บทที่ 2
วรรณคดีที่เกี่ยวข้อง

ในบทนี้จะได้กล่าวถึงเอกสารและงานวิจัยที่เกี่ยวข้อง โดยผู้วิจัยจะแบ่งเนื้อหาสำหรับนำเสนอตามลำดับขั้นดังต่อไปนี้ คือ

1. คลังข้อสอบ
2. ทฤษฎีการตอบสนองต่อข้อสอบ
3. การแจกแจงปกติ
4. งานวิจัยที่เกี่ยวข้อง

คลังข้อสอบ (Item Bank)

เนื่องจากการพัฒนาเกณฑ์ในการเลือกข้อสอบเป็นขั้นตอนที่สำคัญขั้นตอนหนึ่งของการเลือกข้อสอบจากคลังข้อสอบมาประกอบเป็นแบบสอบเพื่อนำไปใช้ตามจุดมุ่งหมายของการสอบ ดังนั้นจึงขอเสนอรายละเอียดเกี่ยวกับคลังข้อสอบดังต่อไปนี้

1. ความเป็นมาของคลังข้อสอบ

คลังข้อสอบเริ่มใช้ในประเทศอังกฤษในช่วงกลางปี ค.ศ. 1960 เป็นการรวมกลุ่มของข้อสอบมารวบรวมไว้ โดยนำมาจัดกระทำจำแนกออกเป็นชนิดและประเภทต่างๆ อย่างเป็นระบบเช่นเดียวกับหนังสือในห้องสมุด ความมุ่งหมายสำคัญของคลังข้อสอบก็เพื่อให้เกิดความคล่องตัวในการดึงข้อสอบเหล่านั้นมาใช้ในการสร้างแบบสอบได้หลายๆ ฟอร์มที่เทียบเท่ากัน นิยามของคลังข้อสอบไม่ใช่แต่เพียงการรวบรวมข้อสอบเข้ามาเก็บไว้อย่างเป็นระบบเท่านั้นแต่คลังข้อสอบเป็นกลุ่มของข้อคำถามจำนวนมากที่มีความสัมพันธ์กันและข้อสอบนี้มีความสะดวกในการนำออกมาใช้ ในการสร้างแบบสอบตามวัตถุประสงค์ ความสะดวกในการ

นำออกมาใช้หมายความว่า ข้อสอบเหล่านั้นได้ถูกจัดกระทำอย่างมีระบบมีดัชนีในการค้นหามีโครงสร้างที่แสดงถึงสิ่งที่ข้อสอบนี้ใช้วัดและตลอดจนข้อมูลอื่นๆ ที่ทำให้เกิดความสะดวกในการเลือกข้อสอบเหล่านี้มาใช้ในการสร้างแบบสอบชุดหนึ่งๆ ข้อมูลเหล่านี้ได้แก่ คุณลักษณะของข้อสอบ เช่น ประเภทของวัตถุประสงค์หรือเนื้อเรื่องที่ข้อสอบวัดและลักษณะในเรื่องค่าสถิติรายข้อ เช่น ค่าความยาก และอำนาจจำแนก จากนิยามของคลังข้อสอบจึงพบประเด็นปัญหาสำคัญของการสร้างคลังข้อสอบว่ามิใช่อยู่ที่การรวบรวมข้อสอบมาเก็บไว้อย่างเป็นระบบเท่านั้น หากแต่เป็นปัญหาที่เกี่ยวเนื่องไปถึงวิธีการในการเลือกข้อสอบออกมาจากคลังข้อสอบอย่างไร จึงจะนำมาสร้างแบบสอบที่มีคุณภาพตามวัตถุประสงค์ของการนำไปใช้และการพิจารณาคัดเลือกข้อสอบดังกล่าวควรจะต้องพิจารณาจากเกณฑ์ใดบ้าง

คลังข้อสอบกำลังเป็นสิ่งที่ทั้งครู ผู้บริหาร ศึกษานิเทศก์ ทั้งในระดับโรงเรียน กลุ่มโรงเรียน เขตการศึกษา ให้ความสนใจและจัดดำเนินการให้มีคลังข้อสอบขึ้น ทั้งนี้ก็ด้วยการเห็นประโยชน์ของคลังข้อสอบที่จะช่วยก่อให้เกิดการสร้างแบบสอบที่เป็นมาตรฐานในการประเมินระดับผลสัมฤทธิ์ในการเรียนของนักเรียน ดังเช่นที่กระทรวงศึกษาธิการได้มอบหมายให้ สำนักทดสอบทางการศึกษา กรมวิชาการ เป็นผู้ดำเนินการ ตรวจสอบผลสัมฤทธิ์ทางการเรียนของนักเรียนในเขตการศึกษาต่างๆ ในระดับชั้นประโยค เพื่อเป็นการควบคุมมาตรฐานในการจัดการศึกษาของโรงเรียนต่างๆ ซึ่งสำนักทดสอบทางการศึกษาก็ได้จัดทำคลังข้อสอบเพื่อรวบรวมข้อสอบที่ดีที่ได้มาจากการคัดสรรข้อสอบที่นำไปทดลองใช้และวิเคราะห์คุณภาพของข้อสอบแล้ว ส่วนในระดับโรงเรียนและกลุ่มโรงเรียนได้มีการจัดทำคลังข้อสอบขึ้นโดยทั่วไป ทั้งนี้ก็เพราะการเห็นประโยชน์ของการใช้คลังข้อสอบสร้างแบบสอบได้รวดเร็วและมีคุณภาพ นอกจากนี้การเก็บข้อสอบที่เป็นระบบจะช่วยให้ครูเกิดแนวความคิดในการเขียนจุดประสงค์การเรียนรู้ ช่วยให้การสอบวัดในแต่ละจุดประสงค์มีความตรง (valid) มากขึ้น ช่วยให้เกิดแนวทางในการสร้างแบบสอบเพื่อนำผลการสอบมาใช้ในการปรับปรุงการเรียนการสอน (formative evaluation) และการสร้างแบบสอบเพื่อนำผลการสอบมาใช้สรุปผลการเรียน (summative evaluation) และทำให้เกิดแนวทางในการสร้างแบบสอบคู่ขนาน (parallel test) เพื่อใช้ในการสอบแก้ตัว หรือการสร้าง

แบบสอบเพื่อการวินิจฉัยผลการเรียนของนักเรียนได้อย่างกว้างขวาง การตื่นตัวในการดำเนินการสร้างคลังข้อสอบมิใช่แต่การเห็นถึงประโยชน์ของคลังข้อสอบเท่านั้นหากแต่เป็นนโยบายในการจัดการศึกษาที่ให้โรงเรียนหรือกลุ่มโรงเรียนได้จัดทำคลังข้อสอบของทุกหมวดวิชา โดยถือเป็นเกณฑ์มาตรฐานสำหรับการประเมินโรงเรียนโดยเฉพาะในเกณฑ์มาตรฐานโรงเรียนมัธยมศึกษา ของกรมสามัญศึกษาที่มีเกณฑ์ว่า โรงเรียนจะผ่านเกณฑ์มาตรฐานได้ต้องมีการวิเคราะห์ข้อสอบโดยจะได้ 3 คะแนน ถ้ามีการวิเคราะห์ข้อสอบทุกหมวดวิชาจะได้ 4 คะแนน และถ้ามีการจัดทำคลังข้อสอบทุกหมวดวิชาจะได้คะแนนเต็ม 5 คะแนน (กนก จันทร์ขจร 2533 : 33) ความพยายามในการสร้างคลังข้อสอบมิใช่แต่ในระดับโรงเรียนเท่านั้น ในส่วนของระดับมหาวิทยาลัยเองก็ได้มีการสนับสนุนให้มีการสร้างคลังข้อสอบขึ้นอย่างจริงจัง ดังเช่น สถาบันภาษา และศูนย์วิจัยการเรียนการสอน คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ได้ให้ทุนสนับสนุนในการสร้างคลังข้อสอบภาษาอังกฤษ (สุพัฒน์ สุกมลสันต์ 2530 : 1) สถาบันวิจัยและพัฒนา มหาวิทยาลัยเกษตรศาสตร์ได้ให้ทุนสนับสนุนในการสร้างคลังข้อสอบเป็นโครงการต่อเนื่องระยะยาวถึง 3 ปี (ภาวิณี ศรีสุขวัฒนานันท์ 2532 : 1) มหาวิทยาลัยสงขลานครินทร์ได้รับทุนสนับสนุนจากสำนักงานการศึกษาเอกชน ให้สร้างคลังข้อสอบวิชาภาษาไทยเพื่อคัดเลือกมาเป็นข้อสอบมาตรฐานในการประเมินนักเรียนในโรงเรียนเอกชนสอนศาสนาอิสลามในจังหวัดชายแดนภาคใต้ ต่อเนื่องถึง 3 ปี (ชิดชนก เชิงเชาว์ 2532 : 1)

คลังข้อสอบแสดงถึงกลุ่มของข้อสอบที่นำมาจัดกระทำจำแนกออกเป็นชนิดและประเภทต่างๆ อย่างเป็นระบบเช่นเดียวกับหนังสือในห้องสมุด โดยมีความมุ่งหมายให้เกิดความคล่องตัวในการดึงข้อสอบเหล่านั้นมาใช้ในการสร้างแบบสอบวัดทางสมอง (mental-test) จริงๆ แล้วคำว่าคลัง (bank) ไม่ค่อยตรงประเด็นกับการจัดกระทำและดำเนินการใช้ข้อสอบในมโนทัศน์ของการสร้างคลังข้อสอบ คำที่ดูจะมีความหมายเหมาะสมมากกว่าควรใช้คำว่า ห้องสมุด (library) (Choppin 1985 : 1)

คำว่า item banks อาจจะเรียกชื่อแตกต่างกันไป ได้แก่ question bank, item pool, item collection, item reservoirs, test item libraries ซึ่งมีความหมายเดียวกัน

เราสามารถนิยามความหมายของคลังข้อสอบ ได้ว่าเป็นกลุ่มของข้อคำถามจำนวนมากโดยข้อคำถามเหล่านี้ถูกจัดขึ้นเพื่อสะดวกในการเลือกนำออกมาใช้เพื่อการสร้างแบบสอบตามวัตถุประสงค์ต่างๆ กล่าวโดยสรุปเมื่อพูดถึงคลังข้อสอบจะหมายถึงการรวบรวมข้อสอบจำนวนมากมาเก็บไว้โดยมีการจัดระบบและกำหนดความหมายทางสถิติกับข้อสอบแต่ละข้ออย่างชัดเจน ไม่ใช่เพียงแต่การรวบรวมข้อสอบจำนวนมากมาเก็บไว้เท่านั้น (Wright and Bell 1984 : 2)

การปรับเทียบมาตรา (calibrate) ข้อสอบ ก็คือกระบวนการทางสถิติที่จะทำให้ข้อสอบแต่ละข้อได้รับการกำหนดค่าความยากให้อยู่ในมาตราอันเดียวกัน เพื่อที่จะได้สามารถเปรียบเทียบผลการวัดได้แม้จะสอบด้วยข้อสอบต่างชุดกัน การปรับเทียบมาตราข้อสอบ มีส่วนที่ประยุกต์มาจากโมเดลความน่าจะเป็น (Probabilistic Model) ซึ่งวัดได้เมื่อนักเรียนคนหนึ่งพยายามที่จะตอบคำถามข้อหนึ่ง (Rasch 1960,1980 อ้างใน Choppin 1985 : 5)โมเดลดังกล่าวนี้ พยายามที่จะแสดงข้อมูลที่เกี่ยวกับขบวนการในการตอบข้อสอบของนักเรียนแต่ละคนอย่างมีระบบ การศึกษาข้อมูลเหล่านี้จะทำให้เกิดความเข้าใจในการแปลความหมายผลการวัด ซึ่งเป็นส่วนที่จำเป็นและสำคัญมากสำหรับขบวนการในการสร้างคลังข้อสอบ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2. ประโยชน์ของคลังข้อสอบ

ปัจจุบันด้วยวิวัฒนาการทางคอมพิวเตอร์ได้เข้ามามีบทบาทในการทำงานในระบบต่างๆ มากขึ้น การใช้คอมพิวเตอร์ทำให้ได้รับความสะดวกและมีประสิทธิภาพสูงในการทำงาน แนวทางหนึ่งในการใช้คอมพิวเตอร์เพื่อเก็บข้อมูลเกี่ยวกับข้อสอบและประมวลผล

ข้อมูลข้อสอบ ตลอดจนนำมาใช้ในกระบวนการสอบอื่นๆ เป็นไปได้มาก คลังข้อสอบที่แต่เดิมเก็บข้อสอบด้วยกระดาษแผ่นเล็กๆ สำหรับข้อสอบแต่ละข้อเข้า แฟ้ม ด้วยแรงคน วิเคราะห์ข้อสอบด้วยมือ จึงมองดูพอที่จะเห็นภาพให้คอมพิวเตอร์เข้ามามีบทบาทในขบวนการดังกล่าวได้ ก่อนที่จะศึกษาต่อไปว่าเราจะนำคอมพิวเตอร์เข้ามาใช้ในคลังข้อสอบได้อย่างไร เราควรมาศึกษาประโยชน์จากการใช้คลังข้อสอบก่อนว่าเราสามารถใช้คลังข้อสอบทำอะไรได้บ้าง

2.1 การนำไปใช้กับครู เมื่อระบบในการสร้างคลังข้อสอบของเราดี ครูสามารถสร้างแบบสอบโดยเลือกข้อสอบจากคลังได้ตามจุดมุ่งหมายที่ต้องการอย่างหลากหลาย โดยไม่มีความจำเป็นอีกต่อไปที่นักเรียนทุกคนจะต้องสอบแบบสอบเดียวกันเพื่อนำผลมาเทียบกัน เราสามารถเลือกข้อสอบจากคลังที่เหมาะกับระดับความสามารถและพัฒนาการของนักเรียน เลือกระดับช่วงความยาก รูปแบบ และเนื้อหาสาระในการวัดนักเรียนแต่ละคน โดยที่เราสามารถแปลความหมายผลการกระทำ (performance) ของนักเรียนเหล่านั้นได้เนื่องจากข้อสอบแต่ละข้อในคลังข้อสอบได้ผ่านการปรับเทียบมาตรามาแล้ว

การจัดระบบคลังข้อสอบให้ดีจะสามารถทำให้ครูสร้างชนิดของแบบสอบที่จะนำมาใช้ได้อย่างกว้างขวางทั้งกับนักเรียนในห้อง ในกลุ่มโรงเรียน และในเขตการศึกษา ตามความมุ่งหมายของการวัดได้อย่างทันท่วงที นอกจากนั้นหากครูต้องการเขียนข้อสอบขึ้นใหม่ ในขณะที่ข้อสอบในคลังข้อสอบไม่เหมาะกับจุดมุ่งหมายของการวัดของเขา เราก็สามารถสร้างข้อสอบขึ้นมาเสริมและปรับข้อสอบนั้นให้เหมาะสมเพื่อเก็บไว้ในคลังต่อไป

ครูสามารถสร้างแบบสอบเพื่อการสำรวจ (survey test) เพื่อการประเมินในเขตการศึกษานั้นๆ โดยสามารถใช้ข้อสอบที่มีช่วงการกระจายความยากได้อย่างกว้างขวาง หรือแม้แต่ในรายละเอียดของแบบสอบที่มีจุดมุ่งหมายเฉพาะ เช่น การวัดผลแบบอิงเกณฑ์ (Criterion-Referenced Measurement) ที่ต้องการใช้การวางแผนสอบเพื่อการสอบรายบุคคล (program admission) และการให้วุฒิบัติ (certification) ครูสามารถมีความเป็นไปได้ที่จะสร้างแบบสอบเพื่อความมุ่งหมายในการสอบที่หลากหลายเหล่านี้

2.2 การนำมาใช้ในการวัดทางจิตมิติ คลังข้อสอบได้นำมาสู่การสร้างข้อสอบในระดับท้องถิ่นที่ไม่ต้องสูญเสียความสามารถในการเปรียบเทียบผลการปฏิบัติของนักเรียนแต่ละคนแต่ละกลุ่ม หรือแม้แต่ผลการปฏิบัติของนักเรียนคนเดียวกันในการสอบด้วยเวลาที่ต่างกัน ข้อสอบที่เก็บไว้ในคลังข้อสอบจะได้รับการนิยามตามหลักสูตรในท้องถิ่นนั้นทั้งในด้านคุณภาพของข้อสอบและความตรงประเด็นในหลักสูตรที่ใช้วัด

ข้อสอบที่มีเนื้อหาร่วมกันจะถูกปรับเทียบมาตราให้อยู่ในตัวแปรเดียวกัน โดยแต่ละข้อสอบจะแสดงตำแหน่งที่ข้อสอบนั้นใช้วัดในตัวแปรนั้นๆ สำหรับระดับความยากต่างๆ และสามารถนำมาเปรียบเทียบกันได้ ซึ่งทำให้เป็นไปได้ในการอ้างอิงไปสู่ความรอบรู้ของนักเรียนหลังจากการถูกวัดแล้ว

ด้วยเหตุที่ข้อสอบในคลังข้อสอบถูกนำมาบริหารและให้คะแนนและสามารถแปลผลได้ตามท้องถิ่นนอกจากนั้นยังไม่มีความจำเป็นที่จะต้องรักษาความลับของข้อสอบ เนื่องจากมีจำนวนข้อสอบจำนวนมากจึงเป็นไปได้ที่จะวิเคราะห์และรายงานผลการตอบของนักเรียนแต่ละคนเป็นรายข้อได้ในทันที จากผลการปรับเทียบมาตราข้อสอบทำให้เกิดความสะดวกในการวิเคราะห์ความสอดคล้อง (consistency) ได้ทันทีโดยครูสามารถที่ให้ความสนใจต่อคำตอบของนักเรียนในส่วนที่ตรงประเด็นกับการวัดมากที่สุด ส่วนของความคลาดเคลื่อนในการตอบของนักเรียนเนื่องด้วยสาเหตุใดๆ ก็ตามครูสามารถวินิจฉัยได้และทราบถึงรายละเอียดของผลการปฏิบัติของนักเรียนแต่ละคน

2.3 การนำไปใช้ในการจัดหลักสูตร คลังข้อสอบสามารถรับเอาข้อสอบใหม่ๆ มาใส่ไว้โดยไม่จำเป็นต้องมีขบวนการทดสอบที่ซับซ้อนก่อน แต่การวิเคราะห์จะอยู่ที่การดึงข้อมูลเกี่ยวกับแบบแผนคำตอบของนักเรียนที่สอดคล้องกับผลการคาดคะเนในกลุ่มนักเรียนแต่ละข้อ โดยที่ข้อสอบใหม่ๆเหล่านี้จะกำหนดจากหลักสูตรที่วางโครงสร้างไว้สำหรับคลังข้อสอบนั้นในลักษณะที่มีเนื้อหาร่วมกัน ข้อสอบใหม่สามารถสร้างขึ้นใหม่เมื่อมีการเปลี่ยนแปลงหรือพัฒนาหลักสูตร ในขั้นแรกข้อสอบเหล่านี้จะถูกนำไปบริหารเมื่อพบว่า มีความสอดคล้องที่จะนำมาเก็บไว้ในคลังข้อสอบข้อสอบนี้จะสามารถนำมาการปรับเทียบมาตราเข้าไว้ในคลังข้อสอบได้ทันที จึงแสดงให้เห็นว่าเนื้อหาของข้อสอบในคลังสามารถพัฒนาให้สอดคล้องกับหลักสูตรที่พัฒนาขึ้นได้ โดยที่ข้อสอบจากคลังจะไม่ติดแน่นอยู่กับเฉพาะหลักสูตรเก่าเท่านั้น ครูสามารถพัฒนาหลักสูตร

วิธีการเรียนการสอนของเขาไปได้ และสามารถเลือกใช้ข้อสอบจากคลังข้อสอบได้ตรงตามความมุ่งหมายอย่างเหมาะสม ลักษณะการใช้คลังข้อสอบจึงเป็นพลวัต (dynamics) ในขณะเดียวกันผลการสอบจะไม่แสดงเพียงแต่ผลสัมฤทธิ์ของนักเรียนแต่ละคน แต่สามารถกำหนดทิศทางของวัตถุประสงค์ในการจัดการเรียนการสอนได้อย่างเหมาะสม

ผลลัพธ์ของการสอบไม่ได้เพียงแต่ประเมินผลการสอนของครูแต่ละคน ครูสามารถเปรียบเทียบผลการสอนของเขากับครูคนอื่นๆ ได้ แม้แบบสอบของเขาจะไม่ใช่ชุดเดียวกันแต่ด้วยข้อสอบในคลังข้อสอบที่นำมาปรับเทียบมาตราให้อยู่ในเสกลร่วมอันเดียวกัน จึงเป็นโอกาสที่ครูจะได้ประเมินยุทธวิธีในการเรียนการสอนและปรับปรุงกระบวนการเรียนการสอนได้เหมาะสมขึ้น

จะพบว่าแบบสอบที่ครูสร้างขึ้นจากคลังข้อสอบไม่เพียงแต่ใช้ในเรื่องการประเมิน แต่ยังเกี่ยวกับการจัดระบบเนื้อหาในหลักสูตรที่สามารถกำหนดโดยการปรับเทียบมาตราของคลังข้อสอบ ซึ่งให้ข้อมูลที่เป็นพื้นฐานและเป็นปรนัยในการพัฒนาและปรับปรุงหลักสูตรต่อไป

2.4 การนำไปใช้กับนักเรียน ในสถานการณ์การสอบพบว่ามีสภาวะการณ์จำนวนมากที่ส่งผลกระทบต่อคำตอบของนักเรียน เช่น บางครั้งนักเรียนเดาคำตอบ หรือทำข้อสอบเลินเล่อ เผลอเลอ นักเรียนเกิดความเบื่อหน่ายในการทำข้อสอบทำไปด้วยความรีบร้อน บางครั้งนักเรียนตอบผิดเนื่องมาจากกระบวนการเรียนการสอนที่ทำให้นักเรียนเข้าใจผิด หรือบางครั้งนักเรียนก็มีลักษณะในการตอบสนองต่อข้อสอบแต่ละข้อผิดแผกกันไป องค์ประกอบเหล่านี้ทำให้เกิดความเข้าใจที่คลาดเคลื่อนต่อคะแนนของนักเรียนได้ การวิเคราะห์อย่างรอบคอบถึงความสอดคล้องในแบบแผนคำตอบของนักเรียนแต่ละคนในทุกๆ การสอบจะช่วยให้แก้ไขความคลาดเคลื่อนเหล่านี้ได้

เมื่อนักเรียนถูกกำหนดให้ทำข้อสอบข้อหนึ่งผลก็คือคำตอบของนักเรียนที่ควรแสดงด้วยความแตกต่างระหว่างความสามารถของนักเรียนและความยากของข้อสอบ โมเดลโลจิสอย่างง่ายสามารถแสดงค่าพารามิเตอร์ระหว่าง ปฏิสัมพันธ์ของนักเรียนกับข้อสอบในส่วนของพารามิเตอร์ที่ตั้งใจจะวัดได้ชัดเจน (Rasch 1960, Wright 1968 อ้างใน

Choppin 1985 : 5) ส่วนผลลัพธ์จากการวิเคราะห์ส่วนที่เหลือ (analysis of residuals) จากโมเดลนี้จะแสดงให้เห็นแหล่งของความคลาดเคลื่อนในการตอบข้อสอบของนักเรียนซึ่งครูสามารถเข้าไปมีส่วนช่วยในการขจัดความคลาดเคลื่อนนั้นในครั้งต่อไป ครูอาจจะดำเนินการสอบซ้ำในสถานการณ์ที่ดีกว่าและนำผลมาเปรียบเทียบกัน ครูควรศึกษาปฏิกิริยาในการสอบของนักเรียนและสภาวะการสอบในแต่ละครั้ง การวิเคราะห์ผลการปฏิบัติของนักเรียนเป็นประจำอย่างคงเส้นคงวา ครูจะสามารถเลือก ทางเลือกที่ดีในการมองปัญหาถึงความตั้งใจการสอบของนักเรียนและทราบสภาพที่แท้จริงของนักเรียนมากขึ้น

นอกจากประโยชน์ของการใช้คลังข้อสอบดังกล่าวมาแล้ว ชอปปิน (Choppin, 1985) ได้กล่าวถึงประโยชน์ของคลังข้อสอบโดยสรุปไว้ดังนี้

1) ทำให้ผู้ใช้ได้เลือกข้อสอบมาสร้างแบบสอบได้หลากหลายตามความต้องการที่เหมาะกับวัตถุประสงค์ที่ตั้งไว้ไม่ว่าจะเป็นเรื่องความยาวหรือเนื้อหาที่จะวัด เหมาะสมกับสภาพของท้องถิ่น โดยไม่สูญเสียในการสรุปอ้างอิง (generalizability) และความคงเส้นคงวาในการวัด

2) คลังข้อสอบทำให้การสร้างข้อสอบคู่ขนานมีความเป็นไปได้มากขึ้น

3) คลังข้อสอบทำให้มีความเป็นไปได้ในการวัดซ้ำบุคคลคนเดียวกันด้วยมิติเดียวกันโดยไม่ต้องใช้ข้อคำถามอันเดียวกัน

4) คลังข้อสอบทำให้ได้ข้อคำถามที่มีคุณภาพมากมายเพื่อนำมาประกอบเข้าเป็นแบบสอบใช้ได้ตลอดปี ทำให้การเลือกใช้ข้อสอบเป็นไปด้วยวิธีการที่มีระบบและมีคุณภาพ

3. การสร้างคลังข้อสอบ

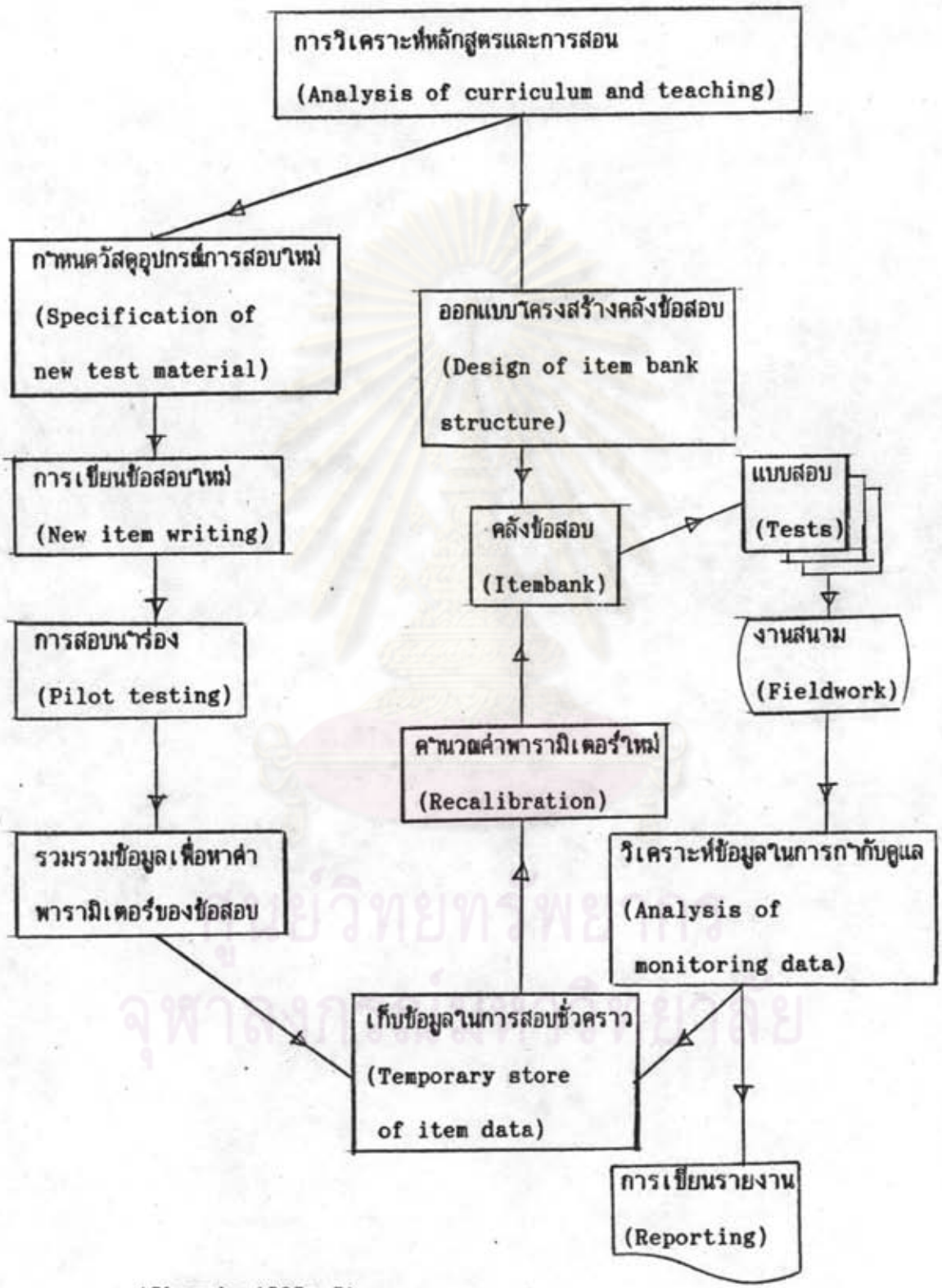
อันดับแรกควรได้กำหนดจุดมุ่งหมายขอบเขตของหลักสูตรที่จะนำข้อสอบมาเก็บไว้ในคลังข้อสอบ ความชัดเจนในการกำหนดหลักสูตร จะช่วยให้สามารถกำหนดข้อสอบที่จะนำมาใช้ในคลังข้อสอบได้อย่างเหมาะสม ข้อสอบแต่ละข้อจะสามารถแสดงถึงความรู้ ทักษะและพฤติกรรมที่นิยามไว้อย่างเหมาะสมนอกจากนั้นการศึกษาว่าใครจะเป็นผู้มาใช้คลังข้อสอบจะช่วยให้การกำหนดเป้าหมายของคลังข้อสอบมีความชัดเจนขึ้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ปัญหาสำคัญของการดำเนินการสร้างคลังข้อสอบก็คือการจัดระบบในคลังข้อสอบซึ่งในส่วนนี้จะสามารถนำเข้ามาใช้ในศูนย์คอมพิวเตอร์ได้ โดยระบบของคลังข้อสอบนี้จะต้องถูกควบคุมโดยบุคคลที่มีความรู้ดีทั้งในด้านหลักสูตรและด้านการวัดผล

จากภาพที่ 3 แสดงว่า ข้อสอบที่จะนำมาเก็บไว้ในคลังข้อสอบจะได้รับการจัดประเภทตามเนื้อหา ความยาก ข้อมูลที่เคยใช้มาแล้วและตัวแปรอื่นๆ ที่ควรได้รับการพิจารณา ชุดของข้อสอบที่จะนำมารวมเข้าเป็นแบบสอบจะถูกดึงมาจากคลังตามวัตถุประสงค์ที่ต้องการโดยผู้เลือกใช้จะกำหนดกฎเกณฑ์เฉพาะ (specific criteria) ในการเลือกข้อสอบ เกณฑ์เหล่านี้ได้แก่ลักษณะเฉพาะของข้อเกี่ยวกับเนื้อเรื่องที่วัด รายละเอียดของช่วงและระดับความยากที่ต้องการ และการไม่เลือกเอาข้อสอบที่เคยใช้อยู่บ่อยๆ มาใช้วัดอีก เมื่อแบบสอบถูกสร้างขึ้น เจ้าหน้าที่ในคลังข้อสอบจะจัดตารางในการแปลผลคะแนนที่จะได้มาเมื่อคะแนนที่วัดมาได้ในแต่ละแบบสอบจะสามารถนำมาแปลความหมายได้ตามมาตราคะแนนนั้นๆ (มาตราคะแนนนี้อาจจะมีหลายชุดถ้ามีการวัดมากกว่า 1 คุณลักษณะแฝง (latent trait)

ภายหลังจากการบริหารแบบสอบและการวิเคราะห์ผลการสอบแล้ว ข้อสอบทั้งหมดหรือบางส่วนจะถูกนำเข้าคลังเพื่อปรับเทียบมาตราอีกครั้ง ผลการกระทำเช่นนี้จะช่วยทำให้มีการแก้ไขความแม่นยำในการวัดได้ดีขึ้น เช่น ค่าความยากของข้อสอบจะสามารถประมาณด้วยค่าที่แม่นยำขึ้น และเป็นการเปิดโอกาสให้มีการตรวจสอบอย่างเป็นระบบในการเปลี่ยนแปลงทางโครงสร้างของความยากของข้อสอบ และมีส่วนช่วยให้ตรวจพบและประเมินหลักสูตรเมื่อมีการเปลี่ยนแปลงคุณลักษณะแฝง เช่น ในวิชาคณิตศาสตร์ที่แสดงผลสัมฤทธิ์ในการเรียนของนักเรียน ผู้เชี่ยวชาญทางด้านเนื้อหาอาจจะสังเกตเนื้อเรื่องที่ขาดหายไปในช่วงเวลาการใช้คลังข้อสอบในระยะต่างๆ จะสามารถตัดสินใจในการสร้างข้อสอบขึ้นมาใหม่ใน คลังที่ขาดหายไปในหลักสูตร ข้อสอบที่สร้างใหม่จะนำไปทดลองเช่นวิธีที่เคยปฏิบัติมาก่อนที่จะนำเข้าสู่คลังข้อสอบต้องเตรียมข้อมูลเพื่อใช้ในการปรับเทียบมาตราในขั้นแรก โดยการนำไปทดสอบกับกลุ่มตัวอย่างนักเรียน ประมาณ 100-300 คน ที่จะนำมาคำนวณหาความยากที่เหมาะสมและภายหลังได้รับการสร้างเสกลผลการสอบแล้ว ข้อสอบใหม่จะนำเข้าไปเก็บไว้ในคลังเพื่อรอการปรับเทียบมาตราใหม่ ต่อไป



(Choppin 1985: 5)

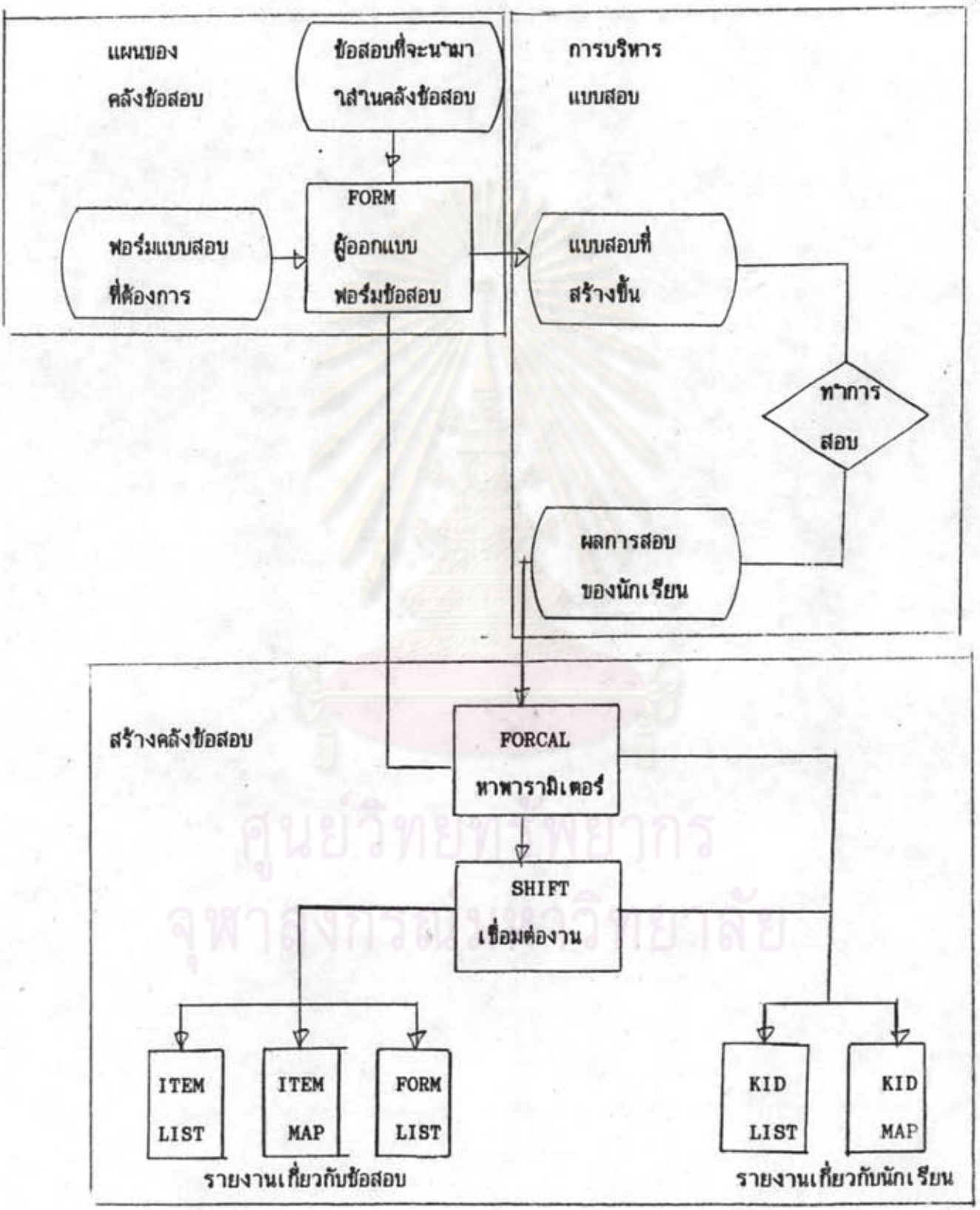
ภาพที่ 3 แสดงระบบในคลังข้อสอบ

สำหรับกระบวนการนำคอมพิวเตอร์เข้ามาใช้ในคลังข้อสอบนั้นได้แสดงไว้ในภาพที่ 2 ในหน้าถัดไป ในการเริ่มสร้างคลังข้อสอบเมื่อกำหนดหลักสูตรแล้ว การรวบรวมข้อสอบในเนื้อหาของหลักสูตรจะต้องกระทำขึ้น ในภาพที่ 2 แสดงให้เห็นชัดเจนถึงขั้นตอนพื้นฐานที่จำเป็นในการสร้างที่รวบรวมข้อสอบในคลังข้อสอบ โดยที่ขั้นตอนเหล่านี้ได้บรรยายในลักษณะที่เป็นชุดของโปรแกรมคอมพิวเตอร์แต่ละโปรแกรมมีการทำงานอย่างไร พอสรุปเป็นรายละเอียดได้ดังนี้

3.1 การวางแผนเกี่ยวกับรูปแบบของข้อสอบ โปรแกรม FORM จะนำมาใช้แยกข้อสอบในระหว่างฟอร์มแบบสอบ เพื่อว่าจะได้แสดงความเกี่ยวพันระหว่างข้อสอบเดียวกันในแต่ละฟอร์ม ซึ่งจะทำให้สามารถคำนวณค่าทางสถิติที่แสดงโครงสร้างที่เชื่อมโยงต่อกันได้ชัดเจนขึ้น เมื่อนำข้อสอบเหล่านี้ไปใช้ในสถานการณ์ต่างๆ

