

สรุป อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ สร้างและพัฒนาการสอบทางคลินิกเชิงโครงสร้างแบบปรนัย พร้อมทั้งพัฒนาวิธีการบริหารการสอบที่เหมาะสม ในการวัดทักษะทางคลินิกสำหรับนักศึกษาพยาบาล ซึ่งทักษะทางคลินิกที่วัดจะยึดตามแนวคิดของ รอส และ คณะ ซึ่งแบ่งทักษะทางคลินิกออกเป็น 3 องค์ประกอบคือ ความสามารถในการสืบค้นข้อมูลทางคลินิก ความสามารถในการแปลผลข้อมูลทางคลินิก ความสามารถในการสังเคราะห์เพื่อนำผลของข้อมูลทางคลินิกไปใช้ ครอบคลุมเนื้อหาเกี่ยวกับการปฏิบัติการพยาบาลในห้องผ่าตัด สร้างเป็นสถานีสอบทั้งสิ้น 10 สถานี โดยเป็นสถานีปฏิบัติการ 9 สถานี และสถานีคำถาม 1 สถานี กลุ่มตัวอย่างที่ใช้ เป็นนักศึกษาพยาบาลของวิทยาลัยพยาบาลบรมราชชนนีสืบปาง จำนวน 42 คน ซึ่งเลือกแบบเจาะจง จากนั้นศึกษาวิธีการบริหารการสอบที่เหมาะสม และจัดดำเนินการสอบ นำผลที่ได้มาวิเคราะห์ ศึกษาคุณภาพของสถานีสอบ โดยพิจารณาความตรงเชิงเนื้อหาของสถานีสอบจากผู้เชี่ยวชาญจำนวน 10 ท่าน และความตรงตามสภาพ (Concurrent Validity) สำหรับความเที่ยง พิจารณาจากค่าความเที่ยงของผู้ตรวจ 2 คน (Interrater Reliability) ความเที่ยงจากการสอบซ้ำ (Test-retest Reliability) ความสอดคล้องภายในสถานี และ ความสอดคล้องในแต่ละสถานี (Internal Consistency) นอกจากนี้ยังพิจารณาหาค่าความสัมพันธ์ระหว่างสถานี (Interstation Correlation) และค่าอำนาจจำแนกของสถานี (Discrimination Power) ได้ผลการวิจัยดังนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ผลการวิจัย

การบริหารการสอบ

1. จัดแบ่งการสอบออกเป็น 2 ส่วน โดย ส่วนแรกทำการสอบในห้องผ่าตัด โดยสอบสถานีปฏิบัติการ 6 สถานี ครั้งที่ 2 ทำการสอบ ที่ห้องปฏิบัติการวิทยาลัยพยาบาล 4 สถานี
2. สถานีปฏิบัติการทุกสถานีใช้เวลาเท่ากันคือสถานีละ 5 นาที ส่วนสถานีคำถามใช้เวลา 10 นาที รวมเวลาที่ใช้ในการสอบในแต่ละส่วนประมาณ 170 - 175 นาที
3. ในการดำเนินการสอบ นักศึกษาจะต้องวนเข้าสอบที่แต่ละจนครบทุกสถานี เมื่อนักศึกษาแต่ละคนสอบเสร็จจะถูกเก็บตัวอยู่ในห้องจนการสอบสิ้นสุดลง
4. ให้ผู้ประเมินสถานีละ 2 คนในการประเมิน โดยเป็นพยาบาลวิชาชีพ 1 คน และอาจารย์พยาบาล 1 คน ให้นำผลการผู้ประเมินทั้งสิ้น 20 คน

คุณภาพของแบบสอบ

1. ความตรงเชิงเนื้อหา จากการตัดสินของผู้เชี่ยวชาญ 10 ท่าน ได้ค่าความสอดคล้องเกิน 0.5 ทุกสถานี
2. ความตรงตามสภาพ เมื่อพิจารณาค่าคะแนนที่ได้จากสถานีปฏิบัติการ กับค่าคะแนนที่ได้จากผลการตัดสินของอาจารย์ในเทศก์ 3 ท่านที่ดูแลนักศึกษาในการฝึกภาคปฏิบัติ พบว่ามีค่าความสัมพันธ์กันในระดับปานกลาง มีค่าเท่ากับ 0.40 อย่างมีนัยสำคัญทางสถิติ ที่ระดับ .05
3. ความเที่ยงระหว่างผู้ประเมิน 2 คน จากการหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน จากคะแนนที่ได้จากการสังเกตของผู้ประเมินคนที่ 1 และผู้ประเมินคนที่ 2 พบว่า มีค่าความสัมพันธ์อยู่ในระดับปานกลาง ถึง สูง คือมีค่าอยู่ระหว่าง 0.61 ถึง 0.93 อย่างมีนัยสำคัญที่ .001
4. ความเที่ยงจากการสอบซ้ำ จากการหาค่าคะแนนเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน จากคะแนนที่ได้ในการสอบครั้งที่ 1 และคะแนนที่ได้ในการสอบครั้งที่ 2 พบว่า ใกล้เคียงกันทุกสถานี
5. ความสอดคล้องภายในสถานี จากการหาค่าสัมประสิทธิ์แอลฟาของครอนบาค ของสถานี ทั้ง 10 สถานี ได้ค่าสัมประสิทธิ์ความเที่ยง .36
6. ค่าความสอดคล้องในแต่ละสถานี มีค่าอยู่ระหว่าง .18 ถึง .93 โดยที่สถานีปฏิบัติการ เรื่องการล้างมือและการเช็ดมือ มีค่าความสอดคล้องสูงสุด และ สถานีคำถามเรื่องการวินิจฉัยปัญหาทางการพยาบาลมีค่าต่ำสุด

