04	-0.05		0.79	0.78			
2. ความรู้	ครั้งที่ 1	0.14	0.06	-0.10	1.24	1.22	0.12	-0.85	.40
	ครั้งที่ 2	0.24	-0.06		1.13	1.09			
3. ทักษะทางปัญญา	ครั้งที่ 1	-0.17	0.10	-0.17	0.57	0.59	0.12	-1.41	.16
	ครั้งที่ 2	0.01	-0.11		0.64	0.64			
4. ความสัมพันธ์ระหว่างบุคคลฯ	ครั้งที่ 1	0.03	0.03	-0.04	1.11	1.09	0.12	-0.36	.72
	ครั้งที่ 2	0.07	-0.03		1.13	1.10			
5. ทักษะการวิเคราะห์เชิงตัวเลข ฯ	ครั้งที่ 1	-0.08	0.08	-0.13	0.85	0.81	0.12	-1.04	.30
	ครั้งที่ 2	0.05	-0.08		0.93	0.89			
6. ทักษะการปฏิบัติทางวิชาชีพ	ครั้งที่ 1	-0.26	0.04	-0.07	1.01	0.94	0.13	-0.57	.57
	และ 2	-0.18	-0.04		1.10	1.03			
7. เอกลักษณ์ของมหาวิทยาลัย	ครั้งที่ 1	-0.04	0.04	-0.07	1.47	1.54	.12	-0.55	.59
	ครั้งที่ 2	0.11	-0.04		1.32	1.43			

p < .05

3.2 ศึกษาพาเซทของสมรรถนะที่ต้องการประเมิน (trait) ใช้ในการพิจารณาความคลาดเคลื่อน แบบฮาโล เป็นการทดสอบสมมติฐานว่างที่ว่าสมรรถนะทุกสมรรถนะอยู่ในระดับความยากเดียวกันทั้งหมด หรือมีความคลาดเคลื่อนแบบฮาโล หรือไม่ หากปฏิเสธสมมติฐานว่างแสดงว่าไม่มีความคลาดเคลื่อนแบบฮาโล

ในการวัดครั้งที่ 1 มีความแตกต่างของความยากในทุกสมรรถนะการประเมินอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($\chi^2 = 14.2$, d.f. = 6, $p = .03$) แสดงว่ามีอย่างน้อย 2 สมรรถนะที่มีความแตกต่างกันของความยาก ค่าเฉลี่ย และช่วงของการประเมินตามสมรรถนะ คือ 0.00 และ 0.40 ตามลำดับ ความแปรปรวนของการประมาณค่าการประเมินตามสมรรถนะค่อนข้างน้อย ($SD = 0.13$) แสดงว่ามีการกระจายตัวของโลจิทความยากในช่วงแคบ ๆ เมื่อพิจารณาสถิติ Separation reliability of separation = 0.51 อยู่ในระดับค่อนข้างต่ำ (ค่าที่ยอมรับได้คือ reliability ≥ 0.50) (ศิริชัย กาญจนวาสี, 2552) แสดงว่าผู้ประเมินแยกความแตกต่างระหว่างการประเมินในแต่ละสมรรถนะได้ค่อนข้างต่ำ ค่าที่ดีคือค่าที่เข้าใกล้ 1 และค่า Separation ratio = 1.03 แสดงว่าการประเมินระดับความยากของแต่ละสมรรถนะมีค่ามากกว่าความคลาดเคลื่อนจากการวัด 1 เท่า และมีค่าไม่เข้าใกล้ศูนย์ ส่วนค่า Separation index มีค่า 1.71 แสดงว่าผู้ประเมินแบ่งระดับความยากง่ายของสมรรถนะได้ประมาณ 2 ระดับ ผลของ Mean Infit และ Mean Outfit คือ 1.00 และ 0.99 แสดงให้เห็นว่าพฤติกรรมกรรมการประเมินของผู้ประเมินมีค่าใกล้เคียงกับพฤติกรรมกรรมการประเมินที่โมเดลคาดการณ์ จึงสรุปได้ว่ามีแนวโน้มจะเกิดความคลาดเคลื่อนแบบฮาลอในระดับกลุ่ม

ในการประเมินครั้งที่ 2 มีความแตกต่างของความยากในทุกสมรรถนะการประเมินอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($\chi^2 = 13.0$, d.f. = 6, $p = .04$) แสดงว่ามีอย่างน้อย 2 สมรรถนะที่มีความแตกต่างกันของความยาก ค่าเฉลี่ย และช่วงของการประเมินตามสมรรถนะ คือ 0.00 และ 0.40 ตามลำดับ ความแปรปรวนของการประมาณค่าการประเมินตามสมรรถนะค่อนข้างน้อย ($SD = 0.12$) แสดงว่ามีการกระจายตัวของโลจิทความยากในช่วงแคบ ๆ เมื่อพิจารณาสถิติ Separation reliability of separation = 0.48 อยู่ในระดับค่อนข้างต่ำ แต่มีค่าใกล้เคียงกับค่าที่ยอมรับได้ (ค่าที่ยอมรับได้คือ reliability ≥ 0.50) (ศิริชัย กาญจนวาสี, 2552) แสดงว่าผู้ประเมินแยกความแตกต่างระหว่างการประเมินในแต่ละสมรรถนะได้ค่อนข้างต่ำ และค่า Separation ratio = 0.95 แสดงว่าการประเมินระดับความยากของแต่ละสมรรถนะมีค่ามากกว่าความคลาดเคลื่อนจากการวัด 1 เท่า และมีค่าไม่เข้าใกล้ศูนย์ ส่วนค่า Separation index มีค่า 1.60 แสดงว่าผู้ประเมินแบ่งระดับความยากง่ายของสมรรถนะได้ประมาณ 2 ระดับ ผลของ Mean Infit และ Mean Outfit คือ 1.01 และ 0.99 แสดงให้เห็นว่าพฤติกรรมกรรมการประเมินของผู้ประเมินมีค่าใกล้เคียงกับพฤติกรรมกรรมการประเมินที่โมเดลคาดการณ์ จึงสรุปได้ว่ามีแนวโน้มที่จะเกิดความคลาดเคลื่อนแบบฮาลอในระดับกลุ่ม ดังแสดงในตารางที่ 43

ตารางที่ 43 สถิติสรุปรวมในระดับกลุ่มสำหรับฟาเซทสมรรถนะ

ค่าสถิติ	ฟาเซทสมรรถนะ	
	การวัดครั้งที่ 1	การวัดครั้งที่ 2
Mean	0.00	0.00
SD	0.13	0.12
N	7	7
Range	0.40	0.43
Infit		
Mean	1.00	1.01
SD	0.28	0.22
Outfit		
Mean	0.99	0.99
SD	0.30	0.24
Reliability of separation	0.51	0.48
Separation index	1.71	1.60
Separation ratio	1.03	0.95
Chi-square	14.2	13.0
<i>p</i>	.03*	.04*

4. เมื่อพิจารณารายบุคคลจาก variable map ของภาพที่ 5 ในสดมภ์ที่ 1 เป็นการกระจายตัวของโลจิทการวัดสำหรับทุกฟาเซท ในช่วงแคบๆ +1 ถึง -1 (ในคำสั่งการวิเคราะห์ด้วยโมดูล FACETS กำหนดให้ ฟาเซทที่ 1 (rater) positive คือค่าบวกแสดงว่าปล่อยคะแนน ค่าลบกดคะแนน ส่วนฟาเซทอื่นกำหนดให้ negative คือ ฟาเซทที่ 2 (ratee) ค่าบวกแสดงว่ามีระดับความสามารถน้อย ค่าลบแสดงว่ามีระดับความสามารถมาก ฟาเซทที่ 3 (trait) ค่าบวกแสดงว่ายาก ค่าลบแสดงว่าง่าย)

4.1 ฟาเซทผู้ประเมิน ในสดมภ์ที่ 3 และ 6 แสดงถึงลำดับของผู้ประเมินที่มีโลจิทในช่วงโลจิทของการกด ปล่อยคะแนน -0.3 ถึง 0.4 ค่าบวกแสดงถึงการปล่อยคะแนน ค่าลบแสดงถึงการกดคะแนน ผู้ประเมินมีแนวโน้มที่จะให้คะแนนการประเมินในระดับ 2 ถึง 3 มากกว่าคะแนนการประเมินอื่นบนมาตราการประเมิน 1 ถึง 4 โดยผู้ประเมินคนที่ 10 มีโลจิทการปล่อยคะแนนมากกว่าผู้ประเมินคนอื่น และคนที่ 3 มีโลจิทการกดคะแนนมากกว่าผู้ประเมินคนอื่น ผู้ประเมินมีโลจิทการปล่อยคะแนนจากมากไปน้อยดังนี้ ผู้ประเมินคนที่ 10, 2,4,5,1,8,9,11,6 และ 7

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้ผู้วิจัยใช้ระเบียบการวิจัย และพัฒนา (Research and Development) มีวัตถุประสงค์เพื่อ 1) พัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล 2) ศึกษาผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล ต่อความรู้ ทักษะ และทักษะการประเมินภาคปฏิบัติ 3) ศึกษาผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลต่อความคลาดเคลื่อนจากผู้ประเมิน

ตัวอย่างที่ใช้ในการศึกษาเป็นอาจารย์ประจำคณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต เลือกรขนาดตัวอย่างจำนวน 2 กลุ่มคือ กลุ่มทดลองใช้โมดูล และกลุ่มใช้โมดูล เลือกตัวอย่างแบบจับคู่ (match) กำหนดเกณฑ์การคัดเลือก (Inclusive criteria) คือ เป็นอาจารย์ที่มีประสบการณ์การสอนในสาขาวิชาเดียวกันหรือใกล้เคียงกัน และมีความพร้อมในการทดลองใช้โมดูลการฝึกอบรมจนเสร็จสิ้นกระบวนการ สุ่มอย่างง่ายทีละคู่ แบ่งเข้ากลุ่มทดลองใช้โมดูล 1 คน และเข้ากลุ่มใช้โมดูลจริง 1 คน จนครบ 6 คู่ ได้ตัวอย่างกลุ่มละ 6 คน คือ กลุ่มทดลองใช้โมดูลมีจำนวน 6 คน และกลุ่มใช้โมดูล 6 คน นอกจากนี้ในกลุ่มใช้โมดูลมีอาจารย์ที่สมัครใจเข้ารับการฝึกอบรมอีก 5 คน รวมเป็น 11 คน และมีกลุ่มผู้ให้ข้อมูลในการสำรวจสภาพปัจจุบันของการประเมินภาคปฏิบัติ ประกอบด้วย 3 ส่วนคือ 1) อาจารย์ประจำ คณะพยาบาลศาสตร์ 4 สถาบันจาก 4 ภาค และอาจารย์ประจำคณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต จำนวน 96 คน 2) ผู้บริหารฝ่ายวิชาการของ 4 สถาบันจาก 4 ภาค และผู้บริหารฝ่ายวิชาการของคณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต รวมเป็น 5 ท่าน และ 3) ผู้ทรงคุณวุฒิภายในคณะพยาบาลศาสตร์จำนวน 7 ท่าน ในการสำรวจความต้องการในการใช้รูปแบบการประเมินภาคปฏิบัติ

เครื่องมือที่ใช้ในการศึกษาสภาพปัจจุบันของการประเมินภาคปฏิบัติ ประกอบด้วย 1) แบบสอบถามสำรวจสภาพปัญหาการประเมินภาคปฏิบัติประกอบด้วยแบบสอบถามแบบมีตัวเลือก และปลายปิดจำนวน 20 ข้อ และส่วนท้ายของแบบสอบถามเป็นคำถามปลายเปิด 2) แบบสัมภาษณ์

แบบกึ่งโครงสร้างในการสัมภาษณ์เชิงลึกผู้บริหารฝ่ายวิชาการ 3) แบบสอบถามสำรวจความต้องการในการใช้รูบรีค จำนวน 9 ข้อ

เครื่องมือที่ใช้ในการศึกษาผลของโมดูล ต่อความรู้ ทักษะคติ และทักษะการประเมินภาคปฏิบัติ รวมถึง ศึกษาความคลาดเคลื่อนจากผู้ประเมิน ประกอบด้วย 1) แบบวัดความรู้ด้านการประเมินภาคปฏิบัติ จำนวน 25 ข้อ 2) แบบวัดทัศนคติเกี่ยวกับการประเมินภาคปฏิบัติจำนวน 10 ข้อ 3) ชุดประเมินทักษะภาคปฏิบัติ ประกอบด้วย รูบรีคในการประเมินภาคปฏิบัติ 7 สมรรถนะหลัก (26 สมรรถนะย่อย) ใช้ร่วมกับสื่อวีดิทัศน์สถานการณ์จำลอง จำนวน 10 ชุด 4) ชุดประเมินความพึงพอใจ ประกอบด้วย แบบวัดความพึงพอใจต่อการฝึกอบรมจำนวน 10 ข้อ และแบบสัมภาษณ์ มีโครงสร้างและ 5) คู่มือฝึกอบรมการเสริมสร้างสมรรถนะการประเมินจำนวน 1 เล่ม

การดำเนินการวิจัยแบ่งเป็น 5 ขั้นตอนคือ ขั้นตอนที่ 1 สำรวจสภาพปัจจุบันของการประเมินภาคปฏิบัติ และการเตรียมผู้ประเมินภาคปฏิบัติรวมถึงความต้องการในการเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล ขั้นตอนที่ 2 การพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล โดยนำข้อมูลที่ได้จากขั้นตอนที่ 1 มาพิจารณาร่วมกับการบูรณาการองค์ความรู้ที่ได้จากการค้นคว้าตำรา เอกสาร งานวิจัยเพื่อการออกแบบโมดูลการฝึกอบรม และมีการตรวจสอบคุณภาพของเครื่องมือทั้งหมดตามลักษณะของข้อมูล ปรับปรุงแก้ไขตามข้อเสนอแนะของผู้ทรงคุณวุฒิ ขั้นตอนที่ 3 ทดลองใช้โมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล จากนั้นจึงมีการตรวจสอบผลของการใช้โมดูล วัดได้จาก ความรู้ ทักษะคติ และทักษะก่อน และหลังการใช้โมดูล และการวัดความพึงพอใจหลังการใช้โมดูล นำผลการตรวจสอบการใช้โมดูล มาปรับปรุงเพื่อนำมาใช้กับตัวอย่างในขั้นตอนต่อไป ขั้นตอนที่ 4 การประเมินผลการใช้โมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล โดยการตรวจสอบผลของการใช้โมดูล โดยการวัดความรู้ ทักษะคติ ทักษะและ ความพึงพอใจหลังการใช้โมดูลการฝึกอบรม รวมถึงการวัดความคลาดเคลื่อนจากการกด ปล้อยคะแนน การให้คะแนนเกาะกลุ่มตรงกลาง ความคลาดเคลื่อนแบบฮาลอ และความเที่ยงระหว่างผู้ประเมิน และขั้นตอนที่ 5 ปรับปรุงโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล และสรุปเป็นโมดูลมาตรฐาน ในการเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล

การวิเคราะห์ข้อมูล แบ่งตามลักษณะของข้อมูลคือ ข้อมูลเชิงปริมาณ ได้แก่ จำนวน ร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าความเที่ยงระหว่างผู้ประเมินใช้ค่า rater agreement index (RAI) และค่า Intra-class correlation coefficient (ICC) การทดสอบความแตกต่างของค่าเฉลี่ยด้วย paired t-test และ F-test วิเคราะห์ ด้วยโปรแกรม SPSS ความคลาดเคลื่อนจากผู้ประเมิน (rater effects) ใช้โมเดลหลายองค์ประกอบของราส์ซ วิเคราะห์ด้วยโปรแกรม FACETS version 3.71.4 วิเคราะห์ข้อมูลเชิงคุณภาพ ด้วยการวิเคราะห์เนื้อหา และสรุปความตามประเด็นสำคัญ

สรุปผลการวิจัย

สรุปผลการวิจัยตามวัตถุประสงค์การวิจัย คือ ตอนที่ 1 ผลการพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล ตอนที่ 2 ผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล ต่อความรู้ ทักษะการประเมินภาคปฏิบัติ และ ตอนที่ 3 ผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลต่อความคลาดเคลื่อนจากผู้ประเมิน มีรายละเอียดดังนี้

ตอนที่ 1 ผลการพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล

การสรุปผลการพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล นำเสนอตามขั้นตอนการพัฒนาดังนี้ (1) การสำรวจสภาพปัญหาการประเมินภาคปฏิบัติ (2) การออกแบบโมดูลการฝึกอบรม (3) การสร้างเครื่องมือ และ ตรวจสอบคุณภาพเครื่องมือที่ใช้ในโมดูล และใช้ในการศึกษาผลของโมดูลการฝึกอบรม ต่อความรู้ ทักษะการประเมินภาคปฏิบัติ และความคลาดเคลื่อนจากผู้ประเมิน

1. สรุปผลการสำรวจสภาพปัจจุบันในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล

การสรุปผลสำรวจสภาพปัจจุบันในการประเมินภาคปฏิบัติ ในประเด็นสำคัญที่นำมาใช้ในการพัฒนาโมดูลดังนี้

1.1 สมรรถนะที่ใช้ในการประเมินภาคปฏิบัติ ส่วนใหญ่ใช้กรอบมาตรฐานคุณวุฒิระดับอุดมศึกษา บางส่วนใช้ร่วมกับสมรรถนะของบัณฑิตตามเอกลักษณ์ของสถาบัน

1.2 เครื่องมือที่ใช้ในการประเมินภาคปฏิบัติมีหลายประเภทในการประเมินครั้งเดียวกัน ส่วนใหญ่เป็นแบบตรวจสอบรายการ รองลงมาเป็นมาตราประมาณค่า ระเบียบพฤติกรรม และ รุบริค ซึ่งบางสถาบันไม่ได้ใช้รุบริคเป็นเครื่องมือหลักในการประเมินเนื่องจากขาดความรู้ความเข้าใจในการพัฒนาเครื่องมือ

1.3 ผู้ให้ข้อมูลส่วนใหญ่เห็นว่าองค์กร และตนเองควรได้รับการเตรียมความพร้อมด้านการประเมินภาคปฏิบัติ เพื่อให้พัฒนาตนเองให้ทันการเปลี่ยนแปลงของศาสตร์ใหม่ ๆ และสร้างความมั่นใจในการประเมิน

1.4 รูปแบบการเตรียมความพร้อมผู้ประเมินส่วนใหญ่ต้องการระบบการปฐมนิเทศโดยใช้อาจารย์พี่เลี้ยง ระบบการสอนงาน และให้คำปรึกษา รองลงมาเป็นการฝึกอบรม การประชุม การใช้เทคโนโลยี และการเข้ามามีส่วนร่วมในกระบวนการประเมิน ตามลำดับ

1.5 ผู้ให้ข้อมูลส่วนใหญ่เลือกวิธีการฝึกอบรมในเรื่อง ความรู้พื้นฐานเกี่ยวกับการวัด และประเมินผล รองลงมาคือ การฝึกการสังเกตพฤติกรรม การฝึกอบรมเพื่อทำความเข้าใจกับความคลาดเคลื่อนจากผู้ประเมิน การฝึกอบรมกรอบของการอ้างอิง และการฝึกอบรมการตัดสินใจตามลำดับ

1.6 ใน 5 สถาบันการศึกษาพยาบาลที่ทำการสำรวจ พบปัญหาคล้ายคลึงกัน คือ ความไม่สอดคล้องกันของผลการประเมิน สาเหตุอาจเกิดจากหลายปัจจัย เช่น ความเป็นอัตตาของผู้ประเมิน หรือการพร้อมรู้ด้านการวัด และประเมินผลของผู้ประเมิน แต่มีแนวทางแก้ปัญหาไปในทิศทางเดียวกันคือ ประชุมปรึกษาหารือ โดยยึดหลักการของการประเมิน และหลักฐานประกอบการประเมิน เช่น บันทึกพฤติกรรมการทำงานของนักศึกษา มาประกอบการพิจารณา

2. สรุปผลการออกแบบโมดูลการฝึกอบรม

การออกแบบโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลในการประเมินภาคปฏิบัตินักศึกษาพยาบาล อาศัยข้อมูลจากการศึกษาสภาพปัญหาของสถาบันการศึกษาพยาบาล ในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล ร่วมกับการบูรณาการข้อมูลจากการค้นคว้า จากตำรา เอกสารทางวิชาการ และงานวิจัยที่เกี่ยวข้อง และการมีส่วนร่วมของผู้มีส่วนได้ส่วนเสีย คือ การสำรวจความต้องการในการใช้รับบริการประเมินภาคปฏิบัติ และการจัดลำดับความยากของกิจกรรมการพยาบาลพื้นฐาน จากผู้ทรงคุณวุฒิภายในคณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต ทำให้ได้โมดูลที่มีหน่วยการเรียนรู้ที่มีเนื้อหา และกิจกรรมในแต่ละหน่วยการเรียนรู้ สอดคล้องกับบริบทของตัวอย่าง จำนวน 5 หน่วยการเรียนรู้ คือ

2.1 หน่วยการเรียนรู้ที่ 1 ความรู้พื้นฐานด้านการประเมินภาคปฏิบัติ มีเนื้อหาที่ทำให้ผู้เข้ารับการฝึกอบรม มีความรู้ความเข้าใจเกี่ยวกับการประเมินภาคปฏิบัติในทุกขั้นตอนของกระบวนการวัดและประเมินผลภาคปฏิบัติ และมีกิจกรรมท้ายบทที่ช่วยยืนยันความเข้าใจที่ถูกต้อง

2.2 หน่วยการเรียนรู้ที่ 2 ความรู้เกี่ยวกับความคลาดเคลื่อนจากผู้ประเมิน มีเนื้อหาที่ทำให้ผู้เข้ารับการฝึกอบรมเข้าใจสาเหตุ การป้องกัน และการตรวจสอบข้อมูลเบื้องต้นที่แสดงถึงการเกิดความคลาดเคลื่อนจากผู้ประเมิน รวมถึงกิจกรรมท้ายบทที่ช่วยสร้างความเข้าใจให้ชัดเจนเกี่ยวกับความคลาดเคลื่อนจากผู้ประเมินแต่ละประเภท

2.3 หน่วยการเรียนรู้ที่ 3 กรอบของการอ้างอิง มีเนื้อหาที่ช่วยสร้างความเข้าใจให้ตรงกันในการใช้เกณฑ์การประเมิน ระหว่างผู้ประเมิน โดยได้เน้นความสำคัญของกิจกรรมในหน่วยการเรียนรู้นี้จะช่วยลดปัญหาในการตัดสินผลการประเมินที่ไม่ตรงกัน ซึ่งในขั้นตอนของการสำรวจสภาพปัญหาพบปัญหานี้เป็นปัญหาสำคัญในทุกสถาบันการศึกษาที่ทำการสำรวจ

2.4 หน่วยการเรียนรู้ที่ 4 การฝึกการสังเกต มีเนื้อหาที่ทำให้ผู้ประเมินทราบถึงเทคนิคในการสังเกตพฤติกรรม รวมถึงการฝึกการบันทึกพฤติกรรมที่สังเกตได้ เพื่อให้ได้ข้อมูลที่มีปริมาณ และคุณภาพเพียงพอสำหรับประกอบการตัดสินผลการประเมิน

2.5 หน่วยการเรียนรู้ที่ 5 การฝึกการตัดสินใจ มีเนื้อหาที่ทำให้เข้าใจธรรมชาติของการตัดสินใจ ซึ่งอาจทำให้เกิดความคลาดเคลื่อนจากผู้ประเมินได้ และการฝึกเทคนิคการควบคุมตนเองให้ตัดสินใจด้วยหลักการ และเหตุผลแทนการตัดสินใจด้วยสัญชาตญาณ

นอกจากนี้ยังมีกิจกรรมการฝึกทักษะการประเมินภายหลังการเรียนรู้ในแต่ละหน่วยการเรียนรู้เสร็จสิ้น เพื่อช่วยให้มีทักษะเพียงพอสำหรับการเป็นผู้ประเมินที่มีคุณภาพ โดยอาศัยการเทียบเคียงกับผลการประเมินของผู้ทรงคุณวุฒิ

3. การสร้างเครื่องมือ และ ตรวจสอบคุณภาพเครื่องมือที่ใช้ในโมดูล และใช้ในการศึกษาผลของโมดูลการฝึกอบรม ต่อความรู้ ทักษะ และทักษะการประเมินภาคปฏิบัติ และความคลาดเคลื่อนจากผู้ประเมิน

มีการพัฒนาเครื่องมือที่ใช้ คือ 1) ชุดการฝึกทักษะประเมินภาคปฏิบัติ ประกอบด้วยรูบริคที่ใช้ในการประเมินทักษะ ร่วมกับวิดีโอทัศนสถานการณ์จำลอง จำนวน 8 ชุด 2) แบบวัดความรู้ด้านการประเมินภาคปฏิบัติ 3) แบบวัดทัศนคติต่อการประเมินภาคปฏิบัติ 4) ชุดประเมินความพึงพอใจ ประกอบด้วย แบบวัดความพึงพอใจต่อการใช้โมดูลการฝึกอบรม และแบบสัมภาษณ์โครงสร้าง 5) คู่มือฝึกอบรมเสริมสร้างสมรรถนะการประเมิน มีการตรวจสอบคุณภาพของเครื่องมือแต่ละประเภทจนได้เครื่องมือที่มีคุณภาพตามมาตรฐาน ตามลำดับต่อไปนี้

3.1 ชุดการฝึกทักษะการประเมินภาคปฏิบัติ ประกอบด้วยรูบริคที่ใช้ในการประเมินทักษะ ร่วมกับ งานที่ใช้ในการประเมินภาคปฏิบัติอยู่ในรูปสื่อวิดีโอทัศนสถานการณ์การปฏิบัติการพยาบาล จำนวน 8 ชุด ใช้ในระยะ ก่อนการฝึกอบรม หลังการฝึกอบรมทันที และหลังการฝึกอบรม 3 เดือน รูบริค มีค่า IOC เท่ากับ 0.57 – 1.00 และ ค่า IOC ของบทวิดีโอทัศน เท่ากับ 0.67 – 1.00 ตรวจสอบความเที่ยงระหว่างผู้ประเมินด้วยค่า Intra-class correlation coefficient :ICC (3,5) มีค่าเท่ากับ 0.45 ถึง 0.80 สรุปได้ว่าวิดีโอทัศนทั้ง 8 ชุด มีความเที่ยงระหว่างผู้ประเมินระดับปานกลางขึ้นไป ถึงระดับสูง และค่า Rater agreement index (RAI) ในรูป Mean RAI มีค่าเท่ากับ 0.84 ถึง 0.95 มีความสอดคล้อง และความสม่ำเสมอในการประเมินระหว่างผู้ประเมินในระดับสูง

3.2 แบบวัดความรู้เกี่ยวกับการประเมินภาคปฏิบัติ มีการตรวจสอบคุณภาพโดยความตรงเชิงเนื้อหาด้วยค่า IOC เท่ากับ 0.54 ถึง 1.00 วิเคราะห์ข้อสอบรายข้อด้วยค่าความยาก (item difficulty: p) มีค่าเท่ากับ 0.31 ถึง 0.79 อำนาจจำแนก (r) เท่ากับ 0.20 ถึง 0.54 และวิเคราะห์คุณภาพชุดข้อสอบด้วยค่า Cronbach's Alpha Coefficient (α) เท่ากับ .72 สรุปได้ว่าได้แบบวัดความรู้ที่ผ่านการตรวจสอบคุณภาพผ่านตามมาตรฐาน

3.3 แบบประเมินทัศนคติต่อการเสริมสร้างสมรรถนะการประเมิน ด้านความตรงเชิงเนื้อหา มีค่า IOC เท่ากับ 0.67 ถึง 1.00 มากกว่า 0.5 ทุกข้อ แสดงว่ามีความสอดคล้องระหว่างข้อคำถามกับวัตถุประสงค์การวัดเป็นอย่างดีทุกข้อ ตรวจสอบความสัมพันธ์รายข้อของแบบวัดด้วยค่าสหสัมพันธ์ (item total correlation) มีค่าอยู่ระหว่าง 0.30 ถึง 0.81 ความเที่ยงของแบบวัดทั้งฉบับใช้วิธีสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Alpha Coefficient: α) เท่ากับ .83 สรุปได้ว่าได้แบบวัดทัศนคติผ่านการตรวจสอบคุณภาพผ่านตามมาตรฐาน

3.4 ชุดประเมินความพึงพอใจ ประกอบด้วย แบบวัดความพึงพอใจในการฝึกอบรม ร่วมกับแบบสัมภาษณ์แบบมีโครงสร้างมีแนวคำถามสอดคล้องตามแบบวัดความพึงพอใจ โดยแบบวัดมีค่า IOC เท่ากับ 0.89 - 1.00 แสดงว่ามีความสอดคล้องระหว่างข้อคำถามกับวัตถุประสงค์การวัดเป็นอย่างดีทุกข้อ

3.5 คู่มือสำหรับอบรมการเสริมสร้างสมรรถนะการประเมินมีการตรวจสอบความตรงเชิงเนื้อหาจากผู้ทรงคุณวุฒิจำนวน 9 ท่าน ใช้ค่า Item Objective Congruence Index (IOC) มีค่าเท่ากับ 0.78 - 0.89. แสดงว่าเนื้อหา มีความสอดคล้องกับความคิดเห็นของผู้ทรงคุณวุฒิ ในระดับสูง

ตอนที่ 2 ผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล ต่อความรู้ ทัศนคติ และทักษะการประเมินภาคปฏิบัติ

การสรุปผลของโมดูลการฝึกอบรม ต่อความรู้ ทัศนคติ และทักษะการประเมินภาคปฏิบัติ แบ่งเป็น 2 ส่วนคือ ส่วนที่ 1 สรุปผลการทดลองใช้โมดูล ส่วนที่ 2 การสรุปผลการใช้โมดูลจริง

1. สรุปผลการทดลองใช้โมดูลการฝึกอบรม

ผลการทดลองใช้โมดูลฯ ในตัวอย่าง จำนวน 6 ส่วนใหญ่เป็นเพศหญิง จำนวน 5 คน เป็นเพศชายจำนวน 1 คน มีระดับการศึกษาปริญญาโท จำนวน 5 คน ปริญญาตรีจำนวน 1 คน สังกัดสาขาวิชา การพยาบาลผู้ใหญ่จำนวน 3 คน รองลงมา สาขาวิชาเด็ก และวัยรุ่น 2 คน และสาขาวิชาการพยาบาลชุมชน 1 คน ตามลำดับ

ผลการทดลองใช้โมดูล ได้ผลการศึกษา 4 ประการคือผลการวัดความรู้ ทัศนคติ และทักษะการประเมินภาคปฏิบัติ และการวัดความพึงพอใจ

1.1 ผลการวัดความรู้ด้านการประเมินภาคปฏิบัติ ก่อน และหลังการใช้โมดูลการฝึกอบรม พบว่า ค่าเฉลี่ยของคะแนนความรู้ด้านการประเมินภาคปฏิบัติหลังการใช้โมดูล สูงกว่าค่าเฉลี่ยของคะแนนความรู้ก่อนการใช้โมดูล อย่างมีนัยสำคัญทางสถิติ ที่ระดับ .05 (t -test = 3.25, p = .02) แสดงว่าการฝึกอบรมทำให้ผู้เข้ารับการอบรมมีความรู้เพิ่มขึ้น

1.2 ผลการวัดทัศนคติต่อการประเมินภาคปฏิบัติ ก่อน และหลังการใช้โมดูลการฝึกอบรม พบว่า ค่าเฉลี่ยของคะแนนทัศนคติด้านการประเมินภาคปฏิบัติ ก่อน และหลังการใช้โมดูลไม่แตกต่างกัน ที่ระดับนัยสำคัญ .05 แสดงว่าการฝึกอบรมไม่ทำให้ทัศนคติของผู้เข้ารับการฝึกอบรมเปลี่ยน

1.3 ผลการวัดทักษะการประเมินภาคปฏิบัติก่อน และหลังการใช้โมดูลการฝึกอบรม พบว่าการใช้โมดูลการฝึกอบรมทำให้พฤติกรรมกรรมการประเมินของผู้เข้ารับการฝึกอบรมเปลี่ยนอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 (Fixed Chi-square = 98.1, d.f. = 5, p = .00) โดยพบว่าทำให้พฤติกรรมกรรมการประเมินของผู้ประเมินคนที่ 1 และคนที่ 6 ดีขึ้น โดยคนที่ 1 มีการปล่อยคะแนนลดลง (target contrast = 0.86, t = 2.16, p = .03) ผู้ประเมินคนที่ 6 มีผลการประเมินใกล้เคียงกับที่โมเดลคาดการณ์มากขึ้น เมื่อเทียบกับก่อนการใช้โมดูลเท่ากับ 1.04 หลังการใช้โมดูลเท่ากับ - 0.11

1.4 ผลการวัดความพึงพอใจของผู้เข้ารับการฝึกอบรมหลังสิ้นสุดการฝึกอบรมทันที ผู้เข้ารับการฝึกอบรมมีความพึงพอใจมากที่สุดในด้าน ประโยชน์ของการฝึกอบรมต่อตัวเอง ประโยชน์ของการฝึกอบรมต่อองค์กร ปริมาณ และคุณภาพของอาหาร และอาหารว่าง มีความพึงพอใจในระดับมาก ด้านความคุ้มค่ากับการเข้ารับการฝึกอบรม ความรู้ ความสามารถของวิทยากร และ ความน่าสนใจของกิจกรรมการฝึกอบรม ความชัดเจนเข้าใจง่ายของเนื้อหาในการฝึกอบรม สถานที่มีความสะดวก สบาย และเหมาะสมสำหรับการฝึกอบรม ความพร้อมของ สื่อสื่อดิจิทัล และมีความพึงพอใจในระดับปานกลางต่อ ระยะเวลาการฝึกอบรม ผลการสัมภาษณ์เฉพาะกลุ่มพบว่ามีความพึงพอใจในความรู้ และทักษะที่ได้รับ แต่ไม่มีความพึงพอใจในระยะเวลาการฝึกโดยเฉพาะการฝึกทักษะการประเมินกับวิธีทัศน์ที่มีระยะเวลานานเกินไป