FORM จะอ่านจำนวนข้อสอบที่ได้รับการปรับเทียบมาตราแล้ว จำนวนข้อสอบที่ต้องการใส่ไว้ในแต่ละฟอร์มของข้อสอบ จำนวนข้อสอบที่ต้องการใส่ไว้มากกว่า 1 ฟอร์ม เพื่อแสดงการเชื่อมต่อ (link) กันระหว่างฟอร์ม ความยากที่คาดหวังในแต่ละข้อ และรูปแบบค่าความยากในฟอร์มซึ่งจะสร้างคู่ขนานกัน FORM จะกำหนดจำนวนข้อที่เชื่อมต่อกันต่อฟอร์ม จำนวนข้อที่เชื่อมต่อกันทั้งหมด และจำนวนฟอร์มที่นำมาเข้าชุดกัน

FORM จะสร้างแฟ้มของคุณลักษณะเฉพาะของข้อสอบจากขบวนการทำงานในคลังข้อสอบจะสร้างบัญชีรายการข้อสอบที่ประกอบด้วยเลขข้อ (identification number) ชื่อข้อ (item name) เลขที่เชื่อมต่อ (link number) ค่าความยากที่คาดหวัง(expected difficulty) คำตอบที่ถูก (correct response) และความสัมพันธ์ระหว่างฟอร์มอื่นเพื่อว่าเมื่อนำข้อสอบในฟอร์มที่คู่ขนานกันจะได้มาศึกษาตรวจสอบได้การแสดงบัญชีข้อสอบโดยจัดแบบนี้จะทำให้เกิดความสะดวกในการตรวจสอบเนื้อหาที่ต่อเนื่องกันและการรวบรวมข้อสอบเข้าสู่ฟอร์มตามที่กำหนด



(Choppin 1985: 7)

ภาพที่ 4 วิธีการสร้างคลังข้อสอบ

3.2 การปรับเทียบฟอร์มของแบบสอบ (Calibrating Test Forms) เมื่อฟอร์มได้รับการวางแผน รวบรวมและนำไปบริหารแล้ว คำตอบของนักเรียนจะถูกบันทึกจะจัดระบบเข้าในระเบียน (record) ของนักเรียนแต่ละคนซึ่งจะมีทั้งเลขประจำตัวนักเรียน ชื่อฟอร์มที่สอบมาแล้ว และลักษณะการตอบสนองต่อข้อสอบ(item response string) ของนักเรียนนี้จะถูกจัดประเภทโดยฟอร์มตั้งแต่แรกและเตรียมพร้อมเพื่อการปรับเทียบมาตราต่อไป ซึ่งโปรแกรม FORCAL จะนำมาใช้ในการปรับเทียบมาตราข้อสอบแต่ละข้อและวิเคราะห์ข้อสอบในแต่ละฟอร์มของนักเรียนแต่ละคน

FORCAL จะประมาณค่าความยากของข้อสอบโดยเป็นอิสระจากกลุ่มตัวอย่างที่สอบ และค่าการกระจายของคะแนนของนักเรียน FORCAL จะประมาณความสามารถของนักเรียนในแต่ละฟอร์ม ค่าความคลาดเคลื่อนมาตรฐาน และรูปแบบคำตอบของนักเรียนด้วยค่าสถิติที่เหมาะสม นักเรียนทุกคนจะถูกวัดและข้อสอบทุกข้อจะถูกปรับเทียบมาตรา ในขั้นนี้ แต่ความสามารถและความยากของข้อที่เป็นลักษณะที่สัมพันธ์กันกับข้อมูลเริ่มแรกของท้องถิ่นที่นิยามมาจากแต่ละฟอร์ม เมื่อต้องการเชื่อมข้อมูลข้ามฟอร์ม ในขั้นนี้ยังไม่สามารถทำได้ ต้องใช้การวิเคราะห์การเชื่อมต่อซึ่งได้มาจากการเทียบเท่าฟอร์ม ทุกฟอร์มให้อยู่ในมาตราเดียวกัน

3.3 การคำนวณเชื่อมข้อสอบระหว่างฟอร์ม (Calculating Test Form Link) เมื่อข้อสอบถูกปรับเทียบมาตรา ในแต่ละฟอร์มยังมีค่าความยากอีกจำนวนมาก ที่จะต้องประมาณสำหรับข้อสอบเหล่านี้ในแต่ละข้อที่อยู่ต่างฟอร์มกัน ข้อสอบที่ปรากฏอยู่มากกว่า 1 ฟอร์ม จะต้องนำมาหาข้อมูลที่เชื่อมต่อ โปรแกรม SHIFT จะนำมาใช้ในการแสดงความแตกต่างคุณลักษณะของข้อสอบที่อยู่ต่างฟอร์ม และคุณลักษณะเฉลี่ยของข้อสอบแต่ละข้อจากทุกฟอร์มที่ข้อนั้นไปปรากฏอยู่

3.4 การแสดงผลจากคลังข้อสอบ (Showing the Resulting Bank) โปรแกรม ITEM LIST จะใช้เพื่อแสดงบัญชีข้อสอบแต่ละข้อโดยลำดับเลขที่ข้อ คุณลักษณะต่างๆ ของข้อ คำตอบที่ถูก ชื่อข้อ ความยากของข้อ

โปรแกรม ITEM MAP จะใช้เพื่อแสดงลักษณะของตัวแปรเป็นกราฟที่แสดงจุดต่างๆ ของข้อสอบในตัวแปรที่วัดตามระดับความยาก การแสดงด้วยกราฟจะขึ้นอยู่กับตัวแปรที่นิยามไว้ในระบบข้อมูล ซึ่งในกรณีนี้ครูสามารถตรวจสอบความสัมพันธ์ระหว่างเนื้อหาของข้อสอบ

และความยากของข้อสอบ เพื่อแสดงขอบเขตลำดับที่ข้อสอบที่มีอยู่ในคลังข้อสอบ ที่นิยามหลักสูตรไว้ว่ามีความสอดคล้องตามคาดหวังไว้หรือไม่ นั่นคือ เป็นการตรวจสอบความตรงเชิงโครงสร้าง (Construct Validity) ของข้อสอบนั่นเอง ข้อมูลนี้จะใช้เป็นกรอบในการเขียนข้อสอบขึ้นใหม่เพื่อเติมช่องว่างในคลังให้เต็มตามที่นิยามไว้ในหลักสูตรและเป็นกรอบในการเลือกข้อสอบใหม่เข้ามาไว้ในคลัง

โปรแกรม FORM LIST นำมาใช้แสดงบัญชีแต่ละฟอร์ม โดยแสดงหมายเลขฟอร์ม ชื่อ จำนวนข้อ ความยากของฟอร์ม ความยากของแต่ละข้อที่อยู่ในฟอร์ม คำตอบที่ถูก ค่าความคลาดเคลื่อนมาตราฐาน ความสะดวกในการใช้โปรแกรมนี้สามารถเลือกให้แสดงรายการข้อมูลในแต่ละฟอร์มทั้งหมดหรือจะกำหนดคุลักษณะบางประการในฟอร์มได้

โปรแกรม KID LIST นำมาใช้แสดงข้อมูลของนักเรียนแต่ละคนได้แก่ เลขประจำตัว (identification) ความสามารถที่วัดได้ ความคลาดเคลื่อนในการตอบข้อสอบ ค่าสถิติที่เหมาะสม (fit statistics) KID LIST จะชี้ว่านักเรียนแต่ละคนมีลักษณะคำตอบที่ไม่เหมาะสมอย่างไร (misfit) โดยแสดงในส่วนท้ายจากค่าที่คาดหวังอื่นๆเพื่อว่าครูและนักเรียนสามารถสังเกตได้ว่าแหล่งความคลาดเคลื่อนในการตอบข้อสอบของนักเรียนเกิดจากการเดาคำตอบ เกิดจากการเฉลยผิดเกิดจากพิมพ์ข้อสอบผิด หรือเกิดจากความเลินเล่อของนักเรียนเอง

โปรแกรม KID MAP นำมาใช้ในการแสดงผลการปฏิบัติของนักเรียนแต่ละคน ในรูปของกราฟ KID MAP จะแสดงภาพในการตอบสนองต่อข้อสอบของนักเรียนแต่ละคน โดยแสดงจุดที่นักเรียนยืนอยู่ในหลักสูตร ความน่าจะเป็นในการตอบข้อสอบของนักเรียนแต่ละข้อ ซึ่งสิ่งนี้จะช่วยให้ครูและนักเรียนได้เห็นภาพของ ผลการปฏิบัติ ของนักเรียนทราบถึงจุดเด่นจุดด้อยในความรอบรู้ตามหลักสูตรของนักเรียนในแต่ละคนได้ชัดเจน

4. สถานการณ์ในการตัดสินใจใช้คลังข้อสอบ

จากการศึกษาลักษณะของหลักการและวิธีการในการสร้างคลังข้อสอบ ซึ่งต้องอาศัยทั้งความรอบรู้ เวลา การลงทุนเพื่อสร้างระบบเหล่านี้ ก่อนที่จะตัดสินใจสร้างคลังข้อสอบ

มิลลแมน และ อาร์ตเตอร์ (Millman and Arter, 1984) ได้แนะนำสถานการณ์ที่จะตัดสินใจไว้โดยสรุปดังนี้

1) แบบสอบที่มีอยู่ไม่ได้รับการยอมรับที่จะนำมาใช้และเพื่อนร่วมทีมมีความต้องการที่จะสร้างแบบสอบขึ้นมาจากคุณลักษณะเฉพาะของท้องถิ่นนั้น

2) การจัดการทดสอบมีบ่อยครั้ง

3) มีความต้องการที่จะสร้างแบบทดสอบขึ้นใช้หลายฟอร์มในแต่ละฉบับ

4) ต้องการใช้ Tailored Tests ในการสอบเด็กเป็นรายบุคคล

5) ผู้ใช้ส่วนมากเต็มใจที่จะมีส่วนร่วมในการสร้างคลังข้อสอบขึ้นเพื่อมีแหล่งที่จะใช้ข้อสอบจากแหล่งเดียวกันมาสร้างแบบทดสอบที่ต้องการ

6) ระบบของคลังข้อสอบซึ่งได้แก่ อาร์ดแวร์และซอฟท์แวร์ (hardware and software) มีพร้อมและมีผู้ใช้ที่มีศักยภาพเพียงพอในการใช้คอมพิวเตอร์ตามวัตถุประสงค์ของคลังข้อสอบ

5. ข้อควรคำนึงของผู้สร้างคลังข้อสอบ

หลังจากที่ได้ศึกษามาแล้วว่า คลังข้อสอบคืออะไร มีประโยชน์อย่างไรจะสร้างคลังข้อสอบอย่างไร และสถานการณ์ใดที่ควรตัดสินใจจัดให้มีคลังข้อสอบขึ้นแล้ว สมมติว่ามีการตัดสินใจจะสร้างคลังข้อสอบขึ้นจริง ยังมีปัญหาตามมาให้ผู้ที่จะสร้างคลังข้อสอบได้ระมัดระวัง เพื่อให้การจัดตั้งคลังข้อสอบเป็นไปตามเจตนารมย์ของการสร้างคลังข้อสอบ ปัญหาดังกล่าวพอสรุปได้ดังนี้

5.1 เกี่ยวกับข้อสอบ

1) การได้มาซึ่งข้อสอบและการพัฒนาข้อสอบ

ก. ท่านจะใช้คลังข้อสอบเก็บข้อสอบของท่านเองหรือเก็บข้อสอบของคนอื่นด้วย ถ้าจะเก็บข้อสอบของท่านเอง ท่านจะวางแผนดำเนินการอย่างไรต่อไป ถ้าจะเก็บข้อสอบของคนอื่น ข้อสอบที่ได้มาจะต้องจัดซื้อมาหรือไม่ จะจัดประเภทข้อสอบเหล่านั้นจะต้องการข้อมูลอะไรบ้าง รูปแบบของข้อสอบเหมาะสมที่จะมาเข้าไว้ในคลังข้อสอบเหมาะในการจัดประเภท และการเลือกใช้ได้สะดวกหรือไม่

ข. รูปแบบของข้อสอบเป็นอย่างไร ข้อสอบที่เป็นคำถามปลายเปิดหรือปลายปิด คำถามแสดงความคิดเห็นวัตถุประสงค์ของการเรียนการสอน รายละเอียดของผลการปฏิบัติงานควรจะนำมาใส่ไว้ในคลังด้วยหรือไม่ ทุกข้อสอบต้องอยู่ในรูปแบบเดียวกัน เช่นเป็นแบบเลือกตอบเหมือนกันหมดหรือจะต้องมีตัวเลือก ก ข ค ง เหมือนกันหรือไม่ จะต้องนำข้อสอบมาปรับเทียบมาตราและทำให้มีความตรงหรือไม่

ค. ข้อสอบที่รวบรวมมาไว้ในคลังข้อสอบจะต้องมีจำนวนเรื่อง (topics) เท่าใดและในแต่ละเรื่องจะแยกเป็นเรื่องย่อยอีกหรือไม่

ง. ขบวนการในการทดลองข้อสอบก่อนการนำไว้ในคลัง ใครจะเป็นผู้ดำเนินการจัดพิมพ์รายงาน และค่าสถิติอะไรจะต้องรวบรวมมา

2) การจัดประเภทข้อสอบ

ก. จะจัดเรียงตามเนื้อหาอย่างไร ใครจะต้องเป็นผู้จัดเตรียมและจัดระบบในการจัดประเภทข้อสอบในคลัง

ข. รายละเอียดของข้อสอบที่จะนำมาจัดเป็นดัชนีมีอะไรบ้าง

ค. จะต้องเก็บข้อมูลอะไรเกี่ยวกับข้อสอบหลังจากการนำไปใช้แล้ว

3) การจัดการ (Management)

ก. จะมีขบวนการอย่างไรที่ทำให้ข้อสอบในคลังมีความเป็นทันสมัยในการเลือกใช้เมื่อเวลาผ่านไป และมีการใช้ข้อสอบบางข้อบ่อยครั้งไป

ข. ข้อที่มีรูปภาพ กราฟ และคุณลักษณะที่ต้องพิมพ์ด้วยมือจะทำอย่างไร

ค. จะจัดข้อสอบที่ใช้ข้อมูลร่วมกับข้ออื่นๆ อย่างไร

5.2 เกี่ยวกับแบบสอบ (Test)

1) การรวบรวมข้อสอบเข้าชุดของแบบสอบ

ก. ผู้สร้างแบบสอบจะกำหนดเลือกข้อสอบได้เองหรือให้คอมพิวเตอร์ช่วยเลือกให้ ถ้าข้อสอบถูกเลือกโดยคอมพิวเตอร์จะใช้กระบวนการเลือกอย่างสุ่มหรือดูว่าครั้งสุดท้ายที่เคยใช้ข้อสอบนี้ใช้มาเมื่อไร หรือใช้ข้อสอบนี้มาบ่อยครั้งเท่าไร

ข. จะต้องกำหนดพารามิเตอร์อะไรของข้อสอบที่จะนำมารวมเข้าในแบบสอบเช่น รูปแบบข้อสอบ ข้อจำกัด ค่าความยากที่ต้องการ การกระจายคะแนนที่คาดหวัง ความเบี่ยงเบนของแบบสอบที่คาดหวัง

ค. การจัดข้อสอบเข้าแบบสอบจะมีลักษณะการสุ่มวางตำแหน่งของข้อและตัวเลือกอย่างไร

ง. ระบบการจัดพิมพ์ข้อสอบที่เลือกมาสร้างแบบสอบ จะดำเนินการอย่างไร จะให้คอมพิวเตอร์พิมพ์ให้ หรือแสดงภาพทางจอเลย

2) การบริหารการให้คะแนนและการรายงาน

ก. จะจัดขบวนการบริหารแบบสอบอย่างไร

ข. จะจัดระบบการให้คะแนนอย่างไร จะต้องใช้สูตรการแก้การเดาหรือไม่ จะมีการประเมินคำตอบของผู้สอบอย่างไร

ค. การรายงานผลการสอบจะต้องรายงานอย่างไร ให้ใครทราบบ้าง ระเบียนบันทึกการตอบจะต้องเก็บไว้หรือไม่ ถ้าเก็บต้องเก็บในเรื่องใดบ้าง และจะเก็บไว้นานเท่าใด

3) การประเมินผลแบบสอบเกี่ยวกับค่าความเที่ยง ค่าความตรง จะต้องเก็บข้อมูลไว้จะต้องเก็บหรือไม่ จะต้องเก็บโดยใคร และจะเอาข้อมูลนั้นมาใช้ทำอะไรได้บ้าง

5.3 ระบบ (System)

1) ใครจะเป็นผู้รับผิดชอบในการจัดการกับระบบการดำเนินงานในคลังข้อสอบ ภายใต้ข้อจำกัดต่างๆ ในการให้ได้มาซึ่งข้อสอบ การจัดดำเนินการใช้ระเบียบการในการใช้และคุณสมบัติของบุคคลที่จะมาใช้คลังข้อสอบ

2) ซอฟท์แวร์และฮาร์ดแวร์ (Software/Hardware)

ก. จะใช้คอมพิวเตอร์ในระบบใดมาใช้ ไมโครคอมพิวเตอร์จะใช้ได้หรือไม่

ข. เมื่อข้อสอบมีจำนวนมากจะต้องแยก แฟ้ม ของผู้ใช้หรือไม่

ค. จะให้ผู้ใช้รู้จักระบบในคลังข้อสอบอย่างไร

ง. ใครจะเป็นผู้รับผิดชอบในการดูแลรักษาอุปกรณ์และเครื่องมือต่างๆ

5.4 การลงทุน (Cost)

1) จะต้องลงทุนในเรื่องใดบ้างในการสร้างคลังข้อสอบ

2) เงินลงทุนที่จะสร้างคลังข้อสอบมีแหล่งที่สนับสนุนเพียงพอหรือไม่

3) เมื่อเปรียบเทียบระบบจากคลังข้อสอบกับระบบอื่นที่เคยปฏิบัติมาในแง่วัตถุประสงค์เดียวกันระบบใดจะคุ้มค่ากว่ากัน

จากการศึกษาเรื่องคลังข้อสอบพบว่าคลังข้อสอบเป็นแหล่งที่มีระบบและให้บริการที่ช่วยสนองวัตถุประสงค์ในการใช้ได้หลากหลาย มิใช่แต่ในเรื่องการสร้างแบบสอบได้ตามวัตถุประสงค์ต่างๆ เฉพาะที่ผู้สร้างต้องการ แต่ขบวนการในการจัดระบบในคลังข้อสอบให้ข้อมูลในการเปลี่ยนแปลงหลักสูตร การตรวจสอบประสิทธิภาพในการเรียนการสอน การวินิจฉัยความสามารถของนักเรียนเป็นรายบุคคล การตรวจสอบผลสัมฤทธิ์ทางการเรียนทั้งในระดับชั้นเรียน โรงเรียน กลุ่มโรงเรียน และระหว่างกลุ่มโรงเรียนด้วยกัน

ในโอกาสที่คอมพิวเตอร์ได้เข้ามามีบทบาทในคลังข้อสอบจึงทำให้การประมวลผลข้อมูล การวิเคราะห์ข้อมูล การจัดระเบียบข้อมูล การรายงานผลข้อมูลเป็นไปได้อย่างมีประสิทธิภาพและทันท่วงที

วิธีการปรับเทียบมาตราข้อสอบให้คุณสมบัติทางสถิติของข้อสอบอยู่ในมาตราเดียวกันทำให้การแปลผลการสอบมีความชัดเจนขึ้น และสามารถเทียบผลการสอบของนักเรียนได้และแม้จะใช้ข้อสอบที่ต่างกันแต่วิธีการปรับเทียบมาตราควรได้มีการตรวจสอบความถูกต้องในการเลือกโมเดล มาวิเคราะห์ได้เหมาะสมในหลักการของการปรับเทียบมาตรานี้เป็นสิ่งที่ผู้รายงานจะต้องศึกษาให้ลึกซึ้งต่อไป

ปัญหาหลักในการสร้างคลังข้อสอบคือปัญหาในการลงทุน มีความเป็นไปได้มากในการเตรียมการวิเคราะห์หลักสูตร การจัดหาข้อสอบ การมีบุคคลที่จะมีความรู้ในการจัดระบบและการใช้คอมพิวเตอร์ในการประมวลผลในระบบคลังข้อสอบ หากมีแหล่งสนับสนุนในการลงทุนแล้วผู้วิจัยมีความเห็นว่า ในอนาคตเราคงมีคลังข้อสอบที่ช่วยให้ครูของเรามีความสะดวกในการจัดการเรียนสอนมากขึ้น และมีเวลาพอที่จะพัฒนาการเรียนการสอนให้มีประสิทธิภาพมากขึ้นกว่าในปัจจุบัน

6. การเลือกข้อสอบจากคลังข้อสอบที่ใช้ในปัจจุบัน

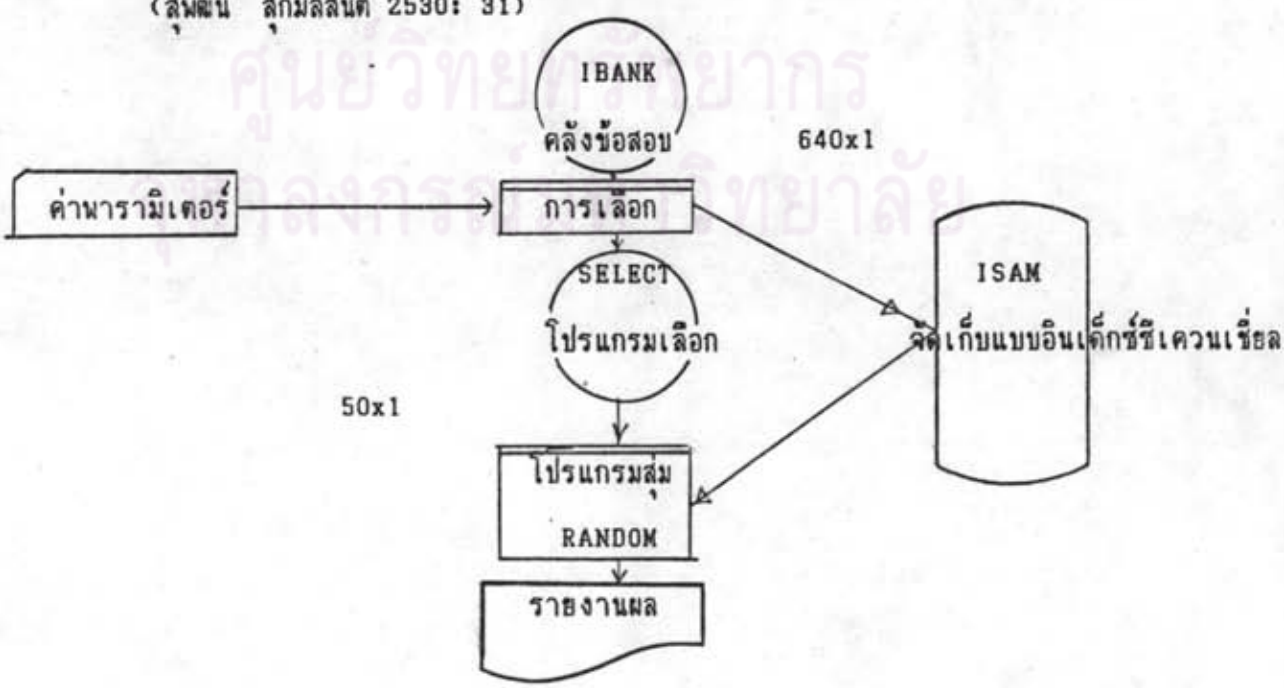
6.1 การเลือกข้อสอบด้วยคอมพิวเตอร์ของ สุพัฒน์ สุกมลสันต์ (2530 : 30-33) ได้เสนอเกณฑ์ในการเลือกข้อสอบเพื่อสร้างแบบสอบชุดใหม่โดยใช้ค่าพารามิเตอร์เป็นเกณฑ์ ดังนี้

1) ระดับของข้อสอบว่าเป็นข้อสอบระดับใด
2) ทักษะที่ต้องการสอบว่าเป็นทักษะใด
3) ทักษะย่อยที่ต้องการคืออะไร
4) ระดับความยากง่ายอยู่ในช่วงใด
5) อำนาจจำแนกตั้งแต่เท่าใด
6) ข้อสอบมีหรือเป็นความเรียงด้วยหรือไม่
7) จำนวนที่ต้องการเท่าใด
8) รายละเอียดของรูปแบบของหัวแบบสอบเป็นอย่างไรบ้าง (เช่น วิชาอะไร สอบเมื่อใด กี่คะแนน ฟอร์มอะไร และของหน่วยงานใด เป็นต้น)

จากเกณฑ์ทั้ง 8 อย่างข้างต้นจะเห็นได้ว่าผู้ใช้ข้อสอบสามารถที่จะกำหนดค่าพารามิเตอร์ต่างๆ ของข้อสอบได้ละเอียดมากน้อยเพียงใด แล้วแต่รายละเอียดของรหัสที่กำหนดให้แต่ละข้อสอบว่ามีมากเพียงใดและการที่จะได้แบบสอบใหม่มีรูปแบบแตกต่างกันได้มากเท่าใด ขึ้นอยู่กับจำนวนของข้อสอบในคลังข้อสอบว่ามีมากน้อยเพียงใด

โดยใช้โปรแกรมชื่อ CIBP ในการสุ่ม ดังภาพที่ 5

(สุพัฒน์ สุกมลสันต์ 2530: 31)



ภาพที่ 5 ขั้นตอนการทำงานของโปรแกรม

และมีขั้นตอนในการทำงานโดยสรุป ดังนี้

(1) อ่านค่าพารามิเตอร์ต่างๆ ที่ผู้ใช้กำหนดขึ้น

(2) เลือกข้อสอบทุกข้อที่สอดคล้องกับเกณฑ์ที่กำหนดขึ้น

(3) นำรหัสประจำข้อสอบทุกอย่างที่กล่าวไว้ในขั้นที่ 2 ไปบันทึกไว้ในส่วนหนึ่งของจานแม่เหล็ก (disk) หรือเทปแม่เหล็ก (tape) ในกรณีที่ไม่อาจใช้จานแม่เหล็กได้หรือไม่ต้องการใช้

6.2 การเลือกข้อสอบด้วยคอมพิวเตอร์ของ พลากร กรพินิทักษ์ (2532: 11) โดยให้ผู้ใช้จะเป็นผู้กำหนดเกณฑ์ในการคัดเลือกข้อสอบจากฐานข้อมูล โดยจะมีรูปแบบชนิดของเกณฑ์กำหนดให้ในจอภาพผู้ใช้จะต้องใส่ค่า (input) เกณฑ์ต่างๆ ตามต้องการโดยไม่จำเป็นต้องใส่ค่าทุกค่าที่กำหนดจะใส่ค่าเฉพาะที่ต้องการเท่านั้น โดยมีรูปแบบของการใช้ตามตัวอย่างข้างล่างดังนี้

เกณฑ์ในการเลือกข้อสอบ

ระดับชั้น . .....วิชา .....

บทที่ .....เรื่องย่อย .....

พฤติกรรมที่ต้องการวัด .....ค่าความยาก (ประเพณีนิยม) .....

ค่าอำนาจจำแนก (ประเพณีนิยม) ....ค่าการยาก (โลจิส) .....

ค่าอำนาจจำแนก (โลจิส) .....ค่าการเดา (โลจิส) .....

เมื่อผู้ใช้ได้กำหนดเกณฑ์ที่ต้องการแล้ว คอมพิวเตอร์จะทำการคัดเลือกข้อสอบที่สอดคล้องกับเกณฑ์ที่ต้องการมาเก็บไว้ในแฟ้มสำรอง เพื่อรอการสุ่มข้อสอบในขั้นต่อไป การสุ่มข้อสอบที่ได้เลือกไว้ตามจำนวนที่ต้องการ ในขั้นตอนนี้ผู้ใช้จะต้องระบุจำนวนข้อสอบที่ต้องการให้คอมพิวเตอร์ทราบเพื่อเป็นการกำหนดจำนวนของการสุ่ม รูปแบบการกระจายการสุ่มเป็นแบบปกติ

กล่าวโดยสรุปในปัจจุบันคลังข้อสอบแม้จะได้รับการยอมรับกันอย่างกว้างขวางในการส่งเสริมสนับสนุนให้มีการวัดและดำเนินการในการสร้างให้เกิดขึ้นในทุกระดับการศึกษา แต่เมื่อพิจารณาถึงการจัดระบบสำหรับคลังข้อสอบแล้ว ยังพบประเด็นปัญหาหลายประการที่ผู้สร้างคลังข้อสอบจำเป็นจะต้องคำนึงถึง อันได้แก่ การได้มาซึ่งข้อสอบที่จะรวบรวมเข้าไว้ในคลังข้อสอบ การจัดระบบข้อสอบอย่างไรจึงสะดวกในการคัดเลือกและนำไปใช้ จะต้องแสดงคุณลักษณะใดของข้อสอบบ้าง และจะเลือกข้อสอบออกมาอย่างไรจึงจะได้ชุดของข้อสอบที่มีคุณภาพเหมาะสมกับวัตถุประสงค์ที่จะนำไปใช้ตามที่ผู้สร้างแบบสอบต้องการ ซึ่งถูกกำหนดไว้ภายหลังจากการบริหารแบบสอบและวิเคราะห์ผลการสอบแล้ว ข้อสอบทั้งหมดหรือบางส่วนจะถูกนำเข้าคลังข้อสอบเพื่อคำนวณค่าสถิติรายข้ออีกครั้ง (recalibration) ซึ่งจะช่วยทำให้ได้สารสนเทศเกี่ยวกับข้อสอบนั้นได้แม่นยำขึ้นและสามารถเลือกไปใช้ได้ตามความเหมาะสมต่อไป ในกรณีที่มีการสร้างข้อสอบใหม่ เช่น เพื่อให้ได้ข้อสอบครอบคลุมหลักสูตรที่มีการเปลี่ยนแปลงไป กระบวนการในการสร้างและการทดลองข้อสอบกับกลุ่มตัวอย่างนักเรียนที่เป็นตัวแทน เพื่อคำนวณคุณลักษณะของข้อสอบซึ่งเป็นการดำเนินการเพื่อให้ได้ข้อสอบมาเก็บไว้ในคลังข้อสอบให้พร้อมที่จะนำไปใช้ในการสร้างแบบสอบต่อไป

จากระบบในการดำเนินการคลังข้อสอบที่กล่าวมาพบว่าคลังข้อสอบเป็นแหล่งรวบรวมข้อสอบที่ผ่านขั้นตอนในการทดลองวิเคราะห์ แก้ไขปรับปรุง และตรวจสอบคุณภาพของข้อสอบมาแล้ว การได้มาซึ่งข้อสอบเก็บไว้ในคลังข้อสอบในปัจจุบันไม่ใช่ปัญหาสำคัญของการสร้างคลังข้อสอบตามโรงเรียนหรือกลุ่มโรงเรียนต่างๆ เพราะได้มีการพยายามสร้างคลังข้อสอบขึ้นโดยรวบรวมข้อสอบที่ครูในโรงเรียนหรือกลุ่มโรงเรียนสร้างไว้มารวมกันด้วยความปรารถนาที่จะทำให้คลังข้อสอบเป็นแหล่งสำคัญในการสร้างแบบสอบที่มีคุณภาพ ปัจจุบันปัญหาสำคัญของคลังข้อสอบอยู่ที่เทคนิควิธีการใช้คลังข้อสอบที่มีอยู่อย่างไรจึงสามารถสร้างแบบสอบได้มีคุณภาพตามที่ผู้สร้างหรือผู้ใช้คลังข้อสอบต้องการ นั่นคือผู้ใช้คลังข้อสอบจะดำเนินการคัดเลือกข้อสอบอย่างไร จึงจะได้กลุ่มข้อสอบที่มีคุณภาพเหมาะกับวัตถุประสงค์ในการนำแบบสอบไปใช้ ซึ่งประเด็นปัญหาดังกล่าวควรได้มีการศึกษาและหาข้อสรุปถึงเทคนิควิธีการที่เหมาะสมในการเลือกข้อสอบจากคลังข้อสอบต่อไป

ทฤษฎีการตอบสนองต่อข้อสอบ (Item Response Theory, IRT)

การวิจัยครั้งนี้ใช้ทฤษฎีการตอบสนองต่อข้อสอบเป็นหลักในการวิเคราะห์ข้อสอบ การหาค่าสารสนเทศ และการเปรียบเทียบประสิทธิภาพสัมพัทธ์จึงขอเสนอความเป็นมาของทฤษฎีอย่างละเอียดดังต่อไปนี้

ทฤษฎีการตอบสนองต่อข้อสอบเป็นพัฒนาการที่มีความสำคัญมากที่สุดทฤษฎีหนึ่งทางจิตมิติ ทฤษฎีการตอบสนองต่อข้อสอบ นำมาใช้มากที่สุดกับคำถามประเภทเลือกตอบ (multiple choice questions) ในแบบสอบความสามารถแม้ว่าทฤษฎีการตอบสนองต่อข้อสอบจะสามารถนำมาใช้ได้ดีเช่นเดียวกันในข้อคำถามที่ให้ตอบอย่างอิสระ (free response)

ถ้าเราดำเนินการสอบด้วยแบบสอบหลายฉบับที่วัดในเรื่องราวเดียวกันกับกลุ่มผู้เข้าสอบกลุ่มหนึ่งเราจะพบว่าคะแนนของผู้สอบกลุ่มเดียวกันอาจจะสูงในแบบสอบฉบับหนึ่ง และคะแนนของผู้สอบกลุ่มเดียวกันนี้จะต่ำในอีกแบบสอบหนึ่ง หรืออาจกล่าวอีกอย่างหนึ่งได้ว่าเราจะพบความคงเส้นคงวาของการปฏิบัติของกลุ่มผู้เข้าสอบในแบบสอบที่ต่างกัน ความคงเส้นคงวานี้ เราอ้างว่าเพราะมีบางสิ่งบางอย่างภายในกลุ่มผู้เข้าสอบที่เป็นเหตุให้เขาได้คะแนนที่คงที่ซึ่งเราเรียกบางสิ่งบางอย่างภายในกลุ่มผู้เข้าสอบที่เป็นเหตุให้เขาได้คะแนนคงที่เช่นนี้ว่าคุณลักษณะทางสมอง (mental trait) คำว่าคุณลักษณะ (trait) ในความเข้าใจทั่วไปเป็นลักษณะ (characteristic) ที่ติดตัวมาแต่กำเนิด และถ่ายทอดทางพันธุกรรมได้ ซึ่งความเข้าใจเช่นนี้เป็นเรื่องที่ไม่ถูกต้อง จริงๆแล้ว trait หมายถึง คุณลักษณะของผู้เข้าสอบที่เป็นต้นเหตุทำให้ผลการปฏิบัติจากแบบสอบคงที่ คุณลักษณะทางสมองไม่สามารถที่จะอ้างอิงได้ถึงลักษณะทางกายภาพ มองเห็นได้เช่นเดียวกับตัวแปรเกิน (intervening variable) ที่ได้ถูกเสนอไว้ว่าเป็นโครงสร้างทางทฤษฎี (hypothetical construct) ดังนั้น คุณลักษณะทางสมองจึงไม่สามารถสังเกตได้โดยตรง เราจึงเรียกอีกอย่างหนึ่งว่า คุณลักษณะแฝง (latent trait)

คุณลักษณะแฝงนี้ได้กำหนดโดยให้ชื่อมาแต่เดิมด้วยอักษรกรีก theta (θ) คำว่า theta (θ), ระดับความสามารถ (ability level), ปริมาณคุณลักษณะ (amount of trait) และปริมาณของความรู้ในเนื้อหาวิชามีความหมายเช่นเดียวกัน

ความมุ่งหมายของการสอบก็คือความรู้ หรือบรรยายถึงความสามารถ (θ) ของผู้เข้าสอบแต่ละคน และเพื่อแสดงตำแหน่งสัมพัทธ์ของผู้เข้าสอบต่าง ๆ บนมาตราความสามารถ (theta Scale) การสอบดังกล่าวก็เพื่อบอกความสามารถ (θ) ของผู้เข้าสอบคนนั้นโดยแสดงด้วยการวัดความสามารถซึ่งแสดงผลด้วยคะแนนจากผลการสอบนั่นเอง ทฤษฎีการตอบสนองต่อข้อสอบ จะเป็นทฤษฎีที่อธิบายให้เห็นถึงความสัมพันธ์ระหว่างความสามารถของผู้เข้าสอบกับพฤติกรรมการตอบสนองต่อข้อสอบนั้น และแสดงความสัมพันธ์ดังกล่าวด้วยฟังก์ชั่นหลายแบบในเชิงคณิตศาสตร์

1. ความเป็นมาของทฤษฎีการตอบสนองต่อข้อสอบ

จุดเริ่มต้นของทฤษฎีคุณลักษณแฝง (Latent Trait Theory) สามารถศึกษาได้จากผลงานของเฟอร์กูสัน (Ferguson) ในปี 1942 และ ลอว์เลย์ (Lawley) ในปี 1943 ทฤษฎีการตอบสนองต่อข้อสอบเป็นเพียงโมเดลหนึ่งในหลายๆ โมเดลของทฤษฎีคุณลักษณะแฝง และราส์ชโมเดล (Rasch Model) ก็เป็นอีกโมเดลหนึ่ง ผลงานที่ตีพิมพ์ในระยะแรกๆ ที่ให้มโนทัศน์คล้ายคลึงกันเป็นผลงานของบรอกเด็น (Brogden) ในปี 1952 ทักเกอร์ (Tucker) ในปี 1946 คาร์รอล (Carrol) ในปี 1950 และครอนบาคกับวาร์ริงตัน(Cronbach and Warrington) ในปี 1952 (Warm 1978 : 19)

ในปี 1952 ลอร์ด (Lord) ได้พิมพ์ผลงานวิจัยเพื่อทำปริญญาเอกซึ่งเขาได้แสดงทฤษฎีการตอบสนองต่อข้อสอบเป็นโมเดลหรือทฤษฎีอย่างชัดเจน ลอร์ดเรียกว่า ทฤษฎีโค้งคุณลักษณะของข้อสอบ (Item Charateristic Curve Theory) ดังนั้น ลอร์ดจึงได้รับการยอมรับว่าเป็นบิดา หรือผู้พบทฤษฎีการตอบสนองต่อข้อสอบ ภายหลังจากการพิมพ์ผลงานวิจัยของเขา ลอร์ดได้หยุดงานเกี่ยวกับ ทฤษฎีการตอบสนองต่อข้อสอบ เป็นเวลา 10 ปี เพราะมีปัญหาในบางประการจากทฤษฎีของเขา

ในปี 1960 ราส์ช (Rasch) ได้ตีพิมพ์ผลงานแสดงโมเดลหนึ่งพารามิเตอร์ กลุ่มตัวอย่างอิสระ (One-Parameter Sample-Free Model) ราส์ชโมเดลนี้ได้รับความสนใจมากและได้นำมาใช้งานสำคัญในช่วงระหว่าง 10 ปีต่อมา โดยได้ถูกนำมาเสนอในอเมริกาโดย