7. ความสัมพันธ์ระหว่างสถานี จากการหาค่าสัมประสิทธิ์สหสัมพันธ์ พบว่าสถานีส่วนใหญ่ ไม่มีค่าความสัมพันธ์กันระหว่างสถานี อย่างมีนัยสำคัญทางสถิติ ที่ระดับ .05 ยกเว้น สถานีปฏิบัติการ เรื่องการตรวจร่างกาย กับ สถานีปฏิบัติการเรื่องเทคนิคปราศจากเชื้อ มีค่าสัมประสิทธิ์สหสัมพันธ์ .54 ที่ระดับนัยสำคัญ .05

8. ค่าอำนาจจำแนกของสถานี จากการหาค่าความสัมพันธ์ระหว่างคะแนนรวมของทุกสถานี กับ คะแนนแต่ละสถานีของผู้ตอบทั้งหมด พบว่ามีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ระหว่าง 0.30 ถึง 0.69 อย่างมีนัยสำคัญทางสถิติ ที่ระดับ .05 โดยที่สถานีปฏิบัติการเรื่องการสวมเสื้อกาวน์และการใส่ถุงมือ มีค่าอำนาจจำแนกสูงสุด ส่วนสถานีคำถามเรื่องการวินิจฉัยปัญหาทางการแพทย์พยาบาลมีค่าอำนาจจำแนกต่ำสุด สำหรับสถานีปฏิบัติการเรื่องการซักประวัติ และ สถานีปฏิบัติการเรื่องการจัดทำเพื่อผ่าตัด ไม่มีค่าอำนาจจำแนก อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

การอภิปรายผล

จากการวิจัยพบว่า ในการจัดการสอบจะแบ่งสอบออกเป็น 2 ส่วนโดยส่วนแรกจะสอบในห้องผ่าตัด และส่วนที่ 2 จะจัดสอบในห้องปฏิบัติการพยาบาลที่วิทยาลัยพยาบาล ที่เป็นเช่นนี้เพราะในการจัดการสอบสถานีที่ใช้จะต้องมีบรรยากาศเหมือนจริง ดังนั้นผู้วิจัยจึงเลือกสอบในห้องผ่าตัด เพราะมีสภาพที่เป็นจริงมากที่สุด เนื่องจากในเวลาราชการ การให้บริการพยาบาลในห้องผ่าตัด จะต้องให้บริการทุกห้อง จึงทำให้ไม่สะดวกในการดำเนินการสอบ ประกอบกับการสอบในเวลาราชการมีบุคลากรที่จะให้เป็นผู้ประเมินมีไม่เพียงพอเพราะส่วนใหญ่จะติดภาระหน้าที่รับผิดชอบ ดังนั้นผู้วิจัยจึงต้องจัดการสอบขึ้นในช่วงวันหยุด เพราะจะทำให้มีความสะดวกสบายในการใช้สถานที่จัดสอบ จากการจัดสอบเป็น 2 ส่วน ทำให้เกิดปัญหาในเรื่องของกลุ่มตัวอย่างบางส่วนไม่สามารถมาทำการสอบได้ครบทั้ง 10 สถานี พบว่า การสอบในส่วนที่ 2 มีนักศึกษาที่ไม่สามารถมาสอบได้ 12 คน ผู้วิจัยจึงได้ให้นักศึกษาที่พร้อมมาสอบแทน จนครบ 30 คน ซึ่งหากในสภาพการณ์ที่มีการนำ การสอบทางคลินิกเชิงโครงสร้างแบบปรนัยไปใช้ในการประเมินผลจริง คงจะไม่เกิดเหตุการณ์เช่น ผู้วิจัยพบ คือการที่ผู้วิจัยไม่สามารถควบคุมให้กลุ่มตัวอย่างมาสอบได้ครบ เพราะการสอบครั้งนี้ไม่มีผลในการตัดสินใจการเรียนการสอน ส่วนรูปแบบการบริหารที่เหมาะสม คือ การให้นักศึกษาเริ่มสอบทีละคนดีกว่าการเริ่มสอบพร้อมกันหลายคน เนื่องจากสถานีสอบที่สร้างขึ้นมีลักษณะเป็นสถานการณ์ที่มีความต่อเนื่อง เช่น การล้างมือควรมาก่อนการใส่เสื้อกาวน์ ดังนั้นการให้นักศึกษาเริ่มสอบทีละคน จะทำให้นักศึกษาจะได้แสดงทักษะได้อย่างต่อเนื่องและคล้ายกับสภาพการเป็นจริงมากที่สุด ส่วนในเรื่องของเวลาที่ใช้ก็ไม่แตกต่างกันมากนัก เป็นที่น่าสังเกตว่าในกรณีที่สถานีสอบแต่ละสถานีไม่ต่อเนื่องกัน และกลุ่มนักศึกษาไม่มาก สถานีสอบน้อย การบริหารการสอบโดยให้นักศึกษาเริ่มสอบพร้อมกันหลายคน จะใช้เวลาในการสอบโดยนักศึกษาทุกคนสามารถผ่านทุกสถานี น้อยกว่า