สรุปการศึกษาในขั้นตอนที่ 3

ผลการทดลองใช้โมดูล ช่วยเพิ่มความรู้ แต่ไม่ได้เกิดการเปลี่ยนแปลงของทัศนคติ ส่วนทักษะการประเมินมีคุณภาพการประเมินดีขึ้นใน ผู้ประเมินคนที่ 1 และคนที่ 6 และนำปัญหาอุปสรรคที่พบในขั้นตอนทดลองใช้โมดูลฯ ไปใช้ในการปรับปรุง เครื่องมือ และกระบวนการฝึกอบรมเพื่อศึกษาต่อในขั้นตอนที่ 4 โดยมีการปรับเนื้อหาการบรรยายในสื่อ Power point ให้กระชับมากขึ้น และมีการปรับปรุงวิธีทัศน์ให้การดำเนินเรื่องมีความกระชับมากขึ้น โดยตัดเนื้อหาปลีกย่อยออกไป แต่

ไม่ได้ตัดส่วนสำคัญที่จำเป็นต้องนำมาใช้ประกอบการให้คะแนนภาคปฏิบัติ โดยวิธีที่ค้นก่อนการปรับแก้มีระยะเวลาเฉลี่ย 30 นาที/ชุด หลังการปรับแก้มีระยะเวลาเฉลี่ย 20 นาที/ชุด

2. การสรุปผลการใช้โมดูลการฝึกอบรมจริง

ผลการใช้โมดูลการฝึกอบรมจริง ในตัวอย่างจำนวน 11 คน ตัวอย่าง ส่วนใหญ่ เป็นเพศหญิง การศึกษาระดับปริญญาโท สังกัดสาขาวิชาการพยาบาลผู้ใหญ่ และการพยาบาลชุมชน มีประสบการณ์การสอน 6 ถึง 10 ปี และไม่เคยเข้ารับการฝึกอบรมเป็นผู้ประเมินภาคปฏิบัติ

ผลการใช้โมดูลได้ผลการศึกษา 5 ประการคือ ผลต่อความรู้ ทักษะ ทักษะ ความพึงพอใจ และความยั่งยืน

1. ผลการวัดความรู้ ด้านการประเมินภาคปฏิบัติ ก่อน และ หลังการใช้โมดูลการฝึกอบรมทันที พบว่าการฝึกอบรมทำให้ความรู้เพิ่มขึ้นกว่าก่อนการฝึกอบรมอย่างมีนัยสำคัญทางสถิติ ($p = .000$)

2. ผลการวัดทัศนคติด้านการประเมินภาคปฏิบัติ ก่อน และ หลังการใช้โมดูลการฝึกอบรมทันที พบว่า ค่าเฉลี่ยของคะแนนทัศนคติต่อการประเมินภาคปฏิบัติไม่แตกต่างกันที่ระดับนัยสำคัญ .05 สรุปได้ว่าโมดูลการฝึกอบรมไม่ส่งผลกระทบต่อเปลี่ยนแปลงของทัศนคติในระยะก่อน และหลังการฝึกอบรม

3. ผลการวัดทักษะด้านการประเมินภาคปฏิบัติหลังการใช้โมดูลต่อทักษะการฝึกอบรม จากการทดสอบสมมติฐานของ target contrast/ joint S.E. ของผู้ประเมินทั้ง 11 คน ไม่พบความแตกต่างของทักษะการประเมินในระยะก่อน และหลังการใช้โมดูลที่ระดับนัยสำคัญ .05 ($t = 0.08 - 1.82$, $p = .07 - .94$)

4. ผลการวัดความพึงพอใจหลังการใช้โมดูลการฝึกอบรม ผู้เข้ารับการฝึกอบรมมีความพึงพอใจมากที่สุด ด้านประโยชน์ของการฝึกอบรมต่อองค์กร ปริมาณ และคุณภาพของอาหาร และอาหารว่าง ประโยชน์ของการฝึกอบรมต่อตนเอง ความคุ้มค่ากับการเข้ารับการฝึกอบรม และความรู้ ความสามารถของผู้ดำเนินการฝึกอบรม และวิทยากร มีความพึงพอใจในระดับมากในประเด็นของ ความน่าสนใจของกิจกรรมการฝึกอบรม ความชัดเจนเข้าใจง่ายของเนื้อหาในการฝึกอบรม สถานที่ มีความสะดวก สบาย และเหมาะสมสำหรับการฝึกอบรมครั้งนี้ ความพร้อมของสื่อสโตนทัศน์ และมีความพึงพอใจในระดับปานกลางต่อประเด็นระยะเวลาการฝึกอบรมมีความเหมาะสม ผลการสัมภาษณ์กลุ่มเฉพาะ มีประเด็นสำคัญ 4 ประเด็นคือ

1) ความรู้ใหม่ที่ได้รับหลังการฝึกอบรม ผู้เข้ารับการฝึกอบรมกล่าวถึงความรู้ที่ได้รับว่าเป็นความรู้ใหม่ที่ทำให้

เข้าใจในหลักการประเมินภาคปฏิบัติที่ถูกต้องได้ลึกซึ้งมากขึ้น 2) ประโยชน์ของการฝึกทักษะการประเมิน บางส่วนของผู้เข้ารับการฝึกอบรมไม่แน่ใจว่าจะนำไปใช้ในสถานการณ์จริงได้ เนื่องจากอาจไม่พบพฤติกรรมที่ต้องการประเมินในระหว่างการฝึกปฏิบัติงานจริง บางส่วนกล่าวว่า จะลองนำความรู้ที่ได้รับ และรูบริคไปทดลองใช้ด้วยตนเอง 3) ทศนคติที่ดีต่อการประเมินภาคปฏิบัติ ผู้เข้ารับการฝึกอบรมกล่าวถึงทศนคติต่อการประเมินภาคปฏิบัติว่า คุณสมบัติสำคัญอย่างหนึ่งของอาจารย์พยาบาลคือการมีทศนคติที่ดีต่อการประเมิน การปฏิบัติงานของนักศึกษาพยาบาล หากมีทศนคติที่ไม่ดีก็ไม่ควรเลือกเป็นอาจารย์พยาบาล 4) สิ่งที่ต้องปรับปรุงในการฝึกอบรมคือ ควรใช้เวลาในการฝึกทักษะการประเมินให้สั้นกว่านี้โดยการปรับให้วิธีทศนคติ มีความกระชับมากกว่านี้

5. ผลของความยั่งยืนของความรู้ ทศนคติ และทักษะการประเมินภาคปฏิบัติหลังการใช้ โมดูลแบ่งได้ 3 ประเด็นดังนี้

5.1 ผลของการใช้โมดูลต่อความยั่งยืนของความรู้ หลังการใช้โมดูลการฝึกอบรม พบว่ามีความยั่งยืนของความรู้เนื่องจากไม่พบความแตกต่างของความแปรปรวนระหว่างหลังการฝึกอบรมทันที และหลังการฝึกอบรม 3 เดือนที่ระดับนัยสำคัญ .05 (mean difference = - 1.91, SE = 0.72, $p = .07$)

5.2 ผลของการใช้โมดูลต่อความยั่งยืนของทศนคติหลังการใช้โมดูลการฝึกอบรม พบว่าการใช้โมดูลไม่มีผลต่อการเปลี่ยนแปลงทศนคติ ทั้งก่อนการใช้ หลังการใช้ทันที และหลังการใช้ 3 เดือน ($F = 2.455, p = .111$)

5.3 ผลของการใช้โมดูลต่อความยั่งยืนของทักษะหลังการใช้โมดูลการฝึกอบรม พบว่าการใช้โมดูลไม่มีผลต่อการเปลี่ยนแปลงของทักษะการประเมินภาคปฏิบัติในผู้ประเมินเป็นส่วนใหญ่ แต่ในผู้ประเมินบางคนมีผลให้คุณภาพการประเมินลดลง ($t = 0.08 - -3.41$)

ตอนที่ 3 ผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล ต่อความคลาดเคลื่อนจากผู้ประเมิน ดังรายละเอียดต่อไปนี้

การสรุปผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลต่อความคลาดเคลื่อนจากผู้ประเมิน คือ การกด ปล่อยคะแนน (Severity/ Leniency effect) การให้คะแนนเกาะกลุ่มตรงกลาง (Central tendency effect) และความคลาดเคลื่อนแบบฮาลโล (Halo effect) จากการประเมินงานภาคปฏิบัติทั้งสิ้น 10 ชุด ในผู้ประเมิน 11 คน ในการ

วัด 2 ครั้งคือ ก่อน และหลังการฝึกอบรมทันที ผลการศึกษานำเสนอเป็น 4 ประเด็นคือ (1) ผลของโมดูลการฝึกอบรม ต่อความคลาดเคลื่อนแบบกต ปล่อยคะแนน (2) ผลของโมดูลการฝึกอบรม ต่อความคลาดเคลื่อนแบบให้คะแนนเกาะกลุ่มตรงกลาง (3) ผลของโมดูลการฝึกอบรม ต่อความคลาดเคลื่อนแบบฮาโล และ (4) ภาพรวมของผลของโมดูลการฝึกอบรมต่อความคลาดเคลื่อนทั้งหมด โดยประเด็นที่ 1 ถึง 3 เป็นการใช่วิธีระดับกลุ่ม และประเด็นที่ 4 เป็นการใช่วิธีระดับบุคคล ดังรายละเอียดต่อไปนี้

1. ผลของโมดูลฯ ต่อความคลาดเคลื่อนแบบกต ปล่อยคะแนนโดยศึกษาจากพาเซทของผู้ประเมิน การวัดครั้งที่ 1 ไม่พบความคลาดเคลื่อนแบบกต ปล่อยคะแนน (reliability of rater separation = 0.60, separation ratio = 1.23, Mean = 0.14, SD = 0.17)

การวัดครั้งที่ 2 ไม่พบความคลาดเคลื่อนแบบกต ปล่อยคะแนน (reliability of rater separation = 0.62, separation ratio = 1.28, Mean = 0.04, SD = 0.18)

เมื่อเปรียบเทียบผลต่างของการประเมินในการวัด 2 ครั้ง ไม่พบความแตกต่างของโลจิต การกต ปล่อยคะแนนที่ระดับนัยสำคัญ .05 ($t = .08 - 1.82, p = .07 - .94$) แสดงว่าการฝึกอบรมไม่ส่งผลต่อการกต ปล่อยคะแนน

2. ผลของโมดูลการฝึกอบรม ต่อความคลาดเคลื่อนแบบให้คะแนนเกาะกลุ่มตรงกลาง

การวัดครั้งที่ 1 พบความแตกต่างของความสามารถของผู้ถูกประเมิน (reliability of rater separation = 0.89, separation ratio = 2.92, Mean = 0.00, SD = 0.33) แสดงว่าไม่พบความคลาดเคลื่อนแบบให้คะแนนเกาะกลุ่มตรงกลาง

การวัดครั้งที่ 2 พบความแตกต่างของความสามารถของผู้ถูกประเมิน (reliability of rater separation = 0.91, separation ratio = 3.17, Mean = 0.00, SD = 0.39) แสดงว่าไม่พบความคลาดเคลื่อนแบบให้คะแนนเกาะกลุ่มตรงกลาง

เมื่อเปรียบเทียบผลต่างของการประเมินในการวัด 2 ครั้ง ไม่พบความแตกต่างของโลจิต ความสามารถของผู้ถูกประเมินที่ระดับนัยสำคัญ .05 ($t = .08 - 1.82, p = .07 - .94$) แสดงว่าการฝึกอบรมไม่ส่งผลต่อการรับรู้ของผู้ประเมินในระดับความสามารถของผู้ถูกประเมิน

3. ผลของโมดูลการฝึกอบรม ต่อความคลาดเคลื่อนแบบฮาลโล

การวัดครั้งที่ 1 พบความแตกต่างของความยากของสมรรถนะค่อนข้างต่ำ (reliability of trait separation = 0.51 separation ratio = 1.03, separation index = 1.71) แสดงว่ามีแนวโน้มจะเกิดความคลาดเคลื่อนแบบฮาลโล

การวัดครั้งที่ 2 พบความแตกต่างของความสามารถของผู้ถูกประเมินค่อนข้างต่ำ (reliability of ratee separation = 0.48, separation ratio = 0.95, separation index = 1.60) แสดงว่ามีแนวโน้มจะเกิดความคลาดเคลื่อนแบบฮาลโล

เมื่อเปรียบเทียบผลต่างของการประเมินในการวัด 2 ครั้ง ไม่พบความแตกต่างของโลจิทความยากของ สมรรถนะที่ระดับนัยสำคัญ .05 ($t = -0.36 - 1.41, p = .16 - .72$) แสดงว่าการฝึกอบรมไม่ส่งผลต่อการรับรู้ของผู้ประเมินต่อระดับความยากของสมรรถนะที่ประเมิน

4. ภาพรวมของผลของโมดูลการฝึกอบรมต่อความคลาดเคลื่อนทั้งหมด

4.1 การกตคะแนนของผู้ประเมิน ผู้ประเมินคนที่ 10 มีการปล่อยคะแนนมากที่สุด มีแนวโน้มจะให้คะแนน 3 บนมาตราการวัด 1- 4 มากที่สุด ผู้ประเมินคนที่ 3 มีการกตคะแนนมากที่สุด มีแนวโน้มจะให้คะแนน อยู่ระหว่าง 1 กับ 2 มากที่สุด ในขณะที่ผู้ประเมินคนที่เหลือมีการกตปล่อยคะแนนใกล้เคียงกัน และมีแนวโน้มจะให้คะแนนอยู่ระหว่าง 2 ถึง 3 มากที่สุด มีการกระจายการกตปล่อยคะแนนไม่มากนัก จึงไม่มีคุณลักษณะของการเกิดความคลาดเคลื่อนแบบกต ปล่อยคะแนน

4.2 ความสามารถของผู้ถูกประเมิน ตามการรับรู้ของผู้ประเมิน ผู้ถูกประเมินคนที่ 10 มีความสามารถมากที่สุด ผู้ถูกประเมินคนที่ 1, 3, 4, 5, 8, 9 และ 4 มีระดับความสามารถปานกลางใกล้เคียงกัน ผู้ถูกประเมินคนที่ 2 และ 7 มีระดับความสามารถน้อยใกล้เคียงกัน ผู้ถูกประเมินคนที่ 6 มีความสามารถน้อยที่สุด มีลักษณะการกระจายกันของความสามารถ จึงไม่มีคุณลักษณะของการเกิดความคลาดเคลื่อนแบบให้คะแนนเกาะกลุ่มตรงกลาง

4.3 ความยากของสมรรถนะ การรับรู้ของผู้ประเมินต่อความยากของสมรรถนะ พบว่าสมรรถนะที่ 2 ยากที่สุด และสมรรถนะที่ 6 ง่ายที่สุด ส่วนสมรรถนะอื่น ๆ มีความยากใกล้เคียงกัน โดยทุกสมรรถนะมีแนวโน้มได้รับคะแนนการประเมินระหว่าง 2 ถึง 3 คะแนนบนมาตรวัด 1 ถึง 4 มีแนวโน้มที่จะมีลักษณะของความคลาดเคลื่อนแบบฮาลโล

อภิปรายผลการวิจัย

การอภิปรายผลการวิจัย ตามวัตถุประสงค์การวิจัย คือ 1) ผลการพัฒนาโมดูลการฝึกอบรม เพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล 2) ผลการศึกษาผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล ต่อความรู้ ทักษะ และทักษะการประเมินภาคปฏิบัติ 3) ผลการศึกษาผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลต่อความคลาดเคลื่อนจากผู้ประเมิน

1. ผลการพัฒนาโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล

การพัฒนาโมดูลการฝึกอบรม เริ่มจากการสำรวจสภาพปัจจุบันในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล ร่วมกับการบูรณาการข้อมูลจากการศึกษา ตำรา เอกสาร งานวิจัยต่าง ๆ มาใช้ประกอบการตัดสินใจเพื่อวางแผนการออกแบบโมดูล ซึ่งจะช่วยให้การวางแผนงานสอดคล้องกับเป้าหมายมากขึ้น (Sarah & Charlene, 2010) โดยใช้ประโยชน์จากทั้งคำถามปลายปิด และคำถามปลายเปิด เนื่องจากข้อมูลจากคำถามปลายปิดไม่พบปัญหาในทางปฏิบัติที่ชัดเจน แต่เมื่อพิจารณาข้อมูลจากคำถามปลายเปิด และการสัมภาษณ์เชิงลึกทำให้ได้ข้อมูลของปัญหาชัดเจนขึ้น (Giorgetti, Prodanof, & Sebastiani, 2004) ปัญหาที่พบคือ ความเข้าใจไม่ตรงกันในเกณฑ์การประเมิน และการตัดสินผลการประเมิน ซึ่งมักเกิดจากความเป็นอัตนัยของผู้ประเมิน การพร่องความรู้ในหลักการวัดและประเมินผลที่ถูกต้อง นอกจากนี้ข้อมูลจากการสำรวจสภาพปัญหาพบว่า ผู้ให้ข้อมูลส่วนใหญ่ต้องการระบบการเตรียมความพร้อมก่อนทำการประเมินภาคปฏิบัติ ด้วยระบบขึ้นฝึกปฏิบัติงาน (Internship) กับครูพี่เลี้ยง ซึ่งเป็นรูปแบบที่มีการประเมินโดยผู้เชี่ยวชาญทำงานใกล้ชิดกับผู้ปฏิบัติงาน ทำให้เกิดความรู้ ทักษะ และความมั่นใจในระดับบุคคล (Taylor-Powell & Boyd, 2008) แต่อาจมีข้อจำกัดในด้านกำลังคนที่บางสถาบันไม่สามารถจัดให้ครูใหม่ขึ้นฝึกปฏิบัติงานพร้อมครูพี่เลี้ยงในการสอนภาคปฏิบัติแก่นักศึกษาได้ ดังการให้สัมภาษณ์ของผู้บริหารฝ่ายวิชาการของสถาบันใหม่แห่งหนึ่งกล่าวว่า “ระบบครูพี่เลี้ยงเป็นเรื่องในอุดมคติ” และวิธีที่ใช้ในการแก้ปัญหาจริงมาคือการฝึกอบรมผู้ประเมิน ซึ่งมีความสำคัญ และจำเป็นต่อการประเมินภาคปฏิบัติ เนื่องจากทำให้ผล

การประเมินมีความเที่ยงมากขึ้น และเป็นการปรับค่าความเที่ยงตรงของผู้ประเมินเพื่อให้สามารถประเมินโดยใช้มาตรฐานเดียวกันมีความเข้าใจตรงกันให้มากที่สุด (Johnson et al., 2009)

หลังการออกแบบโมดูลมีการทดลองใช้ และการใช้งานโมดูล และนำผลการใช้งานมาปรับปรุง เนื้อหาของการฝึกอบรม และวิธีทัศน์ให้กระชับ โมดูลที่ได้รับการพัฒนาแล้วประกอบด้วย หน่วยการเรียนรู้ 5 หน่วย คือ

- (1) ความรู้พื้นฐานเกี่ยวกับการประเมินภาคปฏิบัติ
- (2) การฝึกอบรมเกี่ยวกับความคลาดเคลื่อนจากผู้ประเมิน
- (3) กรอบของการอ้างอิง
- (4) การฝึกทักษะการสังเกต

(5) การฝึกทักษะการตัดสินใจ หน่วยการเรียนรู้ที่ 2 ถึง 5 เป็นโมดูลมาตรฐานที่ใช้ในการฝึกอบรมผู้ประเมิน โดยผู้วิจัยได้ปรับปรุงเนื้อหา และกระบวนการให้สอดคล้องกับบริบทของการวิจัยในครั้งนี้ ส่วนหน่วยการเรียนรู้ที่ 1 เป็นหน่วยการเรียนรู้ที่ผู้วิจัยบรรจุเนื้อหาของความรู้ที่จำเป็นเกี่ยวกับการวัด และประเมินผลการเรียนภาคปฏิบัติในทุกชั้นตอนเพื่อให้สอดคล้องกับความต้องการจากผลการสำรวจสภาพปัญหาที่กล่าวถึงการพร่องความรู้ด้านการวัด และประเมินผลภาคปฏิบัติ และกิจกรรมในโมดูลจะเน้นการฝึกทักษะด้านต่าง ๆ ซึ่งเป็นวิธีมาตรฐานที่ช่วยให้คุณภาพของการฝึกอบรมดีขึ้น (Wolfe & McVay, 2010)

องค์ประกอบในโมดูล แต่ละส่วนที่ได้รับการออกแบบมาให้พร้อมในการใช้งาน โดยแต่ละส่วนสามารถแยกนำไปปฏิบัติได้ สอดคล้องกับความหมายของโมดูลการฝึกอบรม (Cambridge dictionary online, 2016) และสอดคล้องกับการพัฒนาโมดูลมาตรฐานอื่น ๆ (National center on response to intervention, 2012) เช่น การฝึกอบรมในหน่วยการเรียนรู้เรื่อง ความคลาดเคลื่อนจากผู้ประเมิน จะมีเนื้อหาแยกเฉพาะ และมีกิจกรรมให้ทำท้ายบท หรือ หน่วยการเรียนรู้เรื่อง ตัวอย่างเช่น การฝึกอบรมเรื่องกรอบของการอ้างอิง สามารถจัดแยกออกมาฝึกอบรม เช่น การศึกษาของ George และคณะ (2013) ทำการฝึกอบรม ศัลยแพทย์ประจำโรงเรียนแพทย์เพื่อทำการประเมินแพทย์ประจำบ้าน (Resident) ในทักษะด้านศัลยกรรมทั่วไป ด้วยวิธี FOR ระยะเวลาการฝึกอบรมกลุ่มแรก 1 ชั่วโมง และ กลุ่มที่สอง 4 ชั่วโมง โดยการชมวิดีโอทัศน์ที่บันทึกไว้ขณะทำการผ่าตัดจริงจำนวน 10 ชุด และ ประเมินด้วยเครื่องมือประเมินแบบ มาตรฐานประมาณค่า 4 ระดับ ผลการวิจัยพบว่าหลังการฝึกอบรมผู้ประเมินมีร้อยละของความสอดคล้องระหว่างผู้ประเมิน (percentage agreement) ในระดับสูง

ระยะเวลาในการฝึกอบรมที่แตกต่างกันคือ 1 ชั่วโมง และ 4 ชั่วโมง ไม่มีผลต่อความสอดคล้องระหว่างผู้ประเมิน หรือการศึกษาของ Raczyski, Cohen, Engelhard และ Lu (2015) ทำการศึกษาเปรียบเทียบประสิทธิผลของวิธี FOR 2 วิธีคือ อบรมแบบกลุ่ม(Classroom training) ณ ศูนย์การฝึกอบรม และอบรมแบบเรียนรู้ด้วยตนเองแบบออนไลน์ (Self-paced training) ต่อความแม่นยำในการประเมินในการประเมินงานเขียน ผลการศึกษาพบว่าในกลุ่มที่ฝึกอบรมด้วยตนเองในระบบออนไลน์ ให้ความแม่นยำในการประเมินมากกว่ากลุ่มที่ฝึกอบรมเป็นกลุ่ม

การใช้โมดูลอาจใช้ร่วมกันในการฝึกอบรมครั้งเดียวกันได้ เช่น การศึกษาของ Eppich และคณะ (2015) มีวัตถุประสงค์เพื่อพัฒนา 1) พัฒนา โมดูลการฝึกอบรมผู้ประเมินเพื่อประเมินทักษะด้านการทำงานเป็นทีม และทักษะการติดต่อสื่อสาร 2) เพื่อศึกษาความเที่ยงระหว่างผู้ประเมิน โดยใช้วิธีทัศนสถานการณ์จำลอง ของการประเมินผู้ป่วยของแพทย์ ในการฝึกทักษะการประเมิน ที่มีระดับคุณภาพของทักษะที่ระดับต่ำ ปานกลาง และสูง และ 3) เพื่อศึกษาบทบาทของความเที่ยงระหว่างผู้ประเมินจากการสังเกตพฤติกรรมกรรมการประเมิน ผลการศึกษาพบว่า 1) การพัฒนาโมดูลการฝึกอบรมใช้แนวมาตรฐานของการประเมินที่มีความสำคัญระดับสูง (High- stakes assessments) มีโมดูลมาตรฐานจำนวน 4 เรื่องคือ RET, PDT, FOR และ BOT ระยะเวลาทั้งสิ้น 1 วัน ผู้ประเมินประกอบด้วยผู้เชี่ยวชาญ 9 ท่าน มีภูมิหลังเป็นกุมารแพทย์ 5 ท่าน พยาบาลผู้เชี่ยวชาญด้านเดียวกัน 3 ท่าน และนักวิจัยที่เกี่ยวกับด้านบริการสาธารณสุข อีก 1 ท่าน แบ่งช่วงการฝึกอบรมเป็น 3 ช่วงคือ ช่วงที่ 1 ทำการอบรมเรื่อง RET, PDT, FOR และ BOT เป็นระยะเวลา 4 ชั่วโมง ช่วงที่ 2 ใช้ระยะเวลาการฝึกอบรม 4 ชั่วโมง ผู้ประเมินทำการศึกษาคู่มือการฝึกอบรม สังเกตพฤติกรรมที่ต้องการประเมินจากวิธีทัศน โดยแยกกันทดลองประเมิน จำนวน 2 ชุด ระหว่างการฝึกประเมินสามารถปรึกษาทีมงานเกี่ยวกับพฤติกรรมที่ต้องสังเกตในการประเมินทักษะได้ ช่วงที่ 3 ใช้เวลา 2 ชั่วโมง โดยนำผลการประเมินที่ทำให้ช่วงที่ 2 รวมถึงวิธีทัศนทั้งหมด มาประเมินจริง รวมทั้งสิ้น 42 สถานการณ์ ทำการประเมิน 4 ช่วงเวลาคือ 1) ในการฝึกอบรมช่วงที่ 2 และ 2) ในการฝึกอบรมช่วงที่ 3 3) 4 สัปดาห์หลังการฝึกอบรม สุ่มเลือกวิธีทัศน 8 สถานการณ์ ทำการประเมิน และ 4) 16 สัปดาห์ประเมินวิธีทัศนทั้งหมด 42 สถานการณ์ ผลการศึกษาพบว่าสามารถพัฒนาโมดูลการฝึกอบรม และคู่มือการฝึกอบรม และศึกษาความเที่ยงระหว่างผู้ประเมินพบว่าหลังการฝึกอบรมอยู่ในระดับสูง และมีความยั่งยืนของสมรรถนะแม้เวลาผ่านไป 16 สัปดาห์หลังการฝึกอบรม

2. ผลการศึกษาผลของโมดูลการฝึกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล ต่อความรู้ ทักษะ และทักษะการประเมินภาคปฏิบัติ

2.1 ด้านความรู้ ผลการศึกษารูปร่างได้ว่า ค่าเฉลี่ยของคะแนนความรู้ด้านการประเมินภาคปฏิบัติหลังการใช้โมดูลการฝึกอบรม สูงกว่าค่าเฉลี่ยของคะแนนความรู้ก่อนการใช้โมดูลการฝึกอบรมอย่างมีนัยสำคัญทางสถิติ แสดงว่าโมดูลการฝึกอบรมทำให้ผู้เข้ารับการอบรมมีความรู้เพิ่มขึ้น สอดคล้องกับกระบวนการเรียนรู้ KAP model ของ Roger (1962 cited in Chien-Yun et al., 2011) ในการศึกษาของ Chien-Yun, Wan-Fei, Yu-His และ Chia-Hung ได้กล่าวถึงวัตถุประสงค์ของการฝึกอบรมว่ามีวัตถุประสงค์หลักคือ การได้รับความรู้ใหม่ๆเพิ่มขึ้นจะทำให้เกิด การเปลี่ยนแปลงทัศนคติ และเกิดการแสดงพฤติกรรมที่คาดหวัง (Chien-Yun et al., 2011) และ สอดคล้องกับการศึกษาของ Weberschock และคณะ (2005) ทำการฝึกอบรมการเป็นผู้ประเมินของเพื่อนนักศึกษาแพทย์ซึ่งมีการบรรลุเนื้อหาของเวชศาสตร์เชิงประจักษ์ (Evidence-based medicine) เข้าไว้ในหลักสูตรการเรียนเป็นครั้งแรกของเยอรมัน ตัวอย่างเป็นนักศึกษาแพทย์ชั้นปีที่ 3 ของมหาวิทยาลัย Johann Wolfgang Goethe ในประเทศเยอรมัน จำนวน 23 คน ได้เข้าร่วมการสัมมนาเพื่อเตรียมการเป็นผู้ประเมินเพื่อนนักศึกษาแพทย์ ใช้ระยะเวลาในการสัมมนา 4 สัปดาห์ สัปดาห์ละ 1 ครั้ง ๆ ละ 3 ชั่วโมง ผลการประเมินพบว่าตัวอย่างมีความรู้ และทักษะเพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ

ตัวอย่างการศึกษาอื่น ๆ เช่น การศึกษาของ Zhang, He, Wang, Zhang, Ding และ Hua (2015) ที่มีวัตถุประสงค์ของการศึกษาเพื่อประเมินอุบัติการณ์ การเกิดปัสสาวะเล็ด และนำเสนอวิธีการใหม่ในการให้ความรู้ด้านสุขภาพการเจริญพันธุ์ในชุมชน 17 อำเภอของเมืองเซียงไฮ้ของ ประเทศจีน ในตัวอย่างที่ไม่เคยมีการสำรวจปัญหาปัสสาวะเล็ดมาก่อน อายุระหว่าง 18 - 40 ปี จำนวน 2,100 คน ที่มีการสุ่มเข้ากลุ่มทดลอง 1,400 คน และกลุ่มควบคุม 700 คน กลุ่มควบคุมได้รับการสอนสุขศึกษาเรื่องสุขภาพวัยเจริญพันธุ์ตามปกติ กลุ่มทดลองได้รับวิธีการใหม่ในการให้สุขศึกษา ด้านสุขภาพการเจริญพันธุ์ และทำการเก็บข้อมูลด้วยแบบสอบถามก่อนการให้ความรู้ และหลังการให้ความรู้ 6 เดือนผลการศึกษาพบว่ากลุ่มทดลองมีคะแนนเฉลี่ยสูงกว่าก่อนได้รับสุขศึกษาอย่างมีนัยสำคัญทางสถิติที่ระดับ.001 มีทัศนคติเปลี่ยนแปลงไปด้านบวกอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 และมีพฤติกรรมปฏิบัติตัวเปลี่ยนแปลงภายหลัง 6 เดือนของการได้รับสุขศึกษาอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

2.2 ด้านทัศนคติ ผลการวัดทัศนคติต่อการประเมินภาคปฏิบัติ ก่อน และหลังการใช้โมดูล การฝึกอบรม พบว่าค่าเฉลี่ยของคะแนนทัศนคติด้านการประเมินภาคปฏิบัติ ก่อน และหลังการใช้ โมดูลไม่แตกต่างกัน ที่ระดับนัยสำคัญ .05 แสดงว่าการใช้โมดูลการฝึกอบรมไม่ทำให้ทัศนคติของผู้เข้า รับการอบรมเปลี่ยน การที่ทัศนคติไม่เปลี่ยน อาจเนื่องจากการจะต้องอาศัยปัจจัยแวดล้อมที่สำคัญคือ การได้รับความรู้ใหม่ ๆ ในลักษณะที่ซ้ำ ๆ หรือไม่กดดันจะทำให้เกิดทัศนคติในด้านบวก (Cacioppo, Petty, & Crites, 2012) เช่นการศึกษาของ Weberschock และคณะ (2005) หรือ การศึกษาของ Chien-Yun, Wan-Fei, Yu-Hsi และการศึกษาของ Fitriana, Madanijah และ Ekayanti รวมถึงการศึกษาของ Zhang, He, Wang, Zhang, Ding และ Hua ซึ่งเป็นการอบรมให้ ความรู้ในเรื่องใหม่ที่ตัวอย่างยังไม่เคยมีความรู้ในเรื่องนั้นมาก่อน (Chien-Yun, Wan-Fei, Yu-Hsi, & Chia-Hung, 2011; Fitriana, Madanijah, & Ekayanti, 2015; Zhang et al., 2015) จะทำให้เห็น การเปลี่ยนแปลงของทัศนคติที่ชัดเจน (Cacioppo et al., 2012) ในขณะที่การวิจัยในครั้งนี้มีเนื้อหา ทางทฤษฎีที่เป็นเรื่องใหม่สำหรับ ผู้เข้ารับการฝึกอบรมส่วนใหญ่ แต่ด้านทัศนคติต่อการประเมิน ภาคปฏิบัติ หรือทักษะการประเมินภาคปฏิบัติเป็นสิ่งที่ผู้เข้ารับการฝึกอบรมเคยประสบมาแล้ว การ อบรมโดยการให้ความรู้ใหม่สำหรับ ผู้เข้ารับการอบรมจึงมีทัศนคติเดิมเข้ามามีส่วนร่วมในการกำหนด ทิศทางของทัศนคติซึ่งในการวิจัยครั้งนี้ทัศนคติต่อการประเมินภาคปฏิบัติไม่ใช่เรื่องใหม่สำหรับผู้เข้า รับการอบรม