ไรท์ (Wright) นักจิตวิเคราะห์ ณ มหาวิทยาลัยชิคาโก

ในปี 1965 ลอร์ดได้นำเสนอผลงานวิจัยที่สำคัญของเขาซึ่งศึกษาโดยใช้กลุ่มตัวอย่างที่มีขนาดใหญ่กว่า 100,000 คน การศึกษานี้แสดงว่า ปัญหาที่เขาพบในงานของเขาในเวลาอันยาวนานก่อนหน้านี้นั้นไม่เป็นปัญหาจริงๆ และทฤษฎีการตอบสนองต่อข้อสอบเหมาะสมกับแบบสอบแบบเลือกตอบได้จริง จากผลงานของลอร์ดที่เกี่ยวกับทฤษฎีการตอบสนองต่อข้อสอบ จึงเริ่มขึ้นอีกครั้งหนึ่ง

ในปี 1968 ลอร์ดและโนวิค (Novick) ได้พิมพ์ตำราเรียนด้านการวัดทางจิตวิทยาขึ้นเล่มหนึ่งซึ่งบทที่ 17 -20 เขียนโดยเบิร์นบวม (Birnbaum) (Warm 1978: 20) นักสถิติที่เป็นที่รู้จักดีผู้หนึ่ง ในบทที่เบิร์นบวมเขียนเป็นงานที่ให้รายละเอียดในคณิตศาสตร์ของ 2 และ 3 พารามิเตอร์ของนอร์มอลโอไจฟ์ (Normal Ogive) และโลจิสติกโมเดล (Logistic Model)

ต่อมาในปี 1970 เออร์รี (Urry) (Warm 1978: 20) ได้ทำงานวิจัยเพื่อทำปริญญาเอกเสร็จสมบูรณ์โดยเขาได้เปรียบเทียบโมเดลหนึ่ง สอง และสาม พารามิเตอร์ เขาสรุปว่าโมเดลแบบสามพารามิเตอร์บรรยายลักษณะที่แท้จริงของแบบสอบแบบเลือกตอบได้ดีที่สุด ตั้งแต่งานวิจัยของเออร์รีเป็นต้นมา งานวิจัยทั้ง 3 โมเดลคือ หนึ่งพารามิเตอร์ สองพารามิเตอร์ และ สามพารามิเตอร์ ได้เกิดขึ้นมาก แต่โมเดลแบบสามพารามิเตอร์ได้รับความสนใจมากที่สุดในปัจจุบันเพราะว่าเป็นโมเดลบรรยายคุณลักษณะของข้อสอบได้ชัดเจนที่สุด

ในปี 1977 ลอร์ด ได้เปลี่ยนชื่อโมเดลของเขาจากทฤษฎีโค้งคุณลักษณะข้อสอบเป็นทฤษฎีการตอบสนองต่อข้อสอบ (Item Response Theory)

2. ข้อตกลงเบื้องต้น

ในการพิจารณานำทฤษฎีการตอบสนองข้อสอบไปใช้ จำเป็นที่จะต้องทราบถึงข้อตกลงเบื้องต้นของทฤษฎีนี้ดังนี้

1) แบบสอบมิติเดียว (Unidimensional Tests) จากฟังก์ชั่นความน่าจะเป็นในการตอบข้อสอบถูก ณ ระดับความสามารถต่างๆ (θ) ที่แสดงด้วยฟังก์ชันโลจิสติกหรือโอไจฟ์ปกติดังกล่าวมาแล้วก็ตาม ฟังก์ชั่นทั้งสองยืนยันว่าความน่าจะเป็นในการทำข้อสอบแต่ละข้อถูกขึ้น

อยู่กับพารามิเตอร์ของข้อสอบ 3 ตัว ถ้าเราต้องการทราบความสามารถของคนเพื่อกำหนดความน่าจะเป็นที่จะทำข้อสอบถูก ในข้อใดข้อหนึ่งก็จะไม่ช่วยให้ทราบความน่าจะเป็นที่ทำข้อสอบข้ออื่นๆถูกถ้าข้อสอบที่เพิ่มขึ้นเข้าไปไม่ใช่การวัดความสามารถเดียวกับข้อสอบที่กล่าวมาแล้ว ผลการสอบจากข้อสอบเหล่านั้นจะขึ้นอยู่กับคุณลักษณะบางอย่างมากกว่าความสามารถของผู้สอบที่เรามุ่งวัดดังนั้นแบบสอบตามแนวทฤษฎีนี้จึงควรมีเพียงมิติเดียว ซึ่งโดยทั่วไป ความเป็นมิติเดียวนี้หมายถึงข้อสอบทั้งหลายที่วัดเพียงเนื้อเดียวของความรู้หรือความสามารถ ข้อตกลงเกี่ยวกับความเป็นมิติเดียวนี้เป็นเรื่องที่ซับซ้อนและยุ่งยากมากที่สุดของทฤษฎีการตอบสนองต่อข้อสอบในการทดสอบถึงความเป็นมิติเดียวของแบบสอบส่วนใหญ่แล้วจะใช้วิธีการจากการวิเคราะห์ตัวประกอบ เช่น สังเกตค่าออยเกน (eigen value) ของตัวประกอบที่นำมาลงจุดในกราฟกับอันดับของตัวประกอบ ถ้าค่าออยเกนในตัวประกอบแรกใหญ่กว่าตัวประกอบตัวที่ 2 และออยเกนของตัวประกอบที่เหลือมีค่าใกล้เคียงกันจะแสดงถึงความเป็นมิติเดียว เป็นต้น แต่โดยปกติแล้วจากกฎหัวแม่มือ ก็มักจะพบว่าแบบสอบที่มองดูเหมือนเป็นมิติเดียวก็จะเป็นมิติเดียว

2) ความเป็นอิสระของข้อสอบ (Local Independence) หมายถึง ความน่าจะเป็นของผู้เข้าสอบที่จะทำข้อสอบข้อหนึ่งถูก โดยไม่ได้รับผลกระทบมาจากการตอบข้อกระทงอื่นๆในแบบสอบ ซึ่งสามารถแสดงได้ด้วยสมการทางคณิตศาสตร์ดังนี้

$$P(U_i = 1/\theta) = P(U_i = 1/\theta, U_j\ U_k, \ldots) \text{ เมื่อ } i \neq j, k$$

นั่นคือ ความน่าจะเป็นในความสำเร็จในตอบข้อ i ณ ความสามารถ (θ) จะเท่ากับความน่าจะเป็นของความสำเร็จในการตอบข้อ j จากความสามารถ θ โดยไม่เกี่ยวกับความสามารถในการประกอบการในข้อ j หรือ k ใดๆ ซึ่งเราสามารถหาความน่าจะเป็นของความสำเร็จจากข้อสอบทั้งหมดได้เท่ากับผลคูณของความน่าจะเป็นของความสำเร็จที่แยกออกจากกันในแต่ละข้อดังนี้

$$P(U_i = 1, U_j = 1, U_k = 1/\theta) = P(U_i = 1/\theta)P(U_j = 1/\theta)\ p(U_k = 1/\theta)$$

เช่นความน่าจะเป็นในการตอบข้อ 1 ผ่าน = .8 และความน่าจะเป็นในการตอบข้อ 2 ถูกเท่ากับ .9 ถ้าข้อสอบทั้งสองข้อเป็นความเป็นอิสระกันแล้ว ผู้เข้าสอบจะต้องมีความน่าจะเป็น = (.8) (.9) = .72 ในการผ่านทั้ง 2 ข้อ

มีวิธีการอื่นๆ ที่จะแสดง ความเป็นอิสระ ของข้อสอบคือ การพูดว่าความน่าจะเป็นที่ผู้เข้าสอบ A ผ่านข้อสอบที่ 2 (โดยกำหนดให้ผู้เข้าสอบ A ผ่านข้อ 1) เป็นเพียงความน่าจะเป็นที่ผู้เข้าสอบ A ผ่านข้อ 2 คล้ายๆ กับ ความเป็นอิสระ ของผู้เข้าสอบต่างๆ สามารถอธิบายได้ว่า คะแนนของผู้เข้าสอบที่แตกต่างกันเป็นอิสระจากกันโดยกำหนดค่าคะแนนจริงหรือค่าคุณลักษณะแฝงของผู้เข้าสอบมาให้ ยกตัวอย่างถ้าผู้เข้าสอบ A มีความน่าจะเป็น .2 ของการผ่านข้อที่กำหนดไว้ข้อหนึ่ง และผู้เข้าสอบ B มีค่าความน่าจะเป็น .9 ของการผ่านที่ข้อเดียวกัน และถ้าคำตอบของผู้เข้าสอบมีความเป็นอิสระกันแล้ว ความน่าจะเป็นที่ผู้เข้าสอบทั้ง 2 คนผ่านข้อสอบ นั้นคือ (.2) (.9) = .18

ถ้าการกระทำในการตอบข้อสอบข้อหนึ่ง หรือความรู้ในการตอบของข้อสอบข้อหนึ่งมีอิทธิพลต่อการตอบข้อสอบอื่นๆ แล้วข้อสอบนั้นจะไม่เป็นอิสระกัน ยกตัวอย่างเช่น ในการสำรวจทัศนคติเนื้อเรื่องที่ถูกสร้างขึ้น โดยคำตอบที่คำถามบางข้อมีอิทธิพลต่อคำตอบในลำดับต่อมา จะไม่เหมาะสมสำหรับโมเดลคุณลักษณะแฝงนี้ และถ้าการเปลี่ยนลำดับของการบริหารของชุดของข้อสอบมีอิทธิพลต่อการสอบแล้ว ข้อสอบนั้นก็ไม่เป็นอิสระกัน ข้อสอบที่เป็นลูกโซ่ (Chained items) ก็ไม่เป็นอิสระเช่นกัน ข้อสอบที่เป็นลูกโซ่ หมายถึงข้อสอบต่างๆ ซึ่งคำตอบขึ้นอยู่กับคำตอบในข้อที่มาก่อน ในอีกประเด็นหนึ่งที่แสดงถึงความไม่เป็นอิสระก็คือ ถ้าคะแนนของผู้เข้าสอบคนหนึ่งมีอิทธิพลต่อคะแนนของผู้เข้าสอบคนอื่นๆ โดยผ่านการอภิปรายหรือร่วมมือกันให้ได้มาซึ่งคำตอบ จริงๆ แล้วลักษณะของความเป็นอิสระได้มาโดยอัตโนมัติจากความเป็นมิติเดียวนั่นเอง

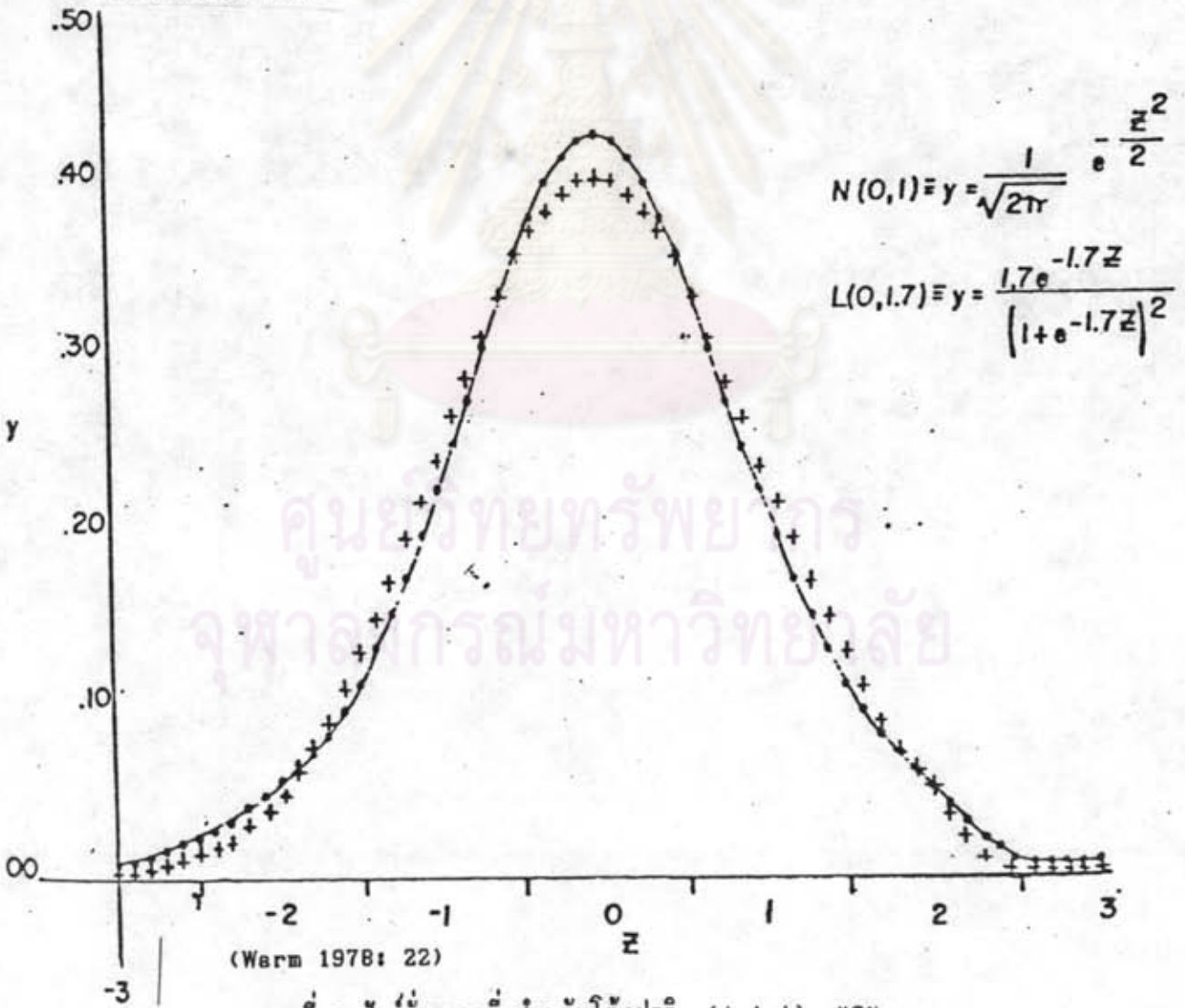
3) โค้งคุณลักษณะของข้อสอบจะเป็นฟังก์ชันแสดงความสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับโอกาสการตอบถูก ซึ่งมีลักษณะเป็นโค้งโลจิส โดยโค้งคุณลักษณะของข้อสอบมีหลายรูปแบบขึ้นอยู่กับโมเดล เช่น โมเดลโลจิสหนึ่งพารามิเตอร์ โมเดลโลจิสสองพารามิเตอร์ โมเดลโลจิสสามพารามิเตอร์ (Lord 1980 : 12-13)

4) ทฤษฎีการตอบสนองต่อข้อสอบอธิบายว่าการที่ผู้สอบจะมีโอกาสที่จะตอบถูกหรือผิดขึ้นอยู่กับความสามารถของผู้สอบเท่านั้น ดังนั้นข้อสอบที่มีลักษณะเป็นข้อสอบวัดความเร็ว (speed test) จะไม่สามารถประยุกต์ใช้กับทฤษฎีนี้ได้ เพราะข้อสอบที่ผู้สอบไม่ได้ทำแต่ได้คะแนนเป็นศูนย์อาจจะเป็นเพราะไม่มีเวลาทำไม่ใช่เพราะทำแล้วตอบผิด

4) ทฤษฎีการตอบสนองต่อข้อสอบอธิบายว่าการที่ผู้สอบจะมีโอกาสที่จะตอบถูกหรือผิดขึ้นอยู่กับความสามารถของผู้สอบเท่านั้น ดังนั้นข้อสอบที่มีลักษณะเป็นข้อสอบวัดความเร็ว (speed test) จะไม่สามารถประยุกต์ใช้กับทฤษฎีนี้ได้ เพราะข้อสอบที่ผู้สอบไม่ได้ทำแต่ได้คะแนนเป็นศูนย์อาจจะเป็นเพราะไม่มีเวลาทำไม่ใช่เพราะทำแล้วตอบผิด

๓. ความเป็นมาของโอไจฟ์ปกติและโอจีฟโลจิสติก

เพื่อที่จะเข้าใจถึงแนวทางทฤษฎีการตอบสนองข้อสอบในการอธิบายคุณลักษณะของข้อสอบจะแสดงแนวความคิดพื้นฐานในเรื่อง โค้งลักษณะของข้อสอบในแบบโอไจฟ์ปกติ และโอจีฟโลจิสติก ต่อไป



(Warm 1978: 22)

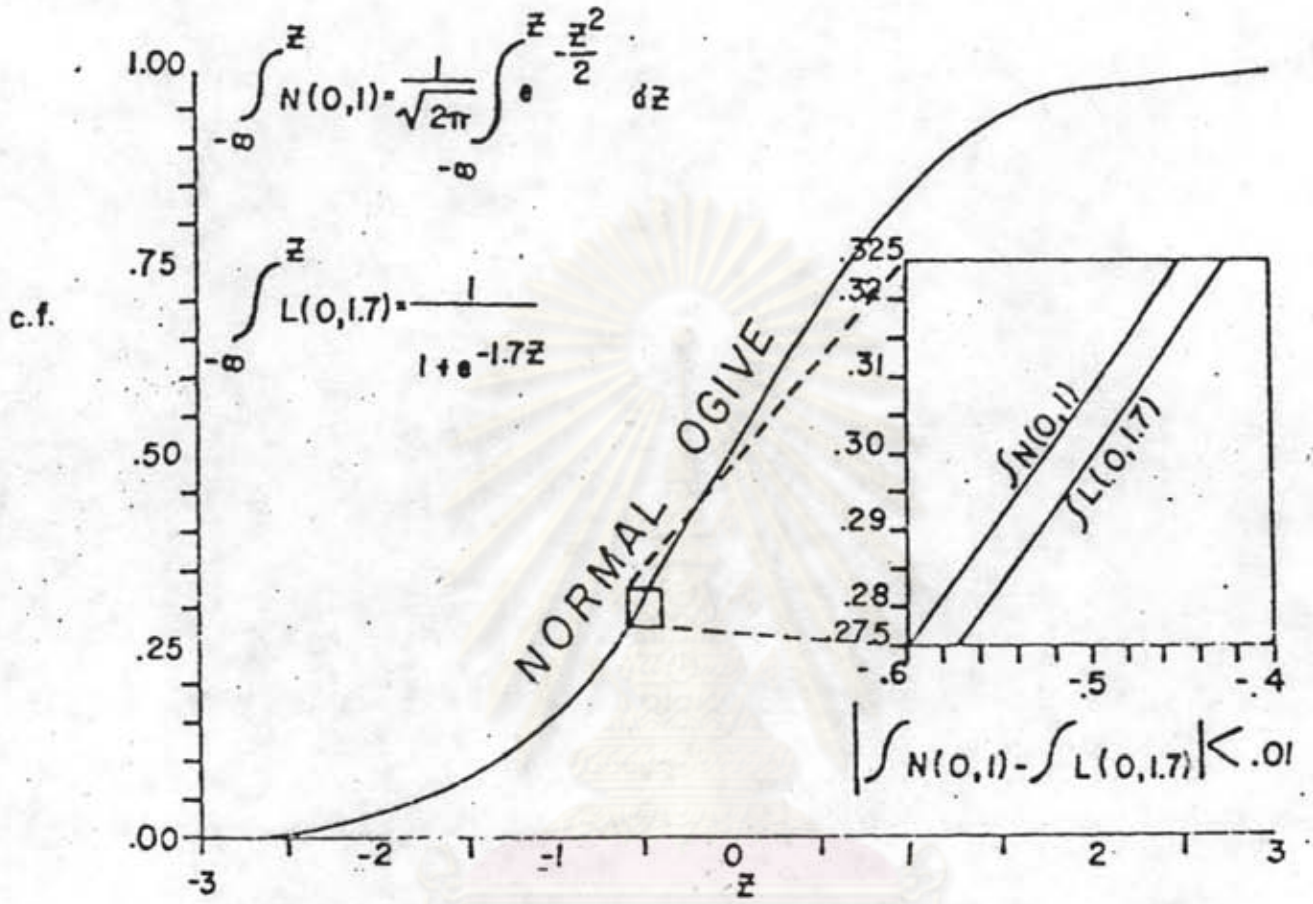
ภาพที่ 6 ฟังก์ชั่นความถี่ สำหรับโค้งปกติ (+ + +) และโค้งโลจิสติก (...)

ในโค้งการแจกแจงปกติดังภาพที่ 6 ส่วนที่มีเครื่องหมายบวก (+ + + + +) จะมีสมบัติคือ มีมัชฌิมเลขคณิต = 0 และส่วนเบี่ยงเบนมาตรฐาน = 1 สูตรของโค้งปกติแสดงในภาพที่ 4 แทนด้วย N (0,1) ลักษณะโค้งการแจงแจงปกติเป็นรูประฆังคว่ำเช่นนี้ เรียกว่าฟังก์ชั่นความถี่ ที่ถูกเรียกว่าฟังก์ชั่นความถี่เพราะแกนตั้ง (ordinate) แทนค่าความถี่ สัดส่วน ร้อยละ หรือความหนาแน่น (density) นั่นเอง ดังนั้นเราจึงเรียกโค้งปกตินี้ว่า ฟังก์ชั่นความถี่ปกติ

โค้งที่อยู่เหนือฟังก์ชันความถี่ปกติในภาพที่ 6 คือโค้งโลจิสติกหรือฟังก์ชันความถี่โลจิสติก แสดงด้วย จุด (.....) ฟังก์ชันความถี่โลจิสติก นี้มีค่ามัชฌิมเลขคณิต = 0 และส่วนเบี่ยงเบนมาตรฐานประมาณ 1.0 สูตรของฟังก์ชันความถี่โลจิสติก แสดงดังภาพที่ 6 แทนด้วย L(0, 1.7) ค่า 1.7 เป็น ค่ายกกำลัง ของสูตรที่ให้ ฟังก์ชันความถี่โลจิสติก เป็นค่าประมาณ ฟังก์ชันความถี่ปกติ ได้ใกล้เคียงมากที่สุดเท่าที่จะเป็นไปได้ ค่าจริงๆคือ 1.6679 ซึ่งประมาณด้วยค่า 1.7 ในการศึกษาบางครั้งค่า 1.7 นำมาแทนกรณีขีดจำกัดบนด้วยอักษร D ตัวอักษร e คือฐานของล็อกธรรมชาติ (natural logarithms) โดยที่ e มีค่าประมาณ 2.718281828

โค้งรูปเอส (S-Shaped Curve) ในภาพที่ 7 แสดงโค้งความถี่สะสมปกติ ซึ่งเรียกโค้งรูปเอสนี้ว่า โอไจฟ์ โค้งนี้แสดงว่าสัดส่วนของพื้นที่โค้งปกติ (ของภาพที่ 6) ที่อยู่ทางด้านซ้ายในแต่ละจุดบนแกนนอน (abaissa) โอไจฟ์ ดังกล่าวนี้ถูกเรียกว่าเป็นฟังก์ชั่นการแจกแจง (distribution function) ที่ถูกเรียกว่าฟังก์ชั่นการแจกแจงเพราะแกนตั้งแทนความถี่สะสม สัดส่วนสะสม เปอร์เซนต์สะสมหรือพื้นที่สะสม ดังนั้นเราจึงเรียกโค้งดังภาพที่ 7 นี้ว่าฟังก์ชั่นการแจกแจงปกติ (normal distribution function) หรือโอไจฟ์ปกติ สูตรของฟังก์ชันความหนาแน่นปกติแสดงในภาพที่ 7 แทนด้วย $\int N(0, 1)$

จากภาพที่ 7 แสดงโค้งโลจิสติก หรือฟังก์ชันความหนาแน่นโลจิสติกสำหรับ ฟังก์ชันความถี่โลจิสติก ในภาพที่ 7 มันไม่สามารถจะมองเห็นได้ เพราะมันเข้าใกล้โอไจฟ์ปกติ มากจนกระทั่ง 2 มาตรานี้อยู่ด้วยกัน โดยที่มีค่าเข้าใกล้มากและได้แสดงส่วนย่อยๆ นี้ โดยนำมาขยายให้ใหญ่ขึ้น 10 เท่า เพื่อว่าจะเห็นชัดเจนขึ้น พื้นที่ที่ขยายออก



(Warm 1978: 24)

ภาพที่ 7 ฟังก์ชั่นการแจกแจง สำหรับฟังก์ชั่นความถี่ของ

N (0,1) และ L (0,.17)

ของโอไจฟ์ทั้งสองนั่นคือส่วนที่ห่างกันจากการพิสูจน์จะพบว่าจุดใดๆ บนแกนนอนของ 2 โอไจฟ์จะมีค่าน้อยกว่า 0.01 ซึ่งแยกต่างหากจากบนแกนตั้ง ข้อมูลนี้ได้แสดงถึงความไม่เท่ากันได้ส่วนขยาย ในภาพที่ 7 สูตรของฟังก์ชันความหนาแน่นโลจิสติกได้รับการนิยามดังภาพที่ 7 แทน ด้วย $\int$ L (0, 1.7)

โอไจฟ์ ที่จะกล่าวในรายงานนี้เกี่ยวข้องกับ โอไจฟ์ปกติ สัญญลักษณะ $\int$ คือ เครื่องหมาย อินทิกรัล (integral sign) สำหรับฟังก์ชันความหนาแน่นปกติแทนด้วย $\int N(0,1)$ ทำให้โอไจฟ์ปกติ ยากที่จะคำนวณด้วยวิธีการทางคณิตศาสตร์และการกำหนดวิธีแห่งตัวเลขเพื่อแก้ปัญหา หรือตารางแสดงค่าต่างๆ

แต่โอไจฟ์โลจิสติกไม่มีเครื่องหมายอินทิกรัลที่ด้านขวาของสมการนิยาม($\int L(0, 1.7)$ ซึ่งได้แสดงไว้ในภาพที่ 7 เป็นฟังก์ชั่นทางพีชคณิตที่บรรยายถึงโอไจฟ์โลจิสติก ซึ่งคำนวณง่ายกว่า

จากภาพที่ 7 แสดงเพียงหนึ่งโอไจฟ์โลจิสติก เท่านั้นจริงๆ แล้ว กลุ่มชุดของโอไจฟ์โลจิสติก (โอไจฟ์ปกติ) มีมากมายนับไม่ถ้วนและแต่ละอันก็มีวิธีการบางอย่างแตกต่างไปจากอีกอันหนึ่ง โอจิฟโลจิสติกมีลักษณะเป็นฟังก์ชั่นเพิ่มอย่างสม่ำเสมอ (strictly monotonic functions) มันเป็นฟังก์ชั่นเพิ่มอย่างสม่ำเสมอเพราะมันจะเพิ่มขึ้นจากซ้ายไปขวา และจะสูงขึ้นเรื่อยๆ ปลายสุดทางซ้ายจะไม่จดแกนนอนเลย และปลายสุดทางขวาจะไม่ลดลง สังเกตได้ว่าจากภาพที่ 7 ช่องระหว่าง -2.0 และ -0.5 บนค่าตามแกนนอนโอไจฟ์จะมีลักษณะหงายขึ้น (concave upward) แต่จุดระหว่าง - 0.5 และ 0.5 โอไจฟ์ นี้จะเปลี่ยนจากหงายขึ้นเป็นหงายลง จุดที่เปลี่ยนโค้งนี้เรียกว่าจุดเปลี่ยนโค้ง (inflection point) จุดนี้เป็นจุดที่ความชัน (slope) ของโอไจฟ์มีค่าสูงสุด จุดนี้สำหรับโอไจฟ์ถูกแสดงด้วยค่าตามแกนตั้งที่ 0.50 และบนแกนนอนที่ค่า 0.0 ด้วยโอจิฟโลจิสติกนี้สามารถแสดงค่าพารามิเตอร์ 3 ค่าได้ 3 วิธีด้วยวิธีการละ 1 ค่า วิธีแรกโอไจฟ์โลจิสติก อาจจะแตกต่างกันในตำแหน่งของค่าบนแกนนอน α จุดเปลี่ยนโค้งภาพที่ 8 แสดงโอไจฟ์โลจิสติก 3 แบบคือ E, F, G ที่มีจุดเปลี่ยนโค้งแตกต่างกันบนแกนนอน โอไจฟ์ทั้ง 3 มีคุณสมบัติอย่างเหมือนกันเพียงแต่ตำแหน่งที่เลื่อนออกไปทางด้านข้าง ตำแหน่งบนแกนนนอน α จุด เปลี่ยนโค้งเรียกว่า พารา-มิเตอร์ b ค่า b ของโอไจฟ์ E, F, และ G ในภาพที่ 8 คือ -.5, 0.0 และ 1.0 ตามลำดับ เพื่อสรุปถึงพารามิเตอร์ b ในฟังก์ชันโอจิฟโลจิสติกจึงจำเป็นต้องแยกพารามิเตอร์ออกจากตัวแปรแกนนอน

จากภาพที่ 6,7 และ 8 เขียนขึ้นโดยแสดงแกนนอนด้วย Z การแสดงนี้เพื่อให้สะดวกในการเข้าใจถึงฟังก์ชันความถี่โลจิสติกและ ฟังก์ชันความหนาแน่น เนื่องจากความ

คุ้นเคยกับคะแนนมาตรฐาน (Z - scores) ตามประเพณีนิยมของการวัด แต่เนื่องจากการวัดในที่นี้เกี่ยวกับมาตราความสามารถจึงถูกเรียกว่า θ ซึ่งในอันดับต่อไปจะได้แสดงค่าตามแกนนอนด้วย θ เมื่อแทน θ ด้วย z ใน ฟังก์ชันโลจิสติกและแทนค่าพารามิเตอร์ b โดยให้ส่วนสูงของโอไจฟ์โลจิสติก จะแสดงด้วยฟังก์ชั่นดังนี้

$$\Psi(\theta)=\left[1+e^{-1.7(\theta-b)}\right]^{-1}$$

บางทีเขียนว่า

$$\Psi(\theta)=\left[1+\exp(-1.7(\theta-b))\right]^{-1}$$

exp หมายถึง e ยกกำลังอะไรก็ตามอยู่ในเครื่องหมายวงเล็บ หลังจากหาค่ายกกำลังแล้ว กรณีข้างบนอักษรกรีก psi (Ψ) นำมาแสดงถึงโค้งโลจิสติก ส่วน phi (ϕ) หมายถึงโอไจฟ์ปกติ

โอไจฟ์โลจิสติก จะมีเส้นกำกับ (asymptotes) สองเส้นซึ่งเส้นกำกับคือเส้นแนวนอนที่โอไจฟ์เข้าไปใกล้ได้มากที่สุดแต่ไม่เคยจะไปถึงเส้นนี้ (เส้นกำกับบนถูกกำหนดด้วยค่าตามแกนตั้งที่ 1.0) ในภาพที่ 7 และ 8 จะพบว่า ส่วนขวาด้านบนของโอไจฟ์โลจิสติกเข้าใกล้ค่า 1.00 บนแกนตั้ง ในภาพที่จะปรากฏว่าสัมผัสกับเส้นแนวนอน 1.00 แต่ที่พูดมาแล้วว่ามันจะไม่ตัดเส้นดังกล่าวที่เดียว เส้นกำกับล่างของโอไจฟ์ ในภาพที่ 7 และ 8 คือแกนนอนที่มีความสูงเป็นศูนย์ ส่วนของโอไจฟ์จะไม่เคยถึงค่า 1.00 ในขีดจำกัดบนและไม่เคยถึงเส้นกำกับล่างในขีดจำกัดล่างนี้เลย

โอไจฟ์โลจิสติก ทั้งหมดในทฤษฎีการตอบสนองต่อข้อสอบมีเส้นกำกับบนที่ 1.00 แต่เส้นกำกับ ล่างทั้งหมดไม่เคยถึง 0.00

ในภาพที่ 9 แสดงโอไจฟ์โลจิสติก 3 แบบ ให้ชื่อว่า H, J, K ซึ่งเหมือนกันยกเว้นเส้นกำกับล่างที่ต่างกัน เส้นกำกับล่างมีค่า α ค่าบนแกนตั้ง .05,. 25 และ .30 ตามลำดับ ค่า b ในแต่ละโอไจฟ์ = 0.0 สังเกตว่า เส้นกำกับ ของทั้ง 3 โอไจฟ์อยู่ที่ 1.00

ภาพที่ 8 โค้งลักษณะข้อสอบ 3 ข้อ (E, F และ G) ที่มีค่า b = -1.5, 0.0, และ 1.0 ตามลำดับ

(Warm 1978: 28)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สังเกตด้วยว่าจุดเปลี่ยนโค้งสำหรับโอไจฟ์ ในภาพที่ 9 อยู่ที่ความสูงแตกต่างกันในความจริงเป็นครั้งทางระหว่างเส้นกำกับบนและล่างของแต่ละโอไจฟ์ ซึ่งจะพบในกรณีเช่นนี้เสมอเส้นกำกับล่างถูกเรียกว่า พารามิเตอร์ c หรือค่า c ซึ่งเป็นพารามิเตอร์อีกตัวหนึ่งในสามพารามิเตอร์ของทฤษฎีการตอบสนองต่อข้อสอบ

ผลของค่า c แสดงในโอไจฟ์ด้วยช่วงค่าแกนตั้งเมื่อขนาดช่วงเล็กลง ช่วงที่ลดลงจะเท่ากับ 1 - c ผลของช่วงตามแกนตั้งที่ถูกลดลงจะไปลดความชันของโอไจฟ์ ที่ทุกๆ จุดบนมาตรา 0 เมื่อค่าอื่นๆ เท่ากัน เราจึงรวมเอาพารามิเตอร์ ในฟังก์ชันโลจิสติก โดยคูณด้วย 1 - c และบวกกับ c ดังนี้

$$\Psi(\theta) = c + (1-c)\left[1+e^{-1.7(\theta-b)}\right]^{-1}$$

ซึ่งเหมือนกับ

$$\Psi(\theta) = c + (1-c)\left[1+\exp(-1.7(\theta-b))\right]^{-1}$$

และ

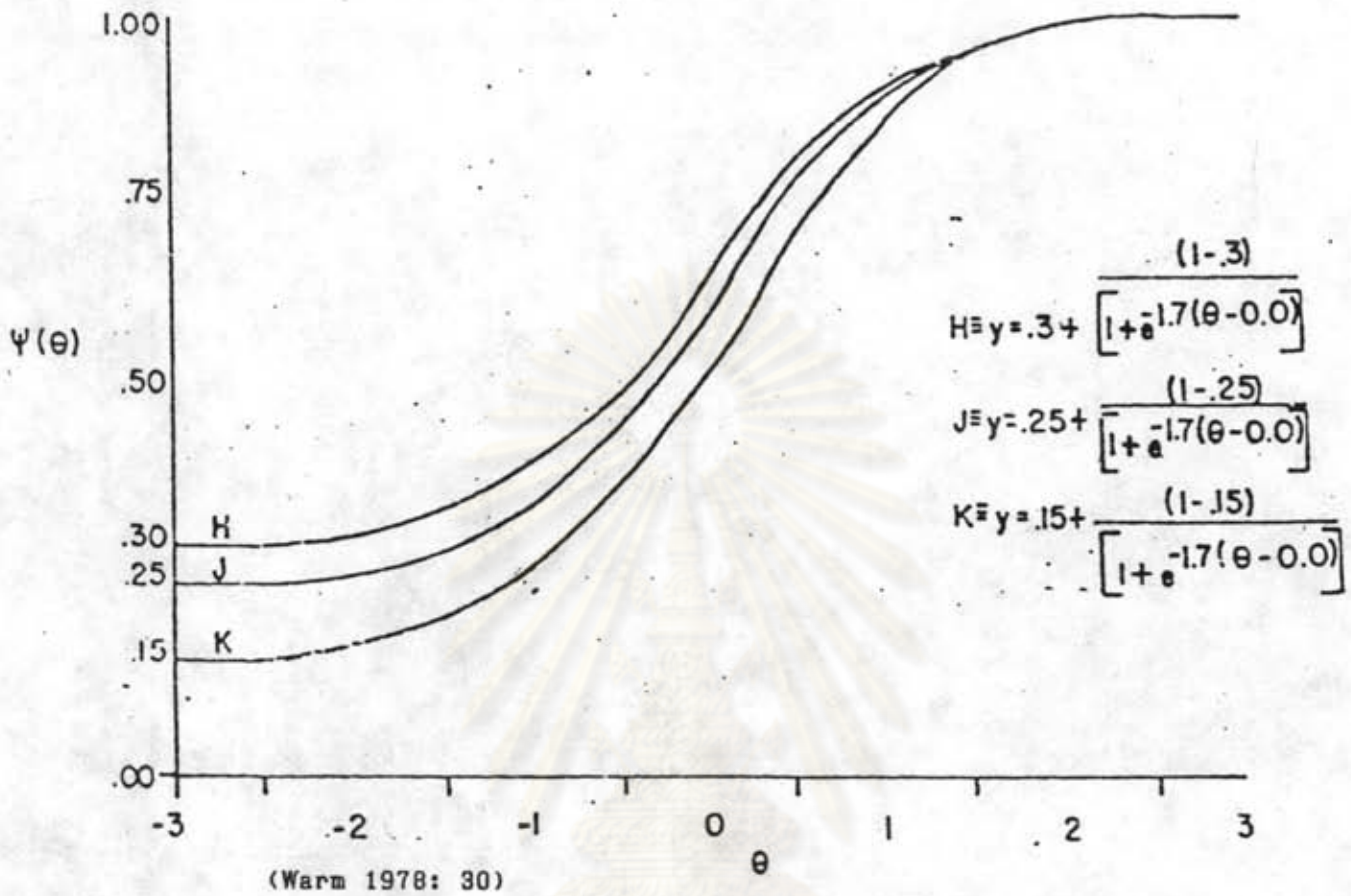
$$\Psi(\theta) = c + \frac{(1-c)}{\left[1+e^{-1.7(\theta-b)}\right]}$$

ค่า c ของ โอไจฟ์ H, J และ K ในภาพที่ 9 คือ .30, .25 และ .15 ตามลำดับ

พารามิเตอร์ตัวที่ 3 ของ ทฤษฎีการตอบสนองต่อข้อสอบ คือ พารามิเตอร์ a หรือค่า a พารามิเตอร์ a มีความสัมพันธ์กับความชันทั้งโอไจฟ์ที่จุดเปลี่ยนโค้งหรือที่ค่า b สำหรับ โมเดลโอไจฟ์ปกติที่มีค่า c = 0.0

$$\sigma = \sqrt{2\pi}\, m \approx 2.5m$$

m คือ ความชันของโอไจฟ์ ที่ค่า b



(Warm 1978: 30)

ภาพที่ 9 โอจีฟโลจิสติก 3 ชุด (H, J และ K) โดยมีค่า b = 0.0 และ c = .30, .25 และ .15 ตามลำดับ