การเริ่มสอบทีละคน สำหรับการเลือกให้ผู้ป่วยสามารถใช้ได้ทั้งผู้ป่วยจริง หรือ ผู้ป่วยจำลองหรือหุ่นจำลองตามความเหมาะสม สำหรับในการสอบครั้งนี้ผู้วิจัยเลือกให้ผู้ป่วยจำลอง และ การใช้หุ่นจำลอง ในการใช้ผู้ป่วยจำลองนั้น ควรจะใช้บุคคลที่มีความเข้าใจในบทบาทของผู้ป่วยที่จะให้แสดงออก และจะต้องหาได้ง่ายไม่ลำบากจนเกินไป สำหรับการวิจัยครั้งนี้ ผู้วิจัยเลือกนักศึกษาที่มีประสบการณ์ในการขึ้นฝึกปฏิบัติจริงในหอผู้ป่วย ซึ่งนักศึกษาเหล่านี้จะมีโอกาสได้พบเห็นผู้ป่วยจริงมาบ้างแล้วทำให้สามารถนึกถึงและแสดงบทบาทได้ใกล้เคียงกับสถานการณ์ที่กำหนด แต่การใช้ผู้ป่วยจำลองมีจุดอ่อนที่บางสถานการณ์ไม่สามารถแสดงอาการผิดปกติได้จริง ยกตัวอย่าง ในสถานีปฏิบัติการเรื่องการตรวจร่างกายที่ผู้วิจัยสร้างขึ้นผู้วิจัยจะให้นักศึกษาตรวจร่างกายผู้ป่วยหลังผ่าตัด ซึ่งในภาวะนั้น ผู้ป่วยจะต้องมีความผิดปกติภายในร่างกายเช่น ชีพจรเต้นเร็ว หรือความดันต่ำ ซึ่งการใช้ผู้ป่วยจำลอง ไม่สามารถทำได้ดังนั้นจึงควรชี้แจงให้นักศึกษาทราบด้วยว่าในการประเมินที่ไม่สามารถทำได้นักศึกษาจะต้องแสดงออกมาเป็นคำพูดเช่น ตรวจดูการเต้นของชีพจร นักศึกษาต้องบอกกับผู้ประเมินด้วยว่า ตรวจดูอย่างไร เช่นชีพจรเต้นเบา หรือเต้นเร็วมีความผิดปกติอย่างไร นักศึกษาจึงจะได้รับการประเมินในขั้นนั้น ส่วนทรัพยากรหรืออุปกรณ์ที่ใช้ในการสอบบางชนิดอาจต้องเสียค่าใช้จ่ายสูงหากให้นักศึกษาปฏิบัติจริงทุกคน เช่น ชุดเสื้อกาวน์ ถุงมือ เครื่องมือ หรือ น้ำยาฟอกผิวหนัง ดังนั้นในการแก้ปัญหา นี้ จะมีการจัดเตรียมอุปกรณ์ในแต่ละสถานีไว้เพียง 2-3 ชุดเท่านั้น โดยจะให้นักศึกษาที่เข้าสอบหมุนเวียนกันใช้ แต่การเตรียมชุดอุปกรณ์ต่างๆให้พร้อมและทันสมัยในแต่ละสถานีได้นั้น จะต้องมียุคคลที่คอยช่วยเหลือในการเตรียมอุปกรณ์ไว้ด้วย โดยบุคคลที่เป็นผู้เตรียมอุปกรณ์ไม่ควรเป็นผู้ประเมินในสถานีนั้นๆ เพราะหากให้ผู้ประเมินเป็นผู้เตรียมจะทำให้เสียเวลาและขาดสมาธิในการประเมินได้ สำหรับสถานีปฏิบัติการเรื่องการฟอกทำความสะอาดผิวหนัง น้ำยาที่ใช้ในขั้นตอนของการฟอกทำความสะอาดมีราคาค่อนข้างแพง ดังนั้นผู้วิจัยจึงให้นักศึกษาแสดงบทบาทสมมติ และใช้หุ่นจำลองคล้ายคนแทนผู้ป่วย โดยได้กำหนดเงื่อนไขในโจทย์คำสั่งว่า "สมมติให้ หุ่นที่เห็นเป็นผู้ป่วยจริง ให้นักศึกษาแสดงการฟอกทำความสะอาดผิวหนังให้ผู้ป่วย" แต่จากสังเกตพบว่า นักศึกษาส่วนใหญ่จะแสดงขั้นตอนการฟอกทำความสะอาด โดยไม่ได้สนใจผู้ป่วยที่เป็นหุ่นจำลอง ทำให้ได้คะแนนน้อยในส่วนที่เป็นเจตคติเช่นการพูด การสัมผัสผู้ป่วย ที่เป็นเช่นนี้อาจเป็นเพราะ การให้แสดงบทบาทสมมติกับหุ่นจำลองมีลักษณะไม่เหมือนจริง นักศึกษาส่วนใหญ่จึงไม่พูดกับหุ่นจำลองดังกล่าว

จากการศึกษารายละเอียดของสถานีปฏิบัติการแต่ละสถานี พบว่า ส่วนใหญ่นักศึกษาจะทำคะแนนได้สูงเกือบทุกสถานีการสอบ ยกเว้น สถานีปฏิบัติการเรื่องการซักประวัติ และการจัดทำเพื่อผ่าตัด นักศึกษาส่วนใหญ่จะได้คะแนนน้อยกว่าเกณฑ์ที่ตั้งไว้ แสดงให้เห็นว่าทักษะการปฏิบัติการพยาบาลส่วนใหญ่ที่นำมาทำสถานีสอบนักศึกษาสามารถปฏิบัติได้ มีนักศึกษาเพียงบางส่วนที่ไม่สามารถปฏิบัติได้ผ่านเกณฑ์ที่กำหนด การกำหนดการผ่านเกณฑ์ในแต่ละสถานีจะทำให้ นักศึกษา