2.3 ด้านทักษะการประเมิน ผลการวัดทักษะการประเมินภาคปฏิบัติก่อน และหลังการ ใช้โมดูลการฝึกอบรม สรุปได้ว่าการฝึกอบรม ไม่มีผลต่อทักษะการประเมินภาคปฏิบัติ ซึ่งสอดคล้อง กับข้อมูลจากการสัมภาษณ์เฉพาะกลุ่มที่กล่าวถึงความชัดเจนของเกณฑ์ และพฤติกรรมของนักศึกษา ในชีวิตทัศน์ ประกอบกับผู้เข้ารับการฝึกอบรมทุกคนมีประสบการณ์ในการทำงานภาคปฏิบัติมาแล้ว ทั้งหมด สำหรับความชัดเจนของรูบรีค นั้นพบว่าสอดคล้องกับการศึกษาของ Barkaoui (2010) และ การศึกษาของ Joe, Harmes และ Hickerson (2011) ที่พบว่าหากมีการพัฒนารูบรีคให้มีความ ละเอียดทำให้เห็นความแตกต่างของสมรรถนะที่ต้องการประเมินในแต่ละระดับชัดเจน และเป็นรูบรีค ประเภทแยกส่วนจะทำให้ผลการประเมินมีความเที่ยงระหว่างผู้ประเมินสูง แม้ผู้ประเมินจะมี ประสบการณ์แตกต่างกันก็ตาม

3. ผลการศึกษาผลของโม่ตุลการฝีกอบรมเพื่อเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลต่อความคลาดเคลื่อนจากผู้ประเมิน

ผลของโม่ตุลการฝีกอบรมต่อความคลาดเคลื่อนจากการ กด ปล่อยคะแนน ต่อคลาดเคลื่อนจากการให้คะแนนเกาะกลุ่มตรงกลาง และ ความคลาดเคลื่อนแบบฮาโล พบว่าการใช้โม่ตุลการฝีกอบรมไม่มีผลต่อการเปลี่ยนแปลงคุณภาพการประเมินเมื่อเปรียบเทียบกับในระยะก่อน และหลังการใช้โม่ตุล และเมื่อตรวจสอบแล้วไม่พบความคลาดเคลื่อนแบบกด ปล่อยคะแนน ไม่พบความคลาดเคลื่อนแบบให้คะแนนเกาะกลุ่มตรงกลาง แต่มีแนวโน้มการเกิดความคลาดเคลื่อนแบบฮาโล เนื่องจากมีช่วงของการแบ่งระดับความยากค่อนข้างแคบ มีค่า reliability of separation และมีค่า separation index ค่อนข้างต่ำ แสดงถึงการประเมินที่มีการกระจายของระดับความยากต่ำ อาจเนื่องจากจำนวนงานที่ใช้ฝีกอบรมเมื่อเทียบผู้ประเมินมีจำนวนน้อยจึงมีผลกระทบต่อขนาดอิทธิพล (effect size) ของการฝีกอบรมต่อคุณภาพการประเมิน สอดคล้องกับการศึกษาของตัวอย่างเช่น การศึกษาของ Cook, Dupras, Beckman, Thomas และ Pankratz (2008) ทำการศึกษาผลของการฝีกอบรมผู้ประเมินต่อการประเมินทางคลินิกโดยใช้เครื่องมือมาตรฐาน Mini-CEX ทำการศึกษาในแพทย์ที่เป็นครูที่เลี้ยงในคลินิกจำนวน 52 คน แบ่งเป็น 3 กลุ่มคือ กลุ่มควบคุม กลุ่มเข้าฝีกอบรมเข้าไป 1 เดือนหลังการทำ pretest และกลุ่มที่เข้าอบรมทันทีหลังการทำ pretest แต่ทำ post test เข้าไป 1 เดือนหลังการฝีกอบรม ทำการฝีกอบรมเชิงปฏิบัติการด้วยวิธี RET, PDT, BOT และ FOR ผ่านการบรรยาย การชมวีดิทัศน์ และการอภิปรายกลุ่มระยะเวลา 1/2 วัน ทำการประเมินวีดิทัศน์ทักษะทางคลินิกของแพทย์ประจำบ้านกับผู้ป่วยมาตรฐานจำนวน 16 ชุด ผลการศึกษาพบว่าการใช้โม่ตุลการฝีกอบรมไม่ได้ช่วยทำให้ความเที่ยงระหว่างผู้ประเมิน หรือความแม่นยำในการประเมินดีขึ้น และพบการเกิดความคลาดเคลื่อนแบบฮาโลด้วย อาจเนื่องจากเลือกตัวอย่างที่มีจำนวนมาก แต่มีงานที่ใช้ประเมินจำนวนน้อย จึงทำให้มีขนาดอิทธิพลต่ำ (small effect)

เหตุผลอีกประการหนึ่งที่ไม่พบการเปลี่ยนแปลงของทักษะการประเมิน อาจเนื่องจากผู้เข้ารับการฝีกอบรมทุกคนมีประสบการณ์ในการทำงานบนหอผู้ป่วยไม่น้อยกว่า 3 ปี และคุ้นเคยกับการประเมินกิจกรรมการพยาบาลพื้นฐานที่ต้องปฏิบัติเป็นประจำ และการใช้รูบรีคที่มีคำอธิบายคุณลักษณะที่ต้องประเมินอย่างชัดเจนจะทำให้ได้ผลการประเมินที่มีคุณภาพ สอดคล้องกับคำสัมภาษณ์กลุ่มเฉพาะหลังการฝีกอบรมทันที ว่ารูบรีคมีความชัดเจน สอดคล้องกับวีดิทัศน์ที่ใช้ฝีก

ประเมิน และพบว่างานวิจัยที่พบว่าการฝึกอบรมช่วยลดความคลาดเคลื่อนจากผู้ประเมิน หรือเพิ่มคุณภาพการประเมินมักทำให้ผู้ไม่มีประสบการณ์ หรือเพื่อคัดเลือกผู้ประเมินที่มีคุณภาพเพียงพอเพื่อการประเมินที่มีผลกระทบมากต่อผู้ถูกประเมิน ซึ่งจำเป็นใช้ระยะเวลาในการฝึกในระดับหนึ่ง ตัวอย่างเช่น การศึกษาของ Davis (2015) ที่ทำการฝึกอบรมครูสอนภาษาอังกฤษจำนวน 20 คนที่มีประสบการณ์การสอน แต่ไม่มีประสบการณ์การประเมินแบบสอบของข้อสอบระดับมาตรฐานของ TOEFL iBT ด้านการพูด ทำการฝึกอบรม โดยศึกษาด้วยตัวเองจากระบบออนไลน์ จากตัวอย่างที่ดีของงานประเมินของผู้เชี่ยวชาญ ในระดับคะแนนรูบริคแบบภาพรวม (holistic) ที่มีตัวอย่างงานตั้งแต่ระดับคะแนน 1 ถึง 4 ทำความเข้าใจนิยามของคำอธิบายระดับคุณภาพของแต่ละระดับ จากนั้นจึงเข้าระบบการฝึกอบรมด้วยวิธี FOR และฝึกประเมิน 3 ครั้ง ครั้งละ 5 – 7 วัน กับงานจำนวน 100 ชุดในรูปวีดีโอคลิปในระบบออนไลน์ และทุกครั้งที่ทำการฝึกประเมินจะมีการเทียบเคียงผลการประเมินกับผลการประเมินของผู้เชี่ยวชาญ หลังการฝึกอบรมทำการตรวจสอบคุณภาพการประเมินด้วยโปรแกรม FACETS พบว่าผู้ประเมินทุกคนมีระดับการกด ปล่อยคะแนนที่สม่ำเสมอ และมีความเที่ยงในการประเมิน มีโลจิกการประเมินเป็นไปที่โมเดลคาดการณ์

ข้อเสนอแนะ

ข้อเสนอแนะสำหรับการนำผลการวิจัยไปใช้

1. สามารถนำโมดูลการฝึกอบรมมาใช้ได้ในสถานการณ์จริง เนื่องจากเครื่องมือที่ผ่านการพัฒนาขึ้นใช้โมดูลมีความสอดคล้องกับเกณฑ์การประเมินผลการเรียนรู้ที่สถานศึกษาพยาบาลในสังกัดสำนักงานคณะกรรมการการอุดมศึกษาทุกสถาบัน ต้องมีการจัดหลักสูตร และจัดการเรียนการสอนให้มีผลลัพธ์ทางการเรียนสอดคล้อง ตามกรอบมาตรฐานคุณวุฒิระดับอุดมศึกษา พ.ศ. 2565 เพื่อให้บัณฑิตในสาขาวิชาพยาบาลศาสตร์มีคุณภาพใกล้เคียงกัน ในการวิจัยครั้งนี้ได้นำเกณฑ์ดังกล่าวร่วมกับเอกลักษณ์ของมหาวิทยาลัยสวนดุสิตมาบูรณาการเข้าไปด้วย ซึ่งหากสถานศึกษาพยาบาลจะนำโมดูลนี้ไปใช้ อาจปรับเปลี่ยนตามเอกลักษณ์ของแต่ละสถานศึกษา หรือตามบริบทตามความเหมาะสม

2. การวิจัยครั้งนี้เป็นกรณีศึกษาสำหรับการพัฒนาโมดูลเสริมสร้างสมรรถนะการประเมินของแต่ละสถานศึกษาได้ เนื่องจากมีขั้นตอนการพัฒนาที่ชัดเจน และเป็นรูปธรรม

3. การนำผลการวิจัยไปใช้อาจต้องปรับเปลี่ยนความยาวของวิดีโอ และการแบ่งสมรรถนะที่ใช้ประเมินเป็นสมรรถนะย่อย ไม่จำเป็นต้องประเมินให้ครบถ้วนในเวลาเดียวกัน จะช่วยกระชับระยะเวลาในการฝึกอบรมไม่ให้เกิดความเหนื่อยล้าเกินไป

ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

1. จากผลการวิจัยที่พบว่าผู้ประเมินทุกคนมีประสบการณ์การปฏิบัติงานจริงบนหอผู้ป่วยมาแล้วจึงทำให้มีทักษะการประเมินที่คุ้นเคย ส่งผลให้ไม่เกิดการเปลี่ยนแปลงของทักษะการประเมินไม่ชัดเจน ดังนั้นควรมีการศึกษาวิจัยเพิ่มเติมในประชากรอาจารย์พยาบาลที่ไม่มีประสบการณ์ในการสอนทั้งภาคทฤษฎี และปฏิบัติ เปรียบเทียบกับพยาบาลที่มีประสบการณ์ในการปฏิบัติงานจริง หรือมีประสบการณ์ในการประเมินภาคปฏิบัติเพื่อพิสูจน์ว่าประสบการณ์ด้านการปฏิบัติงานบนหอผู้ป่วย หรือประสบการณ์การสอนภาคปฏิบัติ มีผลต่อทักษะการประเมินภาคปฏิบัติ

2. ควรมีการวิจัยในสถานการณ์จริงบนหอผู้ป่วยเพื่อประเมินประสิทธิภาพของโมดูลว่าสามารถนำมาใช้งานในทางปฏิบัติได้หรือไม่ โดยก่อน และหลังการฝึกอบรมอาจฝึกการประเมินในสถานการณ์จริง แทนการชมสื่อวิดีโอ ซึ่งเป็นข้อสังเกตที่ได้จากข้อมูลจากการสัมภาษณ์เฉพาะกลุ่ม ผู้เข้ารับการฝึกอบรม ตั้งคำถามถึงความเป็นไปได้ในการนำไปใช้ในสถานการณ์จริงที่ไม่สามารถควบคุมสถานการณ์ต่าง ๆ หรือจัดกระทำให้สถานการณ์ต่าง ๆ ลงตัวได้เช่นเดียวกับการฝึกอบรม นอกจากจะเป็นการทดสอบทักษะการประเมินแล้วยังเป็นการยืนยันคุณภาพของเครื่องมือที่ใช้ในการประเมินทักษะด้วยว่าสามารถนำมาใช้ได้จริงหรือไม่ มีปัญหาอุปสรรคใดในการใช้งาน

3. ในการพัฒนาโมดูลการฝึกอบรมครั้งต่อไปควรเพิ่มปริมาณของงานที่ใช้ในการประเมินเพื่อเป็นการเพิ่มขนาดของอิทธิพลของสิ่งที่ต้องการศึกษาว่าเพียงพอที่จะสนับสนุนประโยชน์ของการใช้โมดูล

4. รูปแบบการฝึกอบรมแบบเผชิญหน้าระหว่างวิทยากร กับผู้เข้ารับการฝึกอบรม จะมีการบริหารจัดการด้านทรัพยากร และด้านงบประมาณหลายประการ เช่น อาคารสถานที่ สื่อโสตทัศนูปกรณ์ อาหารว่าง อาหารกลางวัน การนัดหมายวันเวลาให้ตรงกัน ซึ่งอาจไม่สะดวกในทางปฏิบัติเนื่องจากอาจารย์พยาบาลแต่ละท่านต่างมีภาระหน้าที่ในวันเวลาที่แตกต่างกันไป ดังนั้นการพัฒนารูปแบบการฝึกอบรมในลักษณะออนไลน์ที่ทุกคนสามารถเข้ารับการอบรมตามวันเวลาที่ตนเองสะดวกจึงเป็นอีกทางเลือกหนึ่งที่ควรให้ความสนใจ ใน

ระยะแรกของการพัฒนาอาจต้องใช้งบประมาณในระดับหนึ่ง แต่เมื่อพัฒนาจนสำเร็จแล้ว คาดว่าน่าจะได้ผล การลงทุนที่คุ้มค่าต่อคุณภาพการประเมิน ที่จะเกิดขึ้น

5. การเสริมสร้างทักษะการประเมิน โดยการฝึกอบรมด้วยสื่อต่างๆ นอกเหนือจากวีดิทัศน์สถานการณ์จำลอง ซึ่งไม่เป็นปฏิริยาตอบสนองของผู้เรียนอย่างแท้จริง สื่อในรูปแบบอื่น ๆ เช่น การใช้สถานการณ์จำลองเป็นฐานในการฝึกประเมิน หรือใช้ผู้ปวยมาตรฐาน แทนการเขียนบทวีดิทัศน์ที่จะช่วยให้เห็นปฏิริยาที่สอดคล้องกับสมรรถนะของผู้เรียนมากขึ้น



รายการอ้างอิง

- Abdi, H. (2003). Factor rotations in factor analyses. *Encyclopedia for Research Methods for the Social Sciences*. Sage: Thousand Oaks, CA, 792-795.
- Airasian, P. W. (1991). *Classroom assessment*. New York: McGraw-Hill.
- Athey, T. R., & McIntyre, R. M. (1987). Effect of Rater Training on Rater Accuracy : Levels-of-Processing Theory and Social Facilitation Theory Perspectives. *Journal of Applied Psychology*, 72(4), 567-512.
- Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational behavior and human decision processes*, 50(2), 248-287.
- Barkaoui, K. (2010). Do ESL essay raters' evaluation criteria change with experience? A mixed-methods, cross-sectional study. *TESOL Quarterly*, 31-57.
- Basturk, R. (2008). Applying the many-facet Rasch model to evaluate PowerPoint presentation performance in higher education. *Assessment & Evaluation in Higher Education*, 33(4), 431-444.
- Bernardin, H. J. (1978). Effects of Rater Training on Leniency and Halo Error in Student Rating of Instructor. *Journal of Applied Psychology*, 63(3), 301-309.
- Bernardin, H. J., & Buckley, M. R. (1981). Strategies in Rater Training. *Academy of Management Review*, 6(2), 205-212.
- Bernardin, H. J. (1978). Effects of rater training on leniency and halo errors in student ratings of instructors. *Journal of Applied Psychology*, 63(3), 301-309.
- Bernardin, H. J., & Buckley, M. R. (1981). Strategies in rater training. *Academy of Management Review*, 6(2), 205-212.
- Bernardin, H. J., & Walter, C. S. (1977). Effects of rater training and diary-keeping on psychometric error in ratings. *Journal of Applied Psychology*, 62(1), 64.
- Borman, W. C. (1979). Format and Training Effects on Rating Accuracy and Rater Error. *Journal of Applied Psychology*, 64(4), 410-431.

- Bridge, P. D., & Sawilowsky, S. S. (1999). Increasing physicians' awareness of the impact of statistics on research outcomes: comparative power of the t-test and Wilcoxon rank-sum test in small samples applied research. *Journal of clinical epidemiology*, 52(3), 229-235.
- Burry-Stock, J. A., Shaw, D. G., Laurie, C., & Chissom, B. S. (1996). Rater agreement indexes for performance assessment. *Educational and Psychological Measurement*, 56(2), 251-262.
- Burry-Stock, J. A., Shaw, D. G., Laurie, C., & Chissom, B. S. (1996). Rater agreement indexes for performance assessment. *Educational and Psychological Measurement*, 56(2), 251-262.
- Cacioppo, J. T., Petty, R. E., & Crites, S. L. (2012). Attitude change. In R. V. S. (Ed.), *Encyclopedia of human behavior*. 1, 261-270.
- Calman, L., Watson, R., Norman, I., Redfern, S., & Murrells, T. (2002). Assessing practice of student nurses: methods, preparation of assessors and student views. *Journal of Advanced Nursing*, 38(5), 516-523.
- Cambridge dictionary online. (2016). Definition of module. [on line]. Available from: <http://dictionary.cambridge.org/dictionary/english/module> [23 June, 2016]
- Campbell, M., Julious, S., & Altman, D. (1995). Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. *BMJ: British Medical Journal*, 311(7013), 1145-1148.
- Campitelli, G., & Gobet, F. (2011). Deliberate Practice Necessary But Not Sufficient. *Current Directions in Psychological Science*, 20(5), 280-285.
- Chambers, M. A. (1998). Some issues in the assessment of clinical practice: a review of the literature. *Journal of Clinical Nursing*, 7(3), 201-208.
- Chapman, J. E., Sheidow, A. J., Henggeler, S. W., Halliday-Boykins, C. A., & Cunningham, P. B. (2008). Developing a measure of therapist adherence to

contingency management: An application of the many-facet Rasch model.

Journal of child & adolescent substance abuse, 17(3), 47-68.

Chien-Yun, D., Wan-Fei, C., Yu-Hsi, Y., & Chia-Hung, Y. (2011). A study on modification of knowledge, attitude and practice on vocational high school electronics courses integrated with nanotechnology concept. *International Journal of Thermal and Environmental Engineering*, 4(1), 73-79.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, New Jersey:: L: Erlbaum.

Compton, D. W. (2009). Managing studies versus managing for evaluation capacity building. *New Directions for Evaluation*, 2009(121), 55-69.

Cook, D. A., Dupras, D. M., Beckman, T. J., Thomas, K. G., & Pankratz, V. S. (2008). Effect of Rater Training on Reliability and Accuracy of Mini-CEX Scores: A Randomized, Controlled Trial. *Journal of general internal medicine*, 24(1), 74-79.

Cousins, R. D., Linnemann, J. T., & Tucker, J. (2008). Evaluation of three methods for calculating statistical significance when incorporating a systematic uncertainty into a test of the background-only hypothesis for a Poisson process. Nuclear Instruments and Methods in Physics Research Section A. *Accelerators, Spectrometers, Detectors and Associated Equipment*, 595(2), 480-501.

Davis, L. (2015). The influence of training and experience on rater performance in scoring spoken language. *Language Testing*, 33, 117-135.

Donaldson, J. H., & Gray, M. (2012). Systematic review of grading practice: Is there evidence of grade inflation? *Nurse Education in Practice*, 12(2), 101-114.

Dowell, J., Lynch, B., Till, H., Kumwenda, B., & Husbands, A. (2012). The multiple mini-interview in the UK context: 3 years of experience at Dundee. *Medical teacher*, 34(4), 297-304.

Eckes, T. (2009). Many-facet Rasch measurement. *Reference supplement to the manual for relating language examinations to the Common European*

Framework of Reference for Languages: Learning, teaching, assessment. [online] Available from: <http://www.coe.int/t/dg4/Linguistic/CEF-refSupp-SectionH.pdf> [24, May 2014]

- Ekegren, C., Hart, M., Brown, A., & Gabbe, B. (2016). Inter-rater agreement on assessment of outcome within a trauma registry. *Injury*, 47(1), 130-134.
- Eppich, W., Nannicelli, A. P., Seivert, N. P., Sohn, M. W., Rozenfeld, R., Woods, D. M., & Holl, J. L. (2015). A rater training protocol to assess team performance. *Journal of Continuing Education in the Health Professions*, 35(2), 83-90.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175-191.
- Feldman, M., Lazzara, E. H., Vanderbilt, A. A., & DiazGranados, D. (2012). Rater training to support high-stakes simulation-based assessments. *The Journal of continuing education in the health professions*, 32(4), 279.
- Field, A. (2012). Discovering statistics: repeated measures ANOVA. [On line] Available from: www.discoveringstatistics.com [11, December 2015]
- Fitriana, N., Madanijah, S., & Ekayanti, I. (2015). Analysis of Media Use in the Nutrition Education on Knowledge, Attitude and Practice of the Breakfast Habits on Elementary School Students. *Pakistan Journal of Nutrition*, 14(6), 335.
- Flinn, L., Braham, L., & Nair, R. (2015). How reliable are case formulations? A systematic literature review. *British Journal of Clinical Psychology*, 54(3), 266-290.
- Forjaz, M. J., Ayala, A., Martinez-Martin, P., Dujardin, K., Pontone, G. M., Starkstein, S. E., & Leentjens, A. F. (2015). Is the Parkinson anxiety scale comparable across raters? *Movement Disorders*, 30(4), 545-551.
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2-18.

- Gallagher, P., Smith, T., & Ousey, K. (2012). Problems with competence assessment as it applies to student nurses. *Nurse Education in Practice*, 12(6), 301-303
- García-Iriarte, E., Suarez-Balcazar, Y., Taylor-Ritzler, T., & Luna, M. (2010). A catalyst-for-change approach to evaluation capacity building. *American Journal of Evaluation*, 32: 168-182.
- George, B. C., Teitelbaum, E. N., DaRosa, D. A., Hungness, E. S., Meyerson, S. L., Fryer, J. P., & Zwischenberger, J. B. (2013). Duration of Faculty Training Needed to Ensure Reliable OR Performance Ratings. *Journal of Surgical Education*, 70(6), 703-708.
- Giorgetti, D., Prodanof, I., & Sebastiani, F. (2004). Automated Coding of Open-ended Surveys: Technical and Ethical Issues. *Instituto di Linguistiche Computazionale, CNR, Pisa Italy*.
- Gisev, N., Bell, J. S., & Chen, T. F. (2013). Interrater agreement and interrater reliability: key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy*, 9(3), 330-338.
- Gorman, C. A., & Rentsch, J. R. (2009). Evaluating frame-of-reference rater training effectiveness using performance schema accuracy. *Journal of Applied Psychology*, 94(5), 1336-1344.
- Graham, M., Milanowski, A., & Miller, J. (2012). Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings. [On line] Aviable from: <http://files.eric.ed.gov/fulltext/ED532068.pdf> [1 October, 2015]
- Graig, E. (2007). Building evaluation capacity: five models for funders. [On line] Aviable from: www.usablellc.net/White.../Blended%20Solutions.pdf [12 February, 2013]
- GÜLER, N. (2014). Analysis of open-ended statistics questions with many facet Rasch model. *Eurasian Journal of Educational Research*, 55, 73-90.
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, 8(1), 23.

- Hedge, J. W., & Kavanagh, M. J. (1988). Improving the accuracy of performance evaluations: Comparison of three methods of performance appraiser training. *Journal of Applied Psychology, 73*(1), 68.
- Hibbard, K. M. (1996). *Performance-Based Learning and Assessment. A Teacher's Guide*: ERIC.
- Holmboe, E. S., & Hawkins, R. (2008). Direct observation by faculty. *Practical guide to the evaluation of clinical competence. Philadelphia: Mosby-Elsevier*, 119-129.
- Huffman, D., Thomas, K., & Lawrenz, F. (2008). A Collaborative Immersion Approach to Evaluation Capacity Building. *American Journal of Evaluation, 29*(3), 358-368.
- Iramaneerat, C., & Myford, C. M. (2006). *Rater Effects in Clinical Performance Ratings of Surgery Residents*. Paper presented at the annual meeting of the American Educational Research Association, [on line]. Available from: <http://files.eric.ed.gov/fulltext/ED494158.pdf> [2, November 2015]
- Iramaneerat, C., & Yudkowsky, R. (2007). Rater errors in a clinical skills assessment of medical students. *Evaluation & the Health Professions, 30*(3), 266-283.
- Iramaneerat, C., Yudkowsky, R., Myford, C. M., & Downing, S. M. (2008). Quality control of an OSCE using generalizability theory and many-faceted Rasch measurement. *Advances in Health Sciences Education, 13*(4), 479-493.
- Ivancevich, J. M. (1979). Longitudinal Study of the Effects of Rater Training on Psychometric Error in Rating. *Journal of Applied Psychology, 64*(5), 502-508.
- James, L. R. (1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology, 67*(2), 219.
- James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology, 69*(1), 85.
- Jang, S. Y. (2010). *The development and evaluation of a systematic training program for increasing both rater reliability and rating accuracy*. Doctoral dissertation,

Department of educational and psychology, Collage of education, University of Illinois at Urbana-Champaign.

Joe, J. N., Harmes, J. C., & Hickerson, C. A. (2011). Using verbal reports to explore rater perceptual processes in scoring: A mixed methods application to oral communication assessment. *Assessment in Education: Principles, Policy & Practice*, 18(3), 239-258.

Johnson, R. L., Penny, J. A., & Gordon, B. (2009). *Assessing performance: Designing, scoring, and validating performance tasks*: Guilford Press.

Kane, M. (2011). The errors of our ways. *Journal of Educational Measurement*, 48(1), 12-30.

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs*. San Francisco: Berrett-Koehler.

Knoch, U., Read, J., & von Randow, J. (2007). Re-training writing raters online: How does it compare with face-to-face training? *Assessing Writing*, 12(1), 26-43.

Kobak, K. A., Lipsitz, J. D., & F., A. (2003). Development of a standardized training program for the Hamilton Depression Scale using internet-based technologies: results from a pilot study. *Journal of Psychiatric Research*, 37, 509-515.

Lancaster University. (2013). Evaluation capacity building in widening participation practice. [on line]. Aviable from: www.lancs.ac.uk/fass/.../capacitybuilding/index.php [10 February, 2013]

Lasater, K. (2011). clinical judgment: the last frontier for evaluation. *Nurse Education in Practice*(11), 86-92.

Latham, G. P., Wexley, K.N., Pursell E.D. (1975). Training managers to minimize rating errors in the observation of behavior. *Journal of Applied Psychology*, 60(5), 550-555.

Liao, S. C., Hunt, E. A., & Chen, W. (2010). Comparison between inter-rater reliability and inter-rater agreement in performance assessment. *Annals Academy of Medicine Singapore*, 39(8), 613-618

- Linacre, J. M. (2014). A user's guide to FACETS Rasch-model computer programs program manual 3.71.4 Beaverton. Winsteps.com [on line]. Available from: <http://www.winsteps.com/index.htm> [12 December 2015]
- Linacre, M. L. (1994). *Many-facet rasch measurement*. Chicago: Mesa.
- Lumley, T., & McNamara, T. F. (1995). Rater characteristics and rater bias: Implications for training. *Language Testing*, 12(1), 54-71.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods*, 1(1), 30.
- Murphy, K. R., Jako, R. A., & Anhalt, R. L. (1993). Nature and Consequences of Halo Error: A Critical Analysis. *Journal of Applied Psychology*, 78(2), 218-225.
- Myford, C. M., & Wolfe, E. W. (2000). Strengthening the ties that bind: Improving the linking network in sparsely connected rating designs. *ETS Research Report Series*, 2000(1), i-34. [on line]. Available from: http://www.ets.org/research/policy_research_reports/publications/report/2000/hsdy [16, August 2015]
- Myford, C. M., & Wolfe, E. W. (2003). Detecting and Measuring Rater Effects Using Many-Facet Rasch Measurement: Part I. *Journal of Applied Measurement*, 4(4), 386-422
- Myford, C. M., & Wolfe, E. W. (2004). Detecting and measuring rater effects using Many-facet Rasch measurement: part: II. *Journal of Applied Measurement*, 5(2), 189-227.
- National center on response to intervention. (2012). *Training module facilitator's guide* [on line]. Available from: <http://www.rti4success.org/sites/default/files/Training%20Module%20Facilitator's%20Guide.pdf> [3 June, 2016]
- Neck, C. P., & Barnard, A. W. (1996). Managing Your Mind: What Are You Telling Yourself? *Educational Leadership*, 53(6), 24-27.
- Nitko, A. J. (2007). *Educational assessment of students*: ERIC.

- Niveau, G., Lacasa, M. J., Berclaz, M., & Germond, M. (2015). Inter-rater Reliability of Criteria-Based Content Analysis of Children's Statements of Abuse. *Journal of forensic sciences, 60*(5), 1247-1252.
- Noonan, L. E., & Sulsky, L. M. (2001). Impact of frame-of-reference and behavioral observation training on alternative training effectiveness criteria in a Canadian military sample. *Human performance, 14*(1), 3-26.
- Palm, T. (2008). Performance assessment and authentic assessment: A conceptual analysis of the literature. *Practical Assessment, Research & Evaluation, 13*(4), 1-11.
- Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation, 29*(4) 443-459
- Program development and evaluation, University of Wisconsin, (2008). Extension Cooperative Extension Program Development and Evaluation. [on line]. Available from: <http://fyi.uwex.edu/programdevelopment/developing-programs/> [12, December 2015]
- Pulakos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. *Journal of Applied Psychology, 69*(4), 581-588.
- Raczynski, K. R., Cohen, A. S., Engelhard, G., & Lu, Z. (2015). Comparing the Effectiveness of Self-Paced and Collaborative Frame-of-Reference Training on Rater Accuracy in a Large-Scale Writing Assessment. *Journal of Educational Measurement, 52*(3), 301-318.
- Raymond, M. R., & Houston, W. M. (1990). *Detecting and Correcting for Rater Effects in Performance Assessment*. Paper presented at the the Annual Meetings of the American Educational [Online]. Available from: www.act.org/research/researchers/.../ACT_RR90-14.pdf [21 October, 2015]

- Research Association and the National Council on Measurement in Education, Boston, NA. [Online]. Available from: <http://files.eric.ed.gov/fulltext/ED336429.pdf> [1 February, 2015]
- Reynolds, C. R., Livingston, R. B., & Willson, V. L. (2009). *Measurement and assessment in education*. Upper Saddle River: Pearson Merrill Prentice Hall.
- Roberts, C., Rothnie, I., Zoanetti, N., & Crossley, J. (2010). Should candidate scores be adjusted for interviewer stringency or leniency in the multiple mini-interview? *Medical education, 44*(7), 690-698.
- Roch, S. G., & O'Sullivan, B. J. (2003). Frame of reference rater training issues: recall, time and behavior observation training. *International Journal of Training and Development, 7*(2), 93-107.
- Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczynska, U. (2012). Rater training revisited: An updated meta-analytic review of frame-of-reference training. *Journal of Occupational and Organizational Psychology, 85*(2), 370-395.
- Rudner, L. M. (1992). Reducing Errors Due to the Use of Judges. ERIC/TM Digest.
- Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the Ratings: Assessing the Psychometric Quality of Rating Data. *Psychological Bulletin, 88*(2), 413-428
- Sarah, F., & Charlene, W. (2010). Survey methods and practices. [Online]. Available from: www.statcan.gc.ca/pub/12-587-x/12-587-x2003001-eng.pdf. [2 January, 2016]
- Schaefer, E. (2008). Rater bias patterns in an EFL writing assessment. *Language Testing, 25*(4), 465-493.
- Sebok, S. S., Luu, K., & Klinger, D. A. (2014). Psychometric properties of the multiple mini-interview used for medical admissions: findings from generalizability and Rasch analyses. *Advances in Health Sciences Education, 19*(1), 71-84.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420.

- Steve, W., & Terry, P. (2009). Developing your training modules. [Online]. Available from: from www.excedge.com.au [13 July, 2016]
- Sudweeks, R. R., Reeve, S., & Bradshaw, W. S. (2004). A comparison of generalizability theory and many-facet Rasch measurement in an analysis of college sophomore writing. *Assessing Writing*, 9(3), 239-261.
- Tanner, C. (2006). Thinking like a nurse: A reserach-based model of clinical judgment in nursing. *Journal of nursing education*, 45(6), 204-211.
- Taut, S. (2007). Studying Self-Evaluation Capacity Building in a Large International Development Organization. *American Journal of Evaluation*, 28(1), 45-59.
- Taylor-Powell, E., & Boyd, H. H. (2008). Evaluation capacity building in complex organizations. *New Directions for Evaluation*, 2008(120), 55-69.
- Thornton III, G. C., & Zorich, S. (1980). training to improve observer accuracy. *Journal of Applied Psychology*, 65(3), 351-354.
- Wang, Z., & von Davier, A. A. (2014). Monitoring of Scoring Using the e-rater® Automated Scoring System and Human Raters on a Writing Test. *ETS Research Report Series*, (1), 1-21.
- Warmke, D. L., Billings, R.S. (1979). comparison of training methods for improving the psychometric quality of experimental and administrative performance ratings. *Journal of Applied Psychology*, 64(2), 124-131.
- Weberschock, T. B., Ginn, T. C., Reinhold, J., Strametz, R., Krug, D., Bergold, M., & Schulze, J. (2005). Change in knowledge and skills of Year 3 undergraduates in evidence- based medicine seminars. *Medical Education*, 39(7), 665-671.
- Weigle, S. C. (1998). Using FACETS to model rater training effects. *Language Testing*, 15(2), 263-287.
- Woehr, D. J. (1994). Understanding frame-of-reference training: The impact of training on the recall of performance information. *Journal of Applied Psychology*, 79(4), 525-534.

- Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, 67(3), 189-205.
- Wolfe, E. W., & Dobria, L. (2008). Best practices in quantitative method In J. W. Osborne (Ed.), *Applications of the multifacete rasch model*. Thousand Oaks, CA: Sage.
- Wolfe, E. W., & McVay, A. (2010). Rater effects as a function of rater training context. [on line] Aviable from: [2011, April 12]
- Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. [on line]. Aviable from, <http://www.rasch.org/rmt/rmt83b.htm>. [3 January, 2016]
- Wright, B. D., & Mok, M. M. (2004). An overview of the family of Rasch measurement models. *Introduction to Rasch measurement*, 1-24.
- Xi, X., & Mollaun, P. (2011). Using Raters From India to Score a Large-Scale Speaking Test. *Language Learning*, 61(4), 1222-1256.
- Yue, X. (2011). *Detecting rater centrality effect using simulation methods and Rasch measurement analysis*. Doctoral dissertation, Department of research and evaluation. Faculty of the Virginia Polytechnic institute and state university.
- Zhang, N., He, Y., Wang, J., Zhang, Y., Ding, J., & Hua, K.-q. (2015). Effects of a new community-based reproductive health intervention on knowledge of and attitudes and behaviors toward stress urinary incontinence among young women in Shanghai: a cluster-randomized controlled trial. *International urogynecology journal*, 1-9.
- เอมอร จังศิริพรปกรณ์. (2550). การวัดและประเมินผลการศึกษา. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์ มหาวิทยาลัย.
- คณะพยาบาลศาสตร์, มหาวิทยาลัยราชภัฏสวนดุสิต,. (2554). หลักสูตรพยาบาลศาสตรบัณฑิต.(อัตสำเนา)
- จีราภา คงสมฤทธิ์. (2544). การพัฒนาแบบประเมินการสอนภาคปฏิบัติในคลินิกของอาจารย์พยาบาล. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต, คณะศึกษาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์.

- ชวลิต ชูกำแหง. (2551). การประเมินการเรียนรู้ (พิมพ์ครั้งที่ 2). มหาสารคาม สำนักพิมพ์ มหาวิทยาลัยมหาสารคาม.
- นงลักษณ์ วิรัชชัย. (2555). วิธีการที่ถูกต้องและทันสมัยในการกำหนดขนาดตัวอย่าง. โครงการ Research Zone อาคารศูนย์การเรียนรู้ทางการวิจัย. [ออนไลน์]. แหล่งที่มา: <http://llskill.com/web/files/GPower.pdf> [20 พฤษภาคม 2559]
- นันทนัช อ่อนพวน. (2553). การพัฒนาคู่มือการสร้างรูปรिकเพื่อให้คะแนนสำหรับการประเมินการปฏิบัติงานของผู้เรียน. วิทยานิพนธ์ปริญญาโทบริหารบัณฑิต, สาขาการวัดและประเมินผล การศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์ มหาวิทยาลัย
- น้ำผึ้ง อินทเนตร, องอาจ นัยพัฒน์, ผจจจิต อินทสุวรรณ, และ สุทธิวรรณ พิรศักดิ์โสภณ. (2556). การศึกษาคูณลักษณะของคะแนนแบบทดสอบปลายเปิดวิชาคณิตศาสตร์เมื่อจำนวนผู้ตรวจ และรูปแบบการตรวจให้คะแนนต่างกัน โดยใช้โมเดลการสรุปอ้างอิงและโมเดลหลาย องค์ประกอบของราส์ซ. วารสารการวัดผลการศึกษา, 17(1), 131-142.
- บุญเชิด ภิญโญนนตพงษ์. (2544). การประเมินการเรียนรู้ที่เน้นผู้เรียนเป็นสำคัญ : แนวคิดและ วิธีการ. กรุงเทพฯ ฯ อัมรินทร์พรินต์ติ้ง.
- บุษวรรษ์ แสนปลื้ม. (2556). การศึกษาความสอดคล้องของการระบุความเข้มงวด/ ใจดี และการทำ หน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป ของผู้ตรวจให้คะแนนความสามารถใน การเขียนเรียงความของนักเรียนชั้นประถมศึกษาปีที่ 3. วารสารศิลปการศึกษาศาสตร์วิจัย, 5(2), 335-347.
- ปิยะรัตน์ หยกสุริยันต์. (2549). การพัฒนาแบบประเมินสมรรถนะของพยาบาลวิชาชีพ โรงพยาบาล ราชวิถี. วิทยานิพนธ์ปริญญาโทบริหารบัณฑิต, คณะพยาบาลศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2550). ทฤษฎีการทดสอบแนวใหม่. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์ มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2552). ทฤษฎีการทดสอบแบบดั้งเดิม. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์ มหาวิทยาลัย.
- สภาการพยาบาล. (2554). ข้อบังคับสภาการพยาบาลว่าด้วยการให้ความเห็นชอบหลักสูตรการศึกษา วิชาชีพการพยาบาลและการผดุงครรภ์ ระดับวิชาชีพพ.ศ. ๒๕๕๔. [ออนไลน์]. แหล่งที่มา: www.tnc.or.th/files/2012/06/page-1175/11__17002.pdf [25 พฤษภาคม 2556]

- สำนักงานคณะกรรมการการอุดมศึกษา. (2556). กรอบมาตรฐานคุณวุฒิระดับอุดมศึกษา (TQF: HEd).[ออนไลน์]. แหล่งที่มา: <http://www.mua.go.th/users/tof-hed/news/news6> [9 เมษายน, 2558]
- สุวิมล ว่องวานิช (บรรณาธิการ). (2546). รวมบทความการประเมินผลการเรียนรู้แนวใหม่. กรุงเทพฯ: ศูนย์ตำราและเอกสารทางวิชาการ คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- อนัญญา คูอาริยะกุล. (2554). การพัฒนารูปแบบการประเมินการฝึกภาคปฏิบัติสำหรับนักศึกษาพยาบาลสังกัดสถาบันพระบรมราชชนก. วิทยานิพนธ์ปริญญาคุษฎีบัณฑิต, สาขาวิจัยและประเมินผลการศึกษา ภาควิชาการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยนเรศวร.





ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาคผนวก ก

รายชื่อผู้ทรงคุณวุฒิ

ผู้ทรงคุณวุฒิกลุ่มที่ 1 ผู้ทรงคุณวุฒิภายในคณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิตในการสำรวจความต้องการในการใช้รูปรีค และตรวจสอบความตรงเชิงเนื้อหาของรูปรีคการประเมิน และบทวีดีทัศน์

มีคุณสมบัติดังนี้

1. เป็นผู้ที่มีประสบการณ์ในการเป็นอาจารย์พยาบาลทั้งภาคทฤษฎี และปฏิบัติอย่างน้อย 10 ปีขึ้นไป
2. จบการศึกษาด้านพยาบาลศาสตรบัณฑิต และมีระดับการศึกษาขั้นสูงสุดตั้งแต่ระดับมหาบัณฑิตขึ้นไป
3. เป็นบุคลากรของคณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิตที่มีระยะเวลาในการทำงานภายในคณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิตตั้งแต่ 3 ปีขึ้นไป
4. มีความยินดีให้ความอนุเคราะห์ในการมีส่วนร่วมแสดงความคิดเห็นในการสร้าง และตรวจสอบคุณภาพเครื่องมือ

1. อาจารย์ นิภา ลีสุคนธ์
อาจารย์ประจำสาขาวิชาการพยาบาลชุมชน
2. ผู้ช่วยศาสตราจารย์ ดร. ศรีสุดา วงศ์วิเศษกุล
หัวหน้าสาขาวิชาการพยาบาลชุมชน
3. ผู้ช่วยศาสตราจารย์ นางลักษณ์ เชษฐภักดีจิต
รองคณบดีฝ่ายบริหาร
4. พ.ท. หญิง วชิรา ไกรถิ่น*
อาจารย์ประจำสาขาวิชา การพยาบาลเด็ก และวัยรุ่น
5. อาจารย์ ธนิตา พุ่มท่าอิฐ *
อาจารย์ประจำสาขาวิชาการพยาบาลผู้ใหญ่ และผู้สูงอายุ
6. อาจารย์ ญัฐนาฏ เร้าเสถียร
อาจารย์ประจำสาขาวิชาสุขภาพจิต และการพยาบาลจิตเวช
7. อาจารย์ นิรามัย อูสาหะ*
อาจารย์ประจำสาขาวิชาการพยาบาลมารดา ทารกและการผดุงครรภ์

หมายเหตุ: * ผู้ทรงคุณวุฒิผู้ตรวจสอบความตรงเชิงเนื้อหาของรูปรีค สถานการณ์จำลอง และ บทวีดีทัศน์

ผู้ทรงคุณวุฒิกลุ่มที่ 2

จำนวน 9 ท่าน ทำการตรวจสอบคุณภาพเครื่องมือที่ใช้ในการวิจัย จำนวน 5 ฉบับคือ

- 1) แบบทดสอบความรู้ 2) แบบประเมินทัศนคติ 3) แบบสอบถามประเมินความพึงพอใจหลังการเข้ารับการฝึกอบรม 4) RUBRIC การประเมินภาคปฏิบัติ 5) คู่มือฝึกอบรมการเสริมสร้างสมรรถนะการประเมิน

มีคุณสมบัติดังนี้

1. สำเร็จการศึกษาระดับปริญญาตรีขึ้นไป
2. สำเร็จการศึกษาด้านการศึกษา หรือการวิจัย หรือการวัดและประเมินผลทางการศึกษาในระดับมหาบัณฑิต หรือดุษฎีบัณฑิต
3. สำเร็จการศึกษาด้านพยาบาลศาสตร์ในระดับปริญญาพยาบาลศาสตรบัณฑิต
4. มีความยินดีให้ความอนุเคราะห์ในการตรวจสอบคุณภาพเครื่องมือ

รายชื่อ

1. ดร. เบ็ญจา เตากล้า
คณบดี คณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต
2. ผู้ช่วยศาสตราจารย์ ดร. ประกาย จิโรจน์กุล
รองคณบดีฝ่ายวิชาการ คณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต
3. รองศาสตราจารย์ ดร.ยุวดี วัฒนานนท์
อาจารย์ประจำภาควิชาสูติศาสตร์ นรีเวชวิทยา มหาวิทยาลัยมหิดล
4. ผู้ช่วยศาสตราจารย์ ดร.ณัฐภรณ์ หลาวทอง *
อาจารย์ประจำสาขาการวัดและประเมินผลการศึกษา
ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
5. อาจารย์ ดร.ดวงเนตร ธรรมกุล
รองผู้อำนวยการกลุ่มงานวิจัย และวิชาการ วิทยาลัยพยาบาลบรมราชชนนีนนทบุรี
6. อาจารย์ ดร.รุ่งฤดี กล้าหาญ
อาจารย์ประจำสำนักทดสอบทางการศึกษา และจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ

7. รองศาสตราจารย์ ดร. โชติกา ภาชีผล
อาจารย์ประจำสาขา การวัดและประเมินผลการศึกษา
ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์ มหาวิทยาลัย
8. รองศาสตราจารย์ ดร. งามอาจ นัยพัฒน์ *
อาจารย์ประจำภาควิชาการวัดผลและวิจัยการศึกษา คณะศึกษาศาสตร์มหาวิทยาลัยศรีนครินทรวิโรฒ
9. นาวาอากาศเอกหญิง ดร. กาญจนา ค้ายาตี
รองผู้อำนวยการ กองพัฒนาระบบการบริหารทรัพยากรบุคคล กรมกำลังพลทหาร
กองบัญชาการกองทัพไทย

หมายเหตุ: * ผู้ทรงคุณวุฒิที่ไม่ได้มีคุณสมบัติตามข้อ 3



รายชื่อ ผู้ทรงคุณวุฒิกลุ่มที่ 3

ผู้ทรงคุณวุฒิจำนวน 5 ท่าน ทำการตรวจสอบคุณภาพการใช้สื่อวีดิทัศน์ร่วมกับการใช้รูปрик เพื่อกำหนดเป็นคะแนนมาตรฐานสำหรับใช้ในการฝึกการประเมินร่วมกันรูปริก

มีคุณสมบัติคือ

1. มีประสบการณ์การเรียนการสอนทั้งภาคทฤษฎี และภาคปฏิบัติในด้านพยาบาล ศาสตร์ มาไม่น้อยกว่า 10ปี
2. สำเร็จการศึกษาด้านพยาบาลศาสตร์ในระดับปริญญาพยาบาลศาสตรบัณฑิต
3. สำเร็จการศึกษาระดับปริญญาคุณวุฒิปริญญาขึ้นไป สำหรับผู้ทรงคุณวุฒิที่สำเร็จการศึกษาระดับปริญญาโทบัณฑิต จะต้องมีความรู้ด้านการเรียนการสอนภาคทฤษฎี และด้านปฏิบัติไม่น้อยกว่า 20 ปี
4. มีความยินดีให้ความอนุเคราะห์ในการตรวจสอบคุณภาพเครื่องมือ

รายชื่อผู้ทรงคุณวุฒิ

1. อาจารย์ ดร.สุภาวดี นพรุจจินดา
รองผู้อำนวยการ วิทยาลัยบรมราชชนนี สุพรรณบุรี
2. อาจารย์ นิรามัย อูสาหะ
อาจารย์ประจำสาขาวิชาการพยาบาลมารดาทารก และการผดุงครรภ์ คณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต
3. รองศาสตราจารย์ ดร.ยุวดี วัฒนานนท์
อาจารย์ประจำภาควิชาสูติศาสตร์ นรีเวชวิทยา มหาวิทยาลัย มหิดล
4. ผู้ช่วยศาสตราจารย์ ดร.มนสภรณ์ วิฑูรเมธา
อาจารย์ประจำสาขาวิชาการพยาบาลผู้ใหญ่และผู้สูงอายุ คณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต
5. อาจารย์ ดร. มยุรี บุญทัต
พยาบาลวิชาชีพชำนาญการ (ด้านการสอน)
ภาควิชาการพยาบาลบุคคลทุกช่วงวัย วิทยาลัยพยาบาลบรมราชชนนี ชัยนาท

ภาคผนวก ข

เครื่องมือที่ใช้ในการวิจัย

ตัวอย่างบางส่วนของเครื่องมือที่ใช้ในการวิจัย

1. คู่มือการฝึกอบรมการเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล

คำนำ

การประเมินภาคปฏิบัติของนักศึกษาพยาบาล เป็นกระบวนการในการเตรียมความพร้อมในการผลิตพยาบาลสู่สังคมให้มีคุณภาพ จากการศึกษาเอกสาร หลักฐานเชิงประจักษ์ รวมถึงการสัมภาษณ์ผู้บริหารฝ่ายวิชาการของสถานศึกษาพยาบาล และการพูดคุยแลกเปลี่ยนความคิดเห็นระหว่างอาจารย์พยาบาล พบว่าปัญหาของการประเมินภาคปฏิบัติที่พบเห็นได้บ่อยครั้ง และ เป็นปัจจัยสำคัญที่มีผลต่อคุณภาพการประเมินคือ ความเป็นอึดอัดของอาจารย์ผู้ประเมินภาคปฏิบัติ ดังนั้นจึงเป็นปัญหาที่ควรมีแนวทางการแก้ไขที่ชัดเจน การใช้คู่มือเล่มนี้แม้ว่าอาจไม่ใช่บทสรุปของการแก้ปัญหาที่เบ็ดเสร็จ แต่น่าจะเป็นจุดเริ่มต้นของการแก้ไขปัญหาที่เป็นรูปธรรมมากขึ้น และหวังเป็นอย่างยิ่งว่าการนำคู่มือนี้ไปปฏิบัติจะช่วยให้เกิดประโยชน์ต่อคุณภาพการประเมินภาคปฏิบัติ ในนิสิตนักศึกษาพยาบาลได้ระดับหนึ่ง

ณ โอกาสนี้ผู้เขียนขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. ศิริเดช สุขีวะ อาจารย์ที่ปรึกษา และ ผู้ช่วยศาสตราจารย์ ดร. กมลวรรณ ตังธนากานนท์ อาจารย์ที่ปรึกษาร่วม ที่ให้คำแนะนำให้การสนับสนุนและให้กำลังใจด้วยดีเสมอมา

เนตรรัชนี กมลรัตนานันท์

บทที่ 1 บทนำ

โปรแกรมเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาล ในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล เป็นการประยุกต์ใช้โมเดลสหวิชาการ ในการเสริมสร้างสมรรถนะการประเมิน ของ Preskill และ Boyley มีหัวใจสำคัญอยู่ที่ การใช้แนวคิดของการมีส่วนร่วมของผู้มีส่วนได้ส่วนเสีย และการประเมินกระแสหลัก (mainstreaming evaluation) ซึ่งเป็นการสร้างวัฒนธรรมการประเมินภายในองค์กรทำให้การประเมินเป็นส่วนหนึ่งของชีวิตการทำงานประจำวัน จนมีความกลมกลืนเข้ากับการทำงานจนคุ้นชิน และสู่เป้าหมายสูงสุด คือการสามารถสร้างการประเมินอย่างยั่งยืนภายในองค์กร ซึ่งวัดได้จากการเรียนรู้เกี่ยวกับการประเมิน การเก็บรวบรวมข้อมูล การวิเคราะห์ข้อมูล การแปลความหมายข้อมูล และการใช้ข้อมูลการประเมินเพื่อการตัดสินใจ และนำผลการตัดสินใจไปสู่การลงมือปฏิบัติเพื่อแก้ไขปัญหา หรือพัฒนาองค์กร โดยเลือกใช้กลยุทธ์ ในการดำเนินการคือ การฝึกอบรม (Training) การประชุม (Meeting) การเข้ามามีส่วนร่วมในกระบวนการประเมิน (Involvement in an evaluation process) และ การสอนงาน/ระบบพี่เลี้ยงและการให้คำปรึกษา (Mentoring /coaching) ซึ่งจะประกอบอยู่ในกิจกรรมท้ายบท

จุดประสงค์ของคู่มือ

คู่มือการเสริมสร้างสมรรถนะการประเมินสำหรับอาจารย์พยาบาลเป็นคู่มือสำหรับ

- ☞ อาจารย์ผู้สอนภาคปฏิบัติผู้ทำหน้าที่ประเมินภาคปฏิบัติของนักศึกษาพยาบาล
- ☞ ผู้บริหารฝ่ายวิชาการของสถาบันการศึกษาพยาบาลผู้ทำหน้าที่วางแผน และกำกับติดตามการดำเนินงานด้านการประเมินผลการเรียนรู้ในภาคปฏิบัติของนักศึกษาพยาบาล

มีวัตถุประสงค์คือ

- ☞ เพื่อให้อาจารย์ผู้สอนภาคปฏิบัติ มีความรู้ ความเข้าใจเกี่ยวกับขอบเขต และ เนื้อหาของความรู้พื้นฐานด้านการประเมินภาคปฏิบัติ
- ☞ เพื่อให้อาจารย์ผู้สอนภาคปฏิบัติ เกิดความรู้ความเข้าใจที่ตรงกัน และมีทักษะ ในการใช้วิธีการประเมินภาคปฏิบัติผ่านการฝึกทักษะการประเมินจาก ตัวอย่างงานภาคปฏิบัติ
- ☞ เพื่อให้อาจารย์ผู้สอนภาคปฏิบัติได้พัฒนาตนเองให้เกิดความเที่ยงสูงสุด ระหว่างผู้ประเมิน หลังการฝึกปฏิบัติผ่านงานภาคปฏิบัติ

คำชี้แจงการใช้คู่มือ

จากการสำรวจปัญหาการประเมินภาคปฏิบัติทางการศึกษาในสาขา ที่เกี่ยวข้องกับ วิทยาศาสตร์สุขภาพ รวมถึงการประเมินภาคปฏิบัติทางการศึกษาพยาบาล พบปัญหาคือ ความ ความคลาดเคลื่อนจากตัวผู้ประเมิน จากการใช้ความรู้สึกส่วนตัวของผู้ประเมิน ในการประเมิน ในที่นี้ ผู้ประเมินคืออาจารย์ผู้สอนภาคปฏิบัติ ซึ่งให้ผลการประเมินที่ต่างระดับคุณภาพกัน ในผู้เรียนคน เดียวกัน ทำให้เกิดปัญหาในการประเมินผลการเรียนรู้ และแนวทางในการแก้ปัญหาที่ผ่านการศึกษา มาอย่างต่อเนื่อง และเป็นที่ยอมรับ ในด้านการประเมินภาคปฏิบัติในทุกภาคส่วนคือ การเตรียม ผู้ ประเมิน เทคนิคที่นิยมใช้มากคือ การฝึกรวมผู้ประเมิน สำหรับแนวปฏิบัติในการใช้คู่มือนี้ ใช้เทคนิค การเตรียมผู้ประเมิน โดยประยุกต์ใช้โมเดลการเสริมสร้างสมรรถนะผู้ประเมิน สหวิชาการของPreskill และ Boyley โดยเลือกใช้เทคนิคการฝึกรวมผู้ประเมิน การประชุม การมีส่วนร่วมในกระบวนการ ประเมิน และ การสอนงาน/ระบบพี่เลี้ยง และ การให้คำปรึกษา

.....

คำแนะนำการใช้คู่มือ

ผู้ใช้คู่มือควรเริ่มต้นศึกษาจากบทนำไปตามลำดับที่ละบทเนื่องจากเนื้อหาของแต่ละบทมีความสอดคล้องกัน โดยเริ่มจากการปรับพื้นฐานความรู้ด้านการประเมินภาคปฏิบัติสำหรับอาจารย์พยาบาลที่ยังไม่มีประสบการณ์การสอนภาคปฏิบัติ ไปจนถึงการปรับความรู้ด้านการวัด และประเมินผลการเรียนการสอนภาคปฏิบัติสำหรับอาจารย์ผู้ที่มีประสบการณ์การสอนภาคปฏิบัติ จากนั้นจึงเป็นการทำความเข้าใจ และมีกิจกรรมฝึกทักษะเฉพาะในท้ายบทของแต่ละบท และ บทสุดท้ายจะเป็นการบูรณาการทักษะ ที่ได้ฝึกมาแต่ละทักษะ นำมาประยุกต์ใช้กับสถานการณ์เสมือนจริงผ่านงานภาคปฏิบัติ คือ วัตถุประสงค์สถานการณ์จำลองของนักศึกษาพยาบาลในการขึ้นฝึกปฏิบัติงานบนหอผู้ป่วย และ การปฏิบัติกิจกรรมการพยาบาลพื้นฐาน เนื่องจากเป็นกิจกรรมที่เป็นพื้นฐานของทุกสาขาของการพยาบาล ทำให้สื่อเข้าใจกันได้ไม่ว่าอาจารย์จะมีความเชี่ยวชาญในสาขาใดก็ตาม

บทที่ 2: ความรู้เบื้องต้น
เกี่ยวกับการประเมินภาคปฏิบัติ

จุดประสงค์

เมื่อสิ้นสุดการเรียนรู้ในบทนี้แล้วอาจารย์ผู้สอนภาคปฏิบัติสามารถ

1. อธิบายความหมายของการประเมินภาคปฏิบัติ และอธิบายนิยามสำคัญ ที่ใช้ในการประเมินผลการเรียนรู้ หลักการประเมินผลการเรียนรู้ กระบวนการประเมินภาคปฏิบัติ เครื่องมือที่ใช้ในการประเมินภาคปฏิบัติ การตรวจสอบคุณภาพ ของเครื่องมือ และ คุณภาพของผู้ประเมิน รวมถึงความรู้พื้นฐานในการประเมินภาคปฏิบัติของนักศึกษาพยาบาล
2. นำความรู้ ความเข้าใจที่ได้ ไปสู่การปฏิบัติการประเมินภาคปฏิบัติของนักศึกษาพยาบาล ได้เหมาะสมกับสถานการณ์

บทที่ 3
การสร้างความเข้าใจเรื่องความคลาดเคลื่อนจากผู้ประเมิน
(Rater error training)

จุดประสงค์การเรียนรู้เพื่อให้ อาจารย์ผู้สอนภาคปฏิบัติ

1. เข้าใจความหมาย คุณลักษณะ สาเหตุ และทราบดีวิธีแก้ไข ความคลาดเคลื่อนที่เกิดจากผู้ประเมิน
2. สามารถวินิจฉัยความคลาดเคลื่อนประเภทต่างๆ ที่เกิดจากผู้ประเมินในเบื้องต้นได้
3. สามารถนำความรู้ที่ได้ไปใช้ในการป้องกัน หรือ แก้ไขความคลาดเคลื่อนที่เกิดจากผู้ประเมินได้

.....

บทที่ 4
การฝึกอบรมกรอบของการอ้างอิง
(Frame of Reference training

จุดประสงค์การเรียนรู้เพื่อให้ อาจารย์ผู้สอนภาคปฏิบัติสามารถ

1. มีความรู้ความเข้าใจเกี่ยวกับความหมายคุณลักษณะของมิติของงานภาคปฏิบัติ และสร้างความคุ้นเคยกับมิติที่ใช้ในการประเมินภาคปฏิบัติ และคุณลักษณะของกรอบของการอ้างอิง
2. เกิดความเข้าใจที่ตรงกันเกี่ยวกับระดับคุณภาพของเครื่องมือระหว่าง ผู้ประเมิน และระดับคุณภาพตามมาตรฐาน

.....

บทที่ 5
การพัฒนาทักษะด้านการสังเกต
 (Behavioral observation training)

จุดประสงค์การเรียนรู้เพื่อให้ อาจารย์ผู้สอนภาคปฏิบัติ

1. มีความรู้ความเข้าใจเกี่ยวกับหลักการสังเกตพฤติกรรมกรรมการปฏิบัติงาน
2. สร้างทักษะการสังเกตพฤติกรรมกรรมการปฏิบัติงานได้

ความหมาย

การฝึกการสังเกตพฤติกรรมเป็นกระบวนการ สร้างความแม่นยำให้การประเมิน ด้วยกระบวนการคือ 1) การพัฒนาทักษะด้านการค้นหา 2) การรับรู้ และ 3) การทบทวนความจำต่อพฤติกรรมกรรมการปฏิบัติงาน

.....

บทที่ 6
การฝึกทักษะการตัดสินใจ
 (Decision making training)

จุดประสงค์การเรียนรู้เพื่อให้ อาจารย์ผู้สอนภาคปฏิบัติ

1. มีความรู้ความเข้าใจเกี่ยวกับทฤษฎีพื้นฐานที่ใช้ในการตัดสินใจ ในการประเมินภาคปฏิบัติ
2. สามารถนำแนวคิดพื้นฐานด้านการตัดสินใจไปใช้ในการประเมินผลการปฏิบัติภาคปฏิบัติได้

.....

2. รุบริกการประเมินภาคปฏิบัติตามลำดับของการพัฒนา

รุบริกภาคปฏิบัตินำเสนอผู้ทรงคุณวุฒิภายใน ครั้งที่ 1 ใช้คะแนน 5 ระดับ

นิยามของระดับคะแนน

ระดับคุณภาพที่ 1 below basic หมายถึง การที่นักศึกษาแสดงให้เห็นถึงความรู้ความเข้าใจ และมีทักษะในเรื่องนั้นอย่างจำกัด หรือน้อยมาก หรือไม่เข้าใจในหลักสำคัญหรือสาระสำคัญของเรื่องนั้น

ระดับคุณภาพที่ 2 basic หมายถึง การที่นักศึกษาแสดงให้เห็นถึงความรู้ความเข้าใจและมีทักษะในเรื่องนั้น ในสาระสำคัญบ้างบางส่วน และมีทักษะที่เป็นพื้นฐานโดยปฏิบัติได้ตามที่เห็นจากตัวอย่างมาสถานการณ์ที่ได้รับการฝึกทักษะนั้น

ระดับคุณภาพที่ 3 proficient หมายถึง การที่นักศึกษาแสดงให้เห็นถึงความรู้ความเข้าใจ และมีทักษะในเรื่องนั้นในระดับพื้นฐานที่ควรรู้ (meet standard) ในสาระสำคัญของเรื่องนั้น และมีทักษะที่เป็นพื้นฐานโดยปฏิบัติได้ตามที่เห็นจากตัวอย่างมาสถานการณ์ที่ได้รับการฝึกทักษะนั้น

ระดับคุณภาพที่ 4 accelerated หมายถึง การที่นักศึกษาแสดงให้เห็นถึงความรู้ความเข้าใจ และมีทักษะในเรื่องนั้นทั้งในเรื่องพื้นฐานที่ควรรู้ **ครอบคลุมสาระสำคัญและรายละเอียดอื่นๆอย่างครบถ้วน** และประยุกต์ใช้ความรู้และทักษะอื่นๆที่เกี่ยวข้องได้อย่างยืดหยุ่น (meet standard and use a variety of problem-solving)

ระดับคุณภาพที่ 5 advance หมายถึง การที่นักศึกษาแสดงให้เห็นถึงความรู้ความเข้าใจ และมีทักษะในเรื่องนั้นได้ **ครอบคลุมสาระสำคัญและรายละเอียดอื่นๆอย่างครบถ้วน** และประยุกต์ใช้ความรู้และทักษะอื่นๆที่เกี่ยวข้องได้อย่างยืดหยุ่น และ**คิดวิเคราะห์และตัดสินใจในการปฏิบัติ**ได้ตามหลักการและเหตุผลที่ถูกต้อง

นิยามระดับคุณภาพ 5 ระดับ ใช้สำหรับสมรรถนะข้อที่ 6-7

ระดับ1 dependent ปฏิบัติได้ด้วยตนเองน้อยมากต้องอาศัยการชี้แนะอย่างใกล้ชิด

ระดับ2 supervised ทำกิจกรรมได้เป็นบ้าง ต้องแนะเป็นบางส่วน

ระดับ3 Assisted ทำกิจกรรมได้ด้วยตนเองในสถานการณ์เฉพาะนั้นแต่ไม่สามารถนำไปประยุกต์ใช้ในสถานการณ์อื่นๆ ไม่มีความยืดหยุ่น

ระดับ4 Marginal ทำกิจกรรมได้ด้วยตนเอง เหมาะสมกับสถานการณ์นั้นและสถานการณ์ใกล้เคียง

ระดับ5 independent ทำกิจกรรมได้ด้วยตนเอง และยืดหยุ่นได้กับสถานการณ์ที่หลากหลายระดับ

3. แบบสำรวจความต้องการใช้บริการประเมินที่มีระดับการประเมิน 5 ระดับ 4 ระดับที่ตัดระดับสูงสุดทิ้ง และ 4 ระดับที่ตัดระดับต่ำสุดทิ้ง

บริการประเมินที่ทำโครงร่างเปรียบเทียบเพื่อการตัดสินใจของผู้ทรงภายใน ในการเลือกใช้ระดับคะแนน 5 ระดับ หรือ 4 ระดับแบบตั้งระดับคะแนนสูงสุดออก หรือตัดระดับคะแนนต่ำสุดออก

ตารางสรุปข้อเสนอแนะของผู้ทรงคุณวุฒิในการสำรวจความคิดเห็นครั้งที่ 2

คำชี้แจง ขอความอนุเคราะห์ท่านแสดงความคิดเห็นต่อข้อเสนอแนะต่าง ๆ ซึ่งผู้วิจัยได้รวบรวมในประเด็นสำคัญ จากข้อเสนอแนะจากผู้ทรงคุณวุฒิภายในคณะพยาบาลศาสตร์ จำนวน 7 ท่านในการปรับปรุงบริการจากแบบประเมินภาพรวมของการประเมินภาคปฏิบัติ และแบบประเมินภาคปฏิบัติการพยาบาลพื้นฐาน โดยทำเครื่องหมาย ✓ ในช่องที่ตรงกับความคิดเห็นของท่าน ตามเกณฑ์การพิจารณาดังนี้

1 หมายถึง เห็นด้วยกับความคิดเห็น

0 หมายถึง ไม่แน่ใจกับความคิดเห็น

-1 หมายถึง ไม่เห็นด้วยกับความคิดเห็น (กรณีไม่เห็นด้วย กรุณาให้ข้อคิดเห็นเพิ่มเติม)

ข้อเสนอแนะจากผู้ทรงคุณวุฒิภายในคณะพยาบาลศาสตร์ ในการสำรวจความคิดเห็นครั้งที่ 1	ความคิดเห็น			ข้อคิดเห็น หรือ ข้อเสนอแนะ
	1	0	-1	
1. ควรทบทวนนิยามของแต่ละสมรรถนะใหม่ และตั้งคำ สำคัญของแต่ละสมรรถนะมา หรือการตีความแต่ละ สมรรถนะให้ชัดเจน				
2. สมรรถนะที่ 1 มีความละเอียดมากเกินไป				
3. สมรรถนะที่ 2-7 เกณฑ์ไม่ค่อยตรงประเด็นของ สมรรถนะ ควรตัดออก				
4. ควรปรับเกณฑ์เน้นที่จุดสำคัญของแต่ละสมรรถนะ เช่น เกณฑ์ข้อ 1 แสดงพฤติกรรมสะท้อนคุณธรรมจริยธรรม				

ข้อเสนอแนะจากผู้ทรงคุณวุฒิภายในคณะพยาบาลศาสตร์ ในการสำรวจความคิดเห็นครั้งที่ 1			ความคิดเห็น			ข้อคิดเห็น หรือ ข้อเสนอแนะ
			1	0	-1	
เกณฑ์ข้อ 2 การควบคุมตนเอง เกณฑ์ข้อ 3 ยอมรับความแตกต่าง ความเชื่อและให้เกียรติ เกณฑ์ข้อ 4 วินัย ตรงต่อเวลา เกณฑ์ข้อ 5 จิตอาสา เกณฑ์ข้อ 6 รักษาจรรยาบรรณวิชาชีพ เกณฑ์ข้อ 7 เป็นแบบอย่างที่ดีด้านสุขภาพ เกณฑ์ข้อ 8 พินัยกรรมสิทธิของผู้ป่วย						
กรุณาใช้ตารางระดับคุณภาพด้านล่างนี้ในการประเมินข้อ 5						
ระดับมาตรฐาน เดิม	ระดับมาตรฐานที่ ตัดคะแนนต่ำสุด ที่	ระดับมาตรฐานที่ ตัดคะแนนสูงสุด ที่				
Below basic	ตัดระดับต่ำสุดที่	Below basic				
Basic	basic	Basic				
proficient	proficient	proficient				
Accelerated	Accelerated	Accelerated				
Advance	Advance	ตัดระดับสูงสุดที่				
5. ควรตัดระดับที่ 5 advance ที่เนื่องจากไม่สอดคล้องกับ สมรรถนะของ นศ. ปี 2 และขยับระดับ 4 ขึ้นมาเป็นระดับ						