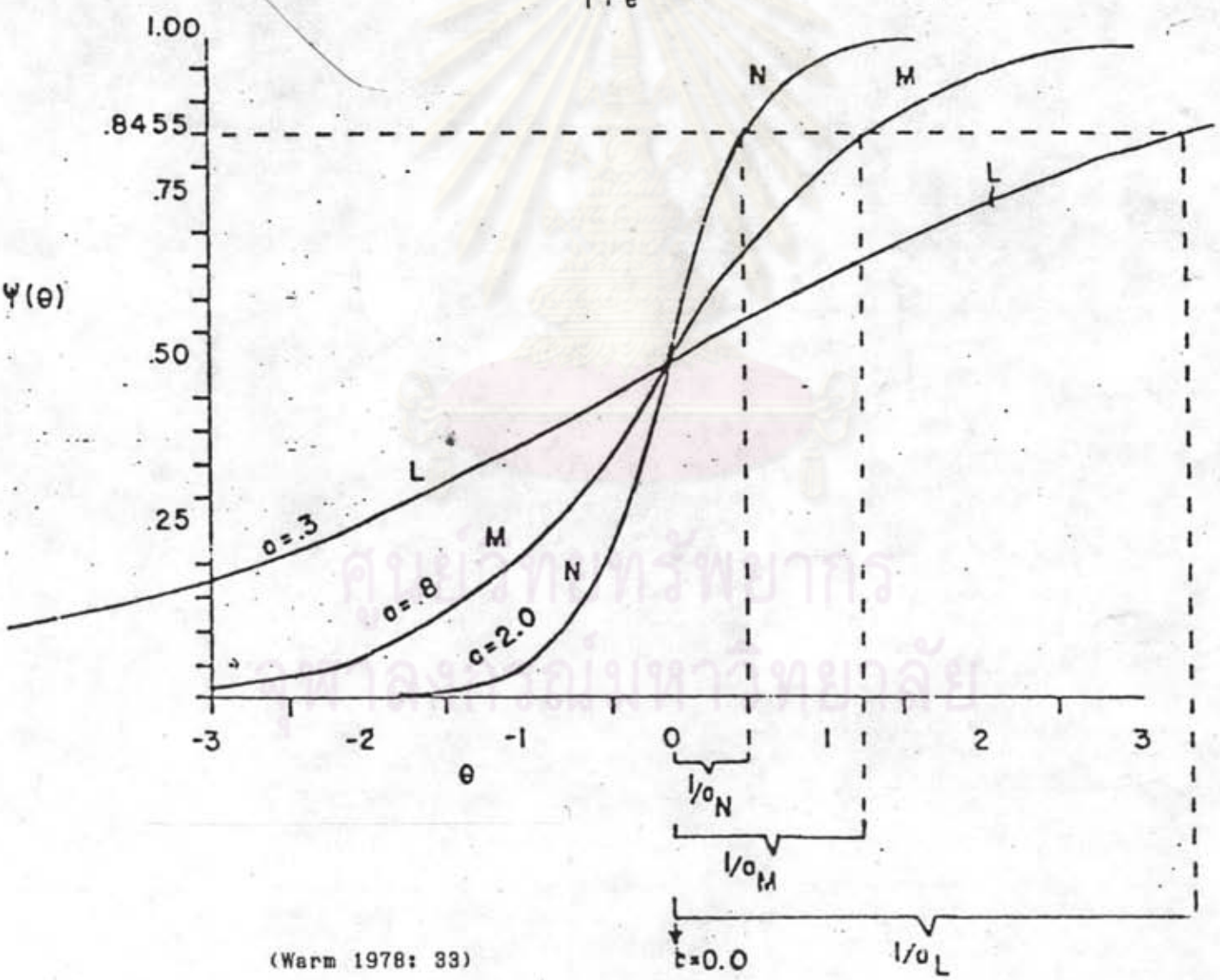
ภาพที่ 10 แสดงโอไจฟ์โลจิสติก 3 แบบ L, M และ N ซึ่งเหมือนกันยกเว้นมีค่า a = .3, .8 และ 2.0 ตามลำดับ และมีค่า b = 0.0, และ c = 0.0 พบว่าเมื่อค่า a ใหญ่ขึ้น โอไจฟ์จะยิ่งชัน โดยเฉพาะอย่างยิ่ง

$$a = \left[\Psi^{-1}(\theta) - b\right]^{-1}$$

ที่ซึ่ง $\Psi^{-1}(\theta)$ = จุดบน θ ณ ความสูงของโอไจฟ์ = c+.8455(1 - c)ค่า -1 ที่แสดงดูเหมือนค่ายกกำลังของ $\Psi^{-1}(\theta)$ แต่ไม่ใช่ค่ายกกำลังเป็นเครื่องหมายแสดงอินเวอร์สของฟังก์ชั่น เป็นฟังก์ชั่นเฉพาะที่ถูกนำมาใช้โดยเริ่มที่จุดใดๆบนแกนนอน ผลของอินเวอร์สจะ

เริ่มที่จุดจุดหนึ่งบนแกนตั้ง (ในกรณีนี้ที่จุด c+.8455(1 - c) และต่อด้วยจุดตามแนวนอนตาม ฟังก์ชั่นตามค่าของจุดบนแนวนอน $\Psi^{-1}(\theta) - 1$ อยู่นอกวงเล็บเป็น ค่ายกกำลัง ซึ่งหมายความถึงส่วนกลับ (reciprocal) จำนวน .8455 เป็นสัดส่วนภายใต้พื้นที่ใต้โค้งฟังก์ชันความถี่โลจิสติกและอยู่ทางด้านซ้ายของค่า Z = 1 ดูภาพที่ 6 ค่า Z มีค่า = 1 เป็นจุดที่สะดวกทางคณิตศาสตร์ที่กำหนดขึ้นอย่างตามอำเภอใจจุดหนึ่ง เมื่อนำพารามิเตอร์ a ใส่เข้าไปในฟังก์ชันโลจิสติก โดยให้เป็นส่วนหนึ่งของค่ายกกำลังของ c จะได้สูตรโอไจฟ์โลจิสติกสามพารามิเตอร์ดังนี้

$$\Psi(\theta) = c + \frac{1-c}{1+e^{-1.7a(\theta-b)}}$$



(Warm 1978: 33)

ภาพที่ 10 โอจีนโลจิสติก 3 ชุด (L, M และ N) โดยมีค่า

b = 0.0, C = .00 a = .3, .8 และ 2.0 ตามลำดับ

4. รูปแบบโมเดลโลจิสติก

โมเดลโลจิสติกถูกพัฒนาขึ้น 3 รูปแบบ คือโมเดลโลจิสติกหนึ่งพารามิเตอร์ (One-Parameter Logistic Model) โมเดลโลจิสติกสองพารามิเตอร์ (Two-Parameter Logistic Model) และโมเดลโลจิสติกสามพารามิเตอร์ (Three-Parameter Logistic Model) ดังนี้

4.1 โมเดลโลจิสติกหนึ่งพารามิเตอร์ โมเดลนี้เบอร์นบวมได้พัฒนาขึ้นในปี 1968 ตรงกับโมเดลของราส์ช (Rasch) ที่เสนอในปี 1960 (Warm 1978 : 19) เป็นโมเดลที่อธิบายข้อสอบด้วยพารามิเตอร์ค่าความยากเพียงค่าเดียว โดยเชื่อว่าโอกาสที่ผู้สอบจะทำข้อสอบได้ถูกหรือไม่นั้นขึ้นอยู่กับความสามารถของผู้สอบกับความยากของข้อสอบ โดยมีฟังก์ชันดังนี้

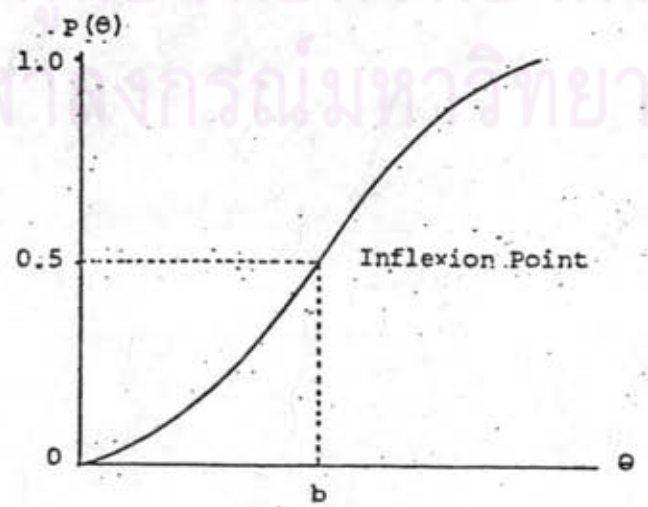
$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1+e^{(\theta - b_i)}}, \quad i = 1, 2, 3, \ldots, n$$

เมื่อ $P_i(\theta)$ คือ โอกาสที่ผู้มีความสามารถ θ จะทำข้อสอบที่ i ได้ถูกต้อง

θ คือ ระดับความสามารถที่แท้จริงของผู้สอบ

b_i คือ ค่าความยากของข้อสอบข้อที่ i

e คือ ค่าคงที่มีค่าเท่ากับ 2.7182818



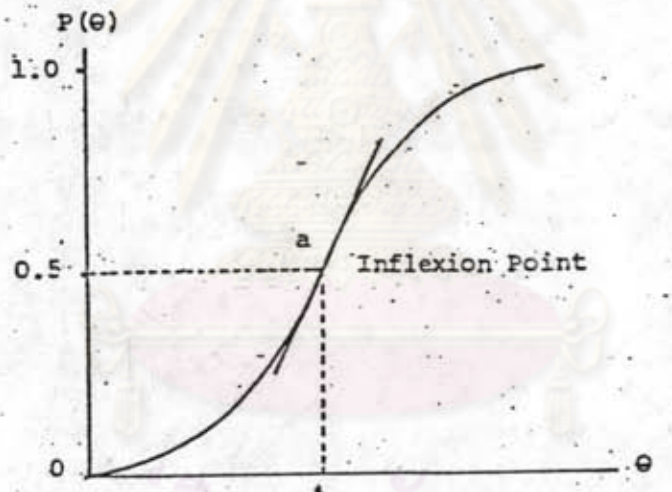
ภาพที่ 11 แสดงค่าความหมายของค่าพารามิเตอร์ของข้อสอบแบบหนึ่งพารามิเตอร์

4.2 โมเดลโลจิสติกสองพารามิเตอร์ เบอร์นบวมได้พัฒนาโมเดลนี้มาจากโมเดลโอไจฟ์ปกติ ซึ่งมีลักษณะใกล้เคียงกันมีฟังก์ชัน ดังนี้ (Hambleton and Cook 1977: 81-82)

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1+e^{Da_i(\theta - b_i)}} \quad ; \quad i = 1, 2, 3, \ldots n$$

a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i

D คือตัวประกอบของเสกลมีค่าเท่ากับ 1.7

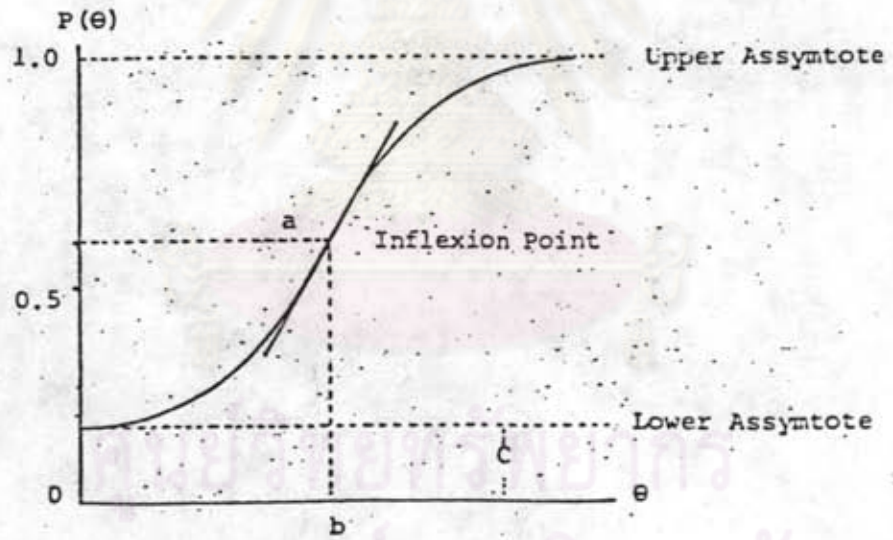


ภาพที่ 12 แสดงความหมายของค่าพารามิเตอร์ของข้อสอบแบบสองพารามิเตอร์

4.3 โมเดลโลจิสติกสามพารามิเตอร์ โมเดลนี้พัฒนาต่อมาจากโมเดลโลจิสติกสองพารามิเตอร์เพื่อใช้กับแบบสอบที่มีอิทธิพลจากการเดาแฝงอยู่ด้วย เช่น ข้อสอบแบบเลือกตอบ โดยมีฟังก์ชันดังนี้

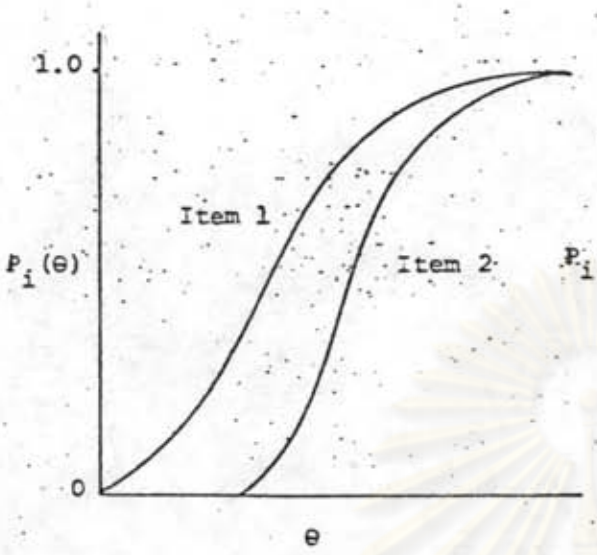
$$P_i(\theta) = c_i + (1-c_i)\frac{e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}}; \quad i = 1,2,3,\ldots n$$

เมื่อ c_i คือ ค่าการเดาของข้อสอบข้อที่ i

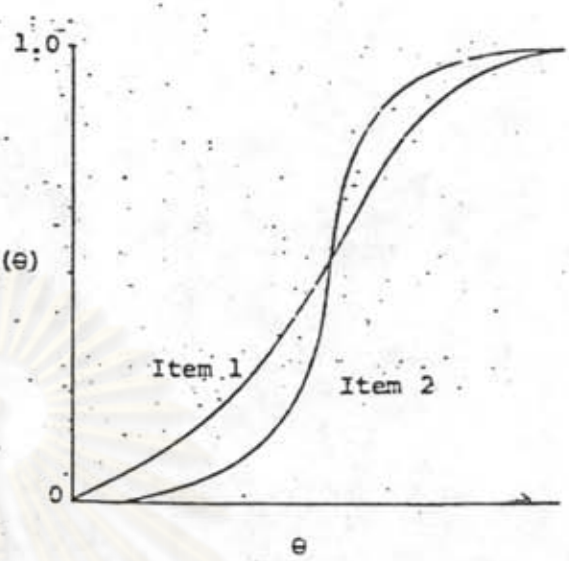


ภาพที่ 13 แสดงความหมายของค่าพารามิเตอร์ของข้อสอบแบบสามพารามิเตอร์

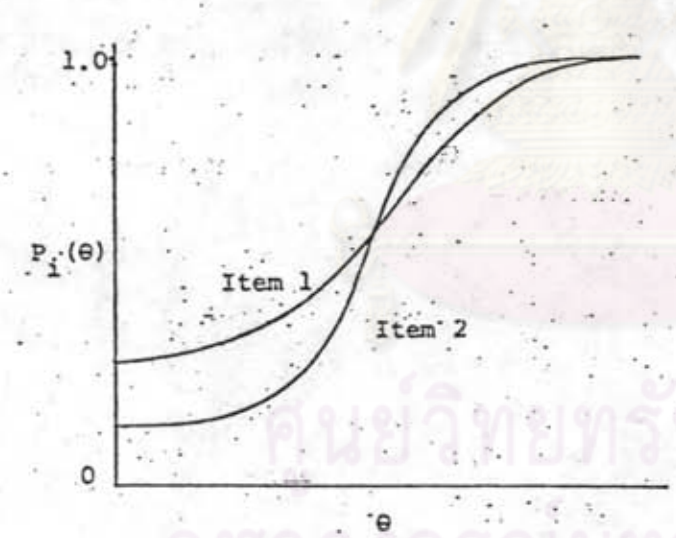
จากโมเดลโลจิสติกทั้งสามโมเดลสามารถเขียนแสดงได้ดังภาพที่ 14 โดย แสดงโค้งคุณลักษณะข้อสอบ ของโมเดลโลจิสติกหนึ่ง สอง สามพารามิเตอร์ (Hambleton and Cook, 1977: 79)



ภาพที่ 14ก โค้งโลจิสติกหนึ่งพารามิเตอร์



ภาพที่ 14ข โค้งโลจิสติกสองพารามิเตอร์



ภาพที่ 14ค โค้งโลจิสติกสามพารามิเตอร์

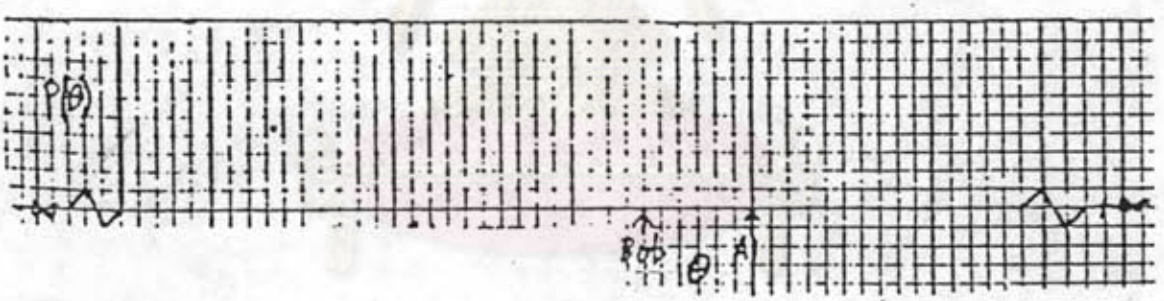
จากภาพที่ 14 แสดงถึงโอกาสที่ผู้ทำข้อสอบได้ถูกต้องขึ้นอยู่กับค่าพารามิเตอร์ ได้แก่ ภาพที่ 14ก แสดงถึงโอกาสที่จะทำข้อสอบถูก ขึ้นอยู่กับค่าความยากเพียงอย่างเดียว โดยถือว่าทุกข้อมีค่าอำนาจจำแนกเท่ากับ 1 และค่าการเดาเท่ากับ 0 จากรูปแสดงว่า ข้อที่ 2 ยากกว่าข้อที่ 1 ภาพที่ 14ข แสดงถึงโอกาสที่จะทำข้อสอบถูก ขึ้นอยู่กับค่าความยากและค่าอำนาจจำแนก จากรูปแสดงว่า ข้อที่ 2 ยากกว่าข้อที่ 1 และข้อที่ 2 จำแนกคนได้ดีกว่าข้อที่ 1

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาพที่ 14ค แสดงถึงโอกาสที่จะทำข้อสอบถูก ขึ้นอยู่กับค่าความยาก ค่าอำนาจจำแนก และค่าการเดา จากรูปแสดงว่า ข้อที่ 1 ยากกว่าข้อที่ 2 และข้อที่ 2 จำแนกคนได้ดีกว่า และมีค่าการเดาน้อยกว่าข้อที่ 1

5. ฟังก์ชันการตอบสนองต่อข้อสอบ (The Item Response Function, IRF)

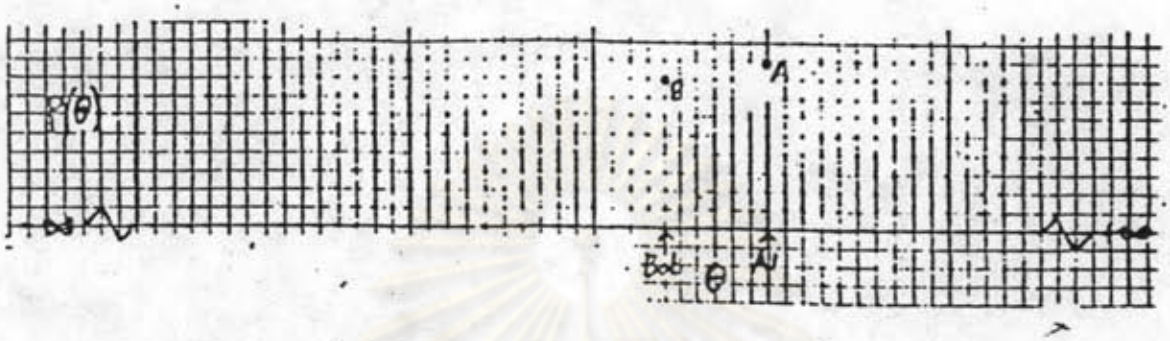
ให้พิจารณาผู้สอบ 2 คน คือ A1 และ Bob ซึ่งมีระดับความสามารถต่างกันคือ θ แตกต่างกันเมื่อกำหนดให้ A1 มี θ มากกว่า Bob หมายความว่าเขาทั้งสองจะถูกกำหนดบนตำแหน่งต่างกันบน θ เสกล ดูภาพที่ 15



ภาพที่ 15 มาตราความสามารถ (θ) กับตำแหน่งสมมติฐานของ A1 และ Bob

โอกาสที่ A1 จะทำข้อสอบข้อ 1 ถูกจะเป็นเท่าใด และโอกาสที่ Bob ทำข้อ 1 ถูกจะเป็นเท่าใด เมื่อก่อนนี้เราจะไม่รู้คำตอบนี้ แต่เดี๋ยวนี้เราทราบได้ว่า A1 มีโอกาสที่จะทำข้อ 1 ถูกมากกว่า Bob เพราะ A1 มีความสามารถ (θ) มากกว่า Bob เมื่อเราได้แสดงความน่าจะเป็นในการตอบถูกในข้อนั้น ณ จุด A และ B ในภาพที่ 16

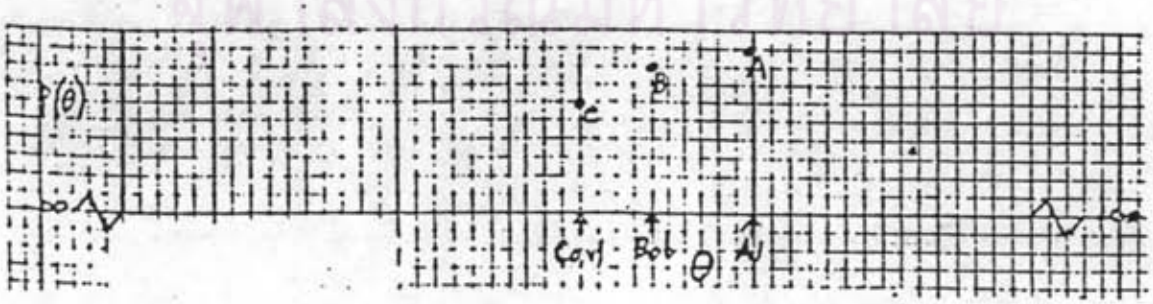
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



ภาพที่ 16 ความน่าจะเป็นของ Al และ Bob ที่ทำข้อ 1 ถูก
ตามค่าฟังก์ชั่นความสามารถของเขา

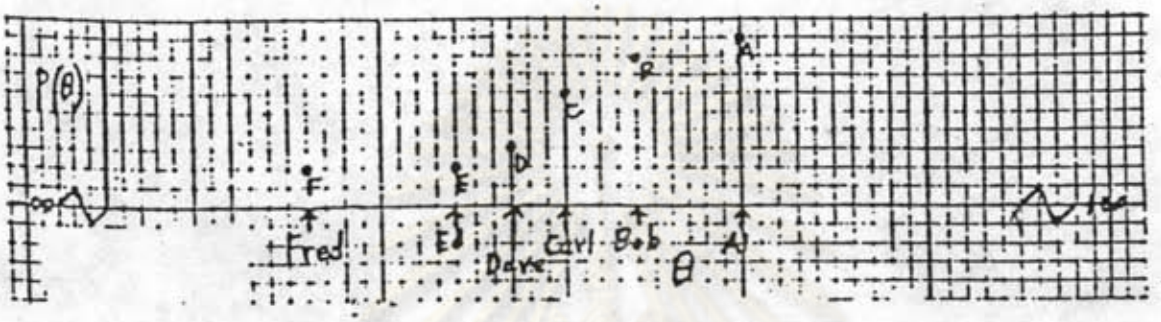
ในการทำเช่นนี้เราจะนิยามค่าแกนตั้งว่าเป็นความน่าจะเป็นที่จะทำข้อสอบข้อนั้นถูก ว่าเป็นฟังก์ชั่นของ θ (ability) ซึ่งอาจจะเขียน $p_i(R|\theta)$ และอ่านว่า "ความน่าจะเป็นในการทำข้อสอบข้อที่ i ถูก ณ ค่า θ นั้น" แต่อาจจะเขียนสั้น ๆ ได้ว่า $P_i(\theta)$ หรืออาจไม่ใส่ตัวแปรย่อย (i) ก็ได้

ต่อมาถ้าเรากำหนดให้ Carl ซึ่งมีความสามารถ (θ) ต่ำว่า Bob Carl จะมีโอกาสน้อยกว่า Bob ที่จะตอบข้อสอบข้อนี้ถูก ดังภาพที่ 17ก



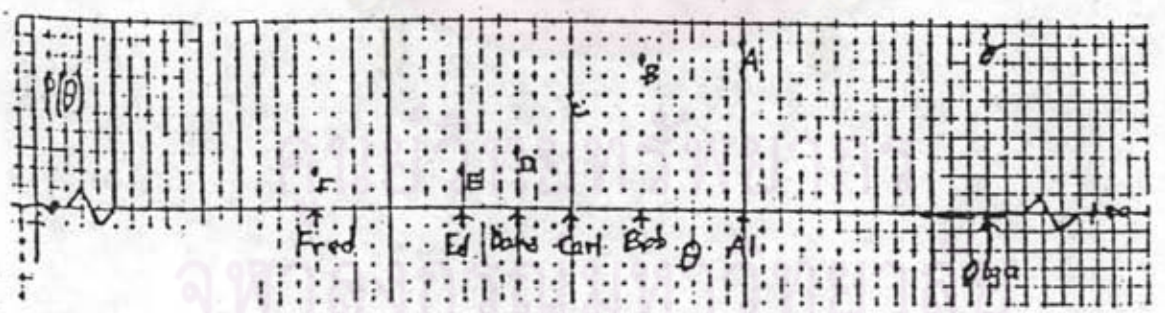
ภาพที่ 17ก ความน่าจะเป็นของ Al, Bob, และ Carl
ที่ทำข้อสอบข้อ 1 ถูก

และเมื่อรวม Dave เข้าไป โดยที่ Ed และ Fred มีความสามารถ (θ) ต่ำกว่า Dave แสดงดังภาพที่ 17ข



ภาพที่ 17ข ความน่าจะเป็นของ Al, Bob, Carl, Dave, Ed, และ Fred ที่ทำข้อ 1 ถูก

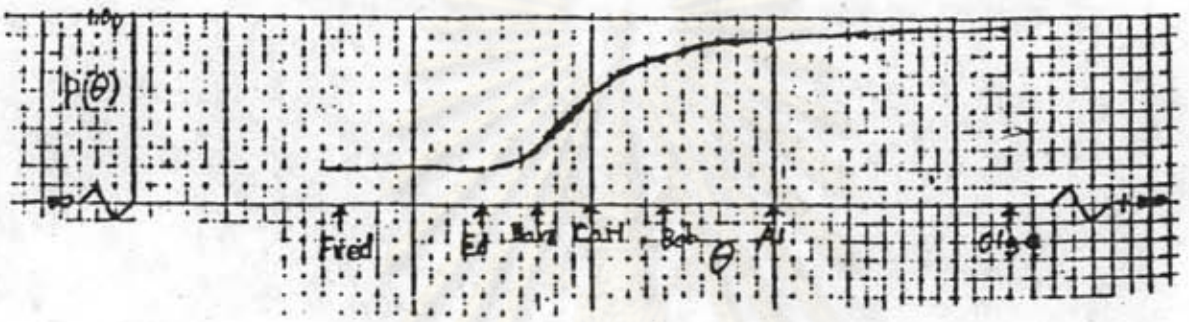
และเมื่อรวม Olga ซึ่งฉลาดที่สุดลงไป จะได้ภาพที่ 17ค



ภาพที่ 17ค ความน่าจะเป็นของ Al, Bob, Carl, Dave, Ed, Fred และ Olga ที่ทำข้อ 1 ถูก

เพราะว่าความน่าจะเป็นที่จะได้รับคำตอบถูกเป็นเพียงฟังก์ชั่นหนึ่งของปริมาณความสามารถ เราสามารถพูดได้ว่าใครก็ตามที่มี θ เดียวกันกับ Al จะมีโอกาสเดียวกันกับ Al ที่จะทำข้อสอบดังกล่าวถูกต้อง และทุกคนซึ่งมี θ เดียวกันกับ Ed จะมีโอกาสเท่ากับ Ed ที่จะทำข้อ

ข้อดังกล่าวถูก เพราะฉะนั้นเราจึงสามารถเชื่อมจุดในภาพที่ 17ค ได้ซึ่งสามารถบอกเราให้ทราบถึง P (θ) ของแต่ละ θ ได้ โค้งนี้เรียกว่าฟังก์ชั่นการตอบสนองต่อข้อสอบ และในปัจจุบันนี้เราเรียกว่าโค้งลักษณะข้อสอบ (Item Characteristic Curve, ICC) ดังภาพที่ 18



ภาพที่ 18 ฟังก์ชั่นการตอบสนองต่อข้อสอบของข้อสอบข้อ 1

เราสามารถได้ข้อมูลต่างๆ จาก ฟังก์ชันการตอบสนองต่อข้อสอบ หรือ โค้งคุณลักษณะของข้อสอบ ได้หลายประการดังนี้

1) โค้งจะมีค่าไม่มากกว่าไป 1.0 เพราะความน่าจะเป็น = 1.0 เป็นสิ่งที่แน่นอน และไม่มีอะไรที่น่าจะเป็นไปได้มากกว่าความแน่นอนดังกล่าวนั้น

2) โค้งจะสูงไม่ถึง 1.0 เพราะในการสอบไม่มีสิ่งใดแน่นอน ดังนั้นโค้งนี้จึงมีเส้นกำกับบนเท่ากับ 1.00

3) ระหว่าง Ed และ Bob โค้งจะสูงขึ้นอย่างรวดเร็ว เพราะจากจุด E ถึงจุด B เป็นระยะระหว่าง θ ของ Ed และ θ ของ Bob ที่สั้น

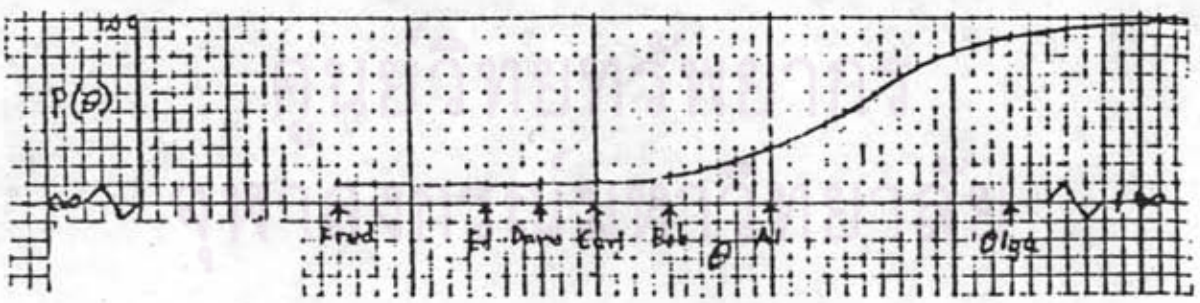
4) โค้งจะต้องสูงขึ้นเสมอ (ไม่เคยเป็นเส้นตามแนวนอนหรือลดต่ำลง) และจะเคลื่อนจากซ้ายไปขวา เพราะในขณะที่ความสามารถเพิ่มขึ้น โอกาสที่จะทำข้อสอบถูกจะเพิ่มขึ้น ดังนั้นโค้งจึงเพิ่มขึ้นอย่างสม่ำเสมอ

5) โค้งไม่มีค่าต่ำสุดเป็น 0.0 เพราะโอกาส 0.0 คือความเป็นไปไม่ได้อย่างสมบูรณ์ ซึ่งไม่มีอะไรที่จะสามารถเป็นไปได้น้อยกว่าความเป็นไปไม่ได้อย่างสมบูรณ์ ดังนั้นโค้ง

จึงไม่มีเส้นกำกับล่างที่ต่ำไปกว่านี้

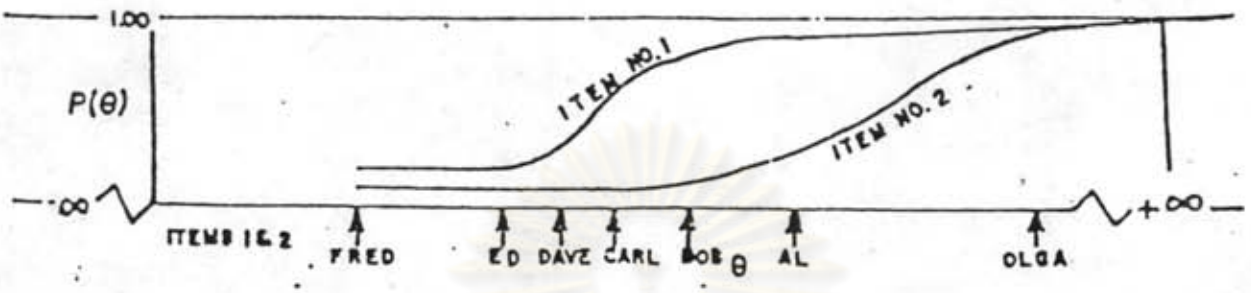
6) เพราะข้อสอบเป็นคำถามประเภทเลือกตอบจึงมีโอกาสที่จะตอบข้อสอบถูก โดยโอกาสไม่ว่าจะมี θ ต่ำอย่างไร ตามประเพณีเราให้โอกาสในการตอบถูกเป็น 1/A เมื่อ A คือ จำนวนตัวเลือกในคำตอบแบบเลือกตอบ ถ้าข้อสอบ 4 ตัว เลือกก็จะมีโอกาสตอบถูก 1/4 = .25 และถ้า 5 ตัวเลือก โอกาสในการตอบถูก 1/5 = .20 ความน่าจะเป็นในการตอบถูกโดยในโอกาสถูกกำหนดขึ้นจากคำถามประเภทเลือกตอบเมื่อใดก็ตาม จึงไม่สามารถจะคาดหมายให้โอกาสในการตอบถูกมีค่าศูนย์ ดังนั้นโค้งในภาพที่ 18 จึงถูกคาดหวังว่า เส้นกำกับล่างจะมากกว่า 0

เมื่อสังเกตข้อสังเกตนี้ทั้งหมดเกี่ยวกับฟังก์ชันการตอบสนองต่อข้อสอบจะพบว่าเป็นจริงในโอไจฟ์ปกติ และโอไจฟ์โลจิสแบบสามพารามิเตอร์ ดังนั้นเราจึงสรุปว่าโอไจฟ์ปกติหรือโอไจฟ์โลจิสอาจจะนำมาใช้บรรยายทฤษฎีการตอบสนองต่อข้อสอบได้ดีมากเราอาจจะใช้ฟังก์ชันโอไจฟ์โลจิส เพื่อบรรยายฟังก์ชันการตอบสนองต่อข้อสอบเชิงคณิตศาสตร์ได้ ถ้าเรารู้ว่าเราได้ลงจุดความน่าจะเป็นในการตอบข้อสอบข้อ 2 ถูก ของ Al Bob, Carl, Dave, Ed, Fred และ Olga เราจะได้ทฤษฎีการตอบสนองต่อข้อสอบหนึ่ง ดังภาพที่ 19



ภาพที่ 19 ฟังก์ชั่นการตอบสนองต่อข้อสอบของข้อ 2

และจากภาพที่ 20 เมื่อนำข้อมูลจากข้อสอบข้อที่ 1 และ 2 มารวมกัน จะพบว่า สำหรับ Olga, Ed และ Fred (และคนอื่นๆ ตามค่า θ ของเขา) ความน่าจะเป็นในการตอบข้อสอบข้อ 1 ถูก($P_1(\theta)$) แต่ข้อสอบข้อ 2 ยากกว่าข้อ 1 สำหรับ Al,

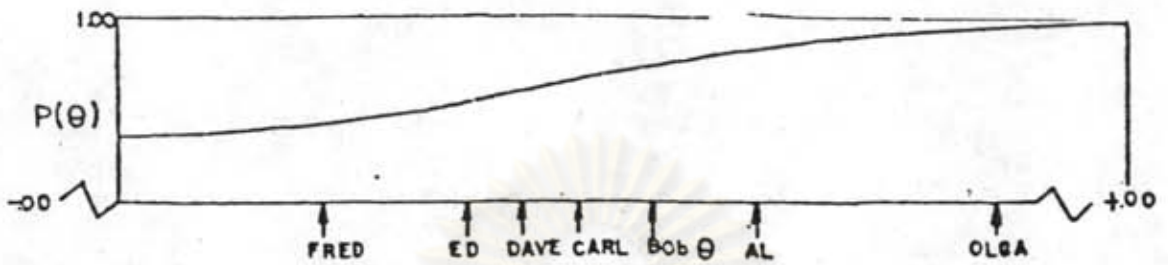


ภาพที่ 20 ฟังก์ชันการตอบสนองต่อข้อสอบของข้อ 1 และ 2

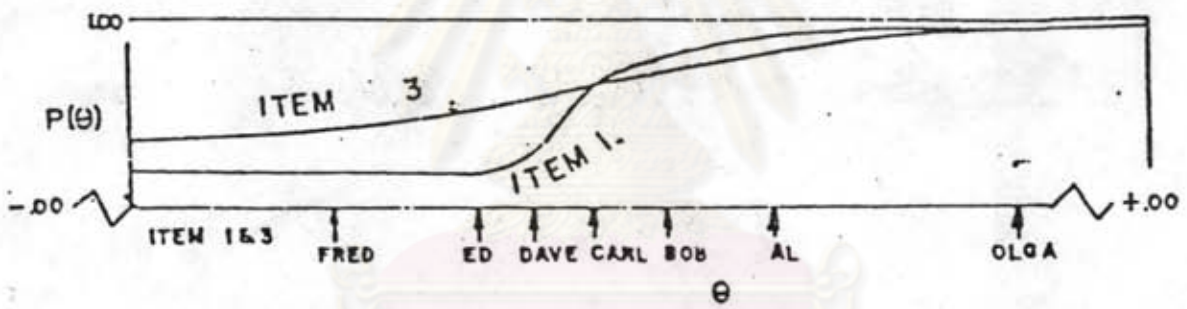
Bob, Carl และ Dave เพราะคนเหล่านี้ความน่าจะเป็นในการตอบข้อสอบข้อ 2 ถูก ($P_2(\theta)$) ต่ำกว่าความน่าจะเป็นในการตอบข้อสอบข้อ 1 ถูก และจะเป็นการยากกับทุกคน ซึ่งมีความสามารถเท่ากับ Al, Bob, Carl หรือ Dvae เราสังเกตได้ว่าโอกาสในการตอบข้อ 2 ถูกของ Bob, Carl, Dave, Ed และ Fred มีค่าประมาณค่าเดียวกัน ข้อสอบข้อ 2 จึงไม่สามารถแยกระหว่างกลุ่มที่ความสามารถเช่นเดียวกับ หรือความสามารถต่ำกว่าข้อสอบที่ยากจะไม่สามารถหาความแตกต่างระหว่างคนที่มีความสามารถต่ำได้ เพราะเขาจะทำข้อสอบนั้นผิดหมด (นอกจากว่าเขาโชคดีที่จะเดาได้ถูก) ข้อสอบที่ง่ายไม่สามารถแยกความแตกต่างคนที่มีความสามารถสูงได้ เพราะเขาจะทำถูกทั้งหมด แบบสอบที่ประกอบด้วยข้อสอบที่มีทฤษฎีการตอบสนองต่อข้อสอบ เช่น ข้อ 2 จะไม่เป็นแบบสอบที่ดีในการวัดความสามารถเชิงสัมพัทธ์ (relative ability) ของบุคคลเช่น Bob, Carl, Dave Ed และ Fred