สามารถทราบข้อบกพร่องในการฝึกปฏิบัติของตนเองได้ ส่วนอาจารย์นิเทศก์หรือผู้สอนก็สามารถวางแผนให้การสอนได้เหมาะสมกับความสามารถของนักศึกษา

สำหรับคุณภาพของแบบสอบพบว่า มีความตรงเชิงเนื้อหา (Content Validity) ซึ่งเป็นการหาคุณภาพของแบบสอบที่สร้างขึ้นว่าสามารถวัดเนื้อหาที่ต้องการวัดได้หรือไม่ โดยจะต้องอาศัยผู้เชี่ยวชาญในการตัดสินใจ ในการหาความตรงเชิงเนื้อหาผู้วิจัยได้นำวัสดุประสงค์เชิงพฤติกรรม สถานการณ์ และแบบประเมินที่ใช้ในแต่ละสถานีให้ผู้เชี่ยวชาญตรวจสอบ จากการหาความตรงตามเนื้อหาโดยให้ผู้เชี่ยวชาญ พบว่ามีค่าความสอดคล้องเกิน 0.5 ขึ้นไปทุกสถานี แสดงว่าแบบสอบที่สร้างขึ้นมีความตรงตามเนื้อหา สามารถวัดทักษะภาคปฏิบัติได้

สำหรับความตรงตามสภาพ (Concurrent Validity) เมื่อพิจารณาค่าคะแนนที่ได้จากสถานีปฏิบัติการกับค่าคะแนนที่ได้จากผลการตัดสินใจของอาจารย์นิเทศก์ 3 ท่าน ที่ดูแลนักศึกษาในการฝึกภาคปฏิบัติ พบว่ามีค่าสัมประสิทธิ์สหสัมพันธ์ 0.40 ที่ระดับนัยสำคัญ .05 แสดงว่าในการสอบทางคลินิกเชิงโครงสร้างแบบปรนัยสามารถวัดทักษะทางคลินิกของนักศึกษาได้ตรงตามความสามารถ โดยเทียบจากค่าคะแนนที่ได้จากผลการตัดสินใจของอาจารย์นิเทศก์ 3 ท่าน ที่ดูแลนักศึกษาในการฝึกภาคปฏิบัติ ที่เป็นเช่นนี้อาจเนื่องมาจากในการประเมินผลภาคปฏิบัติที่อาจารย์นิเทศก์ทั้ง 3 ท่าน ตัดสินนั้น จะเป็นการประเมินในลักษณะภาพรวม ซึ่งจะต้องพิจารณาหลายองค์ประกอบ เช่น ทักษะการปฏิบัติ การทำรายงานเฉพาะกรณี รวมไปถึงการประชุมกลุ่ม ซึ่งเมื่อเปรียบเทียบกับแบบประเมินที่ผู้วิจัยสร้างขึ้นนั้นเป็นเพียงเนื้อหาเพียงบางส่วนเท่านั้น ดังนั้นในการประเมินภาพรวมอาจจะไม่สอดคล้องกับแบบประเมินที่ผู้วิจัยสร้างขึ้น หรืออาจเป็นเพราะในช่วงที่ทำการเก็บข้อมูลกลุ่มตัวอย่างที่ให้อยู่ในระหว่างการเตรียมตัวสอบเพื่อรอรับใบประกอบวิชาชีพอาจทำให้นักศึกษามีความวิตกกังวลทำการสอบได้ไม่เต็มที่ ซึ่งสุทัศน์ สุกมลสันต์ (2530) ได้กล่าวไว้ว่า ความไม่พร้อมทางร่างกายและจิตใจในระหว่างการสอบเป็นปัจจัยข้อหนึ่ง ที่มีผลต่อความตรงของแบบสอบหรือนอกจากนี้ขนาดของกลุ่มตัวอย่างที่มีน้อย และเป็นลักษณะที่เหมือนกันทำให้มีผลต่อความตรงของแบบสอบได้เช่นกัน หรืออีกนัยหนึ่งหากขนาดและลักษณะของกลุ่มตัวอย่างที่ใช้เพิ่มขึ้น แต่พบว่าค่าความตรงตามสภาพมีน้อยอาจแสดงให้เห็นว่า การทำการประเมินทักษะทางคลินิกของนักศึกษาลักษณะนี้ไม่เหมาะสมเมื่อเทียบกับการใช้วิธีการสังเกตของผู้ประเมินจริงตามธรรมชาติก็ได้

สำหรับค่าความเที่ยงของผู้ประเมิน (Interraters Reliability) จากการหาค่าความสัมพันธ
ของคะแนนที่ได้จากผู้ประเมิน 2 คน พบว่ามีค่าความสัมพันธอยู่ระหว่าง 0.61-0.93 อย่างมีนัย
สำคัญทางสถิติที่ .001 ซึ่งถือว่าอยู่ในระดับปานกลางถึงสูง (ทวิวัฒน์ ปิตยานนท์, 2536) แสดงให้
เห็นว่าเครื่องมือที่สร้างขึ้นมีความชัดเจนในเรื่องของข้อรายการที่ใช้ประเมิน ซึ่งสอดคล้องกับงานวิจัย
ของ โรเบิร์ต และ นอร์แมน (Roberts & G. Norman, 1990) ซึ่งได้ค่าความเที่ยงอยู่ระหว่าง 0.80 ถึง
0.99 การที่ค่าความเที่ยงจากผู้ประเมิน 2 คน มีความสัมพันธ์กันสูง อาจเนื่องมาจาก สถานการณ์ที่
กำหนดขึ้นเพื่อให้นักศึกษาแสดงออกนั้นมีความชัดเจน นักศึกษาเข้าใจและแสดงพฤติกรรมออกมา
ได้ตรงกับแบบประเมินที่ใช้ ประกอบกับแบบประเมินที่ใช้มีความชัดเจน ผู้ประเมินอ่านแล้วเข้าใจตรง
กันจึงทำให้การประเมินที่ได้จากผู้ประเมิน 2 คนมีความสอดคล้องกัน