ข้อเสนอแนะจากผู้ทรงคุณวุฒิภายในคณะพยาบาลศาสตร์ ในการสำรวจความคิดเห็นครั้งที่ 1	ความคิดเห็น			ข้อคิดเห็น หรือ ข้อเสนอแนะ
	1	0	-1	
5 ระดับ 1ขึ้นมาเป็นระดับ 2 เพิ่มระดับ 1 ที่มีระดับ below standard มากขึ้น (คะแนนค่อนข้างต่ำด้านลบ มากกว่า)				
6.ควรปรับลดระดับคุณภาพเป็น 4 ระดับ โดยตัด ระดับ 1 ออก แล้วไล่ระดับลดลงจาก 2-5 เป็น 1-4 (คะแนนค่อนข้าง ต่ำด้านบวกมากกว่า)				
7 เห็นด้วยกับจำนวนของระดับคุณภาพเดิม คือ 5 ระดับ				
8 เห็นด้วยกับระดับคุณภาพ 4 ระดับที่ตัดระดับต่ำสุดทิ้ง				
9 เห็นด้วยกับระดับคุณภาพ 4 ระดับที่ตัดระดับสูงสุดทิ้ง				

4. รุบริकरणประเมินฉบับสมบูรณ์ที่ผ่านการตรวจสอบคุณภาพแล้ว

ตัวอย่างรุบริकरणสมรรถนะที่ 1 และสมรรถนะที่ 7

คำชี้แจง แบบประเมินภาคปฏิบัติ โดยใช้รุบริकरणให้คะแนน ให้พิจารณาระดับคุณภาพของแต่ละเกณฑ์ตามสมรรถนะ 7 ข้อ เมื่อท่านพิจารณาคูณลักษณะของงานภาคปฏิบัติที่ทำการประเมินแล้ว ให้ทำเครื่องหมาย ✓ ลงใน ของแต่ละระดับคุณภาพของการประเมินที่ตรงกับความคิดเห็นของท่าน

สมรรถนะ/เกณฑ์ที่ต้องประเมิน	ระดับคุณภาพ			
	ต่ำกว่าพื้นฐาน(คะแนน 1)	พื้นฐาน(คะแนน 2)	มาตรฐาน(คะแนน 3)	สูงกว่ามาตรฐาน(คะแนน 4)
1.คุณธรรม จริยธรรม 1.1 ความซื่อสัตย์ รับผิดชอบมีวินัย	6. - แสดงพฤติกรรมสะท้อน ความซื่อสัตย์น้อยกว่า ร้อยละ 100 ของ เหตุการณ์ - มีความรับผิดชอบ มี วินัย น้อยกว่าร้อยละ 80 ของเหตุการณ์ทั้งหมดของ เวลาขึ้นฝึกกับอาจารย์ ผู้ประเมิน	7. - แสดงพฤติกรรม สะท้อน ความซื่อสัตย์ ร้อยละ 100 ของ เหตุการณ์ - มีความรับผิดชอบ มี วินัย ประมาณร้อยละ 80 ของเหตุการณ์ ทั้งหมดของเวลาขึ้นฝึก กับอาจารย์ผู้ประเมิน	8. - แสดงพฤติกรรมสะท้อน ความซื่อสัตย์ ร้อยละ 100ของเหตุการณ์ - มีความรับผิดชอบ มี วินัยประมาณร้อยละ 90 ของเหตุการณ์ทั้งหมด ของเวลาขึ้นฝึกกับ อาจารย์ ผู้ประเมิน	9. - แสดงพฤติกรรมสะท้อน ความซื่อสัตย์ ร้อยละ 100 ของเหตุการณ์ - มีความรับผิดชอบ มีวินัย ร้อยละ 100 ของเหตุการณ์ ทั้งหมดของเวลาขึ้นฝึกกับ อาจารย์ผู้ประเมิน
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<p>หมายเหตุ</p> <p>เกณฑ์ข้อ 1 หากพบว่าอยู่ในระดับต่ำกว่า 2 ถือว่าไม่ผ่านการพิจารณาการประเมินภาคปฏิบัติ</p> <p>ตัวอย่างพฤติกรรมสะท้อน ความซื่อสัตย์ รับผิดชอบ มีวินัย</p> <ul style="list-style-type: none"> - ความซื่อสัตย์ เช่น ไม่คัดลอกผลงานของผู้อื่น ไม่แจ้ง หรือไม่ใช้ข้อมูลอันเป็นเท็จ หรือไม่บิดเบือนข้อมูลอันจะทำให้เกิดอันตรายต่อผู้ป่วย - ความมีวินัย เช่น ขึ้นปฏิบัติงานตรงเวลา ส่งงานที่ได้รับมอบหมายตรงเวลา - ความรับผิดชอบ เช่น ปฏิบัติงานที่ได้รับมอบหมายเสร็จสิ้นครบถ้วน หรือ ไม่ละทิ้งหน้าที่ 				
1.2 ควบคุมตนเอง	1. ไม่สามารถควบคุมตนเองได้ในสถานการณ์ต่าง ๆ ประมาณร้อยละ 60 ของเหตุการณ์ขึ้นไป	2. ไม่สามารถควบคุมตนเองได้ในสถานการณ์ต่าง ๆ ประมาณร้อยละ 40 ของเหตุการณ์ขึ้นไป	3. ควบคุมตนเองได้ดี ในสถานการณ์ต่าง ๆ ประมาณร้อยละ 80 ของเหตุการณ์ขึ้นไป	4. ควบคุมตนเองได้ดี ในสถานการณ์ต่าง ๆ อย่างสม่ำเสมอ ประมาณร้อยละ 90 ของเหตุการณ์ขึ้นไป
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ตัวอย่าง การควบคุมตนเองในระดับคุณภาพ	1. เช่น สีหน้า แววตา หรือภาษาพูด ภาษา	2. เช่น สีหน้า แววตาหรือภาษาพูด	3. เช่น สีหน้าแววตา หรือภาษาพูด	4. เช่น สีหน้าแววตาหรือภาษาพูด ภาษากาย

สมรรถนะ/เกณฑ์ที่ต้องประเมิน	ระดับคุณภาพ			
	ต่ำกว่าพื้นฐาน(คะแนน 1)	พื้นฐาน(คะแนน 2)	มาตรฐาน(คะแนน 3)	สูงกว่ามาตรฐาน(คะแนน 4)
ต่างๆ	กาย ที่แสดงถึงการใช้อารมณ์อย่างรุนแรง อารมณ์ดีใจ เสียใจ หรือ โกรธเคืองอย่างชัดเจน หรือ ไม่สามารถตัดสินใจได้ว่า สถานการณ์นั้นดีหรือไม่ หรือตัดสินใจไม่ถูกต้องเหมาะสมกับสถานการณ์	ภาษากาย ที่แสดงถึงอารมณ์ ดีใจ เสียใจ หรือโกรธเคือง ลังเล ในการตัดสินใจว่า สถานการณ์นั้นดีหรือไม่ หรือ ตัดสินใจไม่ถูกต้อง เหมาะสมกับสถานการณ์	ภาษากาย ที่ไม่ได้สื่อให้เห็นว่ามีอารมณ์ โกรธ เสียใจ เศร้าใจ อย่างชัดเจน หรือ ตัดสินใจว่า สถานการณ์นั้นดีหรือไม่ ได้เหมาะสมกับสถานการณ์	ที่ไม่ได้สื่อให้เห็นว่ามีอารมณ์ โกรธ เสียใจ เศร้าใจ อย่างชัดเจน หรือ ตัดสินใจว่า สถานการณ์นั้นดีหรือไม่ ได้เหมาะสมกับสถานการณ์
1.3 ยอมรับความแตกต่างของความเชื่อของผู้ใช้บริการหรือญาติ	1. โต้แย้ง ความเชื่อโดยวาจา หรือท่าที่ไม่สุภาพ หรือ ไม่เหมาะสม	2. ลังเล ที่จะตัดสินใจว่าจะโต้แย้ง หรือ ยอมรับความเชื่อ หรือ ไม่แสดง การยอมรับความเชื่อ	3. ยอมรับ ความเชื่อด้วยท่าที่เป็นมิตร แต่ ไม่ได้ให้ข้อมูลที่เป็น ประโยชน์เพื่อช่วยในการตัดสินใจ	4. ยอมรับ ความเชื่อด้วยท่าที่เป็นมิตร โดย ให้ข้อมูลที่เป็นประโยชน์ เพื่อช่วยในการตัดสินใจ
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<p>ตัวอย่าง เช่น</p> <ul style="list-style-type: none"> - เมื่อพบความเชื่อของผู้ใช้บริการ หรือ ญาติที่ขัดแย้งกับการรักษา เช่น หลังผ่าตัดห้ามรับประทานไข่จะช่วยให้แผลหายช้า นักศึกษาโต้แย้งความเชื่อ หรือมีปฏิกิริยาโต้ตอบตามระดับคุณภาพ คือ - การโต้แย้งความเชื่อ เช่น คุณคิดอย่างนี้ไม่ถูกนะ เป็นความคิดที่ไร้สาระ ต้องทานซีดิงจะดี - การลังเลที่จะตัดสินใจ เช่น ไม่ทราบว่าคุณเอาความเชื่อนี้มาจากไหน ไม่แน่ใจว่าจะจริงหรือเปล่า หรือ นิ่งเฉยที่จะออกความคิดเห็น - การยอมรับความเชื่อแต่ไม่ได้ให้ข้อดีข้อเสีย เช่น พูดคุยด้วยสีหน้ายิ้มแย้ม ท่าที่เป็นมิตร ว่า คุณป่าเชื่ออย่างนั้นหรือคะ ก็เคยได้ยินคนไข้หลายคนบอกอย่างนี้เหมือนกันคะ แต่คุณป่าควรทานนะคะ - การยอมรับความเชื่อโดยให้ข้อมูลที่ เป็นประโยชน์ เช่น พูดด้วยสีหน้ายิ้มแย้ม เป็นมิตรว่า หรือคะคุณป่า เคยได้ยินบอยเหมือนกันคะ แต่คุณป่าคะในไข้มีสารอาหารจำพวกโปรตีน ซึ่งมีประโยชน์ในการเสริมสร้างเนื้อเยื่อที่ถูกทำลาย เช่นการผ่าตัด การได้รับโปรตีนน้อย จะทำให้การหายของแผลช้าลงไป ไม่ได้ทำให้การหายของแผลไม่ดี หรือเกิดแผลเป็นแน่นอนคะ เพราะการเกิดแผลเป็น เป็นกลไกการสร้างเนื้อเยื่อมาแทนที่เนื้อเยื่อที่ถูกทำลายโดยธรรมชาติอยู่แล้ว 				
1.4 จิตอาสา	1. แสดงตนว่า ไม่มี ความพร้อม หรือ ปฏิเสธ การร้องขอในการทำหน้าที่จากกลุ่มในการทำหน้าที่เพื่อส่วนรวม	2. แสดงท่าที ลังเล เมื่อมีการร้องขอจากกลุ่มในการทำหน้าที่เพื่อส่วนรวม	3. แสดงท่าที เต็มใจ และ พร้อม เมื่อกลุ่มเลือกให้ทำหน้าที่เพื่อส่วนรวม	4. อาสา ตนเอง เมื่อมีการร้องขอจากกลุ่มในการทำหน้าที่เพื่อส่วนรวม
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<p>ตัวอย่าง เช่น</p> <ul style="list-style-type: none"> - เมื่ออาจารย์เสนอให้มีการคัดเลือกหัวหน้ากลุ่มเพื่อทำการประสานงานกับอาจารย์ในการฝึกปฏิบัติงาน หรือในการสั่งงาน หรือ ส่งงาน นักศึกษาริบอาสาตนเอง หรือพยายามหลบหลีก 				

สมรรถนะ/เกณฑ์ที่ต้องประเมิน	ระดับคุณภาพ			
	ต่ำกว่าพื้นฐาน(คะแนน 1)	พื้นฐาน(คะแนน 2)	มาตรฐาน(คะแนน 3)	สูงกว่ามาตรฐาน(คะแนน 4)
- ช่วงพักรับประทานอาหารกลางวัน นักศึกษา อาสาสมัครติดชอบงานแทนเพื่อน ในระหว่างที่เปลี่ยนให้เพื่อนไปรับประทานอาหารกลางวัน หรือหลบหลีกที่จะทำหน้าที่แทนเพื่อน				
1.5 รักษาจรรยาบรรณวิชาชีพ พินิจลักษณะของผู้ให้บริการ	1. ละเลย หรือ ไม่ตระหนักถึงการรักษาจรรยาบรรณวิชาชีพ หรือ ไม่ให้ข้อมูลที่จำเป็น	2. ร ัก ช า จรรยาบรรณวิชาชีพ เมื่อมีการร้องขอหรือเตือน แต่ ให้ข้อมูลที่จำเป็น น้อยมาก	3. ร ัก ช า จรรยาบรรณวิชาชีพ และให้ข้อมูลเฉพาะที่จำเป็นแต่ยังขาดรายละเอียด	4.รักษาจรรยาบรรณวิชาชีพ และให้ข้อมูลที่จำเป็น และรายละเอียดอย่างครอบคลุม
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ตัวอย่าง - การตระหนักถึงการรักษาจรรยาบรรณหรือให้ข้อมูลที่จำเป็น เช่น แนะนำตนเอง แจ้งว่าจะทำการพยาบาลอะไร บอกเพื่อวัตถุประสงค์ในการทำการพยาบาล กั้นม่านขณะให้การพยาบาล หรือ ขณะอาบน้ำบนเตียงถอดเสื้อผ้าผู้ใช้บริการออกหมด โดยห่ออะไรมาปกปิดส่วนที่ยังไม่ได้ให้การพยาบาลแม้จะปิดม่านอยู่ - การปกป้องสิทธิของผู้ใช้บริการ เช่น ให้ข้อมูลที่จำเป็น เกี่ยวกับสิทธิผู้ใช้บริการ ให้ผู้ใช้บริการทราบ - รักษาจรรยาบรรณและหรือ ให้ข้อมูลที่จำเป็น เช่น แนะนำตนเอง แจ้งวัตถุประสงค์การปฏิบัติ หรืออธิบายขั้นตอนการปฏิบัติเพื่อให้ผู้ป่วยปฏิบัติตัวได้ถูกต้องขณะให้การพยาบาล				
2. ความรู้ เกณฑ์ 2.1 บอกเหตุผล และหลักการสำคัญของการปฏิบัติกิจกรรมการพยาบาลต่างๆ	1. บอกเหตุผล และหลักการสำคัญของการปฏิบัติไม่ได้ หรือ ได้น้อยกว่าร้อยละ 60 ของเนื้อหาของหลักการนั้น	2. บอกเหตุผล และหลักการสำคัญของการปฏิบัติได้ประมาณร้อยละ 60 ของเนื้อหาขึ้นไป	3. บอกเหตุผล และหลักการสำคัญของการปฏิบัติได้ถูกต้องประมาณ ร้อยละ80 ของเนื้อหาขึ้นไป	4. บอกเหตุผล และหลักการสำคัญของการปฏิบัติได้ถูกต้องประมาณร้อยละ90 ของเนื้อหาขึ้นไป
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ตัวอย่าง เมื่อจะให้ยาฉีดเข้ากล้ามเนื้อ นักศึกษาจะต้องบอกหลักการสำคัญของการฉีดยาเข้ากล้ามเนื้อ เช่น ต้องทราบหลัก 6 R ตามหลักการมาตรฐานของการบริหารยา ต้องทราบหลักการให้ยาฉีดทางกล้ามเนื้อ คือ เลือกใช้กระบอกฉีดยาให้เหมาะกับปริมาณยา เลือกใช้เข็มในการดูดยาและฉีดยาได้ถูกต้อง ทราบตำแหน่งบนกระบอกฉีดยาที่ห้ามจับ เลือกตำแหน่งของกล้ามเนื้อที่เหมาะสมกับปริมาณยา และ บอกองศาในการปักเข็มฉีดยาได้ถูกต้องเหมาะสมกับสถานการณ์จริง				
2.2 มีส่วนร่วมในการอภิปรายแสดงถึงการมีความรู้ความเข้าใจในสาระสำคัญของศาสตร์ที่เป็นพื้นฐานชีวิตและวิทยาศาสตร์สุขภาพ	1. ไม่มีส่วนร่วมในการอภิปราย หรือมีส่วนร่วมน้อยกว่า ร้อยละ 60 ของเหตุการณ์ หรือ อภิปรายโดยไม่มีอ้างอิง สาระสำคัญของ	2. มีส่วนร่วมในการอภิปรายประมาณร้อยละ 60 ขึ้นไปของเหตุการณ์ และ อภิปรายโดยอ้างอิงสาระสำคัญของศาสตร์ที่เป็นพื้นฐานของชีวิต	3. มีส่วนร่วมในการอภิปรายประมาณร้อยละ 80 ขึ้นไปของเหตุการณ์ และ อภิปรายโดยอ้างอิง สาระสำคัญของศาสตร์ที่เป็นพื้นฐานของชีวิต	2. มีส่วนร่วมในการอภิปรายทุกครั้ง หรือ ประมาณร้อยละ 90 ของเหตุการณ์ และ อภิปรายโดยอ้างอิง สาระสำคัญของศาสตร์ที่เป็นพื้นฐานของชีวิต วิทยาศาสตร์สุขภาพ

สมรรถนะ/เกณฑ์ที่ต้องประเมิน	ระดับคุณภาพ			
	ต่ำกว่าพื้นฐาน(คะแนน 1)	พื้นฐาน(คะแนน 2)	มาตรฐาน(คะแนน 3)	สูงกว่ามาตรฐาน(คะแนน 4)
และศาสตร์ทางวิชาชีพพยาบาล	ศาสตร์ที่เป็นพื้นฐานของชีวิต วิทยาศาสตร์สุขภาพ และ ศาสตร์ทางวิชาชีพพยาบาล	วิทยาศาสตร์สุขภาพ และ ศาสตร์ทางวิชาชีพพยาบาลได้ ถูกต้อง ประเมินร้อยละ 60 ขึ้นไปของเนื้อหา	วิทยาศาสตร์สุขภาพ และ ศาสตร์ทางวิชาชีพพยาบาลได้ ถูกต้อง ประเมินร้อยละ 80 ขึ้นไปของเนื้อหา	และ ศาสตร์ทางวิชาชีพพยาบาลได้ ถูกต้อง ประเมินร้อยละ 90 ขึ้นไปของเนื้อหา
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ตัวอย่าง ช่วงการทำกิจกรรม quick round/ pre-post conference หรือ clinical teaching นักศึกษาอภิปรายโดยอ้างอิงสาระสำคัญของพื้นฐานด้านวิทยาศาสตร์สุขภาพ และด้านวิชาชีพพยาบาล ที่ถูกต้อง เช่น นักศึกษาตอบว่าคนไข้มีไข้จะต้องเช็ดตัวลดไข้เมื่ออุณหภูมิตั้งแต่ 38.5 องศาเซลเซียสขึ้นไป โดยเช็ดย่อนรอยรุ่มขน จากส่วนปลายเข้าหาลำตัวและเช็ดหยุดตามข้อพับต่างๆที่เป็นจุดที่มีเส้นเลือดมาเลี้ยงมาก ซึ่งนักศึกษาได้ใช้ความรู้ด้านวิทยาศาสตร์สุขภาพ และการพยาบาลพื้นฐานมาอภิปราย				
3. ทักษะทางปัญญา <u>เกณฑ์</u> 3.1 ระบุจุดแข็ง จุดอ่อน และสิ่งที่ต้องพัฒนาตนเองเกี่ยวกับความรู้ และ ทักษะการปฏิบัติงาน	1. ไม่สามารถวิเคราะห์จุดแข็ง จุดอ่อนได้ตรงตามความเป็นจริง และไม่สามารถบอกแนวทางที่เป็นรูปธรรมในการพัฒนาตนเองเกี่ยวกับความรู้ และ ทักษะการปฏิบัติงานได้ จำเป็น ต้อง ได้รับ คำแนะนำเพิ่มเติมทั้งหมด หรือ ตั้งแต่ ร้อยละ 0 ขึ้นไปของประเด็นนี้	2. วิเคราะห์จุดแข็ง จุดอ่อนของตนเองได้ ตรงตามความเป็นจริง ได้เล็กน้อย และสามารถบอกแนวทางที่เป็นรูปธรรมในการพัฒนาตนเองเกี่ยวกับความรู้ และทักษะการปฏิบัติงานได้ ประมาณ ร้อยละ 60 ขึ้นไปของประเด็นนี้ (ควรได้รับ) คำแนะนำเพิ่มเติม ประมาณร้อยละ 40 ของเนื้อหา	3. สามารถวิเคราะห์จุดแข็ง จุดอ่อนของตนเองตามความเป็นจริงเป็นส่วนใหญ่ และสามารถบอกแนวทางที่เป็นรูปธรรมในการพัฒนาตนเองเกี่ยวกับความรู้ และ ทักษะการปฏิบัติงานได้ ประมาณร้อยละ 80 ขึ้นไปของประเด็นนี้ ควรได้รับคำแนะนำเพิ่มเติม ประมาณร้อยละ 20 ของเนื้อหา	4. สามารถวิเคราะห์จุดแข็ง จุดอ่อนของตนเองตามความเป็นจริงได้ และสามารถบอกแนวทางที่เป็นรูปธรรมในการพัฒนาตนเองเกี่ยวกับความรู้ และ ทักษะการปฏิบัติงานได้ ประมาณร้อยละ 90 ขึ้นไปของประเด็นนี้ ควรได้รับ คำแนะนำเพิ่มเติม ประมาณร้อยละ 10 ของเนื้อหา หรือไม่จำเป็น ต้อง ได้รับ คำแนะนำ
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ตัวอย่าง นักศึกษาคนหนึ่ง มักทำของปราศจากเชื้อปนเปื้อนบ่อยๆ และเมื่อจะให้การพยาบาลตามกิจกรรมการพยาบาลพื้นฐานนักศึกษาสามารถตอบหลักการสำคัญของกิจกรรมนั้นๆได้ เช่น บอกหลักการสำคัญของบริหารยา การให้ยากิน ยาฉีด เมื่อเปิดโอกาสให้นักศึกษาสะท้อนคิดเกี่ยวกับเรื่องจุดแข็งจุดอ่อนของตนเอง หรือ จุดที่ต้องพัฒนา นักศึกษาสามารถบอกได้ว่าตนเองมีจุดแข็งจุดอ่อนอย่างไร และ ทราบว่าจะพัฒนาตนเองอย่างไร				

สมรรถนะ/เกณฑ์ที่ต้องประเมิน	ระดับคุณภาพ			
	ต่ำกว่าพื้นฐาน(คะแนน 1)	พื้นฐาน(คะแนน 2)	มาตรฐาน(คะแนน 3)	สูงกว่ามาตรฐาน(คะแนน 4)
3.2 ค้นคว้าได้ตรงประเด็นจากหลากหลายแหล่งข้อมูล มาใช้ในการปฏิบัติ	1. ไม่ค้นคว้าข้อมูล หรือ ค้นคว้าไม่ตรงประเด็น และ ไม่สามารถสรุปผล หรือ วิเคราะห์ผลการค้นคว้าได้ หรือ ไม่สามารถนำข้อมูลที่ได้มาประยุกต์ใช้ในการแก้ปัญหาตามสถานการณ์จริงได้ หรือน้อยกว่าร้อยละ60 ของเนื้อหา	2. ค้นคว้าข้อมูลตรงประเด็นบางส่วนจากแหล่งข้อมูลที่ไม่หลากหลาย ข้อมูลมีความทันสมัยและถูกต้องบางส่วนและสามารถสรุปและวิเคราะห์ผลการค้นคว้าได้บางส่วนหรือสามารถนำข้อมูลที่ได้มาประยุกต์ใช้ในการแก้ปัญหาตามสถานการณ์จริงได้บางส่วน หรือประมาณร้อยละ60 ขึ้นไปของเนื้อหา	3. ค้นคว้าข้อมูลจากแหล่งข้อมูลที่หลากหลาย ตรงประเด็น หรือ ข้อมูลถูกต้องเป็นส่วนใหญ่ และสามารถสรุปและวิเคราะห์ผลการค้นคว้าหรือได้เป็นส่วนใหญ่ หรือสามารถนำข้อมูลที่ได้มาประยุกต์ใช้ในการแก้ปัญหาตามสถานการณ์จริงได้เป็นส่วนใหญ่ หรือประมาณร้อยละ80 ขึ้นไปของเนื้อหา	4. ค้นคว้าข้อมูลหรือนำข้อมูลจากแหล่งข้อมูลที่หลากหลายมาใช้ หรือ ค้นคว้าตรงประเด็นที่กำหนดเป็นส่วนใหญ่ ถูกต้องเป็นส่วนใหญ่ หรือ สามารถวิเคราะห์หรือสรุปผลการค้นคว้าได้เป็นส่วนใหญ่ หรือสามารถนำข้อมูลที่ได้มาประยุกต์ใช้ในการแก้ปัญหาตามสถานการณ์จริงได้เป็นส่วนใหญ่หรือประมาณร้อยละ90 ขึ้นไปของเนื้อหา
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ตัวอย่าง การทำกิจกรรม quick round/ pre-post conference หรือ clinical teaching นักศึกษามีส่วนร่วมในการอภิปรายโดยอาศัยข้อมูลที่ถูกต้องจาก เอกสารประกอบการสอน ตำรา บทความทางวิชาการ หรือ ข้อมูลจากฐานข้อมูลที่เชื่อถือได้ หรือการทำรายงานการวางแผนการพยาบาล มีการนำข้อมูลมาอ้างอิงในการเขียนแผนการพยาบาล				
3.3 วางแผนการพยาบาลตามกระบวนการพยาบาลสอดคล้องกับสถานการณ์ของผู้รับบริการ และแก้ไขปัญหาได้เหมาะสมกับสถานการณ์บนหลักการที่ถูกต้อง	1. นำกระบวนการพยาบาลมาใช้ในการวางแผนการพยาบาล ไม่ถูกต้อง และ วางแผนการพยาบาลไม่สอดคล้องกับสถานการณ์จริง แก้ไขปัญหาได้น้อยมากตามสถานการณ์ หรือ อยู่บนหลักการที่ถูกต้องน้อยกว่าร้อยละ 60 ของกระบวนการทั้งหมด	2. นำกระบวนการพยาบาลมาใช้ในการวางแผนการพยาบาลบางส่วนและถูกต้องบางส่วน หรือ วางแผนการพยาบาลสอดคล้องกับสถานการณ์จริงบางส่วน หรือประมาณร้อยละ 60 ขึ้นไปของกระบวนการที่ถูกต้อง	3. นำกระบวนการพยาบาลมาใช้ในการวางแผนการพยาบาลอย่างครบถ้วนและถูกต้องบางส่วน และวางแผนการพยาบาลสอดคล้องกับสถานการณ์จริงส่วนใหญ่ หรือประมาณร้อยละ 80 ขึ้นไปของกระบวนการที่ถูกต้อง	4. นำกระบวนการพยาบาลมาใช้ในการวางแผนการพยาบาลอย่างครบถ้วนและถูกต้องเป็นส่วนใหญ่ และวางแผนการพยาบาลสอดคล้องกับสถานการณ์จริงที่เป็นปัจจุบันหรือประมาณร้อยละ 90 ขึ้นไปของกระบวนการที่ถูกต้อง
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

สมรรถนะ/เกณฑ์ที่ต้องประเมิน	ระดับคุณภาพ			
	ต่ำกว่าพื้นฐาน(คะแนน 1)	พื้นฐาน(คะแนน 2)	มาตรฐาน(คะแนน 3)	สูงกว่ามาตรฐาน(คะแนน 4)
ตัวอย่าง - ขั้นตอนการประเมิน เช่น นักศึกษาประเมินได้ครอบคลุมทั้งการซักประเมิน ตรวจร่างกาย และการศึกษาผลการตรวจทางห้องปฏิบัติการ หรือ ตรวจพิเศษอื่นๆ - ขั้นตอนการตั้งข้อวินิจฉัย ตั้งข้อวินิจฉัยได้ตรงตามความเป็นจริง เรียงลำดับความสำคัญของปัญหาได้ถูกต้องครอบคลุม - ขั้นตอนการวางแผน กำหนดวัตถุประสงค์ เกณฑ์การประเมิน และกิจกรรมการพยาบาลตรงกับความ เป็นจริง ในเอกสารหรือตำราเรียน แต่ไม่ได้นำข้อมูลสนับสนุนที่แท้จริงของคนไข้มาเขียน หรือ เขียนไม่ครบ ไม่ครอบคลุมข้อมูลที่มีอยู่ - ขั้นตอนการปฏิบัติ นักศึกษาสามารถนำแผนการพยาบาลมาใช้ได้จริง และสอดคล้องกับสถานการณ์จริงของผู้ป่วย - ขั้นตอนการประเมินผลการปฏิบัติการพยาบาลนักศึกษาติดตามประเมินผลการพยาบาลทันเหตุการณ์ และนำผลการประเมินมาใช้ในการวางแผนการพยาบาลครั้งต่อไป				
4 ความสัมพันธ์ระหว่างบุคคล และ ความรับผิดชอบ 4.1 ปฏิสัมพันธ์ต่อบุคลากรในทีมสุขภาพ เพื่อน อาจารย์ หรือ ผู้ใช้บริการ และการปรับตัวเชิงวิชาชีพ	1. ใช้ภาษาพูดและการแสดงออก ต่อผู้อื่น ไม่เหมาะสม หรือเผชิญปัญหาและอุปสรรคได้ไม่เหมาะสมกับสถานการณ์ อยู่เสมอ หรือประมาณร้อยละ 90 ของเหตุการณ์ ทั้งหมด ควรได้รับคำแนะนำทั้งหมด	1. ใช้ภาษาพูดและการแสดงออกต่อผู้อื่น ไม่เหมาะสม และเผชิญปัญหาและอุปสรรคได้เหมาะสมกับสถานการณ์ บางส่วน หรือประมาณร้อยละ 60 ขึ้นไปของเหตุการณ์ ควรได้รับคำแนะนำเป็นส่วนใหญ่	3. ใช้ภาษาพูดและการแสดงออก ต่อผู้อื่นอย่างเหมาะสม และเผชิญปัญหาและอุปสรรคได้เหมาะสมกับสถานการณ์ เป็นส่วนใหญ่ หรือประมาณร้อยละ 80 ขึ้นไปของเหตุการณ์ ควรได้รับคำแนะนำบางส่วน	4. ใช้ภาษาพูดและการแสดงออก ต่อผู้อื่นอย่างเหมาะสม และเผชิญปัญหาและอุปสรรคได้เหมาะสมกับสถานการณ์ ด้วยตนเอง อย่างสม่ำเสมอ ประมาณร้อยละ 90ของเหตุการณ์ อาจได้รับคำแนะนำเล็กน้อย
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ตัวอย่าง - การแสดงออกด้วยภาษากาย หรือภาษาพูดที่สื่อให้เห็นถึงปฏิสัมพันธ์ต่อบุคคลรอบข้าง ในขณะที่ขึ้นฝึกปฏิบัติงาน พฤติกรรมที่เหมาะสม เช่น ยิ้มแย้มแจ่มใสกับคนรอบข้าง หรือ พูดคุยด้วยน้ำเสียงอ่อนโยนไม่แข็งกระด้าง พฤติกรรมที่ไม่เหมาะสม เช่น แสดงสีหน้าเบื่อหน่ายเมื่อต้องให้การพยาบาลผู้ป่วย หรือพูดคุยด้วยน้ำเสียงห้วนๆกับเพื่อน หรือคนรอบข้างขณะขึ้นฝึก - การปรับตัวเชิงวิชาชีพ เช่น เมื่อเกิดเหตุการณ์ที่นักศึกษามีส่วนร่วมในเหตุการณ์ที่มีความขัดแย้งเชิงจริยธรรม เชิงความเชื่อ หรืออื่นๆ กับคนไข้ หรือกับบุคลากรด้านสุขภาพอื่นๆในแหล่งฝึก นักศึกษามีการปรับตัวที่เหมาะสมกับสถานการณ์ ตัวอย่างเช่น นักศึกษาปฏิบัติงานตามหลักการพยาบาลพื้นฐาน ในขณะที่แหล่งฝึกปฏิบัติลำดับขั้นตอนและต้องการให้นักศึกษาปฏิบัติตามเพื่อให้ประหยัดเวลา แต่ผิดหลักการที่ถูกต้อง นักศึกษาที่ตอบสนองอย่างเหมาะสม อาจผัดผ่อน ขอปรึกษาอาจารย์ หรือแก้ปัญหาด้วยตนเอง อาจบอกวาขอหนูทำตามทีหนูเรียนมานะคะเดี๋ยวอาจารย์รู้หนูจะถูกตำหนิ ส่วนนักศึกษาที่ปรับตัวได้ไม่เหมาะสม อาจปฏิบัติตามโดยไม่ขัดขึ้น หรือ เก็บมาครุ่นคิด เคร่งเครียดโดยไม่มีควมพยายามจะแก้ไข้ปัญหา				
4.2 แสดงออกถึงการคิดบวกต่อวิชาชีพ และต่อผู้อื่น	1. แสดงออกถึงการคิดบวกต่อวิชาชีพ และต่อผู้อื่นน้อยกว่า ร้อยละ 60 ของเหตุการณ์	2. แสดงออกถึงการคิดบวกต่อวิชาชีพ และ ต่อ ผู้ อื่น ประมาณร้อยละ 60 ขึ้นไปของเหตุการณ์	3. แสดงออกถึงการคิดบวกต่อวิชาชีพ และต่อผู้อื่น ประมาณร้อยละ80 ขึ้นไปของเหตุการณ์	4. แสดงออกถึงการคิดบวกต่อวิชาชีพ และต่อผู้อื่นอยู่เสมอประมาณร้อยละ 90 ของเหตุการณ์ทั้งหมดขึ้น
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