เมื่อโอกาสในการตอบถูกข้อ 2 ($P_2(\theta)$) ของ Olga มากกว่า Al ดังนั้นข้อสอบข้อ 2 จะแยกความแตกต่างระหว่างคนที่มีความสามารถเหมือน AL และ Olga ได้ ถ้าความแตกต่างอยู่ในช่วงความสามารถตามวัตถุประสงค์ของเราแล้ว แบบสอบฉบับหนึ่งที่รวบรวมมาด้วยข้อสอบเช่นเดียวกับข้อสอบข้อ 2 จะเป็นแบบสอบที่ดี

ข้อ 3 มีฟังก์ชันการตอบสนองต่อข้อสอบ ดังภาพที่ 21 ข้อสอบนี้สูงขึ้นตลอดช่วงมากกว่า ข้อ 1 หรือ ข้อ 2 แต่ความชันน้อยกว่าทุกจุดในระหว่างค่าที่เพิ่มขึ้น ค่าความชันต่ำหมายความว่าข้อ 3 แยกความแตกต่างตลอดช่วงของ θ ได้แต่ทำได้ไม่กี่คนแต่ละค่า θ เฉพาะใดๆ



ภาพที่ 21 ฟังก์ชั่นการตอบสนองต่อข้อสอบของข้อ 3



ภาพที่ 22 ฟังก์ชันการตอบสนองต่อข้อสอบ ของข้อ 1 และ 3

จากภาพที่ 22 แสดงลักษณะของข้อสอบข้อ 1 และ 3 สังเกตพบว่าข้อ 3 ยากกว่าข้อ 1 สำหรับ Bob และ Al แต่ง่ายกว่าสำหรับ Dave Ed และ Fred ความน่าจะเป็นของความยากรายข้อสัมพัทธ์ที่เห็นกลับกันนี้ สำหรับบุคคลที่มีความสามารถแตกต่างกันเป็นผลลัพธ์ของ ฟังก์ชันการตอบสนองต่อข้อสอบอย่างหนึ่งที่น่าประหลาดใจ

เราจะพบว่ายิ่งความชันของฟังก์ชันการตอบสนองต่อข้อสอบมากขึ้น อำนาจจำแนกของข้อนั้นจะจำแนกได้ดีขึ้น แต่ถ้าความชันมีค่าน้อยลงค่าอำนาจจำแนกของข้อสอบนั้นจะน้อยลงด้วย ซึ่งเราได้กล่าวมาแล้วว่า พารามิเตอร์ a ของโอไจฟ์โลจิสได้บรรยายถึงความชันของโค้งของมันเอง ดังนั้นค่า a จึงถูกเรียกว่าดัชนีอำนาจจำแนกของฟังก์ชันการตอบสนองต่อข้อสอบ ค่า a ของ ฟังก์ชันการตอบสนองต่อข้อสอบ ที่มากจะทำให้อำนาจจำแนกของ

ข้อสอบมากขึ้น

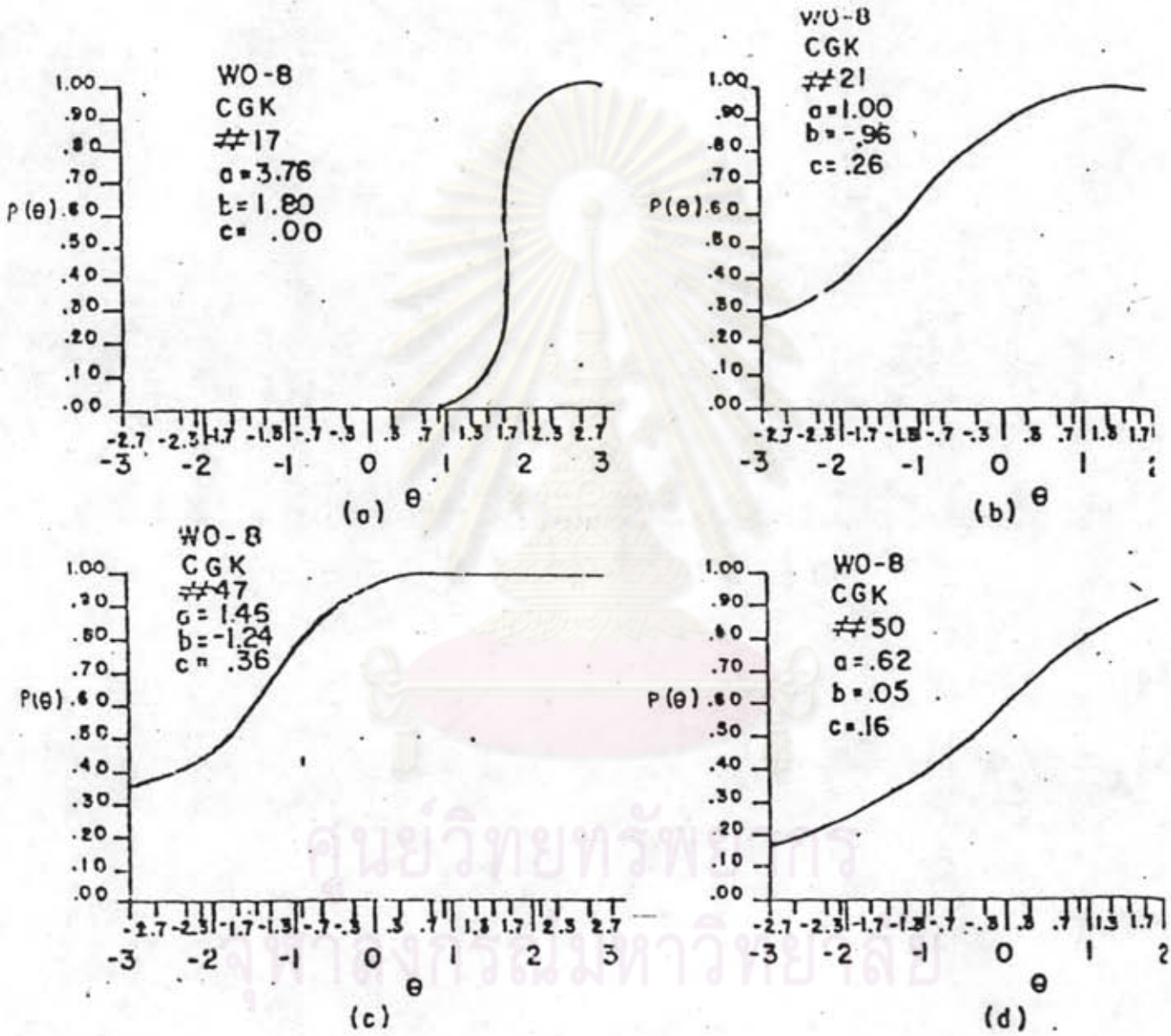
ปรากฏเป็นความจริงว่าการเลื่อนฟังก์ชันการตอบสนองต่อข้อสอบทั้งหมดไปทางซ้ายจะทำให้ข้อสอบง่ายขึ้น และก็เลื่อนไปทางขวาจะทำให้ข้อสอบยากขึ้น การเลื่อนโอไจฟ์โลจิสของฟังก์ชันการตอบสนองต่อข้อสอบ ข้อสอบที่ยากมากขึ้น ค่า b ของฟังก์ชันการตอบสนองต่อข้อสอบจะมีค่าใหญ่ขึ้น

ฟังก์ชันการตอบสนองต่อข้อสอบของข้อ 1, 2 และ 3 มีเส้นกำกับล่างที่แตกต่างกัน เพราะ ฟังก์ชันการตอบสนองต่อข้อสอบ ไม่เคยต่ำไปกว่าเส้นกำกับล่าง ความแตกต่างนี้ในฟังก์ชันการตอบสนองต่อข้อสอบ หมายความถึงว่า ทั้งกระทงดังกล่าวมีความยากแตกต่างกัน แม้คนเข้าสอบจะมีความสามารถต่ำมาก แต่ผู้เข้าสอบที่มีความสามารถต่ำที่แทบจะไม่มีความรู้ที่จะตอบข้อนั้น ดังนั้นเขาจึงต้องเดาความแตกต่างในเส้นกำกับล่างของ ฟังก์ชันการตอบสนองต่อข้อสอบ หมายความว่าคนที่มีความสามารถต่ำมาก จะมีโอกาสดีกว่าในการเดาตัวเลือกที่ถูกจากตัวเลือกที่เหลืออื่นๆ เหตุผลนี้จะอธิบายต่อไปเส้นกำกับล่างของฟังก์ชันการตอบสนองต่อข้อสอบถูกเรียกว่า ดัชนีการเดา (guessing index) หรือบางทีอาจจะเพิ่มเติมมากขึ้นว่าเป็นดัชนีการเดาเทียม (pseudo - guessing index) ของข้อสอบข้อนั้น

ภาพที่ 23 แสดง ฟังก์ชันการตอบสนองต่อข้อสอบของข้อสอบจริงๆ 4 ข้อ จาก The Coast Guard Khowledge section ของ The U.S. Coast Guard Warrant Officer test ข้อ 17 ยากมาก แต่อำนาจจำแนกสูง ข้อนี้มีค่า c = .00 หมายความว่าผู้สอบเกือบทุกคนมีค่า θ ต่ำกว่า 1 ตอบข้อนี้ไม่ถูก ข้อ 17 นี้ไม่ค่อยพบนักใน 2 กรณี ที่ค่า a สูงสุดโด่ง ค่า c เป็น 0.00 อย่างไรก็ตามก็เป็นข้อสอบอุดมคติที่ใช้ได้ในหลายความมุ่งหมาย

ข้อ 21 เป็นข้อง่ายแต่มีค่าอำนาจจำแนกต่ำ ข้อ 47 ค่อนข้างง่ายมากกว่าข้อ 21 แต่มีอำนาจจำแนกดี ข้อ 50 เป็นข้อที่มีความยากปานกลาง และค่าอำนาจจำแนกต่ำ

ไม่ควรเอาฟังก์ชันการตอบสนองต่อข้อสอบไปปนกับโค้งข้อสอบ-แบบสอบ (item-test curve) ซึ่งเป็นคะแนนดิบในแกนนอน แทนด้วย θ โค้งข้อสอบ-แบบสอบจะยุ่งยากจากปัญหาเดียวกันในมาตราที่บิดเบือนไปด้วยจากคะแนนดิบ โค้งข้อสอบ-แบบสอบมีรูปร่างเฉพาะและไม่ได้ไม่เป็นอิสระจากข้อสอบอื่นๆ ในแบบสอบนั้น ความจริงค่าเฉลี่ยของ



(Warm 1978: 48)

ภาพที่ 23 ฟังก์ชันการตอบสนองต่อข้อสอบของข้อสอบจริง 4 ข้อจาก

the Coast Guard Knowledge section of the

U.S. Coast Guard Warrant Officer Test, series 8

โค้งข้อสอบ-แบบสอบและทุกข้อสอบในแบบสอบฉบับหนึ่งเป็นเส้นตรงมีความชันเท่ากับ 1 (ทำมุม 45 องศา) ดังนั้นในหลายๆ ความมุ่งหมายโค้งข้อสอบ-แบบสอบ เป็นประโยชน์น้อย เมื่อนำมาใช้เป็นเครื่องมือในการวิเคราะห์

6. พารามิเตอร์ a, b, และ c

เมื่อกำหนดให้ θ แทนคุณลักษณะ (trait) เช่น ความสามารถ (ability) ทักษะ (skill) ที่กำลังจะถูกวัด สำหรับข้อสอบที่เป็นแบบถูก-ผิด (dichotomous) คือให้คะแนนแบบถูก-ผิดคำตอบถูกให้ 1 คะแนน คำตอบผิดให้ 0 คะแนน ฟังก์ชันการตอบสนองต่อข้อสอบสามารถแสดงได้ง่ายๆ ด้วยความน่าจะเป็น P หรือ P (θ) ของคำตอบถูกต่อข้อสอบข้อนั้น ซึ่งน่าจะเป็นว่า โอกาสในการตอบถูกจะเพิ่มมากขึ้นเมื่อความสามารถของผู้ตอบ (θ) มากขึ้น จากข้อตกลงเช่นนี้ความน่าจะเป็น P (θ) ดังกล่าวจึงสามารถแสดงได้โดยฟังก์ชันโลจิสต์แบบสามพารามิเตอร์ ดังนี้

$$P(\theta) = c + \frac{1 - c}{1 + e^{-1.7a(\theta - b)}}$$

ซึ่ง a, b, c คือ พารามิเตอร์ที่แสดงคุณลักษณะของข้อสอบ

e คือ ค่าคงที่ทางคณิตศาสตร์ มีค่า = 2.71828

โมเดลดังกล่าวนี้ได้ถูกสร้างและพัฒนาโดยเบิร์นบวม พารามิเตอร์ c เป็นความน่าจะเป็นที่คนที่ขาดความสามารถอย่างสมบูรณ์ ($\theta = -\infty$)จะตอบข้อสอบข้อนั้นถูก จึงเรียก c ว่าพารามิเตอร์การเดา หรือระดับคะแนนโอกาสเทียม (pseudo-chance score level) ถ้าข้อไม่สามารถตอ ถูกเลยโดยการเดา ค่า c = 0 พารามิเตอร์ b เป็นพารามิเตอร์ที่แสดงตำแหน่ง จะเป็นตัวกำหนดตำแหน่งของโค้งตามสเกลความสามารถ จึงถูกเรียกว่าค่าความยากของข้อ (item difficulty) ข้อยิ่งยากโค้งจะไปทางขวา เมื่อไม่มีการเดา b จะเป็นระดับความสามารถที่ความน่าจะเป็นที่จะตอบข้อนั้นถูก = .5 เมื่อมีการเดา b คือระดับความสามารถ

ซึ่งความน่าจะเป็นในการตอบถูกเป็นครึ่งหนึ่งของระยะระหว่างค่า c กับ 1.0 พารามิเตอร์ a เป็นสัดส่วนกับความชัน (Slope) ของโค้ง ณ จุดเปลี่ยนโค้ง (inflexion point) ความชันนี้ค่าจริงคือ .425a(1 - c) ดังนั้น a จึงแสดงอำนาจจำแนก (discriminating power) ของข้อสอบ ซึ่งระดับของการตอบสนองรายข้อจะแตกต่างกันไปตามระดับความสามารถ

ฟังก์ชั่นที่แสดงการตอบสนองข้อสอบที่ใช้กันเสมออีกฟังก์ชั่นหนึ่ง คือ ฟังก์ชั่นโอไจฟ์ปกติแบบสามพารามิเตอร์ ดังนี้

$$P \equiv P(\theta) = c + (1-c) \int_{-\infty}^{a(\theta-b)} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt$$

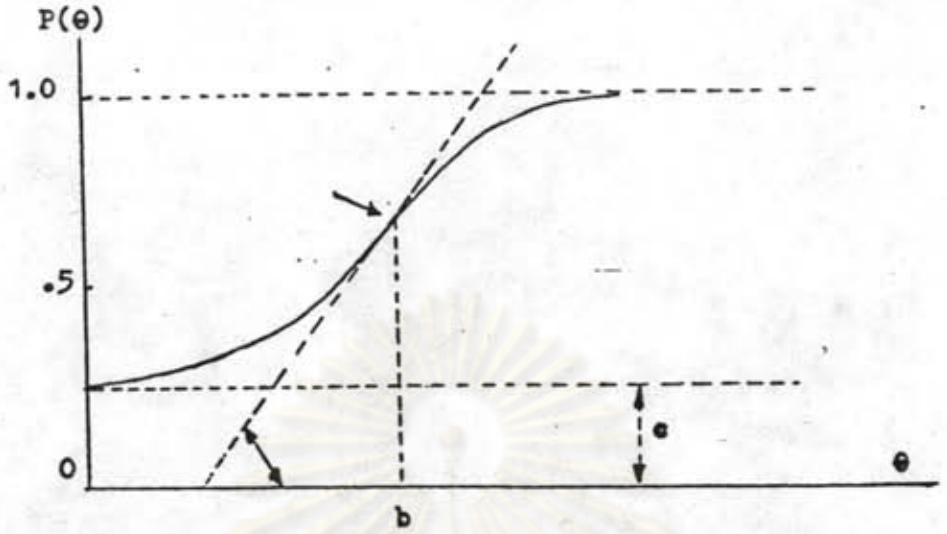
c คือ ความสูงของเส้นกำกับล่าง

b คือ ระดับความสามารถที่จุดเปลี่ยนโค้ง ซึ่งความน่าจะเป็นของการตอบถูก คือ a + c/2

a คือ สัดส่วนกับความชันของโค้งที่จุดเปลี่ยนโค้ง (ความชันนี้จริงๆ มีค่า = a (1 - c) / $\sqrt{2\pi}$

จากที่บรรยายมาแล้วความแตกต่างของฟังก์ชั่นในสมการแรกและสมการหลังมีค่าน้อยกว่า .01 ทุกๆ ชุดของค่าพารามิเตอร์ โดยหลักการทั่วไปผู้สอบที่มีความสามารถสูงไม่ควรจะตอบข้อที่ง่ายผิด แต่ในทางปฏิบัติแล้วอาจจะมีผู้เข้าสอบบางคนเกิดเลินเล่อด้วยเหตุนี้ สำหรับฟังก์ชั่นโลจิสต์ปลายล่างและบนของโค้งจะเข้าใกล้เส้นกำกับของโค้งได้เร็วน้อยกว่าโอไจฟ์ปกติ ในกรณีที่การทำผิดเกิดจากความเลินเล่อดังนั้นฟังก์ชั่นโลจิสต์จึงถูกละเมิดน้อยกว่าโอไจฟ์ปกติ จึงเป็นน่าจะเป็นเหตุผลที่ดีที่งามในทางปฏิบัติจะใช้ฟังก์ชั่นโลจิสติกมากกว่า

สามารถแสดงความหมายของ พารามิเตอร์ ทั้ง 3 ได้ดังภาพที่ 24



ภาพที่ 24 แสดงความหมายของพารามิเตอร์ของข้อสอบ

1) ค่า a เป็นดัชนีอำนาจจำแนกของข้อสอบ ถ้า θ กระจายแบบปกติในโมเดลโอไจฟ์ปกติ ค่า a จะสัมพันธ์กับค่า d ซึ่งสามารถแสดงได้ดังนี้

$$a \approx \frac{d\sqrt{pq}}{\sqrt{(KR\text{-}20)(1-c)^2y^2 - d^2pq}}$$

กำหนด d = ค่า d, คือสหสัมพันธ์ของข้อและแบบสอบทั้งทุกชนิด พอยท์ไบซีเรียล

p = ค่า p, คือสัดส่วนของผู้สอบที่ตอบข้อสอบข้อนั้นถูก

q = 1 - p

KR-20 = ค่าความเที่ยงแบบ Kuder-Richardson สูตร 20

y = ความสูงของ N (0,1) ที่ Z - score ที่จุดตัด p เป็นสัดส่วนของพื้นที่ใต้ N (0,1)

c = ค่า c

$p' = \frac{p - c}{1 - c}$

ค่า a มีความสัมพันธ์กับความชันของฟังก์ชั่นการตอบสนองต่อข้อสอบ และสามารถมีช่วงจาก 0.0 ถึง $+\alpha$ เท่าที่ความชันจะสามมารถมีค่าได้ ความชันค่าติดลบเป็นสิ่งที่เป็นไปได้ แต่ไม่อยู่ในความสนใจในกรณีนี้ ค่า a จากประสบการณ์แสดงว่ามีค่าเฉพาะจาก .5 ถึง 2.5 ช่วงมากสุดจาก 1.0 ถึง 2.0 ข้อที่มีค่าอำนาจจำแนกต่ำสุดตลอดช่วง O เป็นข้อไม่ดี ในขณะนี้ข้อสอบที่มีค่าอำนาจจำแนกรายข้อค่า a สูงๆ ตลอดช่วง θ เล็กๆ จะเป็นข้อที่ดี ข้อกระทงที่มีค่า a ต่ำกว่า .80 เป็นข้อที่ไม่ดีมากๆ และไม่เป็นไปตามความมุ่งหมายมากที่สุด

2) ค่า b เป็นดัชนีความยาก ถ้า θ มีการแจกแจงปกติ ค่า b จะสัมพันธ์กับค่า P ในโมเดลโอไจฟ์ปกติซึ่งแสดงได้ดังนี้

$$b \approx \frac{yz(1-c)\sqrt{KR-20}}{d\sqrt{pq}}$$

กำหนดให้ Z = คะแนน Z ที่จุดตัด P ซึ่งเป็นสัดส่วนในส่วนบนของพื้นที่ใต้โค้ง N (0,1) และสัญญลักษณ์อื่นๆ ได้นิยามไว้แล้วในค่า a ค่า b จะมีช่วงจาก -2.5 ค่า b = 2.5 แสดงว่าข้อนั้นง่ายมาก ส่วนข้อที่มีค่า b = + 2.5 จะยากมาก ข้อที่มีค่า b = 0.0 เป็นข้อที่มีความยากปานกลาง

3) ค่า c เป็นพารามิเตอร์การเดาหรือพารามิเตอร์การเดาเทียม ซึ่งบ่งชี้ว่าความน่าจะเป็นของผู้เข้าสอบที่จะมีความสามารถต่ำมากจะทำข้อสอบนั้นถูก ค่า c ส่วนใหญ่มีช่วงจาก .00 ถึง .40 ข้อกระทง ที่มีค่า c = .30 หรือมากกว่าเป็นข้อที่ไม่ดี ข้อกระทงที่ต้องการจะมีค่า c ประมาณ .20 หรือต่ำกว่า ค่า c ยิ่งต่ำแสดงว่าข้อนั้นยิ่งดีค่า c เป็นศูนย์เป็นเรื่องอุดมคติทีเดียว ค่า c จะมีค่าประมาณ 1/A - .05 เมื่อ A คือจำนวนตัวเลือกในข้อสอบ ดังนั้นข้อสอบ 4 ตัวเลือก จึงมีค่า c ~ .20 (คือ .25 - .05) และข้อสอบ 5 ตัวเลือกจะมีค่า c ~ .15 (คือ .20 - .05)

ข้อที่ไม่มีค่า c เป็น 1/A เพราะผู้สอบไม่ได้ทำดังนั้น จริงๆ การเดาอย่างสุ่มจะเกิดขึ้นเมื่อผู้สอบไม่มีความรู้เลยในการสอบ (ซึ่งเป็นเรื่องกำหนดขึ้นเองในการวิเคราะห์

แนวทฤษฎีการวัดแบบคลาสสิค)

มีคำอธิบายที่แสดงถึงว่าจริงๆ ไม่เกิดการเดาแบบสุ่ม คือ c =/ 1/A ซึ่งลอร์ด ได้แนะนำว่า ผู้เขียนข้อสอบที่ฉลาดมากในการเขียนตัวลวงที่ดึงดูดคนที่เข้าสอบและมีความสามารถต่ำๆ เมื่อผู้เข้าสอบเหล่านี้ไม่รู้คำตอบเขาจะถูกดึงดูดจากตัวลวงมากกว่า คำตอบถูก และทำให้เขาทำผิดมากกว่าถ้าเขาจะเดาแบบสุ่ม

ในโมเดลของราส์ชได้ถือเสียว่าข้อสอบทุกข้อในแบบสอบมีค่า a เท่ากัน และค่า c = 0 ในทุกข้อ

7. โค้งคุณลักษณะของแบบสอบ (The Test Characteristices Curve)

มาตราของความสามารถ (θ) มีลักษณะต่อเนื่องกันแต่เนื่องจากการคำนวณด้วยเครื่องคอมพิวเตอร์แบบระบบตัวเลขเป็นส่วนมาก ค่าความสามารถ (θ) จึงถูกแยกออกเป็นค่าเล็กๆ ช่วงที่แยกออกแต่ละค่าใช้ 0.05 หน่วยของ θ และค่า P (θ) จะคำนวณมากจากค่า θ ที่เพิ่มทีละ 0.05 จาก θ = -5.0 ถึง θ = + 5.0 ช่วงของ θ อาจทำให้กว้างขึ้นและการแบ่งช่วงอาจเล็กกว่า 0.05 ได้ถ้าความมุ่งหมายต้องการความแม่นยำในการประมาณมากขึ้น

จากตาราง 3 แสดงค่า P(θ) จำนวน 17 ค่าในระดับความสามารถ θ ต่างๆ ที่ได้ข้อมูลจากข้อสอบ 4 ข้อ ในภาพที่ 23

ข้อสอบข้อ 1 ให้คะแนนถูกได้ 1 คะแนน ผิดได้ 0 คะแนน คะแนนแบบ ถูก-ผิด เป็นผลมาจากการใช้รูปแบบข้อสอบแบบเลือกตอบ คะแนนแบบ ถูก-ผิด ของผู้เข้าสอบแต่ละคน (0 หรือ 1) ไม่ใช่การวัดความรู้ของเขาที่แม่นยำมาก P(θ) อาจจะถูกอธิบายได้ 2 วิธี P(θ) = .78 หมายความว่า

1) 78 % ของผู้เข้าสอบที่มี θ ตามกำหนดจะทำข้อสอบข้อนั้นถูก

2) ผู้เข้าสอบคนหนึ่งจะทำถูกได้ 78 % ของข้อสอบซึ่ง P(θ) ของเขา =.78 เขาจะได้รับการคาดหวังให้ได้ข้อถูก 78 ข้อ ข้อผิด 22 ข้อ สำหรับ % คะแนนของ 78 % ถ้ามีวิธีบางวิธีที่จะให้เครดิตบางส่วนกับเขา .78 จุด ในแต่ละข้อของข้อสอบ 100 ข้อ แทนคะแนน 0 และ 1 เขาจะได้รับ % คะแนน เป็น 78 % ด้วย นิยามของเครดิตบางส่วนนี้

ตาราง 2 แสดงโอกาสที่จะตอบข้อสอบถูกที่แต่ละระดับความสามารถของผู้สอบ

	P(θ)				
θ	#17	#21	#47	#50	p(0)
-2.7	.00	.30	.38	.20	.88
-2.3	.00	.33	.40	.23	.96
-2.0	.00	.37	.45	.25	1.07
-1.7	.00	.43	.52	.28	1.23
-1.3	.00	.53	.66	.33	1.52
-1.0	.00	.71	.87	.44	2.02
- .7	.00	.62	.77	.48	1.77
- .3	.00	.82	.94	.52	2.28
0	.00	.88	.97	.59	2.44
.3	.00	.92	.99	.65	2.56
.7	.00	.96	.99	.74	2.69
1.0	.01	.97	.99	.79	2.75
1.3	.04	.98	.99	.84	2.85
1.7	.35	.99	.99	.89	3.22
2.0	.78	.99	.99	.91	3.67
2.3	.96	.99	.99	.94	3.88
2.7	.99	.99	.99	.96	3.93

(Warm 1978: 58)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สำหรับข้อสอบข้อหนึ่งขึ้นกับ P(θ) ของเขาได้นำไปสู่แนวความคิดของคะแนนจริง (true score) จากข้อสอบข้อนั้น

บ่อยครั้งที่ไม่พบความจริงว่า ผู้เข้าสอบจะตอบคำถามได้ 100 % หรือ 0 % ข้อสอบแต่ละข้อแบบเลือกตอบ ขาจะได้รับเครดิตเต็มที่ในข้อนั้นคือ 100 % (หรือ 1 เมื่อทำถูก และ 0 เมื่อตอบผิด) ถ้าการวัดความรู้ของเขามีความแม่นยำก็สามารถจะกำหนดระดับที่แน่นอนของความสามารถของเขาได้ P(θ) อาจได้รับการแปลความหมายเช่นเดียวกับการวัดความรู้ของเขา และถูกเรียกอีกอย่างว่าคะแนนจริงของเขาในข้อสอบนั้น ผลจากกการรวมคะแนนจริงรายข้อคือ คะแนนจริงจากแบบสอบของเขา ในกรณีที่การวัดจากแบบสอบไม่มีความคลาดเคลื่อนคะแนนจริงจากแบบสอบก็คือ คะแนนดิบ

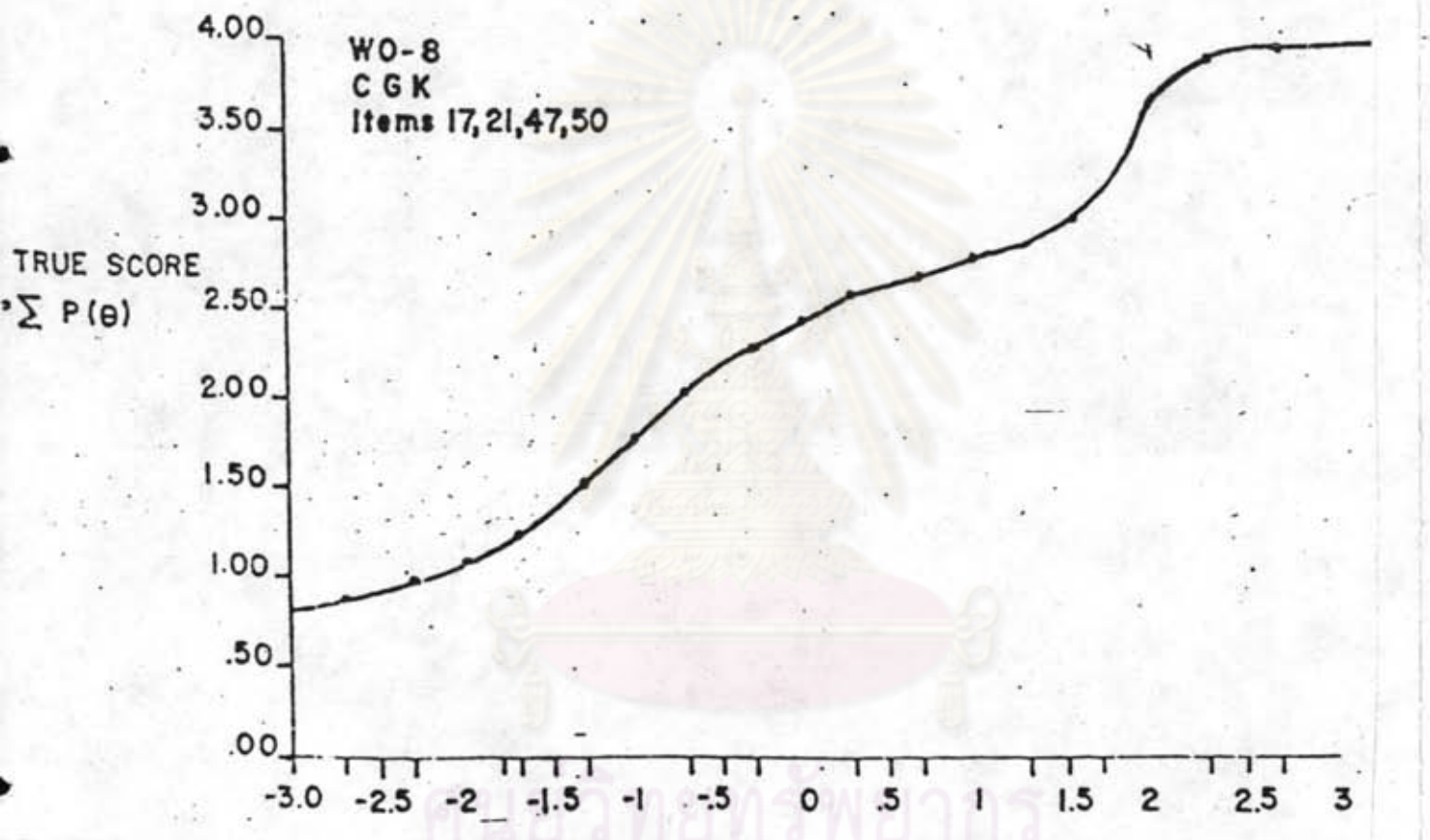
คอลัมน์ทางขวาของตาราง 1 เป็นผลรวมของ p (θ) ของข้อสอบ 4 ข้อในแต่ละข้อตามแต่ละค่า θ ที่กำหนดบนมาตรา p(θ) คือคะแนนจริงของผู้เข้าสอบคนหนึ่ง ณ ความสามารถ θ ที่กำหนดจากแบบสอบที่ประกอบด้วยข้อคำถาม 4 ข้อ

ถ้าเรานำคะแนนจริง p(θ) ในแต่ละค่าความสามารถ (θ) มาเขียนกราฟจะได้กราฟดังภาพที่ 25 ซึ่งแสดงโค้งคุณลักษณะของแบบสอบ สังเกตว่าโค้งคุณลักษณะของแบบสอบไม่เป็นเส้นตรงและไม่เป็นโอไจฟ์แบบสอบแต่ละฉบับจะมี โค้งคุณลักษณะของแบบสอบเองซึ่งได้จากผลรวมของฟังก์ชั่นการตอบสนองต่อแบบสอบของแต่ละข้อสอบจากแบบสอบนั้นนั่นเอง

สิ่งที่น่าสนใจข้อหนึ่งในการใช้โค้งคุณลักษณะของแบบสอบก็คือการกำหนดแจกแจงของคะแนนจริงจากแบบสอบ ในภาพที่ 25 แสดงได้ชัดเจนขึ้น ถ้าความสามารถของผู้เข้าสอบ (θ) มีการแจกแจงปกติ เช่น การแจกแจง θ ด้านล่าง คะแนนจริงของผู้เข้าสอบจะถูกแสดงทางด้านซ้าย การแจกแจงคะแนนจริงถูกพบโดยการฉายภาพ (projecting) ช่วงจากมาตรา θ บน โค้งคุณลักษณะของแบบสอบ นั้น และการแสดงด้วยพื้นที่เดียวกันบนมาตราคะแนนจริง

ภายในอันตรภาคที่ถูกฉายภาพเป็นความสำคัญที่จะให้ข้อสังเกตเกี่ยวกับคะแนนจริง (T) ว่าไม่ใช้คะแนนดิบ (X) คะแนนดิบได้รับการนิยามว่าเป็นคะแนนจริงบวกด้วยความคลาดเคลื่อน (E) (X = T + E) อย่างไรก็ตาม ลอร์ด (Lord, 1969) ได้พบว่า

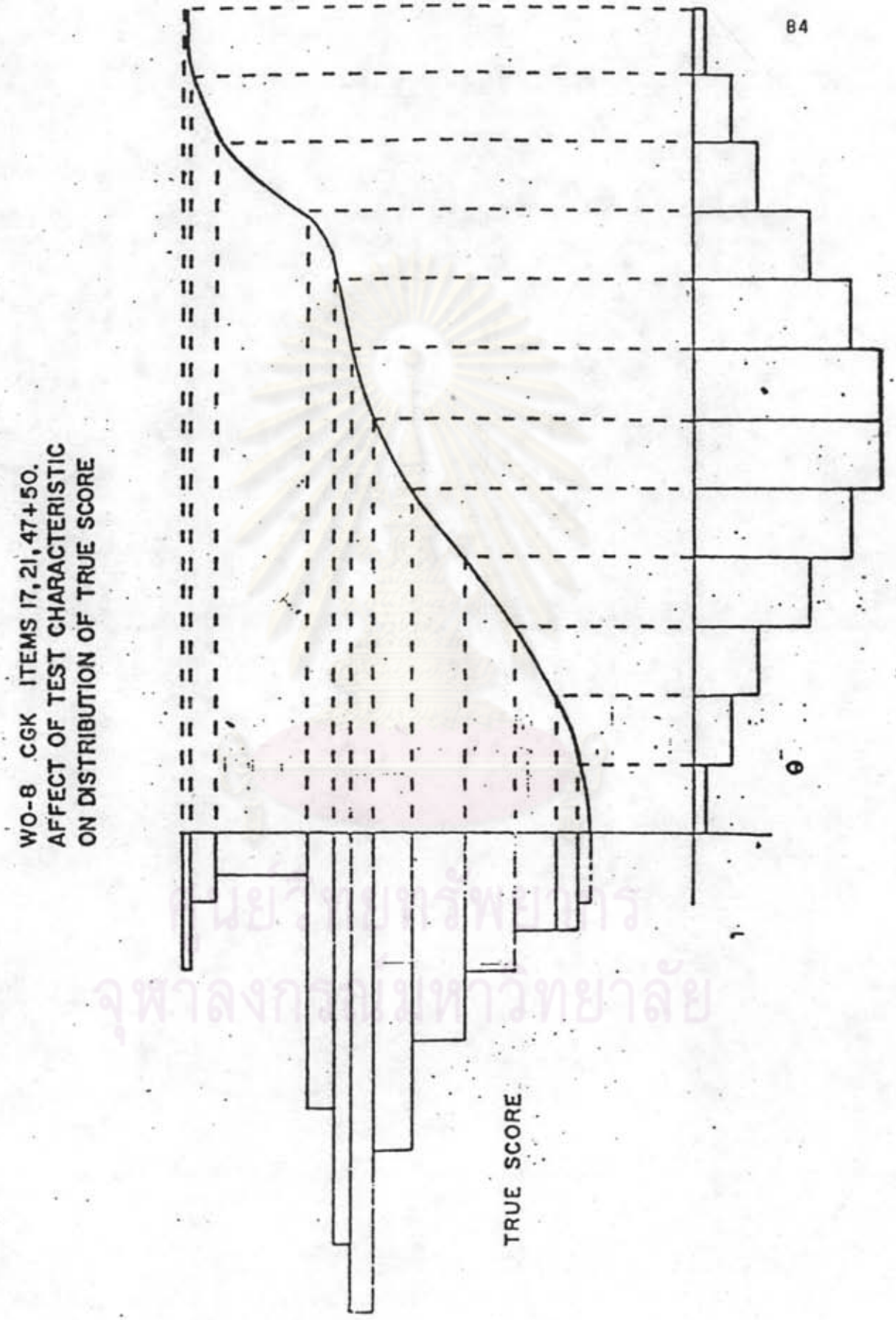
การแจกแจง X จะคล้ายกันกับการแจกแจงของ T แต่ในบางครั้งค่าคะแนนจริงค่าสูงจะมีการแจกแจงแบนต่ำกว่าและค่าคะแนนจริงต่ำบางส่วนจะสูงกว่า การแบนลงเป็นผลเนื่องมาจากความคลาดเคลื่อน



ภาพที่ 25 โค้งคุณลักษณะของแบบสอบที่มีข้อสอบ 4 ข้อ

8. ฟังก์ชั่นสารสนเทศของข้อสอบ (The Item Information Function, IIF)

จากภาพที่ 23a ข้อ 17 จะไม่ช่วยให้เราแยกความแตกต่างระหว่างกลุ่มผู้เข้าสอบผู้ซึ่งมีความสามารถ (θ) น้อยกว่า 1.0 ได้เพราะเขาจะได้รับข้อที่เขาทำผิดหมด ปรากฏว่ามีบางสิ่งบางอย่างเกี่ยวกับข้อ 17 ที่ทำให้ผู้เข้าสอบทั้งหมด ที่มีความสามารถ θ < 1.0 เลือกตัวเลือกที่ผิดสิ่งที่ไม่ใช่สถานการณ์ที่คุ้นเคย แต่ปรากฏจริงๆ ในคำถามนี้ แบบสอบที่แสดงเช่น