ความเที่ยงจากการสอบซ้ำ (Test-retest Reliability) จากการหาค่าคะแนนเฉลี่ยและส่วน
เบี่ยงเบนมาตรฐานของกาสอบครั้งที่ 1 และการสอบครั้งที่ 2 พบว่ามีค่าคะแนนใกล้เคียงกัน
เนื่องจากกลุ่มตัวอย่างที่ทำการสอบซ้ำมีน้อย ทำให้ไม่สามารถหาค่าความสัมพันธระหว่างการสอบ
ครั้งแรก กับการสอบครั้งที่ 2 แต่จากการศึกษาของโรเบิร์ต และ นอร์แมน (Roberts and Norman,
G., 1990) พบว่ามีค่าความสัมพันธอยู่ระหว่าง 0.66 ถึง 0.86 ที่เป็นดังนี้อาจเป็นเพราะเครื่องมือที่ใช้
เป็นแบบสอบที่สามารถกำหนดพฤติกรรมการแสดงออกของผู้สอบได้ โดยการกำหนดสถานการณ์
หรือเงื่อนไขให้ ซึ่งลักษณะเช่นนี้เป็นคุณสมบัติ ที่สำคัญของการสอบทางคลินิกเชิงโครงสร้างแบบ
ปรนัยที่ฮาเดนและคณะได้กล่าวไว้ว่า การสอบนี้สามารถกำหนดจุดประสงค์ไว้ล่วงหน้าได้ ทำให้
สามารถแก้ปัญหาความเที่ยงในการวัดที่เกิดขึ้นในธรรมชาติซึ่งปกติจะไม่สามารถควบคุมได้
(Harden. R.M, 1975) และเนื่องจากเวลาที่ใช้ในแต่ละสถานีใช้เวลาเพียงช่วงสั้นๆ ทำให้สามารถ
สอบซ้ำได้ทันที อาจทำให้ผู้สอบยังสามารถจำพฤติกรรมได้อยู่ จึงทำให้การแสดงพฤติกรรมเป็น
เช่นเดิมซึ่งก็มีผลทำให้ค่าความเที่ยงสูง

ความสอดคล้องภายในสถานี (Internal Consistency) โดยการหาค่าสัมประสิทธิ์แอลฟา
ของครอนบาค ได้ค่าสัมประสิทธิ์เท่ากับ .36 พบว่าค่าที่ได้อยู่ในระดับต่ำ สอดคล้องกับงานวิจัยของ
โรเบิร์ต และ นอร์แมน (Roberts & G. Norman, 1990) ที่ได้ค่า .0198 ซึ่งการที่ค่าที่ได้ต่ำอาจเนื่อง
มาจากสถานีแต่ละสถานี ถูกสร้างขึ้นเพื่อวัดลักษณะกิจกรรมที่แตกต่างกัน ทำให้แต่ละสถานีมีค่าความ
สอดคล้องกันน้อย แต่เมื่อพิจารณาความสอดคล้องภายในของแต่ละสถานีพบว่า มีค่าความสอดคล้อง
ของแต่ละสถานี อยู่ระหว่าง .18 ถึง .91 สถานีที่มีค่าความสอดคล้องสูงนั้นแสดงว่ามีค่าความ
เที่ยงสูง การที่สถานีมีค่าความเที่ยงสูงแสดงว่าข้อรายการในแต่ละสถานีวัดคุณลักษณะเดียวกันจึง
ทำให้ค่าที่ได้มีความสอดคล้องกัน สำหรับสถานีคำถามเรื่องการวินิจฉัยปัญหาทางการแพทย์บาลมี
ค่าความเที่ยงต่ำ อาจเนื่องมาจาก สถานีคำถามนี้จะวัดความสามารถย่อย 3 ส่วน แต่ละส่วนจะมี

จำนวนข้อกระทงที่ใช้วัดน้อย ซึ่งการที่มีจำนวนข้อกระทงที่น้อยจะมีผลทำให้ค่าความเที่ยงต่ำ ซึ่งสอดคล้องกับงานวิจัยของอาภรณ์ ชูดวง (2535) และดังคำกล่าวของเมเรม และ เลแมน (Mehrens and Lahman, 1978:100-103 อ้างถึงในอาภรณ์ ชูดวง, 2535) ที่ว่า "แบบสอบที่มีข้อกระทงมากจะมีค่าความเที่ยงสูงกว่าแบบสอบที่มีข้อกระทงน้อย"