สมรรถนะ/เกณฑ์ที่ต้องประเมิน	ระดับคุณภาพ			
	ต่ำกว่าพื้นฐาน(คะแนน 1)	พื้นฐาน(คะแนน 2)	มาตรฐาน(คะแนน 3)	สูงกว่ามาตรฐาน(คะแนน 4)
ตัวอย่าง การแสดงออกถึงการคิดบวก เช่น เมื่อพบปัญหาอุปสรรคอันเกิดจากการปฏิบัติงาน เช่น การถูกร้องเรียนจากผู้ให้บริการ การเกิดเหตุจากความผิดพลาดในการทำงาน ให้มองว่าปัญหาคือสิ่งเร้าที่กระตุ้นให้เราพัฒนาตนเอง หากไม่พบปัญหาก็กะไม่มีการพัฒนา ไม่ได้ดีพอ หรือหลีกเลี่ยงปัญหา				
4.3 แสดงตนเป็นสมาชิกที่ดีของกลุ่มตามบทบาทของตน	1. แสดงตนเป็นสมาชิกที่ดีของกลุ่ม ตามบทบาทของตน น้อยกว่าร้อยละ 60 ของเหตุการณ์	2. แสดงตนเป็นสมาชิกที่ดีของกลุ่มตามบทบาทของตน ร้อยละ 60 ขึ้นไป	3. แสดงตนเป็นสมาชิกที่ดีของกลุ่มตามบทบาทของตน เป็นส่วนใหญ่ หรือ ร้อยละ 80 ขึ้นไป	4. แสดงตนเป็นสมาชิกที่ดีของกลุ่ม ตามบทบาทของตนร้อยละ 90 ขึ้นไป
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ตัวอย่าง การเป็นสมาชิกที่ดีของกลุ่มตามบทบาท เช่น การร่วมอภิปรายในการทำPre-post conference , การทำหน้าที่สมาชิกตามบทบาทหน้าที่ที่ได้รับมอบหมาย, การมีส่วนร่วมในการรายงานกรณีศึกษาของกลุ่ม ฯ				
4.4 พัฒนาการตนเองอย่างต่อเนื่อง	1. มีการพัฒนาตนเองด้านความรู้ และ ทักษะที่เกี่ยวข้องกับการฝึกปฏิบัติฯ น้อยมาก ไม่เห็นความชัดเจนของการพัฒนา หรือน้อยกว่าร้อยละ 60 ของประเด็นที่ควรพัฒนา	2. มีการพัฒนาตนเองด้านความรู้ และ ทักษะที่เกี่ยวข้องกับการฝึกปฏิบัติฯ บางส่วน เห็นการพัฒนาตนเองได้เล็กน้อย ประมาณร้อยละ 60 ขึ้นไปของประเด็นที่ควรพัฒนา	3. มีการพัฒนาตนเองด้านความรู้ และทักษะที่เกี่ยวข้องกับการฝึกปฏิบัติฯ ได้เป็นส่วนใหญ่ เห็นการพัฒนาตนเองได้บางส่วน ประมาณร้อยละ 80 ขึ้นไป	4. มีการพัฒนาตนเองด้านความรู้ และทักษะที่เกี่ยวข้องกับการฝึกปฏิบัติฯ ได้อย่างต่อเนื่องและสม่ำเสมอ เห็นการพัฒนาตนเองได้ชัดเจน ประมาณร้อยละ 90 ขึ้นไปของประเด็นที่ควรพัฒนา
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ตัวอย่าง การพัฒนาตนเองอย่างต่อเนื่อง เช่น เมื่ออาจารย์แนะนำเรื่องความบกพร่องในการปฏิบัติการพยาบาลการสวนปัสสาวะ ครั้งต่อไปเมื่อมีการปฏิบัติการพยาบาลเดิมนักศึกษามีการพัฒนาทักษะขึ้น ไม่พบข้อบกพร่องจากการปฏิบัติเหมือนครั้งก่อน หรือ เมื่อมีการอภิปรายใน Pre-post conference หรือในแผนการพยาบาลเมื่อมีการชี้แนะข้อบกพร่องต่าง ๆ สามารถพัฒนาจนไม่พบข้อบกพร่องเหล่านั้น โดยพิจารณาการพัฒนาตั้งแต่เริ่มฝึกปฏิบัติงานไปตลอดระยะเวลาการฝึก จนถึงสิ้นสุดการปฏิบัติงาน				
5. ทักษะการวิเคราะห์เชิงตัวเลข การสื่อสาร และการใช้เทคโนโลยีสารสนเทศ	1. ไม่สามารถคำนวณหรือ วิเคราะห์ แผลผลอธิบายความหมายของข้อมูล และนำผลการวิเคราะห์ มา ใช้ เปรียบเทียบ อ้างอิงเพื่อใช้ในการพยาบาล	2. สามารถคำนวณ วิเคราะห์ แผลผล อธิบายความหมายของข้อมูล และนำผลการวิเคราะห์ มา ใช้ เปรียบเทียบ อ้างอิง เพื่อ ใช้ ใน ก า ร	3. สามารถคำนวณ วิเคราะห์ แผลผล อธิบายความหมายของข้อมูล และนำผลการวิเคราะห์ มา ใช้ เปรียบเทียบ อ้างอิงเพื่อใช้ในการพยาบาล	4. สามารถคำนวณ วิเคราะห์ แผลผล อธิบายความหมายของข้อมูล และนำผลการวิเคราะห์ มาใช้เปรียบเทียบ อ้างอิงเพื่อใช้ในการพยาบาลได้อย่างเหมาะสมเป็นส่วน
5.1 การเข้าใจความหมายของตัวเลขและนำข้อมูลตัวเลขมา				

สมรรถนะ/เกณฑ์ที่ต้องประเมิน	ระดับคุณภาพ			
	ต่ำกว่าพื้นฐาน(คะแนน 1)	พื้นฐาน(คะแนน 2)	มาตรฐาน(คะแนน 3)	สูงกว่ามาตรฐาน(คะแนน 4)
ใช้ได้ถูกต้อง	ได้อย่างจำกัด หรือ มากกว่าร้อยละ 60 ของเนื้อหา ต้องการคำแนะนำทั้งหมด	พยายาลได้บ้าง หรือ ประมาณร้อยละ 60 ขึ้นไปของเนื้อหา ต้องการคำแนะนำไม่มากกว่าร้อยละ 40 ของเนื้อหา	ได้อย่างเหมาะสมเป็น บาง ส่วน หรือ ประมาณร้อยละ 80 ขึ้นไปของเนื้อหาโดย ต้องได้รับคำแนะนำ ไม่มากกว่าร้อยละ 20 ของเนื้อหา	ใหญ่ หรือ ประมาณร้อยละ 90 ขึ้นไปของเนื้อหา ควรได้รับคำแนะนำไม่มากกว่าร้อยละ 10 ของเนื้อหา
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<p>ตัวอย่าง</p> <p>การคำนวณปริมาณสารน้ำเป็นหยดต่อนาที การคำนวณปริมาณยา หรือ การบันทึกสารน้ำเข้า-ออกจากร่างกาย เป็นต้น</p> <p>นักศึกษาสามารถคำนวณอัตราการไหลของสารน้ำได้ถูกต้อง หรือสามารถคำนวณปริมาณยาที่ผสมในสารน้ำทางหลอดเลือดดำได้ถูกต้อง หรือการแปลผลการปฏิบัติการที่เป็นตัวเลขได้ถูกต้อง</p>				
5.2 ทักษะการสื่อสาร ด้านการฟัง พูด อ่าน เขียน ภาษาไทย และ ภาษาต่างประเทศ ที่ ถูกต้อง ครบถ้วน	1.ขาดทักษะการสื่อสาร - ภาษาไทย ไม่สามารถ สื่อสารได้ตรงความหมาย เกิดความเข้าใจไม่ตรงกับสารที่ต้องการสื่อ หรือไม่ถูกหลักการใช้ภาษา หลักอักขระวิธีหรือ - ภาษาต่างประเทศ ไม่สามารถ สื่อสารด้วยศัพท์ทางการแพทย์ที่เป็นภาษาอังกฤษได้ถูกต้อง ตรงความหมายเกิดความเข้าใจตรงกัน หรือถูกต้องน้อยกว่าร้อยละ 60 ของ สาระสำคัญ และในภาพรวม น้อยกว่าร้อยละ 60 ของ เหตุการณ์	2. มีทักษะการสื่อสาร - ภาษาไทยสามารถสื่อสารได้ตรงความหมาย เกิดความเข้าใจตรงกับสารที่ต้องการสื่อ และถูกหลักการใช้ภาษา หลักอักขระวิธีตั้งแต่ร้อยละ 60 ของสารที่ ต้องการสื่อ - ภาษาต่างประเทศ สามารถสื่อสารด้วยศัพท์ทางการแพทย์ที่เป็นภาษาอังกฤษได้ ถูกต้อง ตรงความหมายเกิดความเข้าใจตรงกัน หรือถูกต้องตั้งแต่ร้อยละ 50 ของ สาระสำคัญ ขึ้นไป และ ในภาพรวมตั้งแต่ร้อยละ 60 ของ เหตุการณ์	3. มีทักษะการสื่อสาร - ภาษาไทยสามารถสื่อสารได้ตรงความหมาย เกิดความเข้าใจตรงกับสารที่ต้องการสื่อ และถูกหลักการใช้ภาษา หลักอักขระวิธีตั้งแต่ร้อยละ 80 ของสารที่ ต้องการสื่อ - มีทักษะการสื่อสาร ภาษาต่างประเทศ สามารถสื่อสารด้วยศัพท์ทางการแพทย์ที่เป็นภาษาอังกฤษได้ถูกต้อง ตรงความหมายเกิดความเข้าใจตรงกัน หรือถูกต้องตั้งแต่ร้อยละ 70 ของ สาระสำคัญ ขึ้นไป และ ในภาพรวมตั้งแต่ร้อยละ 80 ของ เหตุการณ์	4. มีทักษะการสื่อสาร - ภาษาไทยสามารถสื่อสารได้ตรงความหมาย เกิดความเข้าใจตรงกับสารที่ต้องการสื่อ และถูกหลักการใช้ภาษา หลักอักขระวิธีตั้งแต่ร้อยละ 90 -ของสารที่ ต้องการสื่อ - ภาษาต่างประเทศ สามารถสื่อสารด้วยศัพท์ทางการแพทย์ที่เป็นภาษาอังกฤษได้ถูกต้อง ตรงความหมายเกิดความเข้าใจตรงกัน หรือถูกต้องตั้งแต่ร้อยละ 80 ของ สาระสำคัญ ขึ้นไป และ ในภาพรวมตั้งแต่ร้อยละ 90 ของ เหตุการณ์

สมรรถนะ/เกณฑ์ที่ต้องประเมิน	ระดับคุณภาพ			
	ต่ำกว่าพื้นฐาน(คะแนน 1)	พื้นฐาน(คะแนน 2)	มาตรฐาน(คะแนน 3)	สูงกว่ามาตรฐาน(คะแนน 4)
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<p>ตัวอย่าง</p> <p>เมื่อทำกิจกรรมการ Pre-post conference มีการมอบหมายให้ค้นคว้าความรู้ที่นักศึกษาสามารถอ่านสรุปสาระสำคัญ และนำมาพูดอภิปรายเกี่ยวกับการพยาบาลในผู้ป่วยที่นักศึกษาได้รับมอบหมาย หรือ เมื่อทำกิจกรรมการ Clinical teaching ที่อาจารย์ดำเนินการสิ้นสุดลงนักศึกษา มีความสามารถระดับใด ในการสรุปสาระสำคัญจากการฟัง และสรุปความคิดรวบยอดจากการฟังผ่านการพูดหรือการเขียน</p>				
5.3 การสืบค้นข้อมูลที่เป็นปัจจุบันจากแหล่งต่าง ๆ ทั้งจำนวนเอกสารอ้างอิง แหล่งอ้างอิงความน่าเชื่อถือของแหล่งอ้างอิง	1. การสืบค้นข้อมูลที่ไม่เป็นปัจจุบัน แหล่งข้อมูลที่มีอยู่อย่างจำกัด หรือไม่น่าเชื่อถือ	2. การสืบค้นข้อมูลไม่คอยเป็นปัจจุบัน จากแหล่งข้อมูลที่มีจำกัด แต่มีความน่าเชื่อถือ	3. การสืบค้นข้อมูลที่ค่อนข้างเป็นปัจจุบัน มีความน่าเชื่อถือ และมีความหลากหลาย	4. การสืบค้นข้อมูลที่มีความทันสมัย มีความน่าเชื่อถือ และมีความหลากหลาย
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ตัวอย่าง การสืบค้นในแต่ละระดับคุณภาพ	- การสืบค้นข้อมูลที่ไม่เป็นปัจจุบัน เช่น ใช้ตำราที่พิมพ์ นานเกิน 10 ปีและเป็นความรู้ที่ไม่ทันการเปลี่ยนแปลงของความรู้และการปฏิบัติแล้ว -จำนวนเอกสารอ้างอิง 1-2 เอกสารแหล่งข้อมูลที่มีอย่างจำกัด เช่น ใช้เอกสารอ้างอิงเพียง 1 เล่มหรือไม่น่าเชื่อถือเช่น เว็บไซต์ที่ไม่มีมาตรฐานด้านวิชาการหรือไม่ผ่านการตรวจสอบความถูกต้องทางวิชาการ	- การสืบค้นข้อมูลที่ค่อนข้างเป็นปัจจุบัน เช่น ใช้ตำราที่มีปีที่ตีพิมพ์ 10 ปีขึ้นไปเทียบกับปัจจุบัน แต่เนื้อหาเป็นสิ่งที่ยังพอนำมาปฏิบัติได้ไม่ล่าช้ามากเกินไป - จำนวน เอกสารอ้างอิง 2-3 เอกสาร ไม่มีการใช้บทความทางวิชาการหรือไม่ใช้แหล่งข้อมูลจากสื่ออิเล็กทรอนิกส์ หรือ ใช้ แต่เป็นฐานข้อมูลที่ไม่มีความน่าเชื่อถือ	- การสืบค้นข้อมูลที่เป็นปัจจุบัน หนังสือที่ตีพิมพ์ไม่เกิน 10 ปี หรือวารสารตีพิมพ์ไม่เกิน 5 ปี - จำนวนเอกสารอ้างอิง 3-4 เอกสารขึ้นไป เช่น ใช้เอกสารประกอบการสอน ตำรา หรือเอกสารวิชาการจากสื่ออิเล็กทรอนิกส์ จากฐานข้อมูลที่เชื่อถือได้ คุณภาพของแหล่งข้อมูลจากสื่ออิเล็กทรอนิกส์ เป็นฐานข้อมูลที่มีความน่าเชื่อถือ เช่น ฐานข้อมูลของ WHO, CINAHL, MOSBY ฯ	- การสืบค้นข้อมูลที่ทันสมัย เช่น หนังสือที่ตีพิมพ์ไม่เกิน 5 ปี หรือวารสารตีพิมพ์ไม่เกิน 3 ปี - จำนวนเอกสารอ้างอิงมีตั้งแต่ 4 แหล่งขึ้นไป หรือจากแหล่งข้อมูลที่มีความหลากหลาย เช่น ใช้เอกสารประกอบการสอนในรายวิชานั้น ตำรา หรือเอกสารวิชาการจากสื่ออิเล็กทรอนิกส์ จากฐานข้อมูลที่นำเชื่อถือ เช่นเดียวกับระดับคะแนน 3
6 ทักษะการปฏิบัติทางวิชาชีพ 6.1 ปฏิบัติการพยาบาลอย่างเป็นองค์รวม โดยใช้กระบวนการพยาบาลที่	1. ปฏิบัติการพยาบาลถูกต้องตามกระบวนการพยาบาล และสะท้อนถึงการสร้างเสริมสุขภาพ การป้องกัน การรักษาและการ	2. ปฏิบัติการพยาบาลถูกต้องตามกระบวนการพยาบาล และสะท้อนถึงการสร้างเสริมสุขภาพ การป้องกัน	3. ปฏิบัติการพยาบาลถูกต้องตามกระบวนการพยาบาล และ สะท้อนถึงการสร้างเสริมสุขภาพ การป้องกัน การรักษา	4. ปฏิบัติการพยาบาลถูกต้องตามกระบวนการพยาบาล และสะท้อนถึงการสร้างเสริมสุขภาพ การป้องกัน การรักษาและการ

สมรรถนะ/เกณฑ์ที่ต้องประเมิน	ระดับคุณภาพ			
	ต่ำกว่าพื้นฐาน(คะแนน 1)	พื้นฐาน(คะแนน 2)	มาตรฐาน(คะแนน 3)	สูงกว่ามาตรฐาน(คะแนน 4)
สอดคล้องกับสถานการณ์ และสามารถสะท้อนถึงการสร้างเสริมสุขภาพ การป้องกัน การรักษาและการฟื้นฟูสภาพ	ฟื้นฟูสภาพ ตามสถานการณ์ที่เหมาะสม และถูกต้องด้วยตนเองได้ น้อยน้อยกว่าร้อยละ 60 ของกระบวนการ ขึ้นไป	การรักษาและการฟื้นฟูสภาพ ตามสถานการณ์ที่เหมาะสมและถูกต้องด้วยตนเองได้น้อยกว่า ร้อยละ 60 ของกระบวนการขึ้นไป	และการฟื้นฟูสภาพ ตามสถานการณ์ที่เหมาะสม และถูกต้องได้ ด้วยตนเองประมาณร้อยละ 80 ของกระบวนการขึ้นไป	ฟื้นฟูสภาพ ตามสถานการณ์ที่เหมาะสมและถูกต้องได้ด้วยตนเอง ประมาณร้อยละ 90 ของเหตุการณ์ ขึ้นไป
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6.2 ปฏิบัติการพยาบาลถูกหลักการพยาบาลในกิจกรรมพยาบาล นั้นๆ	1. ปฏิบัติการพยาบาลไม่ถูกหลักการของกิจกรรมการพยาบาลในเรื่อง นั้น ๆ ปฏิบัติด้วยตนเองน้อยมากต้องอาศัยการชี้แนะอย่างใกล้ชิดทุกขั้นตอน หรือปฏิบัติถูกหลักการน้อยกว่าร้อยละ 60 ของกระบวนการลงไป	2. ปฏิบัติการพยาบาลไม่ถูกหลักการของกิจกรรมการพยาบาลในเรื่อง นั้น ๆ ปฏิบัติด้วยตนเองได้บางส่วน ต้องชี้แนะเป็นส่วนใหญ่หรือปฏิบัติถูกหลักการประมาณร้อยละ 60 ของกระบวนการขึ้นไป	3. ปฏิบัติการพยาบาลถูกหลักการของกิจกรรมการพยาบาลในเรื่อง นั้น ๆ เป็นส่วนใหญ่ ต้องชี้แนะเล็กน้อย หรือปฏิบัติถูกหลักการประมาณร้อยละ 80 ของกระบวนการขึ้นไป	4. ปฏิบัติการพยาบาลถูกหลักการของกิจกรรมการพยาบาลในเรื่อง นั้น ๆ อาจไม่ต้องแนะนำหรือชี้แนะน้อยครั้งมาก หรือปฏิบัติถูกหลักการประมาณร้อยละ 90 ของกระบวนการขึ้นไป
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ตัวอย่าง นักศึกษาสามารถใส่คาสายสวนปัสสาวะ ได้ถูกต้องตามขั้นตอนที่สำคัญ และ ตามหลักการให้การพยาบาลด้วยเทคนิคปราศจากเชื้อ หรือสามารถให้ยาฉีดทางกล้ามเนื้อได้ถูกต้องตามหลักการบริการยา				
6.3 ปฏิบัติการพยาบาลถูกหลักการป้องกันการแพร่กระจายเชื้อ	1. ปฏิบัติการพยาบาลไม่ถูกหลักการป้องกันการแพร่กระจายเชื้อ หรือไม่ถูกหลักการน้อยกว่าร้อยละ 60 ของเหตุการณ์	2. ปฏิบัติการพยาบาลไม่ถูกหลักการป้องกันการแพร่กระจายเชื้อ บางส่วน หรือ ปฏิบัติตามหลักการ ประมาณร้อยละ 60 ของเหตุการณ์ขึ้นไป	3. ปฏิบัติการพยาบาลถูกหลักการป้องกันการแพร่กระจายเชื้อ เป็นส่วนใหญ่ หรือ ปฏิบัติตามหลักการ ประมาณร้อยละ 80 ของเหตุการณ์ขึ้นไป	4. ปฏิบัติการพยาบาลถูกหลักการป้องกันการแพร่กระจายเชื้อเป็นประจำ ปฏิบัติถูกต้องตามหลักการประมาณร้อยละ 90 ของเหตุการณ์ ขึ้นไป
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ตัวอย่าง - ทิ้งวัสดุทางการแพทย์ที่มีสิ่งคัดหลั่ง หรือ เลือดของผู้ป่วย ลงในภาชนะรองรับที่ถูกต้องตามชนิดของวัสดุ เช่น วัสดุที่ไม่มีคมทั้งในถุงขยะติดเชื้อสีแดง เข้มทั้งในกระปุกทิ้งเข็ม - ใส่ถุงมือเมื่อต้องปฏิบัติกิจกรรมการพยาบาลได้ถูกต้องตามลักษณะกิจกรรม				

สมรรถนะ/เกณฑ์ที่ต้องประเมิน	ระดับคุณภาพ			
	ต่ำกว่าพื้นฐาน(คะแนน 1)	พื้นฐาน(คะแนน 2)	มาตรฐาน(คะแนน 3)	สูงกว่ามาตรฐาน(คะแนน 4)
6.4 ปฏิบัติการพยาบาลด้วยความเมตตา กรุณา และเอื้ออาทร และทันเหตุการณ์สอดคล้องกับความต้องการของผู้ใช้บริการ	1. ปฏิบัติการพยาบาล ไม่นุ่มนวล ไม่สามารถตอบสนองความต้องการของผู้ใช้บริการได้ ต้องชี้แนะทั้งหมด มีพฤติกรรมปฏิบัติที่ถูกต้องเหมาะสมน้อยกว่าร้อยละ 60	2. ปฏิบัติการพยาบาล ไม่นุ่มนวล ตอบสนองความต้องการของผู้ใช้บริการได้เล็กน้อย ต้องชี้แนะเป็นส่วนใหญ่ มีพฤติกรรมปฏิบัติที่ถูกต้องเหมาะสมประมาณร้อยละ 60 ของเหตุการณ์ขึ้นไป	3. ปฏิบัติการพยาบาล อย่างนุ่มนวล ตอบสนองความต้องการของผู้ใช้บริการได้ทันการณ์บางส่วน ต้องชี้แนะบางส่วน มีพฤติกรรมปฏิบัติที่ถูกต้องเหมาะสมประมาณร้อยละ 80 ของเหตุการณ์ขึ้นไป	4. ปฏิบัติการพยาบาล นุ่มนวล ทันเหตุการณ์ สอดคล้องกับปัญหาความต้องการของผู้ใช้บริการทันการณ์เป็นส่วนใหญ่ อาจไม่ต้องแนะนำ หรือ ชี้แนะเพียงเล็กน้อย มีพฤติกรรมปฏิบัติที่ถูกต้องเหมาะสมประมาณร้อยละ 90 ขึ้นไป
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<p>ตัวอย่างของพฤติกรรมที่แสดงถึงความเมตตากรุณา</p> <p>ขณะให้การพยาบาลนักศึกษาพูดคุยด้วยน้ำเสียง อ่อนโยน กิริยาทางทำนุมนวล จับต้องผู้ป่วยตามความจำเป็นอย่างนุ่มนวล แต่สามารถตอบสนองความต้องการของผู้ป่วยได้ทันเหตุการณ์ เช่น ผู้ป่วยต้องการปัสสาวะแต่มีข้อจำกัดในการเคลื่อนไหว นักศึกษาพูดคุยซักถามด้วยนุมนวลว่าจะนอนปัสสาวะหรือถ้านั่งโหวจะนั่งหรือไม่ เพื่อแสดงความเอื้ออาทรและตอบสนองความต้องการของผู้ป่วย ใส่หมอนนอนอย่างนุ่มนวล และ รวดเร็ว ขณะใส่หมอนนอนนักศึกษาซักถามว่าเจ็บบริเวณที่หมอนนอนกดอยู่หรือไม่ นอนได้ตรงรื่องของหมอนนอนสบายดีหรือไม่ ฯลฯ</p>				
6.5 ปฏิบัติการพยาบาลโดยคำนึงถึงความเป็นปัจเจกบุคคลและความหลากหลายทางวัฒนธรรม	1. ปฏิบัติการพยาบาล โดยละเลยความเป็นปัจเจกบุคคลและความหลากหลายทางวัฒนธรรม ประมาณร้อยละ 80 ของเหตุการณ์ขึ้นไป	2. ปฏิบัติการพยาบาล โดยละเลยความเป็นปัจเจกบุคคลและความหลากหลายทางวัฒนธรรมประมาณร้อยละ 60 ขึ้นไป	3. ปฏิบัติการพยาบาล โดยคำนึงถึงความเป็นปัจเจกบุคคลและความหลากหลายทางวัฒนธรรมประมาณร้อยละ 80 ขึ้นไป	4. ปฏิบัติการพยาบาลโดย คำนึงถึงความเป็นปัจเจกบุคคลและความหลากหลายทางวัฒนธรรม ประมาณร้อยละ 90 ขึ้นไป
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<p>ตัวอย่าง</p> <ul style="list-style-type: none"> - ปฏิบัติการพยาบาลที่คำนึงถึงความเป็นปัจเจกบุคคล เช่น ปฏิบัติการพยาบาลโดยยึดหลักการพยาบาลตามขั้นตอนที่ถูกต้อง แต่ยึดหยุ่นตามข้อจำกัดของผู้ป่วยแต่ละคน เช่น ขณะสวนปัสสาวะเพศหญิงจะต้องจัดทำเพื่อใส่สายสวนให้สะดวกแต่ ผู้ป่วยบอกปวดบริเวณต้นขาข้างหนึ่งบอกว่าเส้นพลิก จะแยกขาข้างหนึ่งไม่ได้ พยาบาลจึงให้ผู้ป่วยแยกข้างที่แยกได้โดยไม่ให้ผู้ป่วยรู้สึกเจ็บปวดมาก - ปฏิบัติการพยาบาลโดยคำนึงถึงความหลากหลายทางวัฒนธรรม เช่น เมื่อรับผู้ป่วยเข้ารักษามีการซักประวัติการนับถือศาสนาในผู้ป่วยนับถือศาสนาอิสลาม ต้องประสานงานเพื่อจัดหาอาหารสำหรับอิสลาม ไม่แสดงท่าทีรำคาญ หรือ ทำให้ผู้อื่นรู้สึกว่าเป็นภาระ ฯลฯ 				
6.6 ปฏิบัติการพยาบาลโดยยึดหลักความปลอดภัย	1. ปฏิบัติการพยาบาลโดย ละเลยเรื่องความปลอดภัย หรือ มีพฤติกรรม	2. ปฏิบัติการพยาบาล โดยละเลยเรื่องความปลอดภัย หรือ มี	3. ปฏิบัติการพยาบาล โดยคำนึงถึงเรื่องความปลอดภัยเป็นส่วนใหญ่	4. ปฏิบัติการพยาบาลโดย คำนึงถึงเรื่องความปลอดภัย อยู่เสมอ หรือ มี

สมรรถนะ/เกณฑ์ที่ต้องประเมิน	ระดับคุณภาพ			
	ต่ำกว่าพื้นฐาน(คะแนน 1)	พื้นฐาน(คะแนน 2)	มาตรฐาน(คะแนน 3)	สูงกว่ามาตรฐาน(คะแนน 4)
	ปฏิบัติที่พึงประสงค์น้อยกว่าร้อยละ 60 ต้องชี้แนะเป็นประจำ	พฤติกรรมการปฏิบัติที่พึงประสงค์ประมาณร้อยละ 60 ขึ้นไป ต้องชี้แนะบ่อยครั้ง	หรือ มีพฤติกรรมการปฏิบัติที่พึงประสงค์ประมาณร้อยละ 80 ขึ้นไป ต้องชี้แนะบางครั้ง	พฤติกรรมการปฏิบัติที่พึงประสงค์ประมาณร้อยละ 90 ขึ้นไป อาจไม่ต้องชี้แนะหรือ ชี้แนะน้อยครั้งมาก
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ตัวอย่างการปฏิบัติการพยาบาลโดยยึดหลักความปลอดภัย - การตั้งไม้กั้นเตียงขึ้นทุกครั้งหลังให้การพยาบาล - การประเมินความสามารถในการทำกิจกรรมของผู้ป่วยเพื่อป้องกันอุบัติเหตุพลัดตกหกล้ม - ไม่วางวัสดุทางการแพทย์ใดๆ ลงบนเตียงคนไข้เพื่อป้องกันการหลั่งมีวัสดุที่อาจเป็นอันตรายของผู้ป่วยไว้บนเตียงคนไข้ เช่น เมื่อต้องให้นยาฉีดทางหลอดเลือดดำ ไม่วางถาดยา หรือ ปลอกเข็มหรือเข็มที่ใช้แล้วไว้บนเตียงคนไข้ ฯ ล ฯ				
7. คุณลักษณะของนักศึกษาตามเอกลักษณ์ของมหาวิทยาลัย คุณลักษณะด้านบุคลิกภาพดี 7.1 แต่งกายถูกระเบียบ สะอาดเรียบร้อย ภูมิรียบไม่ยุ่งเหยิง รองเท้าหุ้มส้น ถูกระเบียบสะอาด ทำทางสง่าในแต่ละอิริยาบถ ทั้งขณะนั่ง ยืนหรือเดิน	1. แสดงพฤติกรรมได้ด้วยตนเองน้อยกว่าร้อยละ 60 ต้องอาศัยการชี้แนะ หรือ ทักท้วงมากกว่าร้อยละ 60 ของเหตุการณ์ขึ้นไป	2. แสดงพฤติกรรมได้ด้วยตนเองได้ประมาณร้อยละ 60ขึ้นไปของเหตุการณ์ ต้องการชี้แนะประมาณร้อยละ 40 ของเหตุการณ์ลงไป	3. แสดงพฤติกรรมได้ด้วยตนเองได้ประมาณร้อยละ 80ขึ้นไปของเหตุการณ์ ต้องการ การชี้แนะประมาณร้อยละ 20 ของเหตุการณ์ลงไป	4. แสดงพฤติกรรมได้ด้วยตนเอง ได้ประมาณร้อยละ 90ขึ้นไปของเหตุการณ์อาจไม่ต้องชี้แนะ หรือชี้แนะประมาณร้อยละ 10 ของเหตุการณ์ลงไป
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ตัวอย่าง - พฤติกรรมที่พึงประสงค์ เช่น นักศึกษาแต่งกายถูกระเบียบตั้งแต่ศีรษะจรดเท้า เสื้อผ้าที่สวมใส่สะอาดเรียบร้อย ภูมิรียบเรียบร้อย การเคลื่อนไหวในแต่ละอิริยาบถสง่างาม - พฤติกรรมที่ไม่พึงประสงค์ เช่น นักศึกษาแต่งกายไม่ถูกระเบียบ ภูมิรียบไม่เรียบร้อย การเคลื่อนไหวในแต่ละอิริยาบถไม่สง่างาม				
7.2 หน้าตาอึมแวม ทำทางเป็นมิตร พุดจาสำเนียงสุภาพ ณะฉาน	1. แสดงพฤติกรรมได้ด้วยตนเองน้อยกว่าร้อยละ60ของเหตุการณ์ ต้องอาศัยการชี้แนะ หรือทักท้วงมากกว่าร้อยละ 60 ของเหตุการณ์ขึ้นไป	2. แสดงพฤติกรรมได้ด้วยตนเองประมาณร้อยละ 60 ของเหตุการณ์ ขึ้นไป ต้องการ การชี้แนะประมาณร้อยละ40 ของ	3. แสดงพฤติกรรมได้ด้วยตนเองประมาณร้อยละ 80 ของเหตุการณ์ ขึ้นไป เป็นธรรมชาติ เหมาะสมกับสถานการณ์ นั้น ต้องการ การชี้แนะ	4. แสดงพฤติกรรมนั้นได้ด้วยตนเอง ประมาณร้อยละ 90 ของเหตุการณ์ ขึ้นไป เป็นธรรมชาติเหมาะสมกับ สถานการณ์ นั้น และ สถานการณ์ใกล้เคียง อาจ