ภาพที่ 26 แสดงการใช้โค้งคุณลักษณะของแบบสอบ เพื่อดูความสัมพันธ์

ระหว่าง การแจกแจงของ θ และคะแนนจริง

ข้อสอบที่ 17 จะไม่ให้อะไรในการแยกความแตกต่างระหว่างผู้เข้าสอบที่มีความสามารถ $\theta<1.0$ เพราะเขาจะได้คะแนนศูนย์ทุกคนจากแบบสอบชุดนี้ และจะไม่ให้ข้อมูลในการแยกความแตกต่างระหว่างผู้สอบนั้น

ข้อ 17 นี้จะไม่ให้ข้อมูลในการแยกความแตกต่างผู้เข้าสอบที่มีความสามารถ (θ) เท่ากับ 2.7 หรือมากกว่าด้วยเพราะผู้เข้าสอบทุกคนจะทำข้อสอบข้อนี้ได้ถูกหมด แบบสอบที่ประกอบขึ้นด้วย ข้อเช่นข้อ 17 ผู้สอบทุกคนที่มีความสามารถ $(\theta) > 2.7$ จะได้คะแนนเต็ม 100

ระหว่าง $\theta = 1.0$ และ $\theta = 2.7$ จะให้ข้อมูลที่แตกต่างกันจาก $\theta = 1.0$ ถึง $\theta = 1.5$ ค่า $p(\theta)$ จะมีค่าจาก $p(\theta = 1.0) = 0.00$ ถึง $p(\theta = 1.5) = .08$ การเปลี่ยนแปลงของ $p(\theta)$ หมายความว่า ข้อสอบจะช่วยแยกระหว่างกลุ่มผู้สอบภายในช่วงความสามารถ (θ) ซึ่งการเปลี่ยนแปลงค่าของ $P(\theta)$ ได้ปรากฏขึ้น ในกรณีเช่นนี้ความแตกต่างระหว่าง $p(\theta) = .08$ มีค่าน้อย ความแตกต่างระหว่าง $p(\theta)$ ใช้สัญญลักษณ์ว่า dp การเปลี่ยนแปลง (dp) จะปรากฏตลอดช่วง $(d\theta)$ ด้วยครึ่งหนึ่งของหน่วยความสามารถ $(1.5 - 1.0)$ อัตราส่วนของ dp กับ $d\theta$ $(dp/d\theta)$ จะเท่ากับค่าเฉลี่ยของความชันของฟังก์ชั่นการตอบสนองต่อข้อสอบ ตลอดช่วงของ $d\theta$ สำหรับช่วงจาก $\theta = 1.0$ ถึง $\theta = 1.5$

$dp/d\theta = .08/.5 = .16$

จาก $\theta = 1.5$ ถึง $\theta = 2.0$ ในข้อ 17 $p(\theta)$ จะเปลี่ยนจาก .08 ถึง .78 ซึ่งเป็นการเปลี่ยนแปลงที่ใหญ่มาก $dp = .70$ $(.78 - .08)$ ในช่วงนี้ และ $dp/d\theta = .70/.50 = 1.40$ ซึ่งเป็นค่าใหญ่มาก ข้อ 17 เป็นข้อที่ดีในการแยกความแตกต่างระหว่างผู้สอบที่มีความสามารถในช่วง $\theta = 1.5$ ถึง $\theta = 2.0$ แบบสอบที่ประกอบด้วยข้อสอบเช่นข้อ 17 จะให้คะแนนประมาณ 8 % ถึง 78 % สำหรับผู้เข้าสอบซึ่งมีความสามารถจาก 1.5 ถึง 2.0 แบบสอบนี้จะให้ข้อมูลที่แยกความแตกต่างมากของผู้เข้าสอบในช่วงของ θ นี้ โดยขยายช่วงของความสามารถของคะแนนแบบสอบออกไป

เราได้เห็นแล้วว่าความชันของ IRT ที่มากขึ้น ข้อมูลของข้อสอบที่จะให้เราเกี่ยวกับผู้สอบในช่วงที่กำลังพิจารณาก็จะมากขึ้นด้วย

ถ้าเราสามารถทำช่วงความสามารถ (θ) ได้ตลอดช่วง ซึ่งเราพบว่าค่าความชัน

จะยิ่งน้อยลงในที่สุด เราจะได้รับความชันของ ฟังก์ชั่นการตอบสนองต่อข้อสอบ ที่จุดหนึ่งซึ่งจะเป็นความชันของ tangent line กับ ฟังก์ชั่นการตอบสนองต่อข้อสอบ ที่จุดเฉพาะจุดหนึ่งของความสามารถ θ ความชันของ ฟังก์ชั่นการตอบสนองต่อข้อสอบ จะวัดด้วยปริมาณข้อมูลเชิงสัมพัทธ์ของข้อสอบที่ได้รับเกี่ยวกับผู้เข้าสอบ ณ จุดนั้น ความชันที่ใหญ่ขึ้นจะให้ข้อมูลที่มากขึ้น ซึ่งเราสามารถหาค่าความชันของโอไจฟ์โลจิสได้ โดยความชันของ ฟังก์ชั่นการตอบสนองต่อข้อสอบ นี้ กำหนดดังนี้

$$p' = \frac{dP}{d\theta} = \frac{1.7a(1-c)e^{1.7a(\theta-b)}}{\left[1+e^{1.7a(\theta-b)}\right]^2}$$

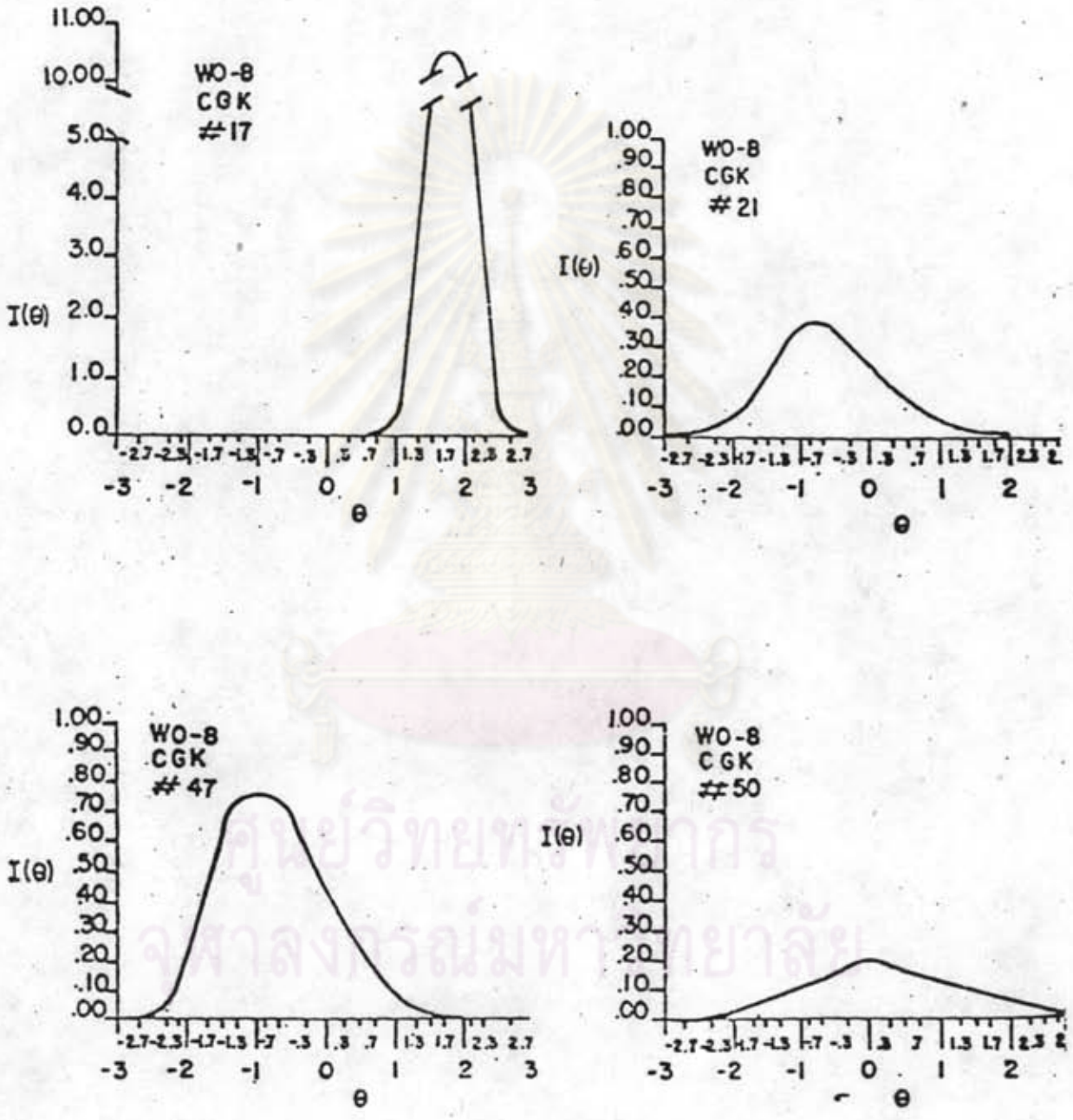
a, b และ c เป็นพารามิเตอร์ของข้อสอบและ θ เป็นจุดซึ่ง dp/dθ คือความชัน ความชันบางครั้งใช้สัญญลักษณ์ว่า p'(θ) หรือ p' ในการคำนวณ p'(θ) ก็คืออนุพันธ์อันดับแรกของ p(θ) เพราะความชัน (p') คือค่าที่ให้ข้อมูลของข้อสอบที่แสดงในแต่ละจุดของเสกลความสามารถ

อย่างไรก็ตาม จากกรณีนี้ ด้วยเหตุผลทางคณิตศาสตร์และสถิติซึ่งเราจะทำต่อไป p'(θ) จะไม่เป็นการวัดข้อมูลที่เหมาะสมและสมบูรณ์ แต่มีฟังก์ชั่นที่สัมพันธ์กันคือ

$$I(\theta,u) = \frac{p'^2}{P(\theta)Q(\theta)} = \frac{(1.7a)^2(1-c)}{\left[c+e^{1.7a(\theta-b)}\right]\left[1+e^{-1.7a(\theta-b)}\right]^2}$$

ค่า p'^2 คือ p' ยกกลังสอง และ Q'(θ) = 1 - p (θ) สังเกตว่ากำลังของ e ทางซ้ายเป็นบวกและทางขวาเป็นลบ

ฟังก์ชั่น I(θ,U) นี้ เรียกว่า ฟังก์ชั่นสารสนเทศข้อสอบ และเขียนแทนด้วย I(θ,U) สูตรข้างต้นดูยุ่งยากกว่า P (θ) แต่ความจริงมีความซับซ้อนขึ้นเพียงเล็กน้อยเท่านั้น เป็นการสะดวกที่จะคำนวณจุดของ I (θ,U) ด้วยการคำนวณด้วยมือ



(Warm 1978: 66)

ภาพที่ 27 ฟังก์ชั่นสารสนเทศของข้อสอบ ของข้อสอบจริง 4 ข้อ

จากภาพที่ 27 แสดง I (θ,U) ของข้อสอบ 4 ข้อ ซึ่ง ฟังก์ชั่นการตอบสนองต่อข้อสอบ ได้แสดงในภาพที่ 28 (ให้สังเกตว่ามาตราตามแนวตั้งของข้อ 17 แตกต่างจากข้ออื่นๆ) ในการเปรียบเทียบฟังก์ชั่นการตอบสนองต่อข้อสอบด้วย ฟังก์ชั่นสารสนเทศของข้อสอบ มีข้อสังเกตถึงความสัมพันธ์ที่สำคัญ 3 อย่าง

1) ฟังก์ชั่นสารสนเทศของข้อสอบมีค่าสูงสุดเมื่อเข้าใกล้ค่าที่ชิ่งโค้งของฟังก์ชั่นการตอบสนองต่อข้อสอบ มีค่าชันที่สุด

2) พื้นที่ใต้โค้งทั้งหมดของฟังก์ชั่นสารสนเทศของข้อสอบเพิ่มขึ้น ขณะที่ค่า a เพิ่มขึ้น

3) พื้นที่ใต้โค้งทั้งหมดของฟังก์ชั่นสารสนเทศของข้อสอบลดลง ขณะที่ค่า c เพิ่มขึ้น

ความจริงที่ว่าข้อมูลทั้งหมด (เช่นพื้นที่ใต้ฟังก์ชั่นสารสนเทศของข้อสอบ) เพิ่มขณะที่ค่า a เพิ่มแสดงความสำคัญของค่า a สำหรับข้อสอบ อย่างไรก็ตามมีผลกระทบอื่นๆ ต่อค่า a ที่สูง ขณะที่ค่า a เพิ่มขึ้น ณ ความกว้างของเสกลความสามารถตลอดช่วงซึ่งให้ข้อมูลลดลงในการแจกแจง ผลกระทบนี้เรียกว่า The Bandwidth Paradox ดังนั้นบางครั้งการประมาณประมาณระหว่างข้อมูลทั้งหมดที่กำหนดโดยข้อสอบและการแจกแจงข้อมูลตลอดช่วงความสามารถ (θ)

สารสนเทศทั้งหมด (A_g) ของข้อสอบ g ถูกกำหนดโดย

$$Ag = \frac{1.7a\,(c \cdot \log c + (1-c))}{1-c} = 1.7a + \frac{1.7ac\log c}{1-c} = 1.7a\left(1 + \frac{c\log c}{1-c}\right)$$

(Warm 1978: 69)

a และ c เป็นพารามิเตอร์ของข้อสอบ และ log c เป็น log ธรรมชาติของ c สังเกตสูตร A_g เราสามารถพบว่าขณะที่ค่า a เพิ่ม A_g จะเพิ่มด้วย ผลปรากฏเช่นนี้เป็นความจริงที่ว่า ขณะที่ c ใกล้ศูนย์ A_g จะใกล้ 1.7a ดังนั้นข้อมูลทั้งหมดสูงสุดของข้อสอบ

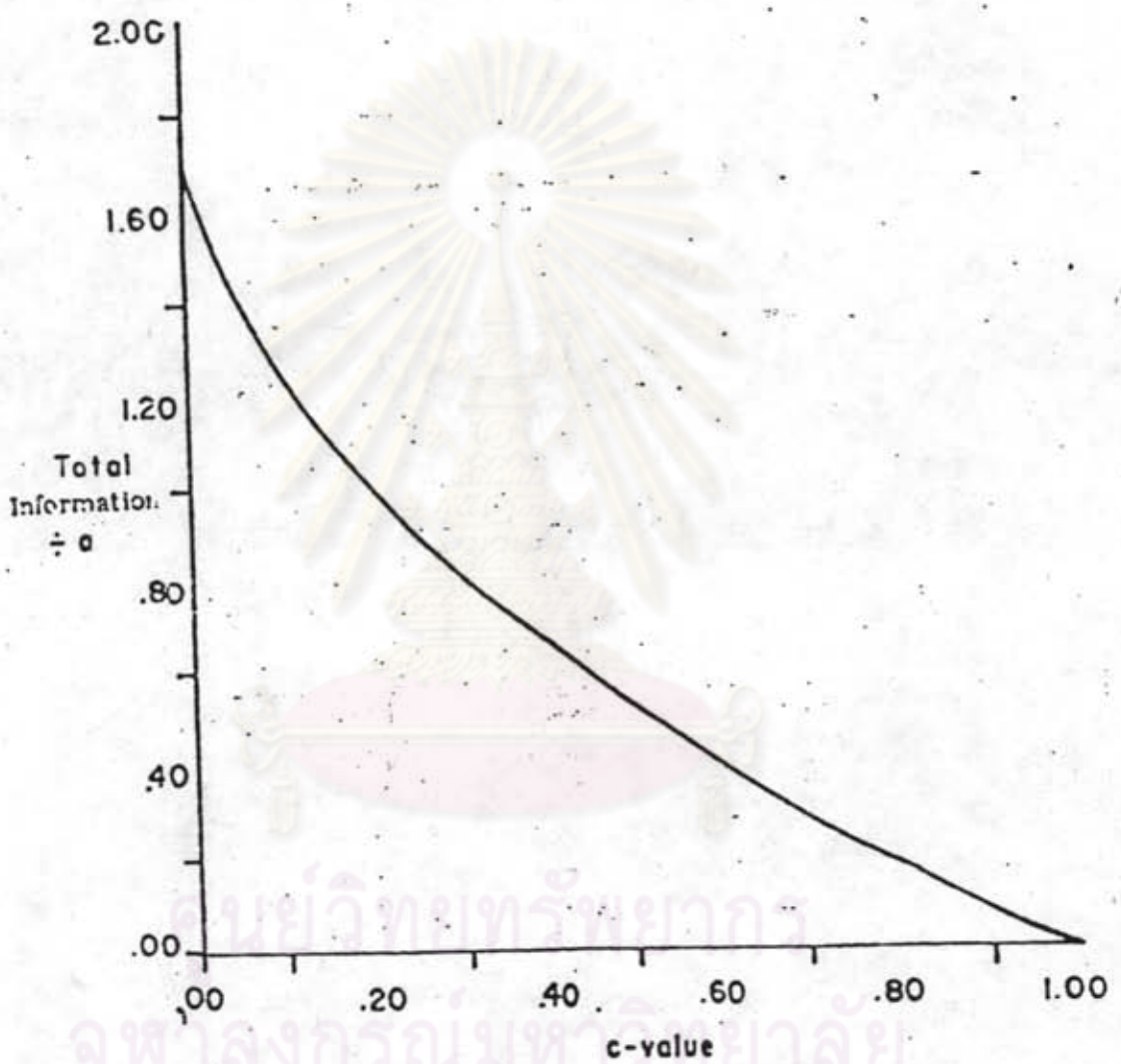
ข้อหนึ่ง จะให้ค่าเป็น 1.7a เป็นที่น่าสังเกตความสัมพันธ์ จากสูตร A_u เมื่อ c เข้าใกล้ 1.00 A_u จะเข้าใกล้ศูนย์ นี้ปรากฏขึ้นเพราะ log c เป็นลบเว้นแต่เมื่อ c = 1 และเพราะว่า c = 1, clog c/(1-c) = -1 ความสัมพันธ์นี้อธิบายถึงผลของค่า c ซึ่งค่า c จะเป็นตัวทำลายข้อมูล ภาพที่ 28 แสดงว่าข้อมูลทั้งหมดลดลงได้อย่างไรในขณะที่ค่า c เพิ่มขึ้น เมื่อกำหนดให้ค่า a คงที่

เพราะค่า b ไม่รวมไว้ในสูตร A_u ค่า b จึงไม่มีผลต่อสารสนเทศทั้งหมด จุดบน θ ซึ่ง ฟังก์ชั่นสารสนเทศของข้อสอบ เป็นค่าสูงสุดไม่ใช่ที่ค่า b เหมือนอย่างที่บางคนคาดหวัง ยกเว้นในกรณีที่ c = 0 จุดบน θ ซึ่งสารสนเทศให้สูงสุดถูกกำหนดโดย

$$\theta_{maxI_{(\theta,u)}} = b + \frac{1}{1.7a}\left[\log(.5+.5\sqrt{1+8c})\right]$$

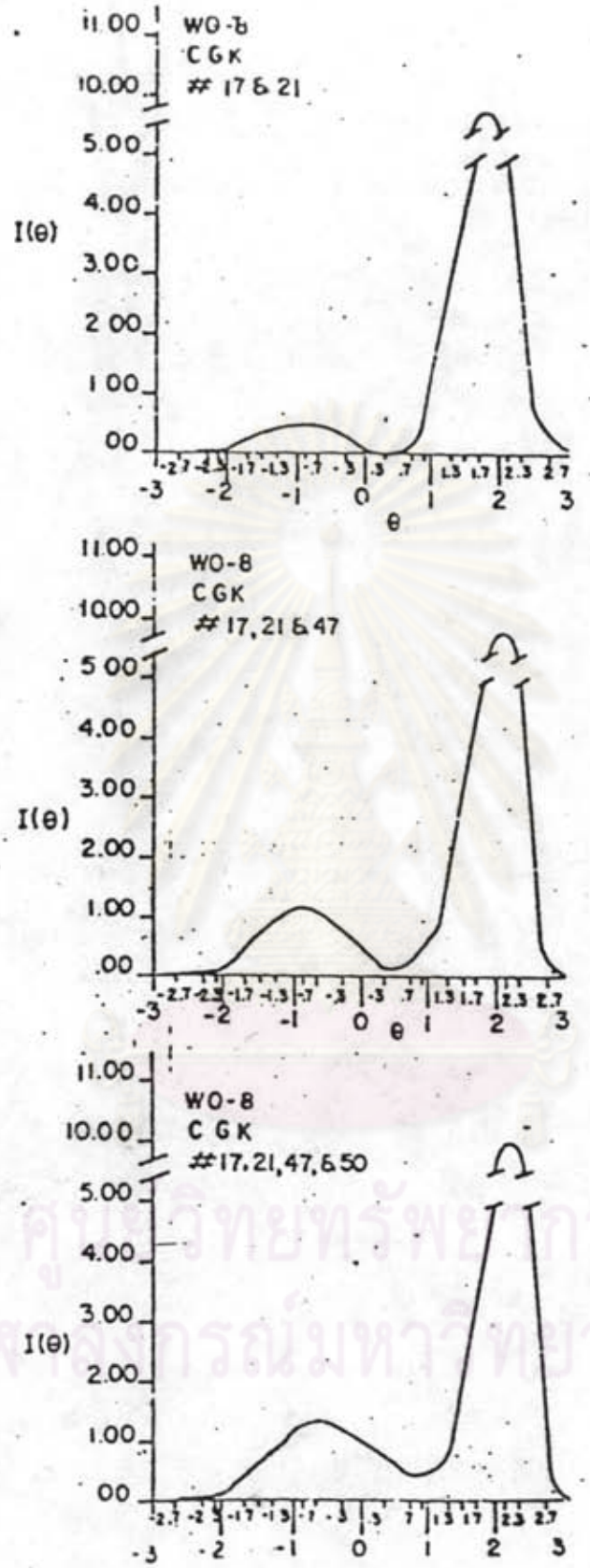
เมื่อ log ในสูตรเป็น log ธรรมชาติ จุดบน θ ซึ่งให้สารสนเทศสูงสุดอยู่ทางขวาของค่า b (ยกเว้นเมื่อ c = 0, มันอยู่ที่ค่า b) แต่ไม่เคยไปไกลกว่าเทอมทางขวามากกว่า .41/a

ฟังก์ชั่นสารสนเทศของข้อสอบ มีลักษณะสมมาตรเมื่อ c = 0 และเบ้ไปทางขวาเมื่อ $c \neq 0$ ยิ่งค่า c มากกว่า 0 ฟังก์ชั่นสารสนเทศของข้อสอบ จะยิ่งเบ้ไปทางขวา การที่โค้งเบ้ไปทางขวาเพราะค่า c ทำลายสารสนเทศในระดับความสามารถ (θ) ต่ำมากกว่าความสามารถระดับสูง ผลลัพธ์นี้เข้าใจได้เพราะว่าผู้สอบที่ความสามารถต่ำจะเดามากกว่าผู้เข้าสอบที่ความสามารถสูงกว่า การเดาทำลายสารสนเทศด้วยเหตุผลนี้ข้อสอบ 5 ตัวเลือกจึงได้รับความนิยมมากกว่า 4 ตัวเลือก



ภาพที่ 28 ความสัมพันธ์ของค่า c กับค่าสารสนเทศทั้งหมดที่ได้จากข้อสอบ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



(Warm 1978: 72)

ภาพที่ 29ก ข และ ค โค้งค่าสารสนเทศของแบบสอบที่ประกอบด้วย ข้อ 17 และข้อ 21 โค้งค่าสารสนเทศของแบบสอบที่ประกอบด้วยข้อ 17, 21 และ 47 และโค้งสารสนเทศของแบบสอบที่ประกอบด้วย ข้อ 17, 21, 27 และ 50 จาก the USCG Warrant Officer Test

9. โค้งสารสนเทศของแบบสอบและโค้งประสิทธิภาพสัมพัทธ์

(The Test Information Curve and Relative Efficiency Curve)

โค้งสารสนเทศแบบสอบคือผลรวมของฟังก์ชั่นสารสนเทศของข้อสอบทุกข้อเข้าด้วยกัน ฟังก์ชั่นสารสนเทศของข้อสอบ ที่ถูกรวมโดยรวมยอดของแต่ละข้อเข้าด้วยกัน ผลรวมฟังก์ชั่นสารสนเทศของข้อสอบ หมายความถึง ความสูง (เช่น ค่านวณของสารสนเทศ) ของ θ ทุกค่า ณ ค่าเฉพาะของ θ ถูกรวมเข้าด้วยกัน เพื่อให้ได้ความสูงของ ฟังก์ชั่นสารสนเทศแบบสอบที่ค่า θ นั้นการกำหนดจุดผลรวมของสารสนเทศข้อสอบที่ค่า θ นี้จะให้ โค้งสารสนเทศแบบสอบ ความสูงของ โค้งสารสนเทศแบบสอบ ที่ θ สามารถเขียนแทนด้วย I(θ)

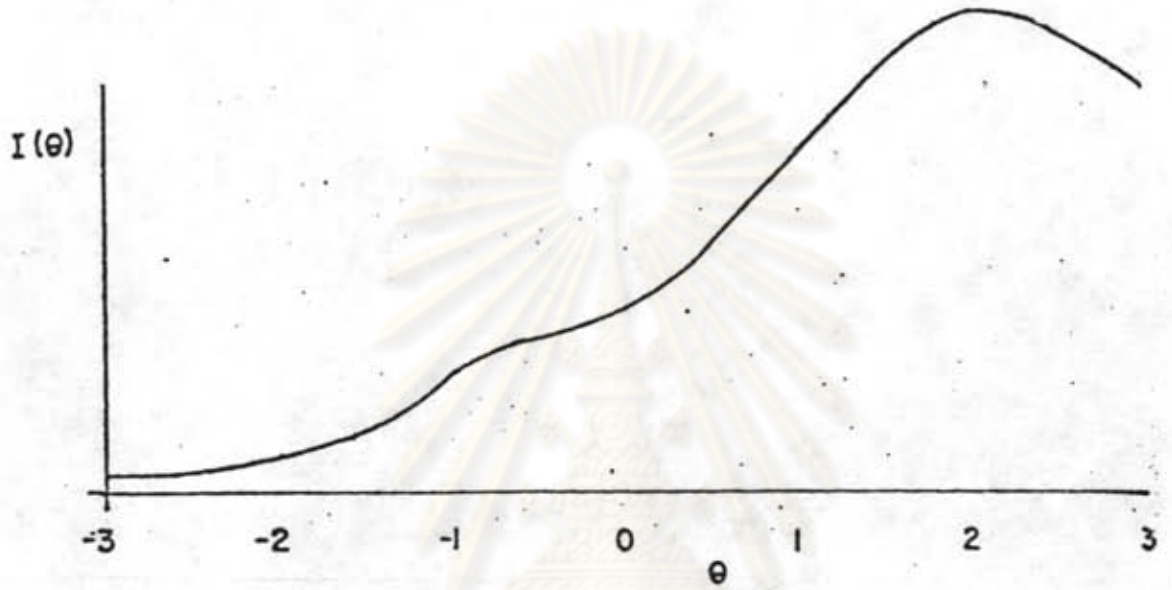
$$I(\theta) = \sum I(\theta, U)$$

ภาพที่ 29ก แสดงผลรวมของ ฟังก์ชั่นสารสนเทศของข้อสอบ ทั้งหมดของข้อสอบข้อ 17 และ 21 เช่น ที่แสดงในภาพที่ 27 และภาพที่ 29ข แสดงฟังก์ชั่นสารสนเทศของข้อสอบของข้อสอบข้อ 47 ที่ถูกรวมเข้ามาด้วยดังภาพที่ 29ก แผนภาพที่ 29ค แสดงฟังก์ชั่นสารสนเทศของข้อสอบของข้อ 50 ที่รวมข้ออื่น ๆ อีก 3 ข้อ แบบสอบที่ประกอบด้วยข้อสอบ 4 ข้อเหล่านี้ จะมีโค้งสารสนเทศแบบสอบ ดังภาพที่ 29ค

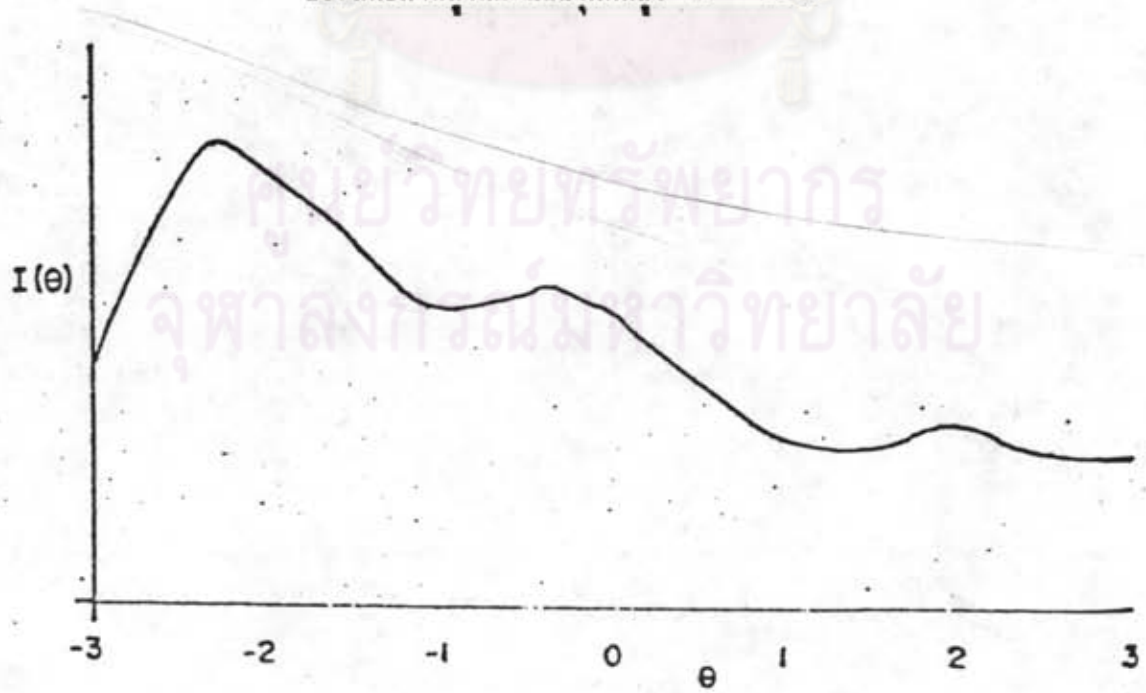
โค้งสารสนเทศแบบสอบ แสดงปริมาณสัมพันธ์ของสารสนเทศที่ถูกกำหนดโดยแบบสอบในแต่ละจุดบน θ สารสนเทศที่เราต้องการขึ้นกับว่าเราจะเอาแบบสอบไปใช้ทำอะไร ถ้าเราต้องการเลือกผู้เข้าสอบจำนวนน้อยๆ จากจำนวนมาก แล้วเราต้องการสารสนเทศมากในระดับความสามารถ (θ) ที่สูง ดังนั้นเราสามารถที่จะตัดสินได้ว่าผู้เข้าสอบคนไหนเป็นคนที่ดีที่สุด ยกตัวอย่างดังภาพที่ 30ค ถ้าเราต้องการเลือกผู้เข้าสอบทุกคนยกเว้นคนจำนวนน้อย แล้วเราก็ต้องการสารสนเทศมากในระดับความสามารถต่ำที่จะสามารถบอกเราได้ว่าผู้เข้าสอบคนไหนมีความสามารถต่ำ (ดังภาพที่ 30ข)

บางครั้งแบบสอบได้ถูกวางแผนมากกว่าความมุ่งหมายเพียงข้อเดียว เช่น ถูกนำมาใช้เพื่อตัดคะแนนออกเป็นสองส่วนในการรับเข้าเรียนในโรงเรียนที่แตกต่างกัน 2 โรง ในกรณีนี้ โค้งสารสนเทศแบบสอบ จะให้สารสนเทศที่ดีที่คะแนนจุดตัดเท่ากับ 2 (ดูภาพที่ 32ก)

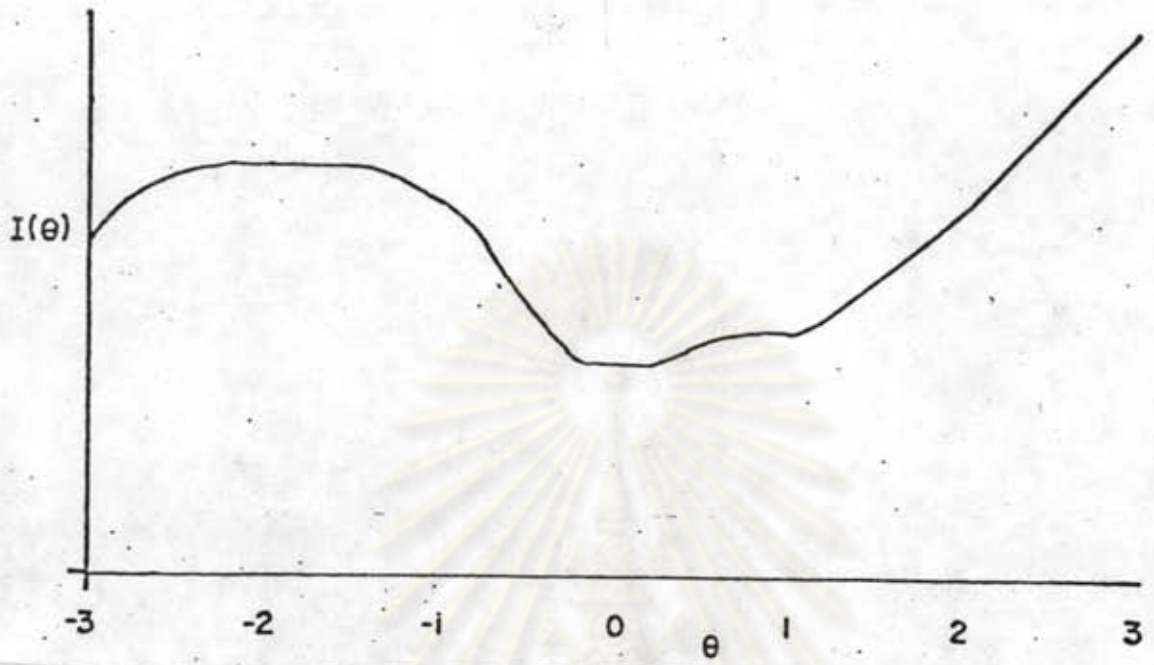
โค้งสารสนเทศแบบสอบ ค่าหนึ่งรูปร่างตามที่ต้องการอาจจะถูกสร้างขึ้น เพื่อ, กำหนดข้อสอบที่มี ฟังก์ชั่นสารสนเทศของข้อสอบ ที่ต้องการจะนำมาเพื่อสร้างโค้งสารสนเทศแบบสอบ



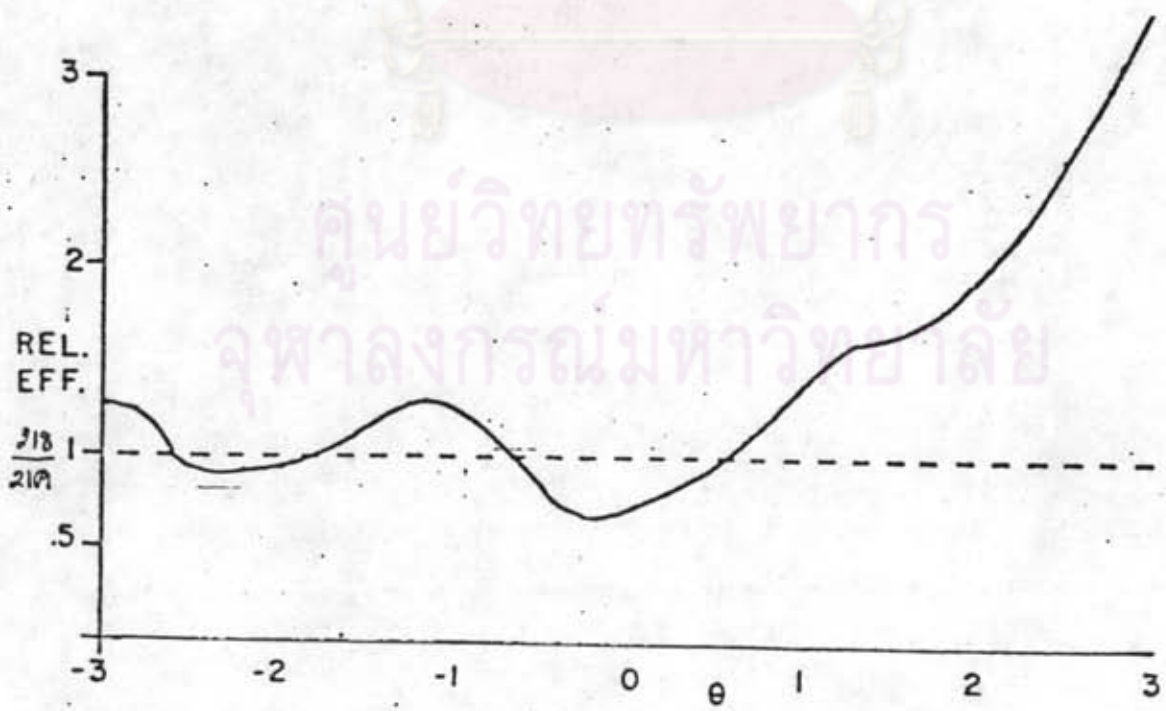
ภาพที่ 30ก โค้งสารสนเทศของแบบสอบสมมติฐานซึ่งมีประสิทธิภาพสูงที่คะแนนจุดตัดสูง (θ = 2.0)



ภาพที่ 30ข โค้งสารสนเทศของแบบสอบสมมติฐานซึ่งมีประสิทธิภาพสูงที่คะแนนจุดตัดต่ำ (θ = -2.3)



ภาพที่ 30ค โค้งสารสนเทศของแบบสอบถามสมมติฐานซึ่งมี
ประสิทธิภาพที่จุดตัดสูงและต่ำ



ภาพที่ 30ง โค้งประสิทธิภาพสัมพันธ์ในการเปรียบเทียบโค้งสารสนเทศ
ของแบบสอบในภาพที่ 30ค กับในภาพที่ 30ข

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

เมื่อเราพร้อมที่จะใช้แบบสอบและต้องการปรับหรือสร้างให้ดีขึ้นเพื่อสนองตามความมุ่งหมายของเรา การเปรียบเทียบชุดข้อสอบใหม่และเก่า ควรจะนำมาใช้โดยการหาโค้งประสิทธิภาพสัมพัทธ์ก็คืออัตราส่วนของโค้งสารสนเทศแบบสอบ อัตราส่วนของ 2 โค้งได้จากการหาร I(θ) จากแบบสอบฉบับหนึ่งด้วย I(θ) จากแบบสอบอื่นที่จุดบน θ หนึ่งๆ ภาพที่ 30ง เป็น โค้งประสิทธิภาพสัมพัทธ์ แสดงการเปรียบเทียบ โค้งสารสนเทศแบบสอบ ในภาพที่ 30ค กับ โค้งสารสนเทศแบบสอบ ในภาพที่ 30ข