ค่าความสัมพันธ์ระหว่างสถานี (Interstation Correlation) จากการหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่ได้ของแต่ละสถานี พบว่า ไม่มีความสัมพันธ์อย่างมีนัยสำคัญ ที่ระดับ .05 ความหมายตามค่าสถิติที่ได้จะแสดงให้เห็นว่าแต่ละสถานีไม่มีความสัมพันธ์กัน หรือแต่ละสถานีมีความสามารถในการวัดเฉพาะทักษะที่แตกต่างกัน ดังนั้นในการจัดการสอบทางคลินิกเชิงโครงสร้างแบบปรนัยเพื่อวัดทักษะทางคลินิกของนักศึกษาพยาบาล ควรนำสถานีสอบทุกสถานีมาใช้ในการประเมินผล สำหรับสถานีปฏิบัติการเรื่องการตรวจร่างกาย กับ สถานีปฏิบัติการเรื่องเทคนิคปราศจากเชื้อ ที่มีค่าความสัมพันธ์กันอยู่ในระดับปานกลางนั้น แสดงให้เห็นว่าสถานีคู่นี้มีความสัมพันธ์กัน อาจจะเลือกเพียงสถานีใดสถานีหนึ่งมาใช้ในการสอบก็ได้ แต่หากพิจารณาถึงเนื้อหาและวัตถุประสงค์ในการประเมินทักษะทางคลินิกของนักศึกษา ผู้วิจัยเห็นว่าไม่ควรนำสถานีใดสถานีหนึ่งออก เพราะจะทำให้การวัดไม่ครอบคลุมทักษะที่ต้องการประเมิน

สำหรับค่าอำนาจจำแนกของแต่ละสถานี พบว่า จากการหาค่าความสัมพันธ์ระหว่างคะแนนรวมของทุกสถานี กับ คะแนนแต่ละสถานีของผู้สอบ ได้ค่าความสัมพันธ์อยู่ระหว่าง 0.30 ถึง 0.69 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สถานีที่มีค่าความสัมพันธ์สูง แสดงว่าสถานีนั้นมีความสอดคล้องเป็นเนื้อเดียวกัน และเป็นสถานีที่มีความสามารถในการจำแนกได้สูงกว่าสถานีที่มีค่าความสัมพันธ์ต่ำ ซึ่งก็จะเป็นสถานีที่จำแนกได้ต่ำ สำหรับสถานีปฏิบัติการเรื่องการรักประวัติ และ สถานีปฏิบัติการเรื่องการจัดทำเพื่อผ่าตัด พบว่า ไม่มีค่าอำนาจจำแนกแสดงว่าสถานีนั้นไม่มีความสามารถในการจำแนกหรือไม่เป็นเนื้อเดียวกันกับสถานีสอบทั้งหมดสามารถตัดออกไปได้ เป็นที่น่าสังเกตว่าการที่ทั้งสองสถานีไม่มีค่าอำนาจจำแนกหรือมีค่าความสัมพันธ์ระหว่างสถานีกับคะแนนรวมต่ำ อาจเนื่องมาจากสถานีนั้นมีความยาก โดยเมื่อพิจารณาจากค่าสถิติที่ได้พบว่า ค่าคะแนนเฉลี่ยของคะแนนสอบของทั้ง 2 สถานีมีค่าค่อนข้างต่ำกว่าเกณฑ์ที่ตั้งไว้แสดงว่าสถานีมีความยาก นักศึกษาส่วนใหญ่ทำไม่ได้ ซึ่งการที่แบบสอบยากหรือง่ายเกินไปก็จะมีผลทำให้ค่าอำนาจจำแนกต่ำ (อนันต์ ศรีโสภณ, 2524) หรืออาจเป็นเพราะจำนวนกลุ่มตัวอย่างที่มีขนาดเล็ก และมีลักษณะกลุ่มใกล้เคียงกัน ซึ่งสอดคล้องกับ บุญชม ศรีสะอาด (2532) ที่กล่าวไว้ว่า การวิเคราะห์หาค่าอำนาจจำแนก ไม่ว่าจะเป็นเครื่องมือชนิดใด หรือหาโดยวิธีใดก็ตาม ควรวิเคราะห์จากผลการสอบ หรือการวัด หรือผลจากการสอบถาม จากคนจำนวนมาก จากหลายห้อง ซึ่งจะทำได้ค่าที่เที่ยงตรง

การวิเคราะห์จากผลการวัดเพียงห้องเดียว อาจเกิดปัญหาที่หาค่าอำนาจจำแนก ไม่ได้หรือได้ค่าอำนาจจำแนกต่ำมาก

นอกจากนี้จากการสอบถามความคิดเห็น เกี่ยวกับการจัดการสอบทางคลินิกเชิงโครงสร้างแบบปรนัย จากผู้ทำการประเมิน จากนักศึกษาผู้ทำการสอบ และจากการสังเกตของผู้วิจัย ได้ข้อสังเกตดังนี้

1. ผู้ทำการประเมินเห็นว่าลักษณะการสอบที่แยกย่อยเป็นแต่ละสถานี สามารถมองเห็นจุดบกพร่องของนักศึกษาได้ยกตัวอย่าง เช่น ในสถานีปฏิบัติการเรื่องการสวมกาวฟันและใส่ถุงมือ พบว่ามีนักศึกษบางคนปฏิบัติผิดขั้นตอน และบางคนปฏิบัติได้ถูกขั้นตอนแต่ผิดหลักปรากฏจากเชื้อคือ มีการปนเปื้อนในระหว่างกาปฏิบัติ สามารถทำให้ทราบข้อบกพร่องที่เกิดขึ้นและทำการแก้ไขได้