สมรรถนะ/เกณฑ์ที่ต้องประเมิน	ระดับคุณภาพ			
	ต่ำกว่าพื้นฐาน(คะแนน 1)	พื้นฐาน(คะแนน 2)	มาตรฐาน(คะแนน 3)	สูงกว่ามาตรฐาน(คะแนน 4)
		เหตุการณ์ลงไป	ประมาณร้อยละ 20 ของเหตุการณ์ลงไป	ต้องชี้แนะประมาณร้อยละ 10 ของเหตุการณ์ลงไป
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ตัวอย่าง - พฤติกรรมที่พึงประสงค์ เช่น นักศึกษาคคนหนึ่ง ยิ้มแย้มแจ่มใส พูดคุยขัดถ้อยขัดคำ ขณะให้คำแนะนำผู้ใช้บริการ หรือการอภิปรายขณะทำ pre-post conference - พฤติกรรมที่ไม่พึงประสงค์ เช่น นักศึกษาคคนหนึ่งทำหน้าที่ไม่คอยยิ้มแย้ม เป็นประจำ พูดเสียงห้วน ไม่มีทางเสียง ฯลฯ				
7.3 ความประณีตพิถีพิถัน	มีหลักฐานแสดงให้เห็นถึงพฤติกรรมที่สะท้อนผลการปฏิบัติงานที่มีความประณีต พิถีพิถันน้อยกว่าร้อยละ 60 ของเนื้องานชิ้นนั้น และน้อยกว่าร้อยละ 60 ของจำนวนชิ้นงานทั้งหมด	มีหลักฐานแสดงให้เห็นถึงผลพฤติกรรมที่สะท้อนการปฏิบัติงานที่มีความประณีต พิถีพิถันประมาณร้อยละ 60 ขึ้นไป ของเนื้องาน ชิ้น นั้น และประมาณร้อยละ 60 ขึ้นไปของจำนวนชิ้นงานทั้งหมด	มีหลักฐานแสดงให้เห็นถึงพฤติกรรมที่สะท้อนผลการปฏิบัติงานที่มีความประณีต พิถีพิถันประมาณตั้งแต่ร้อยละ 80 ขึ้นไปของเนื้องานชิ้นนั้น และประมาณร้อยละ 80 ขึ้นไปของจำนวนชิ้นงานทั้งหมด	มีหลักฐานแสดงให้เห็นถึงพฤติกรรมที่สะท้อนผลการปฏิบัติงานที่มีความประณีต พิถีพิถัน ประมาณร้อยละ 90 ขึ้นไปของเนื้องานชิ้นนั้น และตั้งแต่ร้อยละ 90 ขึ้นไปของจำนวนชิ้นงานทั้งหมด
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ตัวอย่าง หลักฐานที่แสดงถึงความประณีตพิถีพิถันเช่น - ความสะอาดเป็นระเบียบเรียบร้อยของรายงานเช่น แผนการพยาบาล กรณีศึกษา หรือรายงานอื่นๆ ที่ได้รับมอบหมาย คุณภาพของงานที่สะท้อนถึงความพยายามใส่ใจในการทำ - ความเป็นระเบียบเรียบร้อยของผลการปฏิบัติงานที่ได้รับมอบหมาย เช่น จัดสภาพแวดล้อมข้างเตียงได้เป็นระเบียบเรียบร้อย ทำเตียงมีความเรียบร้อย หรือทำกิจกรรมการพยาบาลที่สะท้อนถึงความประณีต เช่น การพันผ้าปิดตาคนไข้ การพันผ้าพันหูกับมือเพื่อเช็ดตัวผู้ป่วยได้เรียบร้อยไม่รุงรัง หรือการติดพลาสติกครอบตัวคนไข้ ทำได้เรียบร้อย เป็นต้น - ความใส่ใจในรายละเอียดของกิจกรรมที่ทำอยู่ที่เกี่ยวข้องโดยตรงกับกิจกรรมการพยาบาลหรือไม่ก็ได้ เช่น การจัดวางสิ่งของต่างๆในห้องconference ของนักศึกษา				

5. บทวิทัศน์ 8 ตอน

ตัวอย่างบทวิทัศน์ตอนที่ 1

ลำดับเหตุการณ์	รายละเอียดเหตุการณ์	ภาพที่ต้องการ/ เวลาที่ใช้	สมรรถนะที่ประเมิน
Shot 1	อาจารย์ 1 คน นักศึกษา 3 คน นั่งล้อมวงรับส่งเวรทางการพยาบาลอยู่กับพยาบาล 1 คน	ภาพมุมกว้างของ การนั่งของอาจารย์ พยาบาล และ นักศึกษา ตัดภาพ มาที่ ภาพนาฬิกาบอก เวลา 8.20 น. เขียนบรรยายเหนือ ภาพ เวลารับเวร 7.30 น. จากนั้น เป็นภาพนักศึกษา นั่งลงบนเก้าอี้ (เวลา 2 – 3 วินาที)	1.1 ความซื่อสัตย์ รับผิดชอบ มีวินัย
Shot 2 ระหว่าง การนั่ง รับ-ส่ง เวร ทาง พยาบาล	มีนักศึกษาชั้นเวรสาย ¹¹ หน้าตา ผมเผ้ายุ่งเหยิง ⁷¹ กำลังเดินมานั่งร่วมวงด้วย ยกมือไหว้อาจารย์และที่ พยาบาลอย่างขอไปที หน้าทำทาง ไม่วิตกทุกซักร้อนกับการมาสาย ⁴¹ และนั่งรับเวรด้วย ท่าที่ง่วงซึม ⁷² “case สุดท้ายเตียง 30 นางจันทร์ แจ่มฟ้า อายุ 29 ปี diagnosis acute appendicitis หลังทำ Appendectomy 2 วัน ตอนนี้ลูกจากเตียงได้ ช่วยเหลือตัวเองได้ มีอาการปวดที่แปล pain score - 4 ฟันตัวได้ดี และมี คนใช้รับเวรจากห้องฉุกเฉินกำลังจะขึ้นมา admission ชื่อนาง กิ่งแก้ว ใจดี อายุ 56 ปี มารพ. ด้วยอาการ 1 วันก่อนหายใจเหนื่อย หน้าบวม ปัสสาวะไม่ออก vital signs อุณหภูมิ 38 องศา เซลเซียส ชีพจร 88 ครั้ง/นาที อัตราการหายใจ 28 ครั้ง/นาที ความดันโลหิต 130/90 mmHg” แพทย์ให้รับไว้วันอน รพ. กำลังจะขึ้นมาคะ เดี่ยวนักศึกษาเตรียม รับใหม่เลยนะคะ”	เปลี่ยนไปโฟกัสภาพ ที่นักศึกษาคนหนึ่ง วิ่ง ขึ้น มา บน หอ ผู้ป่วยบรรยายบน ภาพว่า ” พยาบาลกำลังส่ง เวร คน ไช้ คน สุดท้าย” (30 วินาที)	1.1 ความซื่อสัตย์รับผิดชอบ มี วินัย 4.1 ปฏิสัมพันธ์ต่อ บุคลากรใน ทีมสุขภาพ เพื่อน อาจารย์ หรือ ผู้ใช้บริการและการปรับตัวเชิง วิชาชีพ ปฏิสัมพันธ์ต่อบุคลากรในทีม สุขภาพ เพื่อน อาจารย์ หรือ ผู้ใช้บริการและ การปรับตัวเชิงวิชาชีพ 7.1 แต่งกายถูกระเบียบ สะอาด เรียบร้อย ผมเรียบไม่ยุ่งเหยิง รองเท้าหุ้ม ส้น ถูกระเบียบสะอาด ทำทางส่งไว้ใน แต่ละอิริยาบถ ทั้งขณะนั่ง ยืน หรือเดิน 7.2 หน้าตาอึมแวม ท่าทางเป็น

ลำดับเหตุการณ์	รายละเอียดเหตุการณ์	ภาพที่ต้องการ/ เวลาที่ใช้	สมรรถนะที่ประเมิน
			มิตร ขณะพูดคุยเสียงดังฟังชัด ฉะฉาน ด้วยท่าทางเชื่อมั่นในตัวเอง
Shot 3 ประชุมปรึกษาหารือทางการพยาบาลก่อนการปฏิบัติงาน	<p>บทสนทนา</p> <p>ซัชชา นี่เป็นครั้งที่ 3 ที่เธอมาสายนะ ฐิระเบียบใจใหม่ว่าจะต้องมีการขึ้นฝึกชดเชยวันที่ขาดไป และครูต้องมอบหมายงานเพิ่มเติมเพื่อเป็นการลงโทษ”¹¹</p> <p>นักศึกษาแสดงสีหน้าไม่พอใจ :”อาจารย์ค่ะแต่ 2 ครั้งนั้นหนูไม่สบายนะคะ แล้วรายงานที่ทำอยู่ก็เยอะแล้วนะคะ หนูก็แทบทำไม่ทัน ขอแค่ขึ้นฝึกชดเชยไม่ได้หรือคะ แล้วหนูก็ไม่สบายจริงๆ”¹²</p> <p>อาจารย์: ถ้าไม่สบายจริงๆ ขอใบรับรองแพทย์แนบมาพร้อมในลาป่วยด้วย เพราะตอนขึ้นฝึกวอร์ดที่ผ่านมเธอก็ขาดวอร์ดไม่ใช่หรือ เอละไว้ค่อยมาคุยกันอีกทีเรื่องชดเชยการฝึก”</p> <p>อาจารย์ :”ซัชชาหนูนำเสนอแผนการพยาบาลเลยคะ”</p> <p>นักศึกษา: “คนไข้แคสที่อาจารย์มอบหมายให้กลับบ้านไปแล้วคะ แต่มาไม่ทันเลยไม่ได้ศึกษาแคสใหม่”¹¹</p> <p>อาจารย์: “ถ้าอย่างนั้นนักศึกษาที่รับแคสรับใหม่นี้ดูแลต่อก็แล้วกัน นักศึกษาคิดว่าคนไข้ที่กำลังจะรับใหม่จะวางแผนการพยาบาลอย่างไร”</p> <p>นักศึกษา: นิ่งเงียบไปชั่วขณะและตอบกลับไปว่า: “ก็ยังไม่ได้ศึกษารายละเอียดก็ยังไม่ทราบว่าจะทำอย่างไร”²²</p> <p>อาจารย์: “เอละมาเริ่มกันที่ ว่าคนไข้มีปัญหาอะไรบ้างจะวางแผนให้การพยาบาลอย่างไร”</p> <p>นักศึกษา: “หนูเพิ่งขึ้นวอร์ดก็เลยยังไม่ทันตั้งตัวเลยจดไม่ทันคะ ก็ไม่ทราบจะวางแผนการพยาบาลอย่างไร”³³</p> <p>อาจารย์: “หนูขึ้นมาตอนส่งเวรเคสสุดท้ายพอดีแล้วก็ต่อด้วยการส่งเวรคนไข้รับใหม่รายนี้ไม่ใช่หรือ”</p> <p>นักศึกษา: นิ่งเงียบสีหน้าเคร่งเครียด</p> <p>อาจารย์: ถ้าอย่างนั้นครูให้เวลาหนูไปศึกษาเคสแล้วค่อยมาคุยกันอีกที</p>	<p>ภาพมุมกว้างของการนั่งของอาจารย์และนักศึกษา</p> <p>ตัวหนังสือบรรยายภาพเดิมของ shot 1</p> <p>“Preconference หลังรับเวร” (ตัดภาพไปที่อาจารย์คุยกับนักศึกษา 1 คน)</p> <p>(1 นาที 30 วินาที)</p>	<p>1.1 ความซื่อสัตย์รับผิดชอบ มีวินัย</p> <p>1.2 ควบคุมตนเอง</p> <p>2.2 มีส่วนร่วมในการอภิปรายในช่วงการทำกิจกรรม quick round/ pre-post conference หรือ clinical teaching ที่แสดงถึงการมีความรู้ความเข้าใจในสาระสำคัญของศาสตร์ที่เป็นพื้นฐานชีวิตและวิทยาศาสตร์สุขภาพและศาสตร์ทางวิชาชีพพยาบาล</p> <p>3.2 ค้นคว้าได้</p> <p>ตรงประเด็นจากหลากหลายแหล่งข้อมูลมาใช้ในกรณีปฏิบัติ</p> <p>3.3 วางแผนการพยาบาลตามกระบวนการพยาบาลสอดคล้องกับสถานการณ์ของผู้รับบริการและแก้ไขปัญหาได้เหมาะสมกับสถานการณ์บนหลักการที่ถูกต้อง</p>
Shot 4 มีคนไข้รับใหม่	คนไข้นอนบนเตียง (เมื่อมีคนไข้รับใหม่มา นักศึกษาย้ายเตียงที่จะรับใหม่ โดยอ้างว่ามีเพื่อนทำหน้าที่รับ	(ใช้ภาพเดียวกันทุกสถานการณ์ของ	1.3 ยอมรับความแตกต่างของความเชื่อ

ลำดับเหตุการณ์	รายละเอียดเหตุการณ์	ภาพที่ต้องการ/ เวลาที่ใช้	สมรรถนะที่ประเมิน
กำลังรับผู้ป่วยใหม่เข้ารับการรักษาและทำกิจกรรมการพยาบาลต่างๆ	<p>ใหม่แล้ว)</p> <p>นักศึกษา: “รัชนี มีหน้าที่รับใหม่ให้หนูช่วยเขารับใหม่ดีไหมคะแล้วหนูค่อยเข้าไปซักประวัติทีหลัง”^{4,3} (สีหน้าท่าทางไม่ค่อยมั่นใจในตัวเอง สีหน้าวิตกกังวล พูดเสียงแผ่วเบา)^{7,2}</p> <p>อาจารย์: “หนูต้องดูแลคนไข้เป็นคนที่ต้องศึกษาอยู่แล้วไม่ใช่หรือ ไม่ต้องเกียจเพื่อนทำ รับใหม่ไปเลยจะได้ดูแลตั้งแต่แรก รับ จะได้ซักประวัติ ประเมินร่างกายอย่างละเอียด”^{4,3}</p> <p>นักศึกษาเข้าไปวัดสัญญาณชีพ วัดอุณหภูมิ จับชีพจร หายใจ และความดันโลหิต และซักประวัติแรกรับ (1 นาที) จับภาพนักศึกษากับคนไข้บนเตียง (เป็นนักศึกษากับคนไข้) และจับภาพไปที่นักศึกษาวัดอุณหภูมิ จับชีพจร นับอัตราการหายใจ และ วัดความดันโลหิต</p> <p>อาจารย์: จัดทำศีรษะสูงให้คุณปาก่อน คุณป้าเหนื่อย</p> <p>นักศึกษาซักประวัติหลังประเมินสัญญาณชีพ</p> <p>นักศึกษา : ป้าเอาพวงมาลัยมาไว้ข้างหมอนทำไม (น้ำเสียงเข้ม เหมือนไม่พอใจ)</p> <p>คนไข้: เอาไว้บอกเจ้าที่เจ้าทาง ว่ามาขออาศัยนอน รักษาตัวป่าบอกให้ลูกซื้อที่หน้าโรงพยาบาลตอนรอขึ้นมานอนที่ขึ้นนี้เอง ป้าขอเอาไว้นะ</p> <p>นักศึกษา: ป้าอยากโดนพี่พยาบาลดูหรือ หนูคิดว่าไม่น่าเอาไว้ตรงนี้.^{1,3}</p> <p>คนไข้: ป้าจะเอาไว้ตรงนี้แหละ</p> <p>นักศึกษาแสดงสีหน้าไม่พอใจ: ตามใจป้า งั้นหนูขอซักประวัติต่อ</p> <p>อาจารย์ : เอาอย่างนี้ดีไหมคะคุณป้า วางไว้บนโต๊ะหัวเตียงเดี๋ยวหนูให้นักศึกษาเอาถาดมารองเวลากลับดอกไม้ร่วงจะได้ไม่เลอะเทอะนะคะ และได้กลิ่นหอมอ่อนๆเมื่อไปถึงเตียงข้างๆด้วย</p>	<p>shotนี้) ภาพคนไข้นอนบนเตียง มีพวงมาลัยข้างหมอน นักศึกษายืนอยู่ข้างเตียงคนไข้เตรียมอุปกรณ์วัดเดินมาที่เตียงผู้ป่วยพูดกับผู้ป่วยวัดอุณหภูมิ ความดันชีพจร การหายใจ และ ซัก ประวัติ จากนั้นลงข้อมูลในแฟ้มประวัติผู้ป่วย (2 นาที)</p>	<p>4.3 แสดงตนเป็นสมาชิกที่ดีของกลุ่ม ตามบทบาทของตน</p> <p>1.5 รักษาจรรยาบรรณวิชาชีพ พึงกษสิทธิ์ของผู้ให้บริการ</p> <p>2.1 บอกเหตุผล และ /หรือ บอกหลักการสำคัญของการปฏิบัติ</p> <p>3.3 วางแผน</p> <p>การพยาบาลตามกระบวนการพยาบาลสอดคล้องกับสถานการณ์ของผู้รับบริการ และแก้ไขปัญหาคือได้เหมาะสมกับสถานการณ์บนหลักการที่ถูกต้อง</p> <p>4.1 ปฏิสัมพันธ์ต่อ บุคลากรในทีมสุขภาพ เพื่อน อาจารย์ หรือ ผู้ให้บริการและ การปรับตัวเชิงวิชาชีพ</p> <p>4.2 แสดงออกถึงการคิดบวกต่อวิชาชีพ และต่อผู้อื่น สมรรถนะที่</p> <p>6 ดูตามบทของการพยาบาลการประเมินสัญญาณชีพในระดับคุณภาพ 2</p> <p>7.2 หน้าตาอึมแ่ม ท่าทางเป็นมิตร ขณะพูดคุยเสียงดังฟังชัด ฉะฉาน ด้วยท่าทางเชื่อมั่นในตัวเอง</p>

ลำดับเหตุการณ์	รายละเอียดเหตุการณ์	ภาพที่ต้องการ/ เวลาที่ใช้	สมรรถนะที่ประเมิน
	<p>คนไข้: ก็ดีค่ะ</p> <p>นักศึกษา: ป้าหนูขอซักประวัติเพื่อนำไปใช้ในการวางแผนการพยาบาลให้คุณป้านะคะ คุณป้ามีอาการหายใจเหนื่อย หน้าบวม ปัสสาวะไม่ออก มา 1 วันแล้วใช่ไหม (น้ำเสียง ไม่มีหางเสียงเวลาพูด)</p> <p>คนไข้: ก็บอกหมอไปหมดแล้วเขามาซักตั้งนาน เห็นเขาจดไป ต้องถามซ้ำอีกหรือคะ พุดมากๆ ป้าก็เหนื่อย</p> <p>นักศึกษา: ค่ะๆ ไม่เป็นไรเดี๋ยวนะหนูไปจดเอาที่เพิ่มคนไข้เพิ่ม³³</p> <p>อาจารย์: คุณป้าคะ นักศึกษาจำเป็นต้องขอข้อมูลของคุณป้าอีกครั้งนะคะเพื่อจะได้มั่นใจว่าข้อมูลที่ตรงความเป็นจริงมากที่สุดเพื่อช่วยยืนยันข้อมูลแรกด้วยนะคะว่าไม่มีความผิดพลาดของข้อมูล เป็นประโยชน์กับการนำข้อมูลไปวางแผนการพยาบาล เป็นประโยชน์กับคุณป้านะคะ</p> <p>คนไข้: พยักหน้า</p> <p>อาจารย์: เดียวจะค่อยๆซักประวัติเพิ่มซ้ำๆ นะคะ</p> <p>นักศึกษา: ค่ะคุณป้า ทำท่าพูดซักประวัติต่อ</p>		
Shot 5	<p>นักศึกษา: เดียวหนูจะใส่สายสวนปัสสาวะให้นะ</p> <p>จากนั้นนักศึกษาทำตามขั้นตอนของการสวนปัสสาวะ</p>	<p>นักศึกษาอยู่ข้างเตียงผู้ป่วย ชุมไปที่รถเข็นที่มีเครื่องมือสำหรับสวนปัสสาวะวางอยู่ และภาพนักศึกษายืนสวนปัสสาวะอยู่ที่เตียงผู้ป่วย (2นาที)</p>	<p>สมรรถนะที่ 6 ดูตามบทของการพยาบาลการคาสายสวนปัสสาวะ</p> <p>1.5 รักษาจรรยาบรรณวิชาชีพ พึงทักษสิทธิของผู้ให้บริการ</p>
Shot6	<p>ให้ยาгинปฏิบัติตามขั้นตอนการให้ยาทางปาก</p>	<p>ภาพนักศึกษากำลังตรวจสอบคำสั่งการรักษาในชาร์ทประวัติคนไข้ กับชื่อคนไข้ และ</p>	<p>สมรรถนะที่ 6 ดูตามบทของการพยาบาลการให้ยาทางปาก และประเมิน 7.1, 7.2, 7.3 ขณะ</p>

ลำดับเหตุการณ์	รายละเอียดเหตุการณ์	ภาพที่ต้องการ/ เวลาที่ใช้	สมรรถนะที่ประเมิน
		ตัดภาพแผนที่ นักศึกษา ยืนข้างเตียงพูดคุยและให้ ยา กับ คน ไข้ (1นาที) (ถ่ายแยก เป็น อีก 1 shot แยก ตาม ระดับ คุณภาพถ่ายที่เดียว 4 shot ตามระดับ คุณภาพ 4 ระดับ)	ประเมิน สมรรถนะ6 1.5 รักษาจรรยาบรรณวิชาชีพ พึงกษลิตธิของผู้ใช้บริการ
Shot7	<p>เมื่อให้ยาทางปากเสร็จแล้วอาจารย์ซักถามเกี่ยวกับการดูแลคนไข้ซึ่งเพิ่งได้รับสารน้ำทางหลอดเลือดดำ และออกซิเจน</p> <p>อาจารย์: นักศึกษาคะทำไมคนไข้รายนี้ได้รับออกซิเจน</p> <p>นักศึกษา: คนไข้เหนื่อยค่ะ</p> <p>อาจารย์: เราจะรู้ได้อย่างไรว่าคนไข้เหนื่อย</p> <p>นักศึกษา: คนไข้บอกค่ะ</p> <p>อาจารย์: : แล้วเราดูอาการ หรือ อาการแสดงใดประกอบอีกหรือเปล่า</p> <p>นักศึกษา: เเสบไม่ตอบ²¹</p> <p>อาจารย์: กลับไปอ่านเกี่ยวกับการพยาบาลในการตอบสนองความต้องการด้านการให้ออกซิเจนแล้วนำมาตอบ แล้วเรื่องIV fluidละตอนนี้ ให้5% D/NSS/2 1000 ml vein drip 80 ml/hr จะdrip กี่หยดต่อนาทีลองคำนวณซิ</p> <p>นักศึกษา: 닝ไม่ตอบ⁵¹</p> <p>นักศึกษา: ไม่ทราบาคิดอย่างไร</p> <p>อาจารย์: ก็คิดเทียบบัญญัติไตรยางษิคะ ว่า80มลต่อ 60 นาที1 นาทีได้กัมล แล้วก็เอาจำนวนหยดต่อ1มลไปคูณ แล้วแต่ set ว่าใช้ 20 หยดต่อมล หรือ 15 หยดต่อมล. เข้าใจหรือยัง</p> <p>นักศึกษา: ค่ะ เดี๋ยวหนูลงเอาไปคิดดูคะ</p> <p>นักศึกษาเดินบ่นมากับเพื่อนนักศึกษาอีกคนหนึ่ง</p> <p>นักศึกษาไม่เห็นว่อาจารย์เดินตามหลัง</p> <p>นักศึกษาคคนที่ทำกิจกรรม: รู้ย่่างนี้ไม่เรียนพยาบาลก็ ดี ตื่นสายก็ไม่ได้ คุยกับคนไข้ก็คุยกันไม่รู้เรื่องกวนประสาท⁴²</p>	ภาพนักศึกษา 1 คน ยืนคุยกับอาจารย์1 คน (1 นาที 30 วินาที)	2.1 บอกเหตุผล และ /หรือ บอก หลักการสำคัญของการปฏิบัติ 5.1 การเข้าใจความหมายของตัวเลข และนำข้อมูลตัวเลข มาใช้ได้ถูกต้อง 4.2 แสดงออกถึง การคิดบวกต่อวิชาชีพ และต่อผู้อื่นและ ของความเชื่อ

ลำดับเหตุการณ์	รายละเอียดเหตุการณ์	ภาพที่ต้องการ/ เวลาที่ใช้	สมรรถนะที่ประเมิน
<p>Shot 8</p> <p>ประชุม</p> <p>ปรึกษาหารือ</p> <p>หลังการฝึก</p> <p>ปฏิบัติงาน</p>	<p>อาจารย์: ชี้ขารายงานเคสที่หนูดูแล้ววันนี้เลยคะ</p> <p>นักศึกษา: คนไข้หญิงไทย อายุ 56 ปี มารพ. ด้วยอาการ 1 วันก่อนหายใจเหนื่อย หน้าบวม ปัสสาวะไม่ออก มีไข้ 38.0 องศาเซลเซียส หายใจเหนื่อย 28 ครั้ง/นาที BP 130/90 mmHg</p> <p>อาจารย์: คนไข้มีปัญหาหรือความต้องการที่เราจะตอบสนองได้อย่างไรบ้าง</p> <p>นักศึกษา: เขาเหนื่อยก็เลยต้องช่วยทำกิจวัตรประจำวันที่เขาทำไม่ค่อยได้</p> <p>อาจารย์: ถูกต้องคะ แล้วการพยาบาลที่ช่วยตอบสนองความต้องการด้านร่างกายของเขาคืออะไรคะ</p> <p>นักศึกษา: เจ็บชั่วคราวแล้วตอบว่า เช็ดตัวบนเตียง ป้อนข้าว</p> <p>อาจารย์: หนูก็ลองพิจารณาซิคะว่า อาการ อาการแสดงของคนไข้เป็นอย่างไร แล้วแผนการรักษาเป็นอย่างไร แผนการพยาบาลก็ต้องสอดคล้องกับอาการของคนไข้ และแผนการรักษา คนไข้รายนี้เราจำเป็นต้องตอบสนองความต้องการด้าน ออกซิเจน การขับถ่าย และสารน้ำและอาหารทดแทน การเฝ้าระวังภาวะความดันโลหิตสูงจากค่าความดัน diastolic 90 mmHg และเราก็ให้การพยาบาลไปหมดแล้ว เอาอย่างนี้หนูไปเขียนสรุปใน nursing care plan มาให้ครบตามนี้ แล้วที่ให้ไปค้นเรื่อง ไข้ มาเล่าให้เพื่อนๆ ฟังละ</p> <p>นักศึกษาหยิบเอกสารประกอบการสอนที่ใช้ในการเรียนขึ้นมาเปิดหา³² จากสารบัญและเปิดหน้าที่มีเนื้อหาอ่านค่าควบล้ำไม่ได้ อ่านภาษาต่างประเทศ คำสะกดแบบผิดๆถูกๆ:⁵²²</p> <p>นักศึกษา: ไข้ เป็นกลไกการป้องกันตัวของร่างกาย defense mechanism (ออกเสียง ดีฟ แฟส เมคานิสม) เพาะ (ออกเสียงเพราะไม่ได้ขาดคำควบล้ำ) เป็นสัญญาณเตือนว่าร่างกาย(ออกเสียงร่างไม่ได้)มีความผิดปกติ จากเชื้อโลก ไปกระตุ้นต่อมได้สมอง(ออกเสียง กระผิด)</p> <p>อาจารย์: เตี่ยคะ อ่านผิดกลไกการป้องกันตัวของร่างกายต้องอ่านว่า ดีเฟนเมคานิสมแล้วเวลาพูด ร เรือ ล ลิงหายไปไหนหมดคะ ต้องพูดให้ชัดถ้อยชัดคำนะ คะ แล้วไม่ได้ทำความเข้าใจมาก่อนหรือ ก้มหน้าก้มตาอ่านตลอด เอาหนังสือเอกสารประกอบการสอนของวิชาการพยาบาลพื้นฐานมาเล่มเดียวหรือ ไม่ค้น</p>	<p>อาจารย์ 1 คน นั่งล้อมวงกับนักศึกษา 4 คน</p> <p>(ภาพเดิมจาก shot1) (ตัดภาพไปที่อาจารย์คุยกับนักศึกษา 1 คน) (1 นาที 30 วินาที)</p>	<p>1.4 จิตอาสา</p> <p>5.2.1,5.2.2,5.2.3,5.2.4 ทักษะการสื่อสาร ด้านการฟัง พูด อ่าน เขียน ที่ถูกต้อง ครบถ้วน ทั้งภาษาไทย และภาษาต่างประเทศ</p> <p>5.3 การสืบค้นข้อมูลที่ เป็นปัจจุบัน จากแหล่งต่าง ๆ22 มีส่วนร่วมในการอภิปรายในช่วงการทำกิจกรรม quick round/pre-post conference หรือ clinical teaching ที่แสดงถึงการมีความรู้ความเข้าใจในสาระสำคัญของศาสตร์ที่เป็นพื้นฐานชีวิตและวิทยาศาสตร์สุขภาพและศาสตร์ทางวิชาชีพพยาบาล</p> <p>3.1 ระบุจุดแข็ง จุดอ่อน และสิ่งที่ต้องพัฒนาตนเองเกี่ยวกับความรู้ และ ทักษะ การปฏิบัติงาน</p> <p>3.2 ค้นคว้าได้ตรงประเด็นจากหลากหลายแหล่งข้อมูล มาใช้ในการปฏิบัติ</p> <p>4.1 ปฏิสัมพันธ์ต่อ บุคลากรในทีมสุขภาพ เพื่อน อาจารย์ หรือผู้ใช้บริการ และการปรับตัวเชิงวิชาชีพ</p> <p>4.3 แสดงตนเป็นสมาชิกที่ดีของกลุ่ม ตามบทบาทของตน</p> <p>4.4 พัฒนาตนเองอย่างต่อเนื่อง</p> <p>5.2 ทักษะการสื่อสาร ด้านการฟัง พูด อ่านเขียน ที่ถูกต้อง ครบถ้วน ทั้งภาษาไทยและภาษาต่างประเทศ</p>

ลำดับ เหตุการณ์	รายละเอียดเหตุการณ์	ภาพที่ต้องการ/ เวลาที่ใช้	สมรรถนะที่ประเมิน
	<p>เล่มอื่นๆมาด้วยหรือ”</p> <p>นักศึกษา: “งานเยอะค่ะห้องสมุดก็ปิดเร็ว ก็เลยไม่มี เวลาค้นเพิ่ม“⁵³²¹</p> <p>อาจารย์: ลองสรุปสาระสำคัญเท่าที่หนูได้อ่านมาซิ</p> <p>นักศึกษา: จำไม่ค่อยได้ค่ะ ต้องอ่านถึงจะได้ค่ะ⁵²³</p> <p>อาจารย์: ต่อไปนี้เมื่ออ่านหนังสือเสร็จแล้วลองสรุป ใจความสำคัญเป็นภาษาของเราเองมานะ แล้วครั้งหน้า ครูจะขอ ดู และที่เวิร์ดเขาขอความช่วยเหลือมา เขา ขอให้พวกเราจัดบอร์ดให้ความรู้กับผู้ป่วยและญาติที่ หน้าตึก เนื้อหาเกี่ยวกับการส่งเสริมสุขภาพเพื่อ ป้องกันโรคที่เกิดจากพฤติกรรมสุขภาพไม่ดี ใครพอ ช่วยได้บ้าง ชัชชวาทัง</p> <p>นักศึกษาก็มหันหน้าไม่สบตาครูเมื่อครูเรียกชื่อเงยหน้า พูดว่า: หนูขอตัวนะค่ะ หนูไม่ถนัดเลยคะ</p> <p>อาจารย์: ครูขอจิตอาสาคะ ใครถนัดอะไรก็ทำสิ่งที่ ตัวเองถนัด ไม่ใช่ไม่ถนัดอะไรเลยจะได้ไม่ต้องช่วยไม่ ถูกนะ การอยู่ร่วมกันในสังคม โดยเฉพาะสังคมของพี นังพยาบาลเราจะต้องมึน้ำใจให้กันและกัน แผลงฝึก ให้ความอนุเคราะห์เราได้ฝึก เราก็ควรแสดงน้ำใจ ช่วยเหลือ</p>		

ภาคผนวก ค

คำสั่งที่ใช้ในการวิเคราะห์คุณภาพการประเมิน

1. คำสั่งที่ใช้ในการวิเคราะห์ทักษะการประเมินก่อน และหลังการฝึกอบรมของกลุ่มทดลองใช้โมดูลฯ

```
Title = nurse teacher's trail
Facets = 4 ; rater,task, item, time
unexpected=2
Positive = 1 ; only rater does greater score mean greater measure
Noncentered= 1 ; rater measures float
Vertical =4N,1N,2N,3N,1* ; put the control for the "rulers" in Table 6 here, if
not default
Usort=1,2,3*
Inter-rater = 1, 2, 3 ; rater facet is 1, agreements for facets 2 and 3
Arrange = N,m ; put the order for the measure Table 7 here
Models=
?B,?,?B,?,R4 ; rater and time interaction/bias.
*
Labels=
1,rater
1-6 ; 6 otherwise anonymous examinees
*
2, rater
1-8 ; 10 otherwise anonymous examinees
*
3,time,D; this is a dummy facet, used only for investigating interactions achores at 0
1 = time1,0 ; 2 time,pretest, posttest1, posttest2
2 = time2,0
*
4, item
1 = ethical and moral development
2 = knowledge
3 = cognitive skills
4 = interpersonal skills and responsibility
5 = analytical and communication skills
6 = practical skill for the professional
7 = good personality and precision at work
*
Dvalues =
4,1-7 ; place 1-7 in data facet 3 location
data=
; nurse teacher rating
1,1,1,2,2,2,2,1,2,1
1,2,1,2,3,2,2,3,3,1
1,3,1,3,3,3,3,3,3,4
1,4,1,4,4,4,4,4,4,4
.....
```

2. คำสั่งที่ใช้ในการวิเคราะห์ทักษะการประเมินก่อนการฝึกอบรม หลังการฝึกอบรมทันที และหลังการฝึกอบรม 3 เดือน ของกลุ่มใช้โมดูลต่างๆจริงโดยใช้ ฟาเซทที่ 1 เป็นฟาเซทของผู้ประเมินเป็นหลัก (positive=1)

```

Title = nurse teacher's rating2
Facets = 4 ; rater,ratee, trait, time
unexpected=2
Positive = 1 ; only rater does greater score mean greater measure
Noncentered= 1 ; rater measures float
Vertical =4N,2A,2N,1* ; put the control for the "rulers" in Table 6 here, if not
default
Usort=1,2,3*
Inter-rater = 1, 2, 3 ; rater facet is 1, agreements for facets 2 and 3
Arrange = N,m ; put the order for the measure Table 7 here
Models=
?B,?,?B,?,R4 ; rater and time interaction/bias.
*
Labels=
1,rater
1-11 ; 11 otherwise anonymous examinees
*
2, ratee
1-10 ; 10 otherwise anonymous examinees
*
3,time,D; this is a dummy facet, used only for investigating interactions achores at 0
1 = time1,0 ; 3 time,pretest, posttest1, posttest2
2 = time2,0
3 = time3,0
*
4, trait
1 = ethical and moral development
2 = knowledge
3 = cognitive skills
4 = interpersonal skills and responsibility
5 = analytical and communication skills
6 = practical skill for the professional
7 = good personality and precision at work
*
Dvalues =
4,1-7 ; place 1-7 in data facet 3 location
data=
; nurse teacher rating
1,1,1,1,2,2,4,2,3,4
1,2,1,4,1,2,4,4,4,4.....