ค่าโค้งประสิทธิภาพสัมพัทธ์ที่มากกว่า 1.0 ดังแบบสอบในภาพที่ 30ค (แบบสอบซึ่ง I(θ) เป็นตัวหารของอัตราส่วน โค้งประสิทธิภาพสัมพัทธ์ ดังกล่าว) ดีกว่าแบบสอบในภาพที่ 30ข แต่ถ้าค่าโค้งประสิทธิภาพสัมพัทธ์ ต่ำกว่า 1.0 แบบสอบดังภาพที่ 30ค จะดีกว่า และถ้าโค้งประสิทธิภาพสัมพัทธ์เท่ากับ 1.0 แบบสอบทั้ง 2 ฉบับจะเหมือนกัน

โดยการเริ่มจากแบบสอบเก่า ทำการแทนค่าในข้อสอบและคำนวณหาโค้งประสิทธิภาพสัมพัทธ์ เราจะสามารถทดลองและปรับปรุงแบบสอบเก่าโดยการลองผิดหรือลองถูกให้ได้แบบสอบตามความมุ่งหมายได้

ทุกๆ แบบสอบจะมีความคลาดเคลื่อนรวมอยู่ด้วย ความคลาดเคลื่อนของการประมาณ (The Standard Error of Estimate, SEE) คือ ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนของความสามารถที่ถูกประมาณโดยเฉลี่ย นั่นคือถ้าเราบริหารแบบสอบฉบับหนึ่งกับกลุ่มผู้เข้าสอบที่มีความสามารถคล้ายกันและประมาณค่าความสามารถของเขาจากแบบสอบ ส่วนเบี่ยงเบนมาตรฐานของการประมาณเหล่านี้คือ SEE

ถ้าการประมาณค่า θ ด้วยการประมาณภาวะน่าจะเป็นสูงสุด (maximum likelihood estimate) ค่า SEE ที่ความสามารถเฉพาะ θ จะง่ายในการคำนวณจากโค้งสารสนเทศแบบสอบ SEE จะเท่ากับรากที่สองของส่วนกลับของความสูงของ โค้งสารสนเทศแบบสอบ (I(θ)) เพราะ I(θ) กระจายไปตามเสกลความสามารถเราจะได้ SEE กระจายไปตามค่าเหล่านั้น ถ้า I(θ) มากขึ้น SEE จะลดลง ค่า SEE ที่เล็กจะเป็นสิ่งที่ต้องการมาก

ค่าเฉลี่ย SEE ตลอดผู้เข้าสอบได้นำมาสัมพันธ์กับความเที่ยงของทฤษฎีการวัดแบบคลาสสิค (r_{xx}) เมื่อคะแนนถูกแปลงเป็นมาตรฐานและมีส่วนเบี่ยงเบนมาตรฐานเป็น 1.0

$$r_{xx} = 1 - \frac{\overline{SEE}}{}^{2}$$

ความสัมพันธ์อันนี้อธิบายได้ว่าแบบสอบที่มีความเที่ยงสูงอาจจะเป็นแบบสอบที่ไม่ดีในความมุ่งหมายที่จะนำไปใช้ เพราะมันให้สารสนเทศที่ค่าวิกฤติของ θ ต่ำ ในทำนองเดียวกันแบบสอบที่มีความเที่ยงต่ำอาจจะเป็นแบบสอบที่ดีในความมุ่งหมายบางประการ ถ้าให้สารสนเทศสูงตามที่ต้องการ ดังนั้นความเที่ยงไม่เข้ามาเกี่ยวข้องกับค่าของแบบสอบตามแนวนี้มากนัก

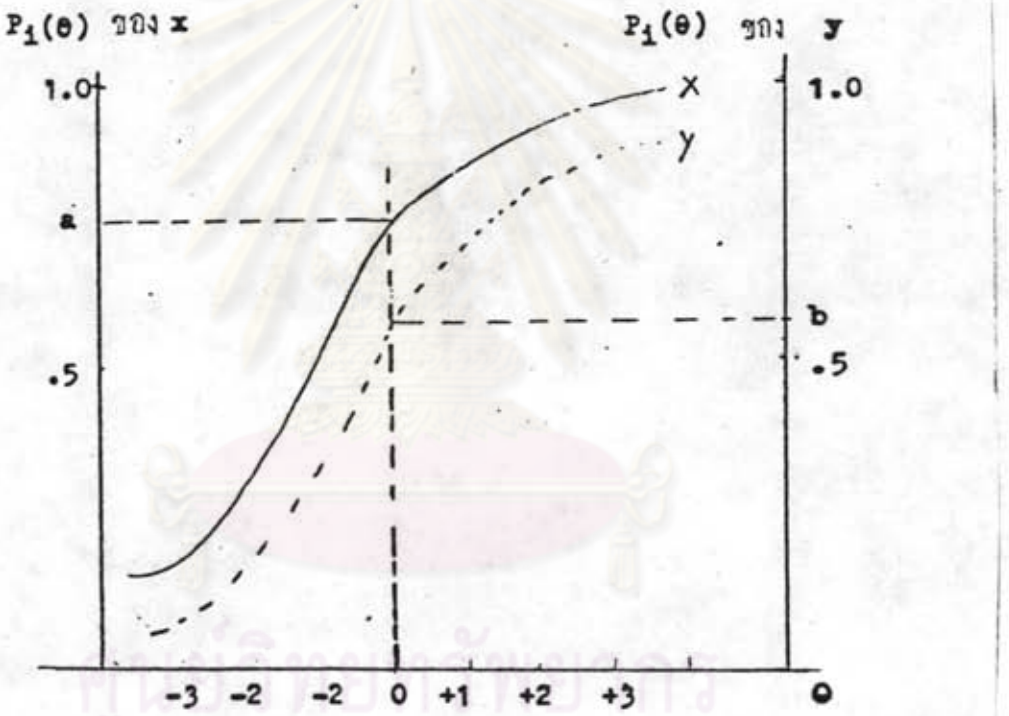
10. การนำทฤษฎีการตอบสนองต่อข้อสอบไปใช้

10.1 การวิเคราะห์ข้อสอบ (Item Analysis) การวิเคราะห์ข้อสอบโดยการนำทฤษฎีการตอบสนองต่อข้อสอบมาใช้ ไม่มีวัตถุประสงค์เพื่อการปรับปรุงข้อสอบ เช่นเดียวกับการวิเคราะห์ข้อสอบตามแนวคลาสสิค แต่การวิเคราะห์ข้อสอบตามแนวทฤษฎีการตอบสนองต่อข้อสอบนี้มุ่งที่จะรู้จักข้อสอบว่ามีคุณสมบัติอย่างไร เหมาะสมกับการสอบวัดแบบใดจะนำเอาโมเดลของ ทฤษฎีการตอบสนองต่อข้อสอบ มาใช้ได้หรือไม่ สำหรับข้อสอบนี้ การที่พฤติกรรมการตอบสนองต่อข้อสอบไม่เป็นไปตามโมเดลที่มาวิเคราะห์เพียงอธิบายได้ว่าพฤติกรรมกาตอบสนองข้อสอบนี้ไม่ได้ฟิตกับโมเดลที่มาวิเคราะห์นั้น เนื่องจากการวิเคราะห์ข้อสอบตามแนวทฤษฎีการตอบสนองต่อข้อสอบ จะทำให้ได้ค่าพารามิเตอร์ของข้อสอบ จึงทำให้สามารถนำคุณลักษณะของข้อสอบที่ได้มาเปรียบเทียบกันได้โดยอิสระ ทั้งนี้เพราะผลการวิเคราะห์ไม่ขึ้นอยู่กับความสามารถของผู้สอบเช่นเดียวกับการวิเคราะห์ในแนวคลาสสิค หลังจากได้ค่าพารามิเตอร์ประจำข้อแล้วเราสามารถนำข้อสอบต่างๆ มารวบรวมเก็บไว้ในคลังข้อสอบต่อไป

10.2 การออกแบบแบบสอบ (Designing Test) จากคุณลักษณะประจำข้อของข้อสอบแต่ละข้อที่รวบรวมไว้ในคลังข้อสอบผู้สร้างแบบสอบ สามารถวางแผนการแบบสอบให้มีคุณสมบัติเหมาะสมกับความมุ่งหมายที่จะนำแบบสอบไปใช้วัด โดยสามารถเลือกข้อสอบที่มีความยากง่าย เหมาะสมกับความสามารถของผู้สอบได้อย่างเหมาะสม ซึ่งสามารถนำไปใช้ได้กับแบบสอบแบบรอบรู้ (mastery testing) แบบลำดับขั้น (Sequential Test) หรือแบบสอบที่เหมาะกับแต่ละคน (Tailored Test)

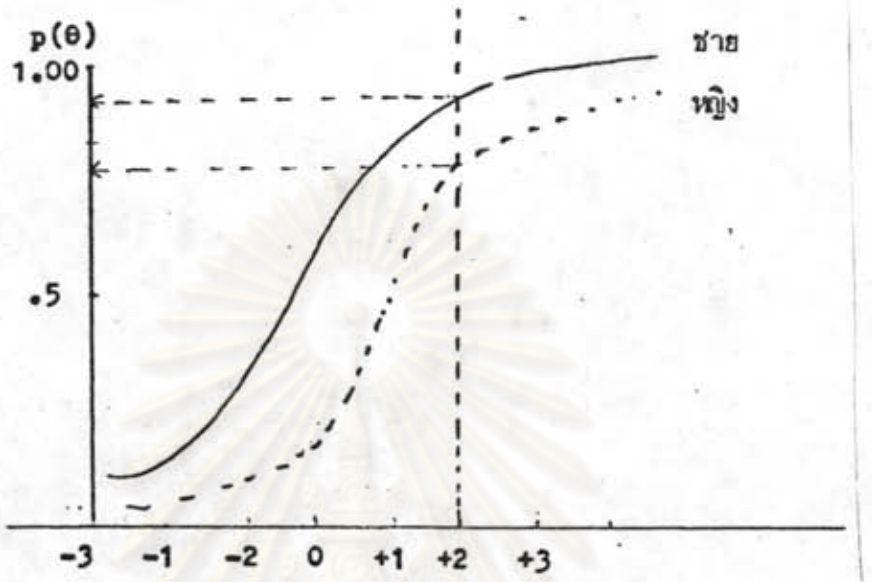
10.3 การเทียบมาตรา (Equating) การเทียบมาตราเป็นการแปลง

คะแนนจากหน่วยหนึ่งไปสู่อีกหน่วยหนึ่งของแบบสอบในเรื่องเดียวกันแต่ต่างฟอร์มกัน เมื่อคำนวณพารามิเตอร์ของข้อสอบได้ ค่าพารามิเตอร์จะอยู่บนมาตราเดียวกันคือ θ วิธีการแปลงมาตราความสามารถของกลุ่มหนึ่งไปยังมาตราความสามารถของอีกกลุ่มหนึ่ง สามารถเป็นไปได้ เพราะค่า b ของข้อไม่แปรเปลี่ยนและเนื่องจากเสกลความสามารถ มีความสัมพันธ์เชิงเส้นตรง จากภาพสามารถแสดงวิธีการเทียบมาตราจากแบบสอบชุด x และ y ดังนี้



จากภาพพอจะเห็นแนวทางการเทียบมาตราได้ว่า ณ θ = 0, โอกาสในการตอบข้อ i ในแบบสอบ x ถูก = a และโอกาสในการตอบข้อ i ในแบบสอบ y ถูก = b สำหรับสารสนเทศจากแบบสอบทั้ง ฉบับจะคำนวณจาก p (θ) ต่อไป

10.4 ความลำเอียงของแบบสอบ (Test Bias) จากโค้งลักษณะของข้อสอบสามารถตรวจสอบได้ว่าข้อสอบดังกล่าวมีความลำเอียงกับผู้สอบในกลุ่มใด เช่น ข้อสอบข้อหนึ่งเมื่อนำไปสอบในกลุ่มเพศชาย และหญิง แล้วได้ลักษณะของโค้งดังภาพ แสดงว่าข้อสอบข้อนี้มีความลำเอียงที่ทำให้เพศชายสามารถมีโอกาสในการตอบได้มากกว่าเพศหญิงทุกระดับมาตราความสามารถ (θ)



กล่าวโดยสรุปแนวทางในการนำทฤษฎีการตอบสนองต่อข้อสอบสามารถนำไปใช้วิธีการอื่นๆ ได้อีก นอกจากที่กล่าวมา เช่น การกำหนดจุดตัดของคะแนนการผ่าน (Setting Minimum Passing Scores) และการเปรียบเทียบประสิทธิภาพของแบบสอบ 2 ฉบับ การประมาณค่าคะแนนจริง ซึ่งเป็นเรื่องราวที่ต้องศึกษาในรายละเอียดต่อไป ทฤษฎีการตอบสนองต่อข้อสอบได้นำมาใช้อย่างกว้างขวางในการวัดทางจิตมิติ และสามารถนำมาอธิบายคุณลักษณะของข้อสอบได้ชัดเจนขึ้นโดยเสริมจุดอ่อนจากการวิเคราะห์แบบสอบแบบคลาสสิคที่ผลการวิเคราะห์ขึ้นอยู่กับความสามารถของผู้สอบ

การศึกษาครั้งนี้ได้เลือกโมเดลโลจิสสามพารามิเตอร์ในการที่จะใช้วิเคราะห์ข้อมูลก็ด้วยเหตุผลดังนี้

1) ค่าคุณลักษณะของข้อสอบที่ได้เป็นค่าคงที่

2) มีค่าบรรยายคุณลักษณะของข้อสอบถึง 3 ตัว ได้แก่ ค่าอำนาจจำแนก ค่าความยาก และค่าการเดา

3) ข้อสรุปในการศึกษาครั้งนี้มุ่งที่จะนำผลไปใช้กับข้อสอบชนิดเลือกตอบ ซึ่งอาจจะมีตัวเลือกไม่เท่ากัน ดังนั้นค่าการเดาก็เป็นคุณลักษณะของข้อสอบประการสำคัญที่จะต้องนำมาศึกษาด้วย

การแจกแจงปกติ

ในการวิจัยครั้งนี้ได้นำเอาโค้งปกติมาตรฐานมาเป็นรูปแบบของโค้งสารสนเทศเป้าหมายของข้อสอบ จึงขอเสนอรายละเอียดของโค้งปกติมาตรฐานดังต่อไปนี้

1. ความสำคัญของการแจกแจงปกติ

การแจกแจงปกติ เป็นพื้นฐานที่สำคัญของทฤษฎีสถิติปัจจุบัน และสามารถนำไปใช้ประโยชน์ได้หลายด้าน เฮส์ (Hay 1965 : 225-227) มู้ดกับเกรย์บิล (Mood และ Graybill 1963 : 156-157) และสนีเดอคอร์กับคอคแรน (Snedecor และ Cochran 1971 : 35) ได้ให้เหตุผลแสดงถึงความสำคัญของการแจกแจงปกติไว้ดังนี้

1) ประชากรที่ได้จากการวัด อาจถือว่ามีการแจกแจงปกติ ทั้งนี้เนื่องมาจากธรรมชาติของการวัด ประชากรดังกล่าวแบ่งเป็น

ก. ข้อมูลทางชีววิทยา เช่น สัดส่วนของจำนวนทารกหญิงและชายที่เกิดในประเทศหนึ่งๆ ในคาบหนึ่งๆ สัดส่วนของพืชและสัตว์ชนิดต่างๆ เมื่อทำการผสมพันธุ์ข้ามชนิด (ตามกฎของเมนเดล)

ข. ข้อมูลทางมนุษยวิทยา เช่น ส่วนสูง น้ำหนัก ดัชนีส่วนศีรษะฯลฯ ของคนกลุ่มใหญ่ที่เป็นเพศเดียวกันและวัยเดียวกัน

ค. ข้อมูลทางเศรษฐกิจและสังคม เช่น อัตราเกิด อัตราตาย อัตราสมรส ในสภาวะปกติ ค่าจ้างและผลิตผลของคนงานที่ทำงานอย่างเดียวกัน

ง. การวัดทางจิตวิทยา เช่น สติปัญญาของคนเมื่อวัดโดยใช้แบบทดสอบมาตรฐาน อัตราเร่งในการสร้างคำสังสรรค์ ระยะเวลาในการมีปฏิกริยาตอบสนอง คะแนนทดสอบทางการศึกษา

จ. ความคลาดเคลื่อนที่เกิดจากการสังเกตหรือการวัด (Errors of observation) เช่น การวัดส่วนสูง การวัดอัตราเร็วในการเคลื่อนไหว การวัดลักษณะทางกายภาพและทางสมอง เป็นต้น ความคลาดเคลื่อนจะทำให้ ตัวแปรเหล่านี้มีค่าเบี่ยงเบนไปจากค่าจริง ทั้งมากและน้อยกว่าค่าจริง ความถี่ของความคลาดเคลื่อนขนาดต่างๆ จะประมาณได้อย่างใกล้เคียงกับโค้งปกติ (Garrett 1964 : 95)

2) ในกรณีที่ประชากรที่ได้จากการวัดแจกแจงไม่ปกติ อาจทำให้แจกแจงเข้าใกล้ปกติได้โดยการเปลี่ยนมาตราวัดอย่างง่าย ที่ใช้กันบ่อยคือ ถอดรากที่สองของค่าตัวแปร หรือเปลี่ยนค่าตัวแปรเป็นค่าล็อก (log)

3) เมื่อถือว่าประชากรแจกแจงปกติ การคำนวณจะทำได้สะดวก ปัญหาทางคณิตศาสตร์สถิติจำนวนมากมายที่แก้ได้ในรูปของการแจกแจงของประชากรปกติเท่านั้น อาศัยคุณสมบัติทางคณิตศาสตร์ที่สำคัญของการแจกแจงปกติ ทำให้นักสถิติสามารถพัฒนาทฤษฎีสถิติขึ้นมาหลายทฤษฎี การแจกแจงปกติจึงเป็น "ต้นกำเนิด (parent)" ของการแจกแจงตามทฤษฎีอื่นๆ ที่สำคัญหลายอย่าง

4) อาจใช้การแจกแจงปกติเป็นค่าประมาณที่ดี (good approximation) ของการแจกแจงตามทฤษฎีอื่นๆ ที่ไม่อาจคำนวณความน่าจะเป็นจากการแจกแจงนั้นๆ โดยตรงได้ หรือคำนวณได้ยาก เช่น เมื่อประชากรทวินาม มีขนาดใหญ่มากๆ การคำนวณความน่าจะเป็นจากการแจกแจงทวินามโดยตรงย่อมทำได้ยาก ก็อาจใช้การแจกแจงปกติเป็นค่าประมาณของการแจกแจงทวินามภายใต้เงื่อนไขบางอย่างได้ แม้ว่าการแจกแจงปกติจะเป็นการแจกแจงแบบต่อเนื่อง ส่วนการแจกแจงทวินามเป็นการแจกแจงแบบไม่ต่อเนื่องก็ตาม

5) การแจกแจงตัวอย่าง (Sampling Distribution) ของค่าสถิติหลายค่า เมื่อกลุ่มตัวอย่างมีขนาดใหญ่ จะเข้าใกล้การแจกแจงปกติ แม้ว่าการแจกแจงของประชากรของกลุ่มตัวอย่างไม่เป็นปกติก็ตาม

2. คุณสมบัติของการแจกแจงปกติ

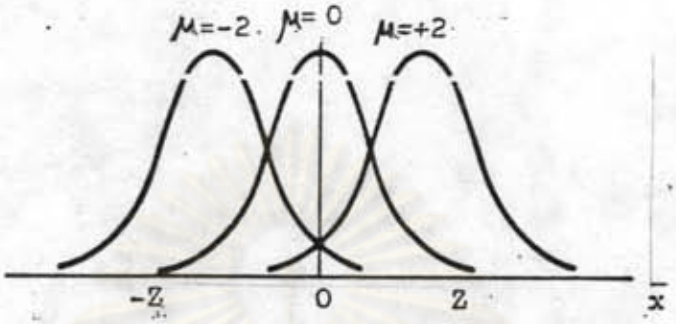
ก. คุณสมบัติทางคณิตศาสตร์

1) ฟังก์ชันความหนาแน่นของการแจกแจงปกติ คือ

$$n(x;\mu,\sigma^2) \quad = \quad f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty$$

โดยมีมัชฌิม = μ; ความแปรปรวน = σ^2 และ μ กับ σ^2 เป็นพารามิเตอร์ของการแจกแจงปกติ

2) เมื่อเปลี่ยนค่า μ รูปโค้งของการแจกแจงจะไม่เปลี่ยน แต่จะเคลื่อนไปตามแกน x ดังภาพที่ 31

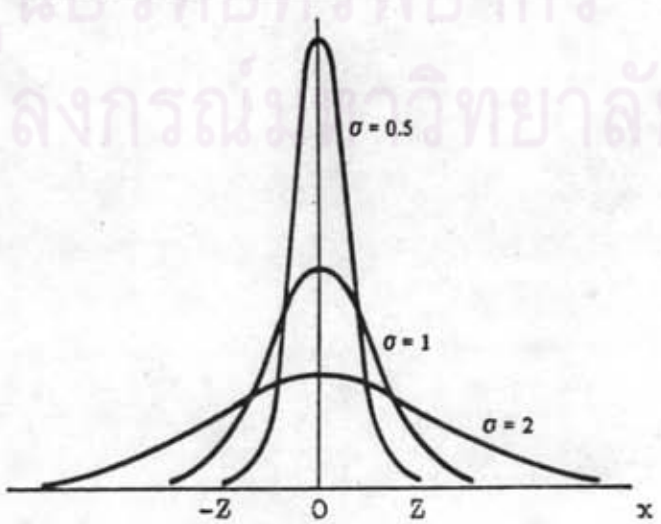


ภาพที่ 31 แสดงการแจกแจงปกติ เมื่อ μ ต่างกัน แต่ σ^2 เท่ากัน

3) เมื่อเปลี่ยนค่า σ^2 รูปโค้งของการแจกแจงจะเปลี่ยน ค่า σ^2 มาก โค้งจะแบนลง ค่า σ^2 น้อยโค้งจะโด่งขึ้น ดังภาพที่ 32

4) ฟังก์ชั่นกำหนดไปตลอดแกน X และมีค่าเป็นบวกสำหรับทุกๆค่าของ X ดังนั้นเส้นโค้งของการแจกแจง (เรียกสั้นๆ ว่า โค้งปกติ) จึงอยู่เหนือแกน X ทั้งหมด

5) ขีดจำกัดของฟังก์ชั่นเมื่อ X เข้าใกล้ $-\alpha$ หรือ $+\alpha$ เป็น 0 แกน X เป็นเส้นที่เข้าใกล้เส้นโค้งตามแนวนอนของกราฟ



ภาพที่ 32 แสดงการแจกแจงปกติ เมื่อ μ เท่ากัน แต่ σ^2 ต่างกัน

6) จากการดิฟเฟอเรนติเอท (Differentiate) ฟังก์ชั่นครั้งที่หนึ่ง

$$f(X) = \frac{(X-u)\, e^{-(x-\mu)^2/2\sigma^2}}{\sigma^3\sqrt{2\pi}}$$

จะได้ว่า $f'(X) = 0$ เมื่อ $X = \mu$

$f'(X) > 0$ เมื่อ $X < \mu$

และ $f'(X) < 0$ เมื่อ $X > \mu$

ดังนั้นค่าออร์ดิเนตที่มากที่สุดคือ $f(x) = 1/(\sqrt{2\pi}\sigma^2)$ และเมื่อ $X<\mu$ หรือ $X>\mu$ ค่าออร์ดิเนตจะน้อยลง แต่จะไม่เป็น 0 เว้นแต่ $(X-\mu)$ เข้าใกล้ α ด้วยเหตุนี้ โค้งปกติจึงเป็นเส้นโค้งยอดเดี่ยว (Unimodal) รูประฆังชิ่งค่อยๆ ลาดลงสู่แกน X และยื่นออกไปทั้งสองข้างอย่างไม่มีที่สิ้นสุด

7) การแจกแจงปกติมีลักษณะสมมาตร (Symmetrical Distribution) รอบ μ ทั้งนี้เพราะตัวแปรอิสระในสมการของฟังก์ชั่นอยู่ใน $(x-\mu)^2$ เท่านั้น ค่า X ที่ต่างกัน 2 ค่าซึ่งมี $|X-\mu|$ เท่ากัน จึงมีความหนาแน่นเหมือนกัน

8) จุดเปลี่ยนโค้ง (Points of Inflection) หาได้จากการดิฟเฟอเรนติเอทฟังก์ชั่นครั้งที่ 2 จะได้

$$f''(x) = \{-\frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/2\sigma^2}\}\{1-\frac{(x-\mu)^2}{\sigma^2}\}$$

$f''(x)$ จะเป็น 0 เมื่อ $x = \mu + \sigma$ และ $x = \mu-\sigma$

และเมื่อ $x = \mu + \sigma$; $f(x) = \frac{1}{\sigma\sqrt{2\pi e}}$

$x = \mu - \sigma$; $f(x) = \frac{1}{\sigma\sqrt{2\pi e}}$

ดังนั้นจุดเปลี่ยนโค้งจะอยู่ที่

$$[\mu-\sigma, \frac{1}{\sigma\sqrt{2\pi e}}] \text{ และ } [\mu+\sigma, \frac{1}{\sigma\sqrt{2\pi e}}]$$

ข. คุณสมบัติทางสถิติ

1) ทุกรายการของข้อมูลจะรวมอยู่ภายใต้เส้นโค้งปกติและอยู่เหนือแกน X

2) มัชฌิม (Mean) มัธยฐาน (Median) และฐานนิยม (Mode) มีค่าเท่ากัน

3) ช่วง μ = 1, 2 และ 3σ จะรวมรายการข้อมูลอยู่ร้อยละ 68.27, 95.45 และ 99.73 ตามลำดับ

4) ช่วง $\mu \pm .6745\sigma$ จะมีรายการข้อมูลอยู่ร้อยละ 50 ช่วงนี้ เรียกว่า Probable Error (P.E.)

5) ส่วนเบี่ยงเบนควอไทล์ (Quartile Deviation) หรือ Semi Inter-quartile Range $\frac{Q_3 - Q_1}{2}$ = ความน่าจะเป็นคลาดเคลื่อน (probable error)

6) ค่า Q_1 และ Q_3 จะมีระยะห่างจากมัธยฐานเท่ากัน

7) ส่วนเบี่ยงเบนเฉลี่ย (Average Deviation, A.D.) = .7979

8) Probable Error (P.E.) = .845 A.D.

9) การเปลี่ยนมาตรวัดแบบเส้นตรง (Linear Transformation of Scale) (Chao 1969 : 177) ถ้า X และ Y เป็นตัวแปรสุ่ม 2 ตัว $Y = aX + b$; $(a \neq 0)$ และ $X \sim n(\mu, \sigma^2)$ ซึ่งหมายความว่า X มีการแจกแจงปกติโดยมีมัชฌิมและความแปรปรวน σ^2 แล้ว $Y \sim n(a\mu + b, a^2\sigma^2)$ ในการเปลี่ยน X ให้เป็นคะแนนมาตรฐานก็เป็นการเปลี่ยนมาตราวัดแบบเส้นตรงของ X โดยให้ $Z = \frac{X - \mu}{\sigma}$ และ $X \sim n(\mu, \sigma^2)$ แล้ว $Z \sim n(0, 1)$

10) Linear combination ของตัวแปรที่มีการแจกแจงปกติ (Chao 1969 : 177-178) ถ้า $Y = \Sigma a_k X_k$ โดยที่ $X_k \sim n(\mu_k, \sigma_k^2)$

X เป็นตัวแปรสุ่มอิสระ และ a เป็นค่าถ่วงน้ำหนัก (ทุกตัวไม่เป็นศูนย์) แล้ว $Y \sim n(\Sigma a_k\mu_k, \Sigma a_k^2\mu_k^2)$ พึงสังเกตว่า มัชฌิมของกลุ่มตัวอย่างสุ่มก็เป็น linear combination ของตัวแปรสุ่มซึ่งรวมกันเป็นกลุ่มตัวอย่างนั้น สมมุติว่า X เป็นมัชฌิมของกลุ่มตัวอย่างสุ่มกลุ่มหนึ่งซึ่งประกอบด้วยสมาชิก n ตัวที่เป็นอิสระ

$$\bar{X} = \Sigma X_i/n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$$

ในที่นี้ 1/n เป็นค่าถ่วงน้ำหนักของค่า X แต่ละตัว ถ้า X มีมัชฌิม u และความแปรปรวน r^2 แล้ว X จะมีมัชฌิม $\Sigma(1/n)\mu_i$ ซึ่งเท่ากับ μ และความแปรปรวน $\Sigma (1/n)^2 r_i^2 = (1/n^2)nr^2 = r^2/n$ นอกจากนี้ ถ้าค่าของ X แจกแจงเป็นปกติ X ก็จะแจกแจงเป็นปกติด้วยนั่นคือ ถ้า $X \sim n(\mu,\sigma^2)$ และ $X = \Sigma X_i/n$ แล้ว $X \sim n(\mu,\sigma^2/n)$ กล่าวอีกนัยหนึ่ง ถ้าสุ่มกลุ่มตัวอย่างซึ่งประกอบด้วยสมาชิกที่เป็นอิสระ n ตัว ได้จากประชากรที่มีการแจกแจงปกติ การแจกแจงตัวอย่าง (Sampling Distribution) ของมัชฌิมจะเป็นปกติด้วย โดยไม่คำนึงถึงขนาดของ n

3 การแจกแจงปกติมาตรฐาน (Standard Normal Distribution)

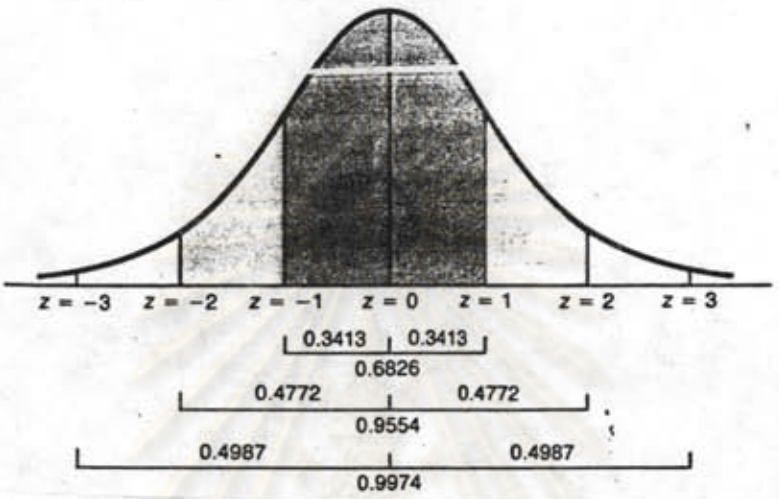
$$n(x;\ \mu,\ \sigma^2) = \frac{1}{\sqrt{2\pi r}} e^{-1/2(x-\mu)^2/\sigma^2} \qquad -\alpha < x < \alpha$$

เป็นฟังก์ชันความหนาแน่นของการแจกแจงปกติซึ่งมีมัชฌิมเท่ากับ u และความแปรปรวนเท่ากับ σ^2 เมื่อให้ $Z = \frac{X - \mu}{\sigma}$ ฟังก์ชันความหนาแน่นข้างบนจะกลายเป็น

$$n(z;\ 0,\ 1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad -\alpha < z < \alpha$$

ซึ่งเป็นฟังก์ชันความหนาแน่นของการแจกแจงปกติมาตรฐาน โดยมีมัชฌิมเท่ากับ 0 และความแปรปรวนเท่ากับ 1

ค่าฟังก์ชั่น N (Z ; 0, 1) สำหรับแต่ละค่าของ z เราไม่จำเป็นต้องคำนวณเอง เพราะนักสถิติได้คำนวณและทำเป็นตารางไว้แล้ว



ภาพที่ 33 แสดงการแจกแจงปกติมาตรฐาน และร้อยละของรายการข้อมูลที่กระจายอยู่ในช่วงต่างๆ รอบมัชฌิม

ในการวิจัยครั้งนี้ได้นำรูปแบบของโค้งปกติมาตรฐานมาเป็นรูปแบบของโค้งสารสนเทศของข้อสอบเป้าหมายสำหรับกลุ่มผู้สอบที่มีความสามารถปานกลาง รูปแบบของโค้งปกติมาตรฐานที่มีค่าเฉลี่ยเท่ากับ 1 มาเป็นรูปแบบของโค้งสารสนเทศของข้อสอบเป้าหมายสำหรับกลุ่มผู้สอบที่มีความสามารถสูง และรูปแบบของโค้งปกติมาตรฐานที่มีค่าเฉลี่ยเท่ากับ -1 มาเป็นรูปแบบของโค้งสารสนเทศของข้อสอบเป้าหมายสำหรับกลุ่มผู้สอบที่มีความสามารถต่ำ

งานวิจัยที่เกี่ยวข้องกับการเลือกข้อสอบ

ในปี 1970 เนลล์ และ แจคสัน (Niel and Jackson, 1970: 651-655) ได้ประเมินยุทธวิธีในการเลือกข้อสอบบุคลิคภาพโดยมีวิธีในการเลือกข้อสอบต่างๆกัน เช่น

เลือกข้อสอบที่มีค่าอำนาจจำแนกสูงโดยพิจารณาจาก ค่าสหสัมพันธ์ไบซีเรียล ค่าสหสัมพันธ์พอยท์-ไบซีเรียล ค่าองค์ประกอบหลัก (Principal Component) ที่มีค่าองค์ประกอบน้ำหนักสูง โดยแปรค่าจำนวนข้อในการประกอบเป็นแบบสอบตั้งแต่ 4 ข้อ จนถึง 40 ข้อ โดยเพิ่มครั้งละ 4 ข้อ แล้วใช้ค่าความเที่ยงจากสูตร KR20 เป็นเกณฑ์ในการตัดสินคุณภาพของแบบสอบ โดยใช้นักศึกษาระดับมหาวิทยาลัย 264 คน ทำข้อสอบทั้ง 88 ข้อจากคลังข้อสอบ(44 ข้อ เป็นข้อความทางบวกและ อีก 44 ข้อเป็นข้อความทางลบ) สรุปผลที่น่าสนใจได้ว่าถ้าจำนวนข้อสอบยิ่งมากขึ้น ไม่ว่าจะใช้วิธีการเลือกข้อสอบแบบใดก็จะให้ค่าความเที่ยงแบบ KR20 ใกล้เคียงกันโดยเฉพาะเมื่อจำนวนข้อสอบประมาณ 40 ข้อขึ้นไป

ในปี 1977 เนโว (Nevo, 1977: 847-852) ได้เปรียบเทียบวิธีการเลือกข้อสอบ 2 วิธีคือ ITRS (Item Test-Retest Stability) กับ IIC (Item Internal Consistency) โดยเกณฑ์ในการพิจารณาก็คือ ความสอดคล้องภายในของแบบสอบ (Test Internal Consistency) และความคงที่ของการสอบซ้ำ (Retest Stability) ใช้กลุ่มตัวอย่างนักศึกษาปีที่ 1 จำนวน 100 คนทำแบบสอบ 4 แบบ คือ

1. V (Vocabulary Test) เป็นแบบสอบเลือกตอบจำนวน 40 ข้อ
2. G (General Information Test) เป็นแบบสอบเลือกตอบจำนวน 33 ข้อ
3. AI (Achievement via Independence Scale) เป็นข้อสอบถูกผิดจำนวน 21 ข้อ
4. IE (Intellectual Efficiency Scale) เป็นข้อสอบถูกผิดจำนวน 28 ข้อ

ได้ผลสรุปว่าการใช้วิธี ITRS จะให้ความคงที่ในการสอบซ้ำใกล้เคียงกับวิธี IIC ยกเว้นแบบสอบ V ครึ่งฉบับ และให้ค่าความสอดคล้องภายในน้อยมากสำหรับแบบสอบ AI และ IE

ต่อมา คุกและแฮมเบิลตัน (Cook and Hambleton, 1978 cited in Hambleton and Swaninathan, 1985: 240-253) ได้ศึกษาถึงวิธีการเลือกข้อสอบ 5 แบบ จากข้อสอบวัดผลสัมฤทธิ์จำนวน 200 ข้อโดยเลือกมา 30 ข้อ ด้วยวิธีการต่างๆ โดยใช้เกณฑ์ในการตัดสินคุณภาพจากการเปรียบเทียบสารสนเทศของแบบสอบ (Test Information Function) ที่ระดับความสามารถต่างๆ วิธีการเลือกข้อสอบที่ใช้ ได้แก่

1) วิธีสุ่ม (Random) เป็นวิธีการเลือกข้อสอบ 30 ข้อ ด้วยตารางเลขสุ่มจากกลุ่มข้อสอบทั้งหมด

2) วิธีมาตรฐาน (Standard) วิธีนี้ใช้ค่าสถิติรายข้อแบบคลาสสิค คือค่าความยากและอำนาจจำแนกเป็นเกณฑ์ในการคัดเลือก ข้อสอบได้ถูกเลือกออกมาจากค่าความยากที่อยู่ในช่วง .3 ถึง .7 และข้อสอบที่เฉพาะมีค่าอำนาจจำแนกสูงในช่วง .53 และ .70 จะถูกเลือกออกมา

3) วิธีค่าความยากปานกลาง (Middle Difficulty) ข้อสอบ 30 ข้อที่ให้ค่าสารสนเทศสูงสุดที่ระดับความสามารถ(θ) เป็น 0 จะถูกเลือกออกมาจากกลุ่มข้อสอบทั้งหมด

4) วิธีช่วงบนและช่วงล่าง (Up and Down) วิธีนี้ประกอบด้วยกระบวนการ 3 ขั้นตอนที่กระทำซ้ำๆกันจนกระทั่งได้ข้อสอบ 30 ข้อที่ต้องการ ขั้นตอนแรก เกี่ยวกับการเลือกข้อสอบจากกลุ่มข้อสอบที่ให้สารสนเทศสูงสุดที่ระดับความสามารถ -1.0 และ ขั้นที่สอง เกี่ยวกับการเลือกข้อที่ให้สารสนเทศสูงสุดที่ความสามารถระดับ 0.0 และขั้นตอนที่สาม คือ การเลือกข้อสอบที่ให้สารสนเทศสูงสุดที่ระดับความสามารถ +1.0 ทั้ง 3 ขั้นตอนนี้ทำซ้ำจนได้ข้อสอบ 30 ข้อที่จะให้ได้แบบสอบที่ให้สารสนเทศสูงสุด ครอบคลุมแต่ละส่วนของสเกลความสามารถที่ซึ่งมีจำนวนผู้สอบจำนวนมากที่จะได้รับการคาดหวังให้อยู่ในตำแหน่งของเสกลความสามารถนี้