2. ในระหว่างดำเนินการสอบ พบว่า ในบางสถานีเช่น สถานีปฏิบัติการเรื่องเทคนิคปรากฏจากเชื้อ นักศึกษาส่วนใหญ่จะทำเสร็จก่อนเวลา ก็ไม่ควรอนุญาตให้นักศึกษาออกจากสถานีไปก่อน เพราะการให้นักศึกษาออกไปรอในสถานีต่อไป นักศึกษาอาจมีโอกาสได้สังเกตการสอบก่อนและเตรียมตัวล่วงหน้าได้ นอกจากนี้ในระหว่างดำเนินการสอบจะต้องดำเนินการสอบติดต่อกันตลอดเวลาสอบ หากต้องสอบเป็นระยะเวลาานอาจต้องพิจารณาช่วงเวลาการพักให้ผู้ประเมินด้วย ซึ่งอาจจะเป็นเวลาสั้นๆ ประมาณ 5 ถึง 10 นาที

3. จากการซักถามความรู้สึกของนักศึกษาเกี่ยวกับการสอบ นักศึกษาส่วนใหญ่รู้สึกว่าการจัดสอบนี้มีบรรยากาศที่เหมือนจริงมาก และนักศึกษามีความรู้สึกลึกซึ้งตื้นตันในการเข้าสอบแต่ละสถานี จนบางครั้งนักศึกษาคงลืมสถานการณ์หรือคำสั่งที่ให้นักศึกษาปฏิบัติจนทำให้ได้คะแนนในสถานีนั้นต่ำ และนอกจากนี้ในการบริหารการสอบ จะต้องให้นักศึกษาที่สอบเสร็จนั่งรอในห้องที่จัดไว้ให้เพื่อป้องกันการรั่วของข้อสอบ พบว่านักศึกษบางคนบอกว่ารู้สึกเบื่อในช่วงเวลาที่ต้องนั่งรอเพื่อนๆ ดังนั้นในการบริหารการสอบอาจจัดให้มีข้อสอบที่มีความหลากหลายในแต่ละสถานี แต่สามารถวัดในเรื่องเดียวกัน เพื่อแก้ไขในเรื่องของการนั่งรอภายหลังจากสอบเสร็จ

4. จากสภาพการประเมินผลภาคปฏิบัติที่เป็นอยู่ในปัจจุบัน เป็นการประเมินผลที่ให้นักศึกษาแสดงพฤติกรรมในสถานการณ์จริง ซึ่งการประเมินในสภาพจริงมีจุดอ่อนอยู่ที่ ผู้วัดไม่สามารถควบคุมเหตุการณ์ที่เกิดขึ้นได้ สภาพการปฏิบัติการอาจมีการเปลี่ยนแปลงทุกครั้งที่มีการสอบ ทำให้การวัดทักษะการปฏิบัติงานของผู้เรียนทำให้อุปกรณ์ได้ยาก (สุวิมล วงศ์วานิช, 2534) ผลของการวัดอาจไม่คงเส้นคงวา เนื่องจากองค์ประกอบภายนอกเข้ามามีอิทธิพลต่อการวัด ทำให้

สถานการณ์เปลี่ยนแปลงจากที่ควรจะเป็น และนอกจากนี้ยังขึ้นอยู่กับภาระของอาจารย์นิเทศก์ หรือพยาบาลบนหอผู้ป่วย ซึ่งจะต้องให้การดูแลสังเกตอย่างใกล้ชิดเพื่อให้ได้ข้อมูลที่เป็นจริง แต่ในความเป็นจริงแล้วในบางสถานการณ์ นักศึกษาบางคนอาจมีโอกาสดำเนินการหรือความคิด ในการแก้ปัญหาในการปฏิบัติการพยาบาลให้อาจารย์เห็น หรือ นักศึกษาบางคนอาจไม่มีโอกาสได้ แสดงให้อาจารย์ได้เห็นเพราะไม่เกิดสถานการณ์เช่นนั้นให้ได้ปฏิบัติ ดังนั้นในการประเมินผลโดยการ สังเกตพฤติกรรมกรรมการแสดงออกของผู้เรียนในแบบเดิมที่ปฏิบัติอยู่ ก็อาจทำให้การประเมินไม่สามารถ ประเมินความสามารถที่แท้จริงของผู้เรียน แต่การดำเนินการสร้างและพัฒนาการสอบคลินิกเชิงโครง สร้างแบบปรนัย ซึ่งผู้วิจัยเห็นว่าการสอบชนิดนี้เป็นการสอบที่ สามารถที่จะแก้ไขจุดอ่อนของการ ประเมินผลภาคปฏิบัติแบบเดิมได้ เพราะในการสอบนักศึกษาที่จะต้องปฏิบัติจริงเช่นกัน แต่ภายใต้ เงื่อนไขและสภาวะการณ์ที่ผู้ประเมินสามารถกำหนดขึ้นมาได้ว่าการวัดสิ่งใด ทำให้สามารถ จัดการสอบได้ตรงตามความต้องการที่จะวัด นอกจากนี้การจัดการสอบภายใต้เงื่อนไขที่กำหนด สามารถให้นักศึกษาแสดงพฤติกรรมซ้ำได้อีก ผู้วัดสามารถประเมินทักษะที่ต้องการวัดซ้ำๆ ได้ และ ผู้เรียนก็สามารถแสดงพฤติกรรมออกมาให้เห็นได้ตามความสามารถที่แท้จริง นอกจากนี้ในการสอบ ทางคลินิกเชิงโครงสร้างแบบปรนัยอาจสามารถช่วยแก้ไขปัญหาในเรื่องของความไม่พอเพียงของ ผู้ป่วยในการฝึกปฏิบัติจริง ดังนั้นในการประเมินผลภาคปฏิบัติอาจนำการสอบประเภทนี้มาใช้แทน การประเมินในสภาพจริงในสถานการณ์ซึ่งมีความซับซ้อนยุ่งยาก หรือสถานการณ์ที่ขาดแคลนผู้ป่วย อาจจะไม่เหมาะสมมากกว่า แต่ก็ไม่ได้หมายความว่า การสอบนี้สามารถใช้ได้ในทุกสถานการณ์ ขึ้นอยู่ กับความเหมาะสมของแต่ละบริบทของหน่วยงาน หรือแต่ละเนื้อหาที่ใช้