```

3. คำสั่งที่ใช้ในการวิเคราะห์ทักษะการประเมินก่อนการฝึกอบรม เป็นการวัดครั้งที่ 1

ของกลุ่มใช้โมดูลฯจริงโดยใช้ ฟาเซทที่ 1 เป็นฟาเซทของผู้ประเมินเป็นหลัก (positive=1) ส่วนฟาเซทผู้ถูกประเมิน และฟาเซทสมรรถนะมีผลเป็นลบ เพื่อศึกษาผลของโมดูลต่อความคลาดเคลื่อนแบบกด ปล่อยคะแนน ความคลาดเคลื่อนแบบให้คะแนนเกาะกลุ่มตรงกลาง และความคลาดเคลื่อนแบบฮาโลในการวัดครั้งที่ 1

```

itle = nurse teacher's time 1
Facets = 3          ; rater,ratee, trait
unexpected=2
Positive = 1       ; only rater does greater score mean greater measure
Noncentered= 1    ; rater measures float
Vertical =         ; put the control for the "rulers" in Table 6 here, if not default
Arrange = N,m     ; put the order for the measure Table 7 here
Models=
?,?,?,R4; rater and time interaction/bias.
*
Labels=
1,rater
1-11 ; 11 otherwise anonymous examinees
*
2, ratee
1-10 ; 10 otherwise anonymous examinees
*
3, trait
1 = ethical and moral development
2 = knowledge
3 = cognitive skills
4 = interpersonal skills and responsibility
5 = analytical and communication skills
6 = practical skill for the professional
7 = good personality and precision at work
*
Dvalues =
3,1-7 ; place 1-7 in data facet 3 location
data=
; nurse teacher rating
1,1,1,2,2,4,2,3,4
1,2,4,1,2,4,4,4,4
1,3,3,4,4,1,4,2,2
1,4,4,1,2,4,4,4,1
.....

```

4. คำสั่งที่ใช้ในการวิเคราะห์ทักษะการประเมินหลังการฝึกอบรมทันที เป็นการวัดครั้งที่ 2

ของกลุ่มใช้โมดูลฯจริงโดยใช้ ฟาเซทที่ 1 เป็นฟาเซทของผู้ประเมินเป็นหลัก (positive=1) ส่วนฟาเซทผู้ถูกประเมิน และฟาเซทสมรรถนะมีผลเป็นลบ เพื่อศึกษาผลของโมดูลต่อความคลาดเคลื่อนแบบกด ปล่อยคะแนน ความคลาดเคลื่อนแบบให้คะแนนเกาะกลุ่มตรงกลาง และความคลาดเคลื่อนแบบฮาโลในการวัดครั้งที่ 2

```
Title = nurse teacher's time 2
Facets = 3 ; rater,task, item
unexpected=2
Positive = 1 ; only rater does greater score mean greater measure
Noncentered= 1 ; rater measures float
Vertical = ; put the control for the "rulers" in Table 6 here, if not default
Arrange = N,m ; put the order for the measure Table 7 here
Models=
?,?,?,R4; rater and time interaction/bias.
*
Labels=
1,rater
1-11 ; 11 otherwise anonymous examinees
*
2, task
1-10 ; 10 otherwise anonymous examinees
*
3, item
1 = ethical and moral development
2 = knowledge
3 = cognitive skills
4 = interpersonal skills and responsibility
5 = analytical and communication skills
6 = practical skill for the professional
7 = good personality and precision at work
*
Dvalues =
3,1-7 ; place 1-7 in data facet 3 location
data=
; nurse teacher rating

1,1,1,2,2,4,2,3,3
1,2,4,1,2,4,3,4,3
1,3,2,3,3,1,4,2,2
1,4,3,1,2,3,3,4,1
1,5,4,1,4,1,4,1,4
1,6,1,1,1,1,1,3,3
1,7,4,3,4,1,4,3,1
.....
```


ผนวก ง

ผลการวิเคราะห์ข้อมูล

1 ผลการวิเคราะห์ทักษะการประเมินก่อน และหลังการฝึกอบรมของกลุ่มทดลองใช้โมดูลฯ

1.1 การวิเคราะห์ผลของฟาเซทผู้ประเมิน

Table 7.1.1 rater Measurement Report (arranged by N).)

Total Correlation PtMea	Total Count PtExp	Obsvd Average N rater	Fair (M) Average	Model Measure	Infit S.E.	Outfit MnSq ZStd	Estim. MnSq ZStd	Discrm	
.86	342	112	3.05	3.01	1.92	.19	1.05 .3	1.07 .5	.92
.87	303	112	2.71	2.57	.60	.18	.95 -.3	.95 -.3	1.01
.82	337	112	3.01	2.95	1.73	.19	1.11 .8	1.13 .9	.90
.87	334	112	2.98	2.91	1.63	.19	1.05 .4	1.00 .0	.98
.86	310	112	2.77	2.65	.83	.18	.88 -.9	1.02 .1	1.05
.83	351	112	3.13	3.13	2.26	.20	.82 -1.3	.72 -2.0	1.27
.93		6 6							
.86	329.5	112.0	2.94	2.87	1.50	.19	.98 -.2	.98 -.1	
.04	17.2	.0	.15	.20	.59	.01	.10 .8	.13 .9	
.04	18.9	.0	.17	.22	.65	.01	.11 .9	.14 1.0	
Model, Populn: RMSE .19 Adj (True) S.D. .56 Separation 2.97 Strata 4.29 Reliability .90 Model, Sample: RMSE .19 Adj (True) S.D. .62 Separation 3.28 Strata 4.71 Reliability .92 Model, Fixed (all same) chi-square: 60.2 d.f.: 5 significance (probability): .00 Model, Random (normal) chi-square: 4.6 d.f.: 4 significance (probability): .33									

1.2 ปฏิสัมพันธ์ระหว่างฟาเซทผู้ประเมินในการวัด 2 ครั้ง

Bias/Interaction: 1. rater, 3. Time

-----+-----													
-----+-----													
-----+-----													
Target	Target		Obs-Exp	Context	Target		Obs-Exp	Context	Target				
Joint	Welch												
N	r	Measr	S.E.	Average	N	time	Measr	S.E.	Average				
S.E.	t	d.f.	Prob.						Contrast				
-----+-----													
-----+-----													
1	1	2.65	.30	.16	1	time1	1.29	.26	-.16	2	time2	1.36	.39
3.46	109	.0008											
2	2	-.06	.25	-.19	1	time1	1.29	.26	.19	2	time2	-1.35	.36
-3.72	109	.0003											
3	3	1.78	.27	.01	1	time1	1.71	.27	-.01	2	time2	.07	.38
.19	109	.8498											
4	4	1.49	.26	-.03	1	time1	1.78	.27	.04	2	time2	-.28	.38
-.75	109	.4535											
5	5	.58	.25	-.07	1	time1	1.09	.26	.07	2	time2	-.51	.36
-1.43	109	.1551											
6	6	2.83	.30	.12	1	time1	1.78	.27	-.11	2	time2	1.05	.40
2.59	109	.0108											
-----+-----													

2. ผลการวิเคราะห์ทักษะการประเมินก่อนการฝึกอบรม หลังการฝึกอบรมทันที และหลังการฝึกอบรม
 3 เดือนของกลุ่มใช้โมดูลฯจริงพิจารณาตามฟาเซท

2.1 ฟาเซทของผู้ประเมิน เป็นฟาเซทหลัก (positive=1) แปลผลเป็นบวก ค่ามากดี ค่าน้อยไม่ดี

Table 7.1.2 rater Measurement Report (arranged by mN).

Total Correlation		Total Score	Total Count	Obsvd Average	Fair (M) Average	Model Measure	Infit S.E.	Outfit MnSq	ZStd	Discrm	Estim. PtMea
608	210	2.90	2.95	.32	.06	1.12	1.6	1.10	1.1	1.08	.35
.35	10										
597	210	2.84	2.90	.28	.06	.98	-.2	1.01	.1	.89	.34
.36	4										
593	210	2.82	2.88	.26	.06	1.22	3.0	1.32	3.5	1.05	.29
.36	2										
580	210	2.76	2.81	.21	.06	.92	-1.2	.93	-.9	.92	.37
.37	5										
579	210	2.76	2.80	.21	.06	1.14	2.0	1.19	2.2	1.20	.37
.37	8										
576	210	2.74	2.79	.20	.06	1.09	1.4	1.11	1.3	1.34	.42
.37	1										
566	210	2.70	2.74	.16	.06	.95	-.7	.98	-.2	.91	.37
.37	11										
559	210	2.66	2.70	.13	.06	.86	-2.3	.88	-1.6	.81	.40
.38	9										
552	210	2.63	2.66	.10	.06	1.12	1.9	1.15	2.0	1.34	.40
.38	7										
540	210	2.57	2.60	.06	.06	.90	-1.6	.93	-.9	1.00	.41
.38	6										
483	210	2.30	2.29	-.16	.06	.67	-5.6	.69	-4.6	.51	.36
.40	3										
566.6	210.0	2.70	2.74	.16	.06	1.00	-.2	1.03	.2		.37
Mean (Count: 11)											
32.6	.0	.16	.17	.13	.00	.15	2.4	.16	2.2		.04
S.D. (Population)											
34.2	.0	.16	.18	.13	.00	.16	2.5	.17	2.3		.04
S.D. (Sample)											
Model, Populn: RMSE		.06	Adj (True)	S.D. .11	Separation	1.76	Strata	2.68			
Reliability		.76									
Model, Sample: RMSE		.06	Adj (True)	S.D. .12	Separation	1.88	Strata	2.83			
Reliability		.78									
Model, Fixed (all same)		chi-square: 44.9	d.f.: 10	significance (probability): .00							
Model, Random (normal)		chi-square: 8.2	d.f.: 9	significance (probability): .52							

2.2 ฟาเซทของผู้ถูกประเมิน แผลผลเป็น negative คือค่ามากไม่ดี ค่าน้อยดี

Table 7.2.1 ratee Measurement Report (arranged by N).

-----+-----												
-----+-----												
Total		Total	Obsvd	Fair (M)	Model		Infit	Outfit		Estim.		
Correlation												
Score	Count	Average	Average	Measure	S.E.	MnSq	ZStd	MnSq	ZStd	Discrm	PtMea	
PtExp	Nu task											
-----+-----												
571	231	2.47	2.47	.19	.06	.82	-3.7	.81	-3.8	1.31	.47	
.23	1 1											
710	231	3.07	3.09	-.28	.06	.93	-.9	.88	-1.5	1.31	.56	
.21	2 2											
590	231	2.55	2.56	.13	.06	1.01	.1	1.01	.1	-.20	-.07	
.23	3 3											
630	231	2.73	2.74	.00	.06	.99	-.1	1.03	.5	1.23	.29	
.23	4 4											
555	231	2.40	2.40	.25	.06	.92	-1.5	.91	-1.6	1.00	.28	
.23	5 5											
473	231	2.05	2.03	.52	.06	.85	-2.3	.82	-2.7	1.46	.64	
.22	6 6											
704	231	3.05	3.07	-.26	.06	1.25	3.2	1.33	3.8	.93	-.02	
.22	7 7											
611	231	2.65	2.65	-.06	.06	1.33	5.5	1.33	5.4	.93	-.02	
.23	8 8											
576	231	2.49	2.49	.18	.06	.88	-2.3	.88	-2.2	.23	.11	
.23	9 9											
813	231	3.52	3.54	-.80	.08	.98	.0	1.26	1.7	.91	-.15	
.16	10 10											
-----+-----												
623.3	231.0	2.70	2.70	.00	.06	1.00	-.2	1.03	.0		.21	
Mean (Count: 10)												
91.4	.0	.40	.41	.35	.01	.16	2.6	.20	2.8		.27	
S.D. (Population)												
96.3	.0	.42	.43	.37	.01	.17	2.8	.21	3.0		.28	
S.D. (Sample)												
-----+-----												
-----+-----												
Model, Populn:		RMSE	.06	Adj (True)	S.D.	.34	Separation	5.57	Strata	7.76		
Reliability		.97										
Model, Sample:		RMSE	.06	Adj (True)	S.D.	.36	Separation	5.88	Strata	8.17		
Reliability		.97										
Model, Fixed (all same) chi-square: 246.5 d.f.: 9 significance (probability): .00												
Model, Random (normal) chi-square: 8.7 d.f.: 8 significance (probability): .37												
-----+-----												

2.3 ฟาเซทของสมรรถนะ แปลผลเป็น negative คือค่ามากไม่ดี ค่าน้อยดี

Table 7.3.1 trait Measurement Report (arranged by N).

Total Score Item	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq ZStd	Outfit MnSq ZStd	Discrm	Estim.	Correlation	PtMea	PtExp	N
868	330	2.63	2.66	.06	.05	.80 -4.2	.79 -3.9	1.27	.51	.37			1
787	330	2.38	2.39	.25	.05	1.11 2.0	1.06 1.0	1.20	.39	.38			2
887	330	2.69	2.72	.01	.05	.63 -8.4	.62 -7.4	.89	.52	.36			3
859	330	2.60	2.63	.08	.05	1.11 2.1	1.09 1.5	1.29	.41	.37			4
876	330	2.65	2.69	.04	.05	.91 -1.9	.87 -2.3	1.22	.45	.37			5
959	330	2.91	2.96	-.17	.05	.98 -.2	.92 -1.2	1.13	.40	.34			6
997	330	3.02	3.08	-.27	.05	1.48 6.8	1.84 9.0	.36	-.16	.33			7
890.4 (Count: 7)	330.0	2.70	2.73	.00	.05	1.00 -.5	1.03 -.5		.36				Mean
63.8 (Population)	.0	.19	.21	.16	.00	.25 4.6	.36 4.8		.22				S.D.
68.9 (Sample)	.0	.21	.23	.17	.00	.27 5.0	.39 5.2		.23				S.D.
Model, Populn: RMSE .05 Adj (True) S.D. .15 Separation 3.03 Strata 4.38 Reliability .90 Model, Sample: RMSE .05 Adj (True) S.D. .17 Separation 3.30 Strata 4.74 Reliability .92 Model, Fixed (all same) chi-square: 69.0 d.f.: 6 significance (probability): .00 Model, Random (normal) chi-square: 5.5 d.f.: 5 significance (probability): .36													

2.4 ปฏิสัมพันธ์ระหว่างฟาเซทของผู้ประเมินกับการวัด 3 ครั้ง

Target Nu ra	Target Measr	Target S.E.	Obs-Exp Average	Context N time	Target Measr	Target S.E.	Obs-Exp Average	Context N time	Target Contrast	Target S.E.	Joint t	Welch d.f.	Welch Prob.
1 1	.23	.11	.04	1 time1	.08	.11	-.14	2 time2	.15	.15	.99	137	.3232
1 1	.23	.11	.04	1 time1	.28	.11	.10	3 time3	-.05	.15	-.31	137	.7573
1 1	.08	.11	-.14	2 time2	.28	.11	.10	3 time3	-.20	.15	-1.30	137	.1960
2 2	.22	.11	-.05	1 time1	.18	.11	-.10	2 time2	.04	.15	.23	137	.8182
2 2	.22	.11	-.05	1 time1	.39	.11	.15	3 time3	-.17	.16	-1.09	137	.2764
2 2	.18	.11	-.10	2 time2	.39	.11	.15	3 time3	-.21	.16	-1.32	137	.1886
3 3	-.22	.11	-.07	1 time1	-.41	.11	-.29	2 time2	.19	.16	1.18	137	.2391
3 3	-.22	.11	-.07	1 time1	.13	.11	.36	3 time3	-.35	.15	-2.27	137	.0248
3 3	-.41	.11	-.29	2 time2	.13	.11	.36	3 time3	-.54	.16	-3.41	137	.0009
4 4	.20	.11	-.10	1 time1	.14	.11	-.17	2 time2	.06	.15	.38	137	.7030
4 4	.20	.11	-.10	1 time1	.52	.12	.27	3 time3	-.33	.16	-2.05	137	.0427
4 4	.14	.11	-.17	2 time2	.52	.12	.27	3 time3	-.39	.16	-2.42	137	.0169
5 5	.23	.11	.02	1 time1	.22	.11	.01	2 time2	.01	.15	.08	137	.9387
5 5	.23	.11	.02	1 time1	.18	.11	-.03	3 time3	.05	.15	.31	137	.7591
5 5	.22	.11	.01	2 time2	.18	.11	-.03	3 time3	.04	.15	.23	137	.8182
6 6	-.04	.11	-.11	1 time1	.00	.11	-.07	2 time2	-.03	.15	-.23	137	.8203
6 6	-.04	.11	-.11	1 time1	.21	.11	.19	3 time3	-.24	.15	-1.59	137	.1138
6 6	.00	.11	-.07	2 time2	.21	.11	.19	3 time3	-.21	.15	-1.37	137	.1741
7 7	.02	.11	-.10	1 time1	.01	.11	-.11	2 time2	.01	.15	.08	138	.9397
7 7	.02	.11	-.10	1 time1	.28	.11	.21	3 time3	-.26	.15	-1.67	137	.0963
7 7	.01	.11	-.11	2 time2	.28	.11	.21	3 time3	-.27	.15	-1.75	137	.0825
8 8	.13	.11	-.10	1 time1	.10	.11	-.13	2 time2	.02	.15	.15	137	.8793
8 8	.13	.11	-.10	1 time1	.40	.11	.23	3 time3	-.28	.16	-1.78	137	.0777
8 8	.10	.11	-.13	2 time2	.40	.11	.23	3 time3	-.30	.16	-1.93	137	.0561
9 9	.26	.11	.15	1 time1	-.02	.11	-.19	2 time2	.28	.15	1.82	137	.0708
9 9	.26	.11	.15	1 time1	.16	.11	.04	3 time3	.09	.15	.61	137	.5400
9 9	-.02	.11	-.19	2 time2	.16	.11	.04	3 time3	-.18	.15	-1.21	137	.2273
10 10	.47	.12	.16	1 time1	.28	.11	-.05	2 time2	.19	.16	1.19	137	.2363
10 10	.47	.12	.16	1 time1	.23	.11	-.11	3 time3	.24	.16	1.50	137	.1369
10 10	.28	.11	-.05	2 time2	.23	.11	-.11	3 time3	.05	.15	.31	137	.7573
11 11	.15	.11	-.01	1 time1	-.01	.11	-.21	2 time2	.16	.15	1.06	137	.2904
11 11	.15	.11	-.01	1 time1	.34	.11	.22	3 time3	-.19	.16	-1.23	137	.2192
11 11	-.01	.11	-.21	2 time2	.34	.11	.22	3 time3	-.35	.15	-2.28	+	-----

3. ผลการวิเคราะห์ทักษะการประเมินก่อนการฝึกอบรม เป็นการวัดครั้งที่ 1

ของกลุ่มใช้โมดูลฯจริงโดยใช้ ฟาเซทที่ 1 เป็นฟาเซทของผู้ประเมินเป็นหลัก (positive=1) ส่วนฟาเซทผู้ถูกประเมิน และฟาเซทสมรรถนะมีผลเป็นลบ เพื่อศึกษาผลของโมดูลต่อความคลาดเคลื่อนแบบกด ปล่อยคะแนน ความคลาดเคลื่อนแบบให้คะแนนเกาะกลุ่มตรงกลาง และความคลาดเคลื่อนแบบฮาโลในการวัดครั้งที่ 1

Table 7.1.1 rater Measurement Report (arranged by mN) (ฟาเซทผู้ประเมิน)

Total Correlation	Total Score	Obsvd Count	Fair(Average)	Model Measure	Infit S.E.	Outfit MnSq	Estim. ZStd	Discrm	PtMea			
.33	214	70	3.06	3.11	.47	.12	1.10	.7	1.08	.5	.89	.28
	52.9	30.8	10	10								
.35	197	70	2.81	2.85	.25	.11	.89	-.8	.88	-.8	.63	.26
	48.4	30.4	9	9								
.36	178	63	2.83	2.83	.23	.12	.77	-1.9	.75	-1.8	.99	.44
	56.5	30.9	5	5								
.35	212	77	2.75	2.81	.22	.10	1.31	2.6	1.28	2.1	1.34	.34
	62.8	30.0	1	1								
.35	194	70	2.77	2.81	.21	.11	1.19	1.5	1.17	1.2	1.21	.36
	56.1	30.3	2	2								
.36	192	70	2.74	2.78	.19	.11	1.11	.9	1.08	.6	.90	.30
	64.4	30.3	4	4								
.36	188	70	2.69	2.71	.14	.11	.99	.0	.98	-.1	.94	.35
	62.1	30.1	11	11								
.36	186	70	2.66	2.68	.12	.11	1.13	1.1	1.09	.7	1.28	.39
	64.0	30.0	8	8								
.37	177	70	2.53	2.54	.01	.11	1.07	.6	1.06	.5	1.21	.38
	63.3	29.6	7	7								
.37	172	70	2.46	2.46	-.05	.11	.83	-1.5	.83	-1.4	.97	.40
	61.1	29.3	6	6								
.37	156	70	2.23	2.21	-.24	.11	.64	-3.4	.64	-3.2	.78	.39
	47.7	28.2	3	3								
	187.8	70.0	2.68	2.71	.14	.11	1.00	.0	.98	-.2		.35
		Mean (Count: 11)										
	16.2	3.0	.21	.23	.17	.00	.19	1.7	.18	1.5		.05
		S.D. (Population)										
	17.0	3.1	.22	.24	.18	.00	.20	1.8	.19	1.5		.05
		S.D. (Sample)										
Model, Populn: RMSE .11 Adj (True) S.D. .14 Separation 1.23 Strata 1.97												
Reliability (not inter-rater) .60												
Model, Sample: RMSE .11 Adj (True) S.D. .15 Separation 1.33 Strata 2.10												
Reliability (not inter-rater) .64												
Model, Fixed (all same) chi-square: 26.8 d.f.: 10 significance (probability): .00												
Model, Random (normal) chi-square: 7.3 d.f.: 9 significance (probability): .61												
Inter-Rater agreement opportunities: 3843 Exact agreements: 2236 = 58.2% Expected: 1152.9 = 30.0%												

Table 7.3.1 trait Measurement Report (arranged by mN).)(ฟ้าเขตสมรรถนะ)

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+																
Total		Total	Obsvd	Fair(M)	Model		Infit	Outfit		Estim.						
Correlation																
Score		Count	Average	Average	Measure	S.E.	MnSq	ZStd	MnSq	ZStd	Discrm	PtMea				
PtExp		N item														
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+																
	270	110	2.45	2.46		.19	.09		1.24	2.5	1.22	2.1		1.13		.33
.38 2 knowledge																
	283	110	2.57	2.59		.09	.09		1.47	4.6	1.54	4.8		-.60		-.22
.38 7 good personality and precision at work																
	285	110	2.59	2.61		.08	.09		1.11	1.1	1.09	.8		1.05		.36
.38 4 interpersonal skills and responsibility																
	296	110	2.69	2.72		.00	.09		.77	-2.7	.74	-2.8		1.65		.63
.37 1 ethical and moral development																
	299	110	2.72	2.75		-.03	.09		.85	-1.7	.81	-1.9		1.46		.55
.37 5 analytical and communication skills																
	311	110	2.83	2.87		-.12	.09		.57	-5.3	.59	-4.2		.94		.54
.37 3 cognitive skills																
	322	110	2.93	2.98		-.21	.09		1.01	.1	.94	-.4		1.18		.43
.36 6 practical skill for the professional																
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+																
	295.1	110.0	2.68	2.71		.00	.09		1.00	-.2	.99	-.2				.37
Mean (Count: 7)																
	16.3	.0	.15	.16		.13	.00		.28	3.1	.30	2.9				.26
S.D. (Population)																
	17.7	.0	.16	.18		.14	.00		.31	3.4	.32	3.1				.28
S.D. (Sample)																
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+																
Model, Populn: RMSE .09 Adj (True) S.D. .09 Separation 1.03 Strata 1.71																
Reliability .51																
Model, Sample: RMSE .09 Adj (True) S.D. .10 Separation 1.18 Strata 1.91																
Reliability .58																
Model, Fixed (all same) chi-square: 14.2 d.f.: 6 significance (probability): .03																
Model, Random (normal) chi-square: 4.2 d.f.: 5 significance (probability): .52																

4. ผลการวิเคราะห์ทักษะการประเมินหลังการฝึกอบรมทันที

ของกลุ่มใช้โมดูลฯจริงโดยใช้ ฟาเซทที่ 1 เป็นฟาเซทของผู้ประเมินเป็นหลัก (positive=1) ส่วนฟาเซทผู้ถูกประเมิน และฟาเซทสมรรถนะมีผลเป็นลบ เพื่อศึกษาผลของโมดูลต่อความคลาดเคลื่อนแบบกด ปล่อยคะแนน ความคลาดเคลื่อนแบบให้คะแนนเกาะกลุ่มตรงกลาง และความคลาดเคลื่อนแบบฮาโลในการวัดครั้งที่ 2

Table 7.1.1 rater Measurement Report (arranged by mN).)(ฟาเซทผู้ประเมิน)

Total Correlation	Total Score	Total Count	Obsvd Average	Fair(M) Average	Model Measure	Infit S.E.	Outfit ZStd	Estim. MnSq	Discrm ZStd	PtMea		
PtExp	Obs %	Exp %	Nu rater									
.37	199	70	2.84	2.89	.27	.11	1.33	2.4	1.27	1.8	1.40	.40
	60.3	30.3	10	10								
.39	176	63	2.79	2.83	.22	.12	.82	-1.4	.82	-1.2	.75	.40
	62.1	30.9	5	5								
.37	191	70	2.73	2.76	.17	.11	1.28	2.2	1.29	2.0	1.23	.34
	61.1	30.4	2	2								
.37	187	70	2.67	2.70	.12	.11	.94	-0.5	.93	-0.5	.74	.33
	66.7	30.4	4	4								
.38	184	70	2.63	2.65	.09	.11	1.16	1.3	1.15	1.2	1.06	.34
	64.9	30.4	8	8								
.36	200	77	2.60	2.62	.07	.10	1.06	.6	1.03	.3	1.23	.38
	68.3	30.1	1	1								
.38	176	70	2.51	2.52	-.01	.11	1.28	2.3	1.27	2.0	1.34	.34
	63.4	30.3	7	7								
.38	175	70	2.50	2.51	-.02	.11	.92	-0.7	.91	-0.7	.75	.34
	60.1	30.3	6	6								
.38	174	70	2.49	2.49	-.03	.11	.82	-1.7	.80	-1.6	.88	.41
	65.9	30.3	11	11								
.38	173	70	2.47	2.47	-.04	.11	.79	-2.0	.79	-1.8	.82	.44
	65.6	30.3	9	9								
.37	141	70	2.01	1.96	-.42	.11	.64	-3.0	.64	-2.6	.81	.38
	39.6	29.1	3	3								
	179.6	70.0	2.57	2.58	.04	.11	1.00	.0	.99	-.1		.37
			Mean (Count: 11)									
	15.4	3.0	.21	.24	.18	.00	.22	1.9	.22	1.6		.04
			S.D. (Population)									
	16.2	3.1	.22	.25	.19	.00	.23	2.0	.23	1.7		.04
			S.D. (Sample)									
Model, Populn: RMSE .11 Adj (True) S.D. .14 Separation 1.28 Strata 2.04												
Reliability (not inter-rater) .62												
Model, Sample: RMSE .11 Adj (True) S.D. .15 Separation 1.38 Strata 2.17												
Reliability (not inter-rater) .66												
Model, Fixed (all same) chi-square: 27.1 d.f.: 10 significance (probability): .00												
Model, Random (normal) chi-square: 7.3 d.f.: 9 significance (probability): .60												
Inter-Rater agreement opportunities: 3843 Exact agreements: 2370 = 61.7% Expected: 1162.4 = 30.2%												

Table 7.3.1 item Measurement Report (arranged by mN).)(พาเขตสมรรถนะ)

Total Correlation	Total Count	Obsvd Average	Fair(M) Average	Model Measure	Infit S.E.	Outfit MnSq ZStd	Estim. MnSq ZStd	Discrm	PtMea	
PtExp	N item									
.39	256	110	2.33	2.31	.20	.09	1.13 1.4	1.09 .8	1.25	.41
	2 knowledge									
.39	274	110	2.49	2.49	.06	.09	1.32 3.3	1.43 3.7	-.46	-.14
	7 good personality and precision at work									
.39	279	110	2.54	2.55	.03	.09	1.13 1.4	1.10 .9	1.26	.43
	4 interpersonal skills and responsibility									
.39	282	110	2.56	2.58	.00	.09	.93 -.7	.89 -1.0	1.30	.49
	5 analytical and communication skills									
.39	284	110	2.58	2.60	-.01	.09	.79 -2.5	.78 -2.3	1.48	.58
	1 ethical and moral development									
.39	288	110	2.62	2.64	-.04	.09	.64 -4.6	.64 -3.9	1.07	.59
	3 cognitive skills									
.38	313	110	2.85	2.89	-.23	.09	1.10 1.0	1.03 .3	1.03	.34
	6 practical skill for the professional									
	282.3	110.0	2.57	2.58	.00	.09	1.01 -.1	.99 -.2		.39
	Mean (Count: 7)									
	15.8	.0	.14	.16	.12	.00	.22 2.5	.24 2.3		.23
	S.D. (Population)									
	17.1	.0	.16	.18	.13	.00	.24 2.7	.25 2.5		.25
	S.D. (Sample)									
Model, Populn: RMSE .09 Adj (True) S.D. .08 Separation .95 Strata 1.60 Reliability .48										
Model, Sample: RMSE .09 Adj (True) S.D. .10 Separation 1.11 Strata 1.81 Reliability .55										
Model, Fixed (all same) chi-square: 13.0 d.f.: 6 significance (probability): .04										
Model, Random (normal) chi-square: 4.1 d.f.: 5 significance (probability): .53										

ภาคผนวก จ

ภาพกิจกรรมของการพัฒนาโมดูล และการนำโมดูลการฝึกอบรมไปใช้



ภาพที่ 1 การประชุมปรึกษาหารือของผู้ทรงคุณวุฒิในการให้คะแนนรูบริค ร่วมกับวีดิทัศน์สถานการณ์จำลองการขึ้นฝึกปฏิบัติงานของนักศึกษาพยาบาล ในวันที่ 11 กรกฎาคม 2558 ณ ห้องประชุม 1 คณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต



ภาพที่ 2 การทดลองใช้โมดูลการฝึกอบรมในกลุ่มทดลองใช้โมดูลในตัวอย่าง 6 คน ในวันที่ 20 -21 กรกฎาคม 2558



ภาพที่ 3 กิจกรรมการฝึกอบรมโดยใช้โมดูลการเสริมสร้างสมรรถนะการประเมินในตัวอย่าง 11 คน
ในวันที่ 28 -29 กรกฎาคม 2558



ประวัติผู้เขียนวิทยานิพนธ์

นางเนตรรัชนี กมลรัตนานันท์ เกิดวันที่ 29 พฤษภาคม พ.ศ. 2511 ที่จังหวัด แพร่ ประวัติการศึกษา จบการศึกษา ประกาศนียบัตรพยาบาลศาสตร์ และผดุงครรภ์ชั้นสูง (เทียบเท่าปริญญาตรี) พ.ศ. 2534 จากวิทยาลัยพยาบาลลำปาง สาธารณสุขศาสตร์บัณฑิต (อาชีวอนามัยและความปลอดภัย) พ.ศ. 2542 จากมหาวิทยาลัยสุโขทัยธรรมมาธิราช พยาบาลศาสตร์มหาบัณฑิต (การพยาบาลเวชปฏิบัติครอบครัว) พ.ศ. 2550 จาก มหาวิทยาลัยบูรพา ประวัติการทำงาน ปฏิบัติงานตำแหน่งพยาบาลวิชาชีพ งานห้องคลอด โรงพยาบาลลำปาง ตั้งแต่ พ.ศ. 2534-2539 ตำแหน่ง พยาบาลวิชาชีพ งานห้องคลอดโรงพยาบาลชลบุรี ตั้งแต่ พ.ศ. 2539-2547 ตำแหน่ง อาจารย์สังกัดคณะพยาบาลศาสตร์ มหาวิทยาลัยบูรพา พ.ศ. 2547-2548 อาจารย์ประจำสาขาวิชา การพยาบาลมารดาทารก และการผดุงครรภ์ คณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต พ.ศ. 2549 ถึงปัจจุบัน สถานที่ทำงาน คณะพยาบาลศาสตร์มหาวิทยาลัยสวนดุสิต 204/3 อาคารเฉลิมพระเกียรติ 50 พรรษามหาวิจิราลงกรณ ถนนสีรินธร แขวงบางพลัด เขตบางพลัด กรุงเทพมหานคร