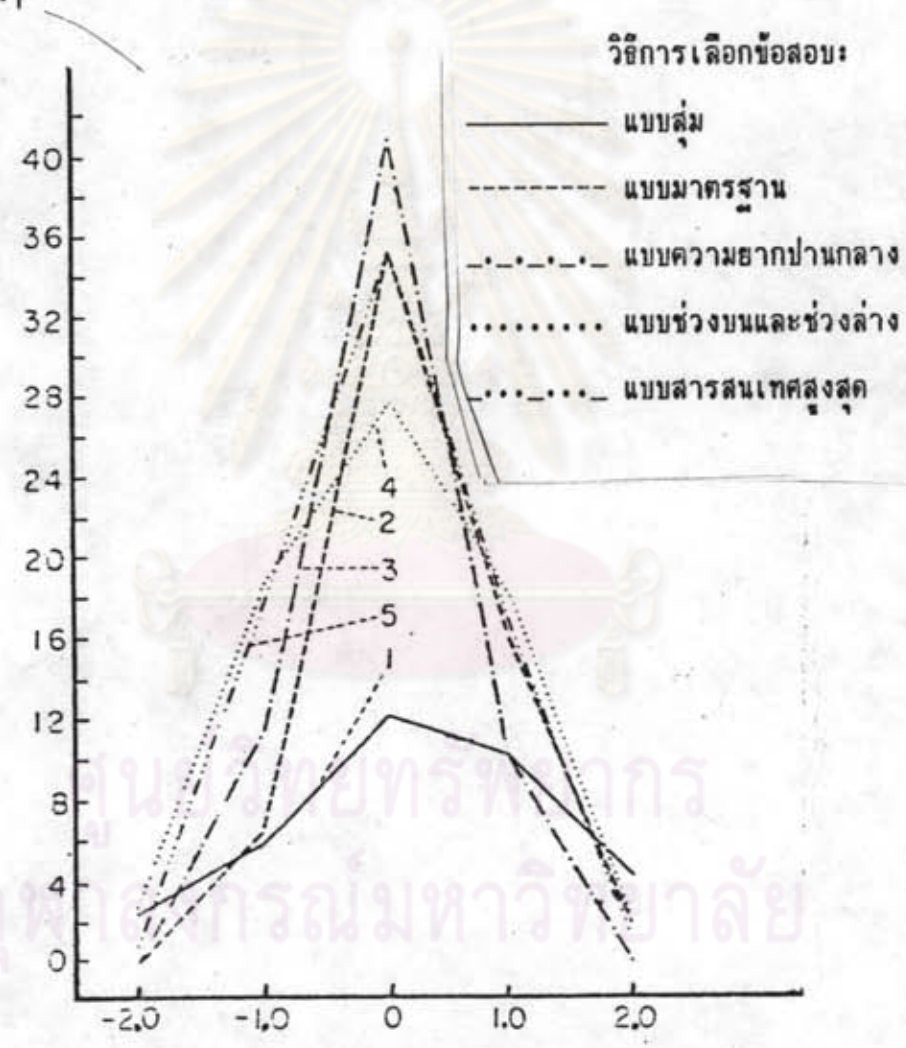
5) วิธีการให้ได้สารสนเทศสูงสุด (Maximum Information) วิธีการเลือกข้อสอบวิธีนี้จะใช้การเฉลี่ยค่าสารสนเทศที่ได้มาจากข้อสอบ 200 ข้อ ในกลุ่มข้อสอบทั้งหมดในระดับความสามารถ 3 ระดับ คือ -1.0, 0.0 และ 1.0 ข้อสอบ 30 ข้อที่ให้ค่าเฉลี่ยของปริมาณสารสนเทศสูงสุดที่ค่าความสามารถ ทั้ง 3 ระดับ จะถูกเลือกออกมา

แสดงผลของฟังก์ชั่นสารสนเทศจากวิธีการทั้ง 5 วิธี ดังภาพที่ 34

จากผลการวิจัยครั้งนี้ให้ข้อค้นพบว่าวิธีการเลือกข้อสอบจากกลุ่มข้อสอบโดยการสุ่มจะให้สารสนเทศน้อยกว่าวิธีอื่นๆ ที่ระดับความสามารถ -1, 0 และ +1 และพบว่าฟังก์ชั่นสารสนเทศของแบบสอบด้วยวิธีการนี้เป็นแบบฐานนิยมเดี่ยว (unimodal) ที่ให้สารสนเทศสูงสุดในจุดกลางของการแจกแจงความสามารถซึ่งเป็นคุณลักษณะของข้อสอบทั้งหมด

วิธีการมาตรฐานให้ฟังก์ชั่นสารสนเทศแบบสอบสูงสุดที่ความสามารถตรงจุดกลางของการแจกแจงของความสามารถซึ่งให้ปริมาณสารสนเทศสูงกว่าวิธีการสุ่ม สารสนเทศที่ได้มาที่ระดับความสามารถ +1 จะสูงกว่าวิธีสุ่ม ส่วนที่ระดับความสามารถ -1 วิธีการสุ่มและวิธีมาตรฐานมีความแตกต่างกันน้อยมาก

วิธีค่าความยากปานกลางให้ สารสนเทศสูงสุดที่ระดับความสามารถ (θ) เป็น 0 สูงกว่าวิธีอื่นๆ



(Cook and Hambleton 1987b: 246)

ภาพที่ 34 โค้งสารสนเทศแบบสอบจากวิธีการเลือกข้อสอบห้าวิธี (แบบสอบละ 30 ข้อ)

วิธีช่วงบนและช่วงล่าง ให้สารสนเทศต่ำสุดที่ระดับความสามารถ 0.0 ยกเว้นเมื่อเทียบกับวิธีสุ่ม แต่ให้สารสนเทศสูงมากกว่าวิธีอื่นๆ ที่ระดับความสามารถ +1.0 และ -1.0

วิธีการให้ได้สารสนเทศสูงสุดได้ให้ปริมาณของสารสนเทศที่น่าพอใจที่ระดับความสามารถ 3 ระดับ คือ -1.0, 0.0 และ +1.0 วิธีนี้ไม่ได้ให้สารสนเทศมากนักที่ระดับความสามารถ 0.0 เช่นเดียวกับวิธียากปานกลาง แม้ว่าจะให้สารสนเทศมากขึ้นที่ระดับความสามารถใกล้เคียงกันกับระดับความสามารถ +1 และ -1 มากกว่าวิธีอื่นๆ ซึ่งคาดหวังว่าจะเกิดกับวิธีช่วงบนและช่วงล่าง

ประเด็นที่น่าสนใจคือ ปริมาณของความซ้ำซ้อนกันที่เกิดขึ้นจากวิธีการแต่ละวิธี ปรากฏว่าวิธีสุ่มและวิธีช่วงบนและช่วงล่างมีข้อสอบที่ซ้ำกันน้อยสุด คือ 4 ข้อ ในขณะที่มีความซ้ำซ้อนกันสูงในวิธีมาตรฐานและวิธีค่าความยากปานกลางและระหว่างวิธีช่วงบนช่วงล่างกับวิธีการให้ได้สารสนเทศสูงสุดซึ่งวิธีการทั้งคู่ซ้ำซ้อนกันด้วยข้อสอบ 19 ข้อ หรือประมาณร้อยละ 63.3

การดึงข้อสอบจากกลุ่มข้อสอบตามวิธีของคุกและคณะนี้ใช้เกณฑ์ของสถิติรายข้อ ทั้งสถิติที่ได้จากการวิเคราะห์ตามแนวคลาสสิคและแนวทฤษฎีการตอบสนองต่อข้อสอบ ส่วนในวิธีการมาตรฐานและเกณฑ์สถิติรายข้อที่ได้จากการวิเคราะห์ตามแนวทฤษฎีการตอบสนองต่อข้อสอบในวิธีที่ 3, 4 และ 5 คือวิธีค่าความยากปานกลาง วิธีช่วงบนช่วงล่าง และวิธีการให้ได้สารสนเทศสูงสุดตามลำดับ ซึ่งพบถึงความซ้ำซ้อนในการเลือกข้อสอบจากเกณฑ์ ค่าสถิติแบบเดิมกับวิธีความยากปานกลางประมาณ ร้อยละ 63.3 อย่างไรก็ตามเกณฑ์ที่คุกและคณะใช้ในการเปรียบเทียบวิธีการเลือกข้อสอบได้ใช้ฟังก์ชั่นสารสนเทศแบบสอบตามแนวของทฤษฎีการตอบสนองต่อข้อสอบเป็นเกณฑ์ในการเปรียบเทียบคุณภาพของแบบสอบที่สร้างขึ้นจากวิธีการเลือกข้อสอบวิธีต่างๆ ซึ่งลอร์ด(Lord, 1974d อ้างใน Hambleton and Swaminathan 1985: 239) ได้เสนอให้ใช้ประสิทธิภาพเชิงสัมพัทธ์เป็นเกณฑ์ในการพิจารณาการวางแผนการเลือกข้อสอบ

ในปี 1980 แฮทตัน (Hatton, 1980) ได้ทำวิทยานิพนธ์ปริญญาเอกเปรียบเทียบการวิเคราะห์ข้อสอบแบบคุณลักษณะแฝง (Latent Trait) และแบบคลาสสิคในการพัฒนาแบบสอบผลสัมฤทธิ์ ได้ลงข้อสรุปว่าการใช้ทฤษฎีคุณลักษณะแฝงจะให้ค่าสถิติรายข้อที่ไม่แปรเปลี่ยน(invariant) แม้ว่าจะใช้ผู้สอบที่ต่างกันและการประเมินความสามารถของผู้สอบก็ไม่

แปรเปลี่ยนแม้ว่าจะใช้ข้อสอบที่ต่างกัน ได้ทำการศึกษาจากนักเรียนเกรด 7 จำนวน 7,460 คน ทำข้อสอบคณิตศาสตร์ 40 ข้อ และนักเรียนเกรด 9 จำนวน 6,975 คน ทำข้อสอบการอ่าน 22 ข้อ โดยสุ่มมาทำข้อสอบกลุ่มละ 2000 คน โดยมีวิธีการดังนี้

ขั้นที่ 1 เลือกข้อสอบที่มีค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนกับโมเดลของราส์ชน้อยที่สุดมาก่อน ส่วนอีกครั้งเลือกโดยโมเดลคลาสสิค

ขั้นที่ 2 ตรวจสอบความถูกต้องโดยใช้ ค่าความเที่ยง ความคลาดเคลื่อนมาตรฐานของการวัด ค่าความยาก ค่าสหสัมพันธ์แบบไบซีเรียล ประสิทธิภาพสัมพัทธ์

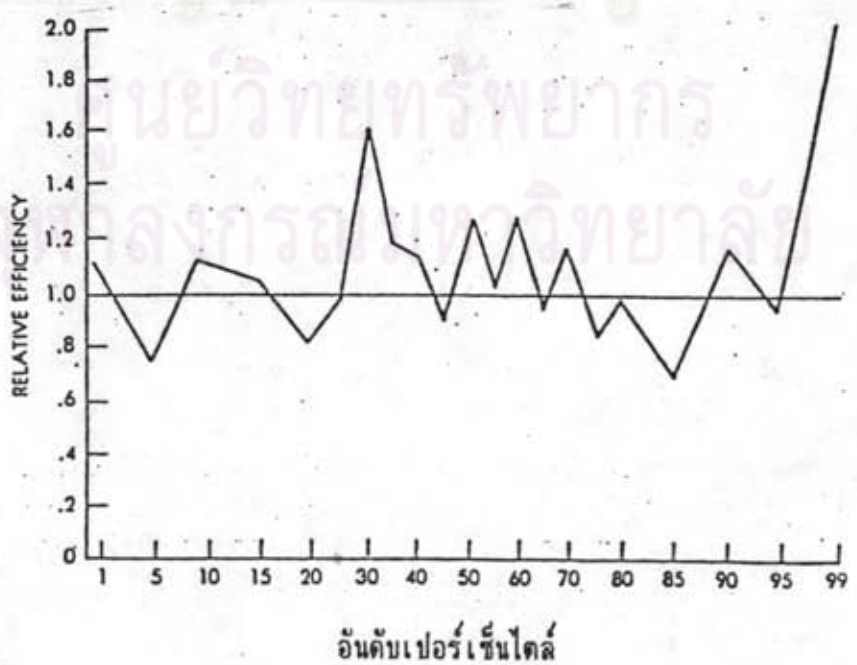
จากการศึกษาข้างต้นพบว่าทฤษฎีคุณลักษณะแฝงและทฤษฎีคลาสสิคในการพัฒนาแบบสอบผลสัมฤทธิ์ไม่แตกต่างกันในเรื่องของความแม่นยำ (precision) และประสิทธิภาพ (efficiency) แต่ทฤษฎีคุณลักษณะแฝงน่าสนใจกว่าที่ค่าพารามิเตอร์ของข้อสอบมีคุณสมบัติไม่แปรเปลี่ยนไปตามกลุ่มข้อสอบ

ในปี 1981 ฮาลาไดนา และ รอยด์ (Haladyna and Roid, 1981:1-38) ได้เปรียบเทียบวิธีการเลือกข้อสอบในการสร้างแบบสอบอิงเกณฑ์โดยวิธีการเลือกข้อสอบ 2 วิธี คือ แนวทฤษฎีคลาสสิคสุ่มจากโดเมนที่นิยามไว้ดีแล้วของข้อสอบ กับแนวทฤษฎีคุณลักษณะแฝงที่ใช้ค่าความยากของข้อสอบสอดคล้องกับระดับผลสัมฤทธิ์ของนักเรียนแล้วนำมาเปรียบเทียบค่าความคลาดเคลื่อนของการวัด ผลปรากฏว่าการใช้โมเดลราส์ชทำให้ได้ค่าความคลาดเคลื่อนของการวัดต่ำกว่า และความยาวของแบบสอบสัมพันธ์กับความคลาดเคลื่อนของการวัดไม่เป็นเชิงเส้นตรงและความคลาดเคลื่อนจะลดลงมากที่สุดระหว่างข้อสอบ 10-20 ข้อ

ในปี 1982 เบนสัน (Benson, 1982: 21-30) ได้ศึกษาเปรียบเทียบประสิทธิภาพวิธีการเลือกข้อสอบในการวัดของโมเดลหนึ่งพารามิเตอร์กับสามพารามิเตอร์โลจิสติก โดยโมเดลหนึ่งพารามิเตอร์ใช้วิเคราะห์ด้วยโปรแกรม BICAL แล้วเลือกข้อที่พิตกับโมเดลมากที่สุดหรือค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานน้อยที่สุดจำนวน 40 ข้อ ส่วนโมเดลสามพารามิเตอร์ใช้วิเคราะห์ด้วยโปรแกรม LOGIST เลือกข้อที่มีค่าอำนาจจำแนก (a) มากที่สุดมา 40 ข้อ ผลลัพธ์จากการศึกษาดูได้จากตาราง 3 จะเห็นว่าค่าเฉลี่ยของแบบสอบทั้ง 2 ต่างกันอย่างมีนัยสำคัญทางสถิติ ($p < .05$) และโมเดลหนึ่งพารามิเตอร์ง่ายกว่า แต่ค่าคลาดเคลื่อนมาตรฐานของการวัดของโมเดลหนึ่งพารามิเตอร์จะน้อยกว่า การประมาณค่าความสอดคล้องภายในใกล้เคียงกัน แต่ประสิทธิภาพสัมพัทธ์ของโมเดลสามพารามิเตอร์เปรียบเทียบกับโมเดลหนึ่งพารามิเตอร์ยังสรุปไม่ได้ชัดเจนนัก ดังภาพที่ 35

ตาราง 3 สถิติบรรยายการประมาณค่าพารามิเตอร์ข้อสอบและทั้งแบบสอบ โดยใช้โมเดลโลจิสสามพารามิเตอร์ (ค่าในวงเล็บคือส่วนเบี่ยงเบนมาตรฐาน)

	โมเดลหนึ่งพารามิเตอร์	โมเดลสามพารามิเตอร์
ค่าเฉลี่ยความยากของข้อสอบ	.01(.98)	.74(.87)
พิสัยค่าความยากของข้อสอบ	-2.36 - 1.85	-2.17 - 2.13
ค่าเฉลี่ยค่าอำนาจจำแนกของข้อสอบ	-	1.26(.37)
พิสัยค่าอำนาจจำแนกของข้อสอบ		.81 - 2.00
ค่าเฉลี่ยค่าการเดาของข้อสอบ	-	.22(.14)
พิสัยค่าการเดาของข้อสอบ		.11 - .43
ค่าเฉลี่ยของแบบสอบ (40 ข้อ)	20.53(6.09)	18.64(6.85)
ความสอดคล้องภายใน	.83	.83
ความคลาดเคลื่อนมาตรฐานของการวัด	2.51	2.82



ภาพที่ 35 ประสิทธิภาพสัมพันธ์ของโมเดลสามพารามิเตอร์เปรียบเทียบกับหนึ่งพารามิเตอร์

จะเห็นว่าตรงอันดับของเปอร์เซ็นไตล์กลางๆ ประสิทธิภาพสัมพัทธ์ของโมเดลสามพารามิเตอร์จะสูงกว่าแต่ที่เปอร์เซ็นไตล์น้อยๆและมากๆ ประสิทธิภาพสัมพัทธ์โมเดลสามพารามิเตอร์จะสูงกว่า

ในปี 1983 แฮมเบิลตัน และ เดอกรุจเตอร์ (Hambleton and De Gruijter, 1983: 355-367) ได้ศึกษาการประยุกต์ใช้โมเดลการตอบสนองต่อข้อสอบในการเลือกข้อสอบเพื่อสร้างแบบสอบแบบอิงเกณฑ์ และได้สรุปว่าถ้าคลังข้อสอบมีค่าพารามิเตอร์ของข้อสอบใกล้เคียงกัน (homogeneous) เหมาะในการใช้เลือกข้อสอบโดยวิธีการสุ่ม ในการเลือกข้อสอบแบบอิงเกณฑ์จะเลือกข้อสอบที่มีค่าอำนาจจำแนกได้ดีที่คะแนนจุดตัด (cutoff score) ของเสกลคะแนนโดเมน (domain score scale) สำหรับคลังข้อสอบที่มีค่าพารามิเตอร์ของข้อสอบต่างกันมากๆ (heterogeneous) ความตรงในการจำแนกกลุ่มผู้สอบรอบรู้กับไม่รอบรู้ต่ำเพราะว่าข้อสอบบางข้อมีฟังก์ชั่นที่เหมาะสมไม่ตรงกับคะแนนจุดตัด ตัวอย่างเช่น ข้อสอบอาจจะยากมากหรือง่ายมาก หรือมีค่าอำนาจจำแนกต่ำถูกเลือกขึ้นมาแต่อย่างไรก็ตามข้อสอบที่ไม่ดีพวกนี้อาจจะเหมาะสมกับคะแนนจุดตัดที่แตกต่างออกไปก็ได้ และได้แยกอธิบายเป็น 3 ประเด็น คือ

1) ความไม่เหมาะสมของค่าสถิติของข้อสอบแบบคลาสสิค ค่าสถิติของข้อสอบแบบคลาสสิคก็คือ ค่าความยาก (p) และค่าอำนาจจำแนก (r) ค่าความยากจะมีค่าบนเสกล [0,1] และนิยามจากประชากรผู้สอบ คะแนนโดเมนของผู้สอบ (τ) ก็มีค่าบนเสกล [0,1] แต่นิยามจากประชากรข้อสอบ ดังนั้นธรรมชาติของการอ้างอิงไปยังทั้ง 2 เสกลแตกต่างกันโดยสิ้นเชิง

2) โมเดลการตอบสนองต่อข้อสอบในการเลือกข้อสอบ จุดเด่นก็คือค่าสถิติของข้อสอบอยู่บนเสกลเดียวกันระหว่างความสามารถของผู้สอบและคะแนนจุดตัดที่กำหนด ดังนั้นสามารถเลือกข้อสอบที่ค่าอำนาจจำแนกสูงสุดที่บริเวณคะแนนจุดตัด

3) วิธีการเลือกข้อสอบที่เหมาะสมมีขั้นตอน ดังนี้

1) เตรียมคลังข้อสอบขนาดใหญ่ที่มีข้อสอบที่มีความตรง

2) หาพารามิเตอร์ของข้อสอบด้วยกลุ่มตัวอย่างขนาดใหญ่

3) ทดสอบความพิดของข้อสอบกับโมเดล เลือกข้อสอบที่พิดกับโมเดล

4) เลือกคะแนนจุดตัดและค่าที่อยู่ในขอบเขตของเสกลของคะแนนโดเมน

5) แปลง π_1, π_2 และ π_n ไปเป็น θ_1, θ_2 และ θ_n ตามลำดับ

$$\pi = \frac{1}{m} \Sigma p_i(\theta)$$

6) กำหนดค่า p^*

$$p_m = \max(p_1, p_n) < p^*$$

7) เลือกข้อสอบที่ให้ค่าสารสนเทศสูงสุดที่ θ_2

8) แปลง θ_1, θ_2 และ θ_n ไปเป็น π_1^*, π_2^* และ π_n^* ตามลำดับ โดยใช้โค้งคุณลักษณะของแบบสอบ (Test Characteristic Curve) ประกอบในการเลือกข้อสอบ k ข้อ คำนวณ p_1, p_n และ p_m กับจุดตัดหลายๆจุดพิจารณาตัวเลขจำนวนเต็มที่ใกล้ค่า $\Sigma p_i(\theta_2)$

9) ถ้า $p_m > p^*$ เลือกข้อสอบข้อต่อไปที่ทำให้ $I(\theta_2)$ มากที่สุด

10) ถ้า $p_m < p^*$ หยุดเลือกเพราะว่าถึงเกณฑ์ที่กำหนดแล้ว

และทำการศึกษาถึงการเลือกข้อสอบที่คลังข้อสอบเป็นแบบเอกพันธ์ใช้ค่า b ค่าความยากเป็นตัวแปร ส่วนคลังข้อสอบที่เป็นวิวิธพันธ์ต้องใช้ทั้งค่า b และ a เพราะค่าความยากอย่างเดียวจะไม่ทำให้ได้ $I_i(\theta)$ ที่ต้องการไว้ได้ทั้งหมด

ในปี 1983 ฮาร์เวล (Harwell, 1983 : 129) ทำวิทยานิพนธ์ปริญญาเอกเปรียบเทียบวิธีการเลือกข้อสอบ 2 วิธีในการวัดแนวอิงเกณฑ์ โดยมีวิธีการเลือกข้อสอบแบบการแปลงคะแนนอย่างง่าย (simple change score) และวิธีคุณลักษณะแฝง (Latent Trait Method) แล้วใช้ค่าความเที่ยงของแบบสอบเป็นเกณฑ์ ข้อมูลที่ใช้ได้จากการจำลองแบบซึ่งเปรียบเทียบภายใต้เงื่อนไขที่เกี่ยวกับสถานการณ์ในชั้นเรียน ได้แก่ จำนวนข้อ จำนวนวิชา ความแตกต่างของค่าเฉลี่ยและจุดตัดแล้ว ทำการเลือกข้อสอบและตรวจสอบความเที่ยงจนได้ความเที่ยงสูงสุด พบว่าวิธีการเลือกข้อสอบทั้ง 2 วิธีให้ค่าความเที่ยงแตกต่างกันอย่างมีนัยสำคัญทางสถิติ

ในปี 1983 ออยส์เชอร์ (Oescher, 1983) ได้ทำวิทยานิพนธ์ปริญญาเอกเรื่องการวิเคราะห์เปรียบเทียบวิธีการเลือกข้อสอบ 3 วิธี โดยเน้นการพัฒนาแบบสอบผลสัมฤทธิ์ที่

เหมาะกับการในไปใช้ตามจุดมุ่งหมายที่ต่างกัน แบบสอบ 3 ชุดย่อย ชุดละ 30 ข้อ ถูกเลือกจากคลังข้อสอบ จำนวน 110 ข้อ เนื้อหาเกี่ยวกับการบวกและการลบ นำไปใช้กับนักศึกษามหาวิทยาลัยชั้นปีที่ 1 จำนวน 371 คน โดย

แบบสอบชุดที่ 1 ใช้วิธีโมเดลคลาสสิคกำหนดเกณฑ์ในการเลือกข้อสอบโดยค่าอำนาจจำแนกสูงสุด

แบบสอบชุดที่ 2 ใช้เกณฑ์คุณลักษณะแฝงโดยใช้ค่าความยากโมเดลของราส์ชที่อธิบายค่าความยากได้ดีที่สุดคือฟิตกับโมเดลมากที่สุด

แบบสอบชุดที่ 3 ใช้วิธีการสุ่ม

แล้วนำแบบสอบทั้ง 3 ชุดมาเปรียบเทียบโดยกำหนดนักศึกษาลงบนเสกลผลสัมฤทธิ์และความเป็นตัวแทนของโดเมนของคลังข้อสอบเมื่อเปลี่ยนแปลงกลุ่มผู้สอบ ผลปรากฏว่าทั้งชุดที่ 1 และ 2 ให้ผลต่างจากชุดที่ 3 ในการนำไปสอบกับนักเรียนต่างกลุ่ม ส่วนความเป็นตัวแทนของโดเมนแตกต่างกันในกลุ่มผู้สอบที่มีความสามารถต่ำ

ในปี 1984 ซิลวา (Silva, 1984: 155) ทำวิทยานิพนธ์ปริญญาเอกเปรียบเทียบวิธีการเลือกข้อสอบแนวคลาสสิคและแนวการตอบสนองต่อข้อสอบสำหรับแบบสอบอิงเกณฑ์ โดยแนวคลาสสิค มี 3 วิธี คือ ดัชนีคอกซ์และเวกัส (Cox-Vages Index) สหสัมพันธ์พอยท์ไบซีเรียล และสัมประสิทธิ์ฟี ส่วนแนวการตอบสนองต่อข้อสอบ มี 2 วิธี คือ การประมาณค่าสารสนเทศของข้อสอบแบบโมเดลหนึ่งพารามิเตอร์ และสามพารามิเตอร์ โดยมีพลวิจัยคือ นักเรียนเกรด 1 และ 2 จำนวน 945 คน เกรด 4-6 จำนวน 1,796 คน แบบสอบที่ใช้เป็นแบบสอบการอ่าน จากการวิเคราะห์ความแปรปรวนพบว่าโมเดลการตอบสนองต่อข้อสอบจะจำแนกการเรียนรู้ของนักเรียนได้ดีกว่าโมเดลคลาสสิค และยังพบว่าเมื่อทำการประมาณค่าพารามิเตอร์ (calibration) โมเดลสามพารามิเตอร์จะให้ค่าอำนาจจำแนกดีกว่าและค่าความยากเหมาะสมกว่า และมีข้อสอบเพียงจำนวนน้อยที่ไม่ฟิตกับโมเดล

ในปี 1987 แฮมเบิลตัน (Hambleton, 1987) ได้เสนอวิธีการเลือกข้อสอบที่เหมาะสมในการประชุมประจำปีของสมาคมวิจัยการศึกษาแห่งอเมริกา โดยเปรียบเทียบวิธีการเลือกข้อสอบแนวทฤษฎีการตอบสนองต่อข้อสอบ 2 วิธี คือ วิธีออฟทิมอล (Optimal) และ วิธีออฟทิมอลตามเนื้อหา (Content Optimal) กับแนวที่ไม่ใช่ทฤษฎีการตอบสนองต่อ

ข้อสอบ 2 วิธี คือ แบบสุ่ม และแบบคลาสสิค โดยทำการทดลองกับข้อสอบเพื่อการออกวุฒิบัตร 4 แบบ ใช้ข้อสอบ 20 ข้อ จากคลังข้อสอบจำนวน 250 ข้อ มีเนื้อหาเกี่ยวกับสุขภาพ แล้วพิจารณาที่จุดตัด 3 ระดับ คือ 65% 70% และ 75% ผลปรากฏว่า วิธีออฟทิมอลจะให้สารสนเทศใกล้เคียงกับจุดตัดมากกว่าแบบสุ่ม วิธีออฟทิมอลตามเนื้อหาใกล้เคียงกับออฟทิมอล และวิธีแนวทฤษฎีการตอบสนองต่อข้อสอบสามารถปรับปรุงความถูกต้องในการตัดสินใจได้มากกว่า

งานวิจัยที่เกี่ยวข้องกับการเลือกข้อสอบที่ได้ดำเนินการไปในระยะเริ่มแรกที่กล่าวมานี้ส่วนใหญ่ ได้ทำการเปรียบเทียบวิธีการเลือกข้อสอบตามแนวคลาสสิคโดยใช้ค่าความเที่ยงเป็นเกณฑ์ในการเปรียบเทียบ เช่น งานวิจัยของเนโว (1977: 847-852) ในระยะต่อมาได้มีการเปรียบเทียบวิธีการเลือกข้อสอบโดยใช้ทั้งตามแนวคลาสสิคและตามแนวทฤษฎีการตอบสนองต่อข้อสอบ เช่น งานวิจัยของฮาลาไดนา และ รอยด์ (1981: 1-38) ออยส์เซอร์ (1983) ชิลวา(1984: 155) คุก และแอมเบิลตัน (1985: 240-253) และ แอม เบิลตัน (1987) เป็นต้น

วิธีการเลือกข้อสอบที่นำมาใช้ในการเปรียบเทียบตามแนวคลาสสิค ส่วนใหญ่จะใช้การเลือกข้อสอบตามค่าอำนาจจำแนกที่คำนวณจากวิธีสหสัมพันธ์แบบพอยท์ไบซีเรียล (Point -Biserial Correlation) วิธีสหสัมพันธ์แบบไบซีเรียล (Biserial Correlation) โดยใช้ค่าสูงสุด นอกจากนี้ยังมีการเลือกข้อสอบตามค่าความยาก วิธีมาตรฐานและวิธีการสุ่ม โดยใช้เกณฑ์ในการเปรียบเทียบคือ ค่าความเที่ยงทั้งแบบความสอดคล้องภายในและแบบสอบซ้ำ และดัชนีความไว (Sensitivity Index) การเลือกข้อสอบตามแนวทฤษฎีคลาสสิคที่กล่าวมามีจุดอ่อนหลายประการได้แก่ ค่าสถิติของข้อสอบแปรเปลี่ยนไปตามกลุ่มผู้สอบ ซึ่งทำให้ไม่แน่ใจว่าค่าสถิติที่จะนำไปใช้เป็นค่าสถิติที่ถูกต้อง ค่าความยาก (p) มีค่าอยู่บนเสกล [0, 1] โดยนิยามจากประชากรผู้สอบ ส่วนคะแนนของผู้สอบ (x) ก็มีค่าอยู่บนเสกล [0, 1] แต่นิยามจากประชากรข้อสอบ ดังนั้นการที่จะอ้างอิงไปถึง 2 เสกลย่อมเป็นไปไม่ได้ (Hambleton and De Gruijter, 1983: 355-367) จึงทำให้การเลือกข้อสอบในแนวทฤษฎีคลาสสิคไม่เหมาะที่จะนำไปใช้กับกลุ่มผู้สอบที่มีความสามารถต่างกันและเป็นไปได้ยากที่จะใช้กับข้อสอบแบบอิงเกณฑ์ จากการพัฒนาการของทฤษฎีการตอบสนองต่อข้อสอบจะทำให้ได้ค่าสถิติประจำข้อสอบที่คงที่แม้ว่า

จะประมาณจากผู้สอบที่ต่างกลุ่มกัน และการประมาณความสามารถของผู้สอบก็คงที่แม้ว่าใช้ข้อสอบต่างชุดกัน (Hatton, 1980) ทำให้ทฤษฎีการตอบสนองต่อข้อสอบเป็นทฤษฎีที่สามารถนำไปใช้เป็นพื้นฐานในการพัฒนาเกณฑ์ในการเลือกข้อสอบได้ และจากการศึกษาที่ผ่านมาไม่ว่าจะเป็นของคุก และ แฮมเบิลตัน (1978) แฮทตัน (1980) ฮาลาไดนา และรอยด์ (1981) ออยส์เซอร์ (1983) และ ชิลวา (1984) ซึ่งทุกคนก็ได้เปรียบเทียบวิธีการเลือกข้อสอบตามแนวทฤษฎีคลาสสิคและทฤษฎีการตอบสนองต่อข้อสอบ และลงข้อสรุปว่าการเลือกข้อสอบแนวทฤษฎีการตอบสนองต่อข้อสอบให้ประสิทธิภาพเท่าเทียมกันหรือสูงกว่าการเลือกข้อสอบตามแนวทฤษฎีคลาสสิค

ส่วนในแนวทฤษฎีการตอบสนองต่อข้อสอบ วิธีการเลือกข้อสอบที่ใช้นำมาเปรียบเทียบได้แก่ วิธีการเลือกตามค่าความยาก (b) ปานกลาง เลือกข้อที่มีค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานจากโมเดลน้อยที่สุด วิธีช่วงบนและช่วงล่าง วิธีการให้ได้ค่าสารสนเทศสูงสุดสำหรับเกณฑ์ที่ใช้ในการเปรียบเทียบ คือประสิทธิภาพเชิงสัมพัทธ์ของโค้งสารสนเทศของแบบสอบ ทฤษฎีการตอบสนองต่อข้อสอบที่แพร่หลายอยู่ในปัจจุบันมีหลายโมเดลได้แก่ โมเดลโลจิสหนึ่งพารามิเตอร์ แสดงความสัมพันธ์ในรูปฟังก์ชั่นของโอกาสที่ผู้มีความสามารถระดับใดๆ ทำข้อสอบถูกขึ้นอยู่กับค่าความยากของข้อสอบเพียงอย่างเดียว ส่วนโมเดลโลจิสสองพารามิเตอร์ แสดงความสัมพันธ์ในรูปฟังก์ชั่นของโอกาสที่ผู้มีความสามารถระดับใดๆทำข้อสอบถูกขึ้นอยู่กับค่าความยากและค่าอำนาจจำแนกของข้อสอบ สำหรับโมเดลโลจิสสามพารามิเตอร์ แสดงความสัมพันธ์ในรูปฟังก์ชั่นของโอกาสที่ผู้มีความสามารถระดับใดๆทำข้อสอบถูกขึ้นอยู่กับค่าความยากและค่าอำนาจจำแนกและค่าการเดาของข้อสอบ การอธิบายคุณลักษณะของข้อสอบได้ด้วยพารามิเตอร์หลายๆตัวย่อมจะให้ค่าที่เหมาะสมได้มากกว่า จากการศึกษาของเบลสัน (1982: 21-30) เปรียบเทียบประสิทธิภาพของแบบสอบที่เลือกข้อสอบจากโมเดลโลจิสหนึ่งพารามิเตอร์และโมเดลโลจิสสามพารามิเตอร์ ยังไม่ได้ข้อสรุปที่ชัดเจนแต่พอจะเห็นว่าที่ความสามารถระดับกลางๆของกลุ่ม ประสิทธิภาพเชิงสัมพัทธ์ของการเลือกข้อสอบโดยใช้โมเดลโลจิสสามพารามิเตอร์สูงกว่า ประกอบกับข้อสอบวัดผลสัมฤทธิ์ที่ผู้วิจัยมุ่งนำผลการวิจัยไปใช้เป็นข้อสอบแบบเลือกตอบจำนวนตัวเลือกน่าจะมีผลต่อโอกาสในการเดาคำตอบ จึงใช้โมเดลโลจิสสามพารามิเตอร์ที่มีค่าพารามิเตอร์การเดามาใช้ในการศึกษา

จากการศึกษาเกี่ยวกับจำนวนข้อสอบของ ฮาลาไดนา และ รอยด์ (Haladyna and Roid, 1981:1-38) พบว่าแบบสอบที่มีข้อสอบในช่วง 10-20 ข้อ จะมีค่าเฉลี่ยของความคลาดเลื่อนมาตรฐานแตกต่างกันมากที่สุด

จากที่กล่าวมาข้างต้นจะเห็นว่าส่วนใหญ่การวิจัยจะเน้นไปในเรื่องการหาวิธีการเลือกข้อสอบให้เหมาะกับจุดประสงค์ของการนำข้อสอบไปใช้ เช่น ออยส์เซอร์ (1983) แต่ก็ยังไม่ได้ข้อสรุปที่ชัดเจน ในเรื่องจำนวนข้อสอบ เนลล์ และ แจคสัน (1970) ได้สรุปว่าไม่ว่าจะใช้วิธีการเลือกข้อสอบแบบใดถ้าใช้จำนวนข้อสอบมากขึ้นก็จะมีคุณภาพใกล้เคียงกัน โดยเฉพาะเมื่อแบบสอบประกอบด้วยข้อสอบตั้งแต่ 40 ข้อขึ้นไป และงานวิจัยของ ฮาลาไดนา และรอยด์ (1981) ได้เสริมว่าช่วงการเพิ่มของจำนวนข้อสอบระหว่าง 10 ถึง 20 ข้อ จะให้ค่าความคลาดเคลื่อนในการวัดลดลงมากที่สุด ส่วนในเรื่องทฤษฎีการสอบที่นำมาใช้ค่อนข้างชัดเจนว่าทฤษฎีการตอบสนองต่อข้อสอบให้ค่าคุณลักษณะของข้อสอบที่ไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ แม้ว่าการศึกษาของ เบนสัน (1982) ยังสรุปผลข้อค้นพบไม่ได้แน่นอน แต่แฮทตัน (1980) ซิลวา (1984) และแฮมเบิลตัน (1987) สรุปตรงกัน และเกณฑ์ที่ชัดเจนในการพิจารณาคุณภาพของแบบสอบก็คือค่าสารสนเทศของข้อสอบ คุก และแฮมเบิลตัน (1978) พบว่าการเลือกข้อสอบโดยสุ่มจะให้ค่าสารสนเทศต่ำที่สุด แม้ว่าการวิจัยที่ผ่านมาได้มีการศึกษาเกี่ยวกับวิธีการคัดเลือกข้อสอบเป็นจำนวนมากไม่ว่าจะอิงทฤษฎีคลาสสิคหรือทฤษฎีการตอบสนองต่อข้อสอบแต่ก็ยังไม่ได้ข้อค้นพบที่เด่นชัดแต่ประการใดจุดอ่อนของวิธีการศึกษาที่ผ่านมาอาจจะเป็นเพราะไม่ได้นำองค์ประกอบด้านความสามารถของผู้สอบเข้าไปพิจารณาเพราะว่าการคัดเลือกข้อสอบโดยวิธีใดๆ ที่จะทำให้เหมาะกับการนำไปใช้กับผู้สอบทุกระดับความสามารถนั้นย่อมเป็นไปได้ยาก งานวิจัยที่ผ่านมาที่ได้กล่าวถึงความสัมพันธ์ระหว่างวิธีการเลือกข้อสอบกับความสามารถของผู้สอบได้กล่าวไว้โดยอ้อม เช่น งานวิจัยของฮาลาไดนา และ รอยด์ (1981) ที่ใช้วิธีการเลือกข้อสอบที่มีค่าความยากในทฤษฎีการตอบสนองต่อข้อสอบสอดคล้องกับระดับผลสัมฤทธิ์ทางการเรียนของนักเรียน และงานวิจัยของ แฮมเบิลตัน และ เดอกรุจเตอร์ (1983) ที่ได้แนะนำว่าข้อสอบที่ไม่ดีสำหรับคะแนนจุดตัดบางจุดอาจจะเหมาะสมกับคะแนนจุดตัดที่แตกต่างออกไปก็ได้ ผู้วิจัยจึงสนใจที่จะพัฒนาเกณฑ์ในการเลือกข้อสอบที่เหมาะกับความสามารถของผู้สอบโดยคำนึงถึงความแตกต่างของบุคคลนั่นคือข้อสอบบางข้อเหมาะกับผู้สอบบางคน ซึ่งผู้วิจัยคาดว่าจะได้เกณฑ์ที่ชัดเจนในการเลือกข้อสอบที่เหมาะสมกับผู้สอบต่อไป