5. ในการสร้างสถานีสอบในการวัดพฤติกรรมที่ซับซ้อน ความยุ่งยากและความซับซ้อนใน การวัดและการประเมินก็มีมาก ดังนั้นในการจะสร้างสถานการณ์ หรือแบบประเมินจึงต้องอาศัย ผู้มีประสบการณ์ ความชำนาญในเนื้อหา และ ผู้ที่มีประสบการณ์และมีทักษะในการออกข้อสอบ จึงจะทำให้ข้อสอบที่สร้างขึ้นสามารถวัดทักษะที่ซับซ้อนได้ การที่จะเป็นผู้เชี่ยวชาญในเนื้อหา และ ทักษะการออกข้อสอบนั้นจะต้องอาศัยความรู้และความสามารถสูง ดังนั้นในการสร้างและ พัฒนาการสอบในการวัดระดับสูงจึงเป็นเรื่องยากที่จะทำได้ การใช้ผู้เชี่ยวชาญจึงมีความสำคัญมาก ในการสร้างแบบสอบประเภทนี้

6. นอกจากนี้ในเรื่องของการให้คะแนน การประเมินผลพฤติกรรมเพียง การทำ หรือ ไม่ทำ เหมาะในการที่จะประเมินเพื่อตัดสินผล (summative evaluation) แต่หากต้องการใช้ผลการประเมิน เพื่อวินิจฉัยข้อบกพร่อง อาจกำหนดรูปแบบการให้คะแนนเป็นแบบมาตราประมาณค่าที่มีระดับ ความถูกต้องมากน้อย ซึ่งการทำเช่นนี้อาจจะทำให้ผลการวัดที่ได้วินิจฉัยความบกพร่องจากการเรียน ได้อีกด้วย

ข้อเสนอแนะ

ข้อเสนอแนะในการนำไปใช้

เนื่องจากในการสอบทางคลินิกเชิงโครงสร้างแบบปรนัยที่สร้างขึ้นมีทั้งหมด 10 สถานี จำนวนสถานีที่ค่อนข้างมาก การดำเนินการสอบต้องใช้บุคลากรมาก หากมีผู้ที่ต้องการจะนำไปใช้ อาจไม่จำเป็นต้องนำไปใช้ทั้งหมด สามารถที่จะนำบางสถานีไปใช้ได้ โดยให้ท่านพิจารณาเลือก สถานีที่ไม่มีความสัมพันธ์ระหว่างสถานี หรือสถานีที่มีค่าความสอดคล้องภายในสูง หรือสถานีที่มี ค่าอำนาจจำแนกสูง ไปใช้ในการสอบได้

ข้อเสนอแนะในการทำวิจัย

1. เนื่องจากการวิจัยในครั้งนี้มีความจำกัดในเรื่องของขนาดของกลุ่มตัวอย่าง และลักษณะ ของกลุ่มตัวอย่างที่ใกล้เคียงกัน ทำให้ค่าที่ได้ อาจมีความคลาดเคลื่อน ดังนั้นจึงควรมีการศึกษาและ พัฒนาการสอบทางคลินิกเชิงโครงสร้างแบบปรนัย โดยใช้กลุ่มตัวอย่างที่มากขึ้นและมีความแตกต่างกันมากขึ้น
2. ควรมีการการสร้างและพัฒนารสอบทางคลินิกเชิงโครงสร้างแบบปรนัย ในการวัด ทักษะทางคลินิก ของนักศึกษาพยาบาลในเนื้อหาและบริบทอื่นๆ เพื่อยืนยันว่าการสอบนี้สามารถวัด ทักษะทางคลินิกของนักศึกษาได้จริง และ มีประสิทธิภาพเพียงพอในการประเมินผลภาคปฏิบัติ
3. เนื้อหาเกี่ยวกับการฝึกปฏิบัติการพยาบาลในห้องผ่าตัด ซึ่งมีทักษะการปฏิบัติค่อนข้าง มาก ดังนั้นสถานีการสอบที่สร้างขึ้นส่วนใหญ่จะเป็นสถานีปฏิบัติการ ส่วนสถานีคำถามจะมีเพียง สถานีเดียว ดังนั้นอาจจะมีการสร้างและพัฒนารสอบทางคลินิกเชิงโครงสร้างแบบปรนัยที่มีสถานี ทั้ง 2 ท่อๆกัน และศึกษาคุณภาพที่ได้จากสถานีทั้ง 2
4. การสอบทางคลินิกเชิงโครงสร้างแบบปรนัยในการวัดทักษะทางคลินิกของนักศึกษา พยาบาลยังมีการศึกษากันน้อย ดังนั้นจึงควรจะมีศึกษาถึงปัจจัยหรืออิทธิพลอื่นๆที่น่าจะมีผลต่อ คุณภาพของแบบสอบ