บทที่ 2

วรรณคดีที่เกี่ยวข้อง

สำหรับวรรณคดีที่เกี่ยวข้องในงานวิจัยฉบับนี้ ผู้วิจัยได้นำเสนอไว้ 2 ประเด็นหลัก ดังนี้

1. มโนทัศน์ที่เกี่ยวกับแบบสอบอิงโดเมน
2. มโนทัศน์เกี่ยวกับโมเดลโลจีสติก 3 พารามิเตอร์
3. การใช้โมเดลโลจีสติกในการพัฒนาแบบสอบอิงโดเมน
4. งานวิจัยที่เกี่ยวข้อง

1. มโนทัศน์เกี่ยวกับแบบสอบอิงโดเมน

1.1 ความเป็นมาและแนวคิดทั่วไปเกี่ยวกับแบบสอบอิงโดเมน

จากจุดอ่อนของแบบสอบอิงกลุ่ม ซึ่งสร้างขึ้นด้วยวัตถุประสงค์ที่จะวัดความรู้ความสามารถของผู้สอบโดยการนำไปเปรียบเทียบกันภายในกลุ่ม ดังนั้นในการแปลความหมายของคะแนนจากแบบสอบอิงกลุ่ม จึงไม่ช่วยให้ทราบถึงระดับความสามารถที่แท้จริงของผู้สอบ และในการประเมินผลสัมฤทธิ์ทางการเรียนในปัจจุบันมีความต้องการที่จะบ่งชี้ถึงระดับผลสัมฤทธิ์ทางการเรียนหรือศักยภาพของผู้เรียนแต่ละคนในเรื่องใดเรื่องหนึ่งว่ามีอยู่ในระดับใด อันเป็นข้อมูลที่จำเป็นต่อการปรับปรุงการเรียนการสอน ดังนั้นนักวัดผลจึงได้เสนอแนวคิดที่ต่างไปจากแนวคิดของแบบสอบอิงกลุ่ม โดยเริ่มจากแนวคิดของฟลานากาน ในปี 1951 (Flanagan 1951 อ้างถึงใน พิศิษฐ ตัณฑวณิช 2529 : 27) ที่ว่า

"คะแนนจากแบบสอบจะมีความหมายและมีคุณค่าก็ต่อเมื่อคะแนนเหล่านั้นสามารถแปลความหมายได้ในเทอมของศักยภาพ ความหมาย และความสำเร็จในกิจกรรมของการศึกษาที่มีความสำคัญ"

ต่อมาในปี 1963 เกลเซอร์ (Glaser 1963 : 519 ได้เสนอแนวคิดเกี่ยวกับ "การวัดในระบบอิงเกณฑ์" โดยให้ความหมายไว้ว่า

"...ถ้าการทดสอบจัดกระทำเพื่อบ่งชี้ถึงระดับความสามารถของผู้สอบว่าอยู่ตรงจุดใดในช่วงของความต่อเนื่องของความรู้ รวมทั้งมีการตรวจสอบว่า ผู้สอบมีความสามารถในระดับใดเมื่อเทียบกับเกณฑ์มาตรฐานที่กำหนดขึ้นก็จะเรียกระบบดังกล่าวนี้ได้ว่าเป็นระบบของการวัดผลแบบอิงเกณฑ์"

และในปี 1978 โปแฟม (Popham 1978 : 93) ได้ให้คำจำกัดความของแบบสอบอิงเกณฑ์ว่า หมายถึง

"แบบสอบที่ใช้เพื่อค้นหาสมรรถภาพของบุคคลภายใต้ขอบเขตของเนื้อหาที่ได้รับการกำหนดชัดแจ้ง"

จะเห็นว่าประเด็นสำคัญของแนวคิดที่นักวัดผลได้เสนอไว้ดังกล่าวก็คือ ต้องการให้การแปลความหมายของคะแนนที่ได้จากการวัดผลสัมฤทธิ์ทางการเรียนเป็นไปในลักษณะของการนำมาเทียบกับเกณฑ์มาตรฐาน โดยแฮมเบิลตัน และสวามินาธาน (Hambleton and Swaminathan 1978 : 3) ได้กล่าวสรุปถึงคำว่า "เกณฑ์" ว่าอาจหมายถึง มาตรฐานการกระทำ (performance standard)ระดับความรอบรู้ขั้นต่ำ (minimum proficiency level) หรือ คะแนนจุดตัด (cut-off score) เป็นต้น ซึ่งต่างไปจากแนวคิดเดิมที่เน้นการเปรียบเทียบกันภายในกลุ่ม

มิลแมน (Millman 1974 : 17-21) กล่าวว่า ในหลาย ๆ กรณีของการวัดแบบอิงเกณฑ์ ถ้าจุดเน้นของการพัฒนาแบบสอบอยู่ที่การกำหนดขอบเขตภาระงานไว้ชัดแจ้งและการสุ่มตัวอย่างข้อสอบให้มีความเป็นตัวแทนของโดเมนนั้น จะเรียกแบบสอบนี้ว่า แบบสอบอิงโดเมน

บุคคลที่ได้รับการยอมรับว่าเป็นผู้บุกเบิกเรื่องของการทดสอบอิงโดเมน ทั้งในเชิงนิยามและระบบการพัฒนาแบบสอบอิงโดเมน คือ ไฮฟ์ลี (Well Hively) แห่งมหาวิทยาลัยมินเนโซต้า โดยในปี 1966 ไฮฟ์ลีได้เริ่มโครงการ MINNEMAST ซึ่งเป็นโครงการทดสอบอิงโดเมนในวิชาวิทยาศาสตร์ขึ้น ทั้งนี้ไฮฟ์ลีได้ชี้ถึงปัญหาของการนำวัตถุประสงค์เชิงพฤติกรรม (Behavioral Objectives : BO) ไปใช้ในการเรียนการสอนว่า อาจบ่งไปที่พฤติกรรมปลีกย่อย ไม่ครอบคลุมเนื้อหาและไม่ให้ความหมายในเชิงอ้างอิงไปสู่สัดส่วนของเนื้อหาทั้งหมดแต่อย่างใด และจากปัญหาของแบบสอบมาตรฐานที่พัฒนาขึ้นโดยสถาบันต่าง ๆ เช่น

016019

ETS (Educational Testing Service) ซึ่งเป็นแบบสอบที่ไม่อิงหลักสูตร ทำให้ไม่สามารถบอกถึงผลสัมฤทธิ์ทางการเรียนที่จัดขึ้นตามหลักสูตรหนึ่งหลักสูตรใดได้ และนอกจากนี้ยังมีปัญหาของแบบสอบอิงกลุ่มที่ไม่ช่วยให้ทราบถึงระดับความสามารถที่แท้จริงของผู้สอบ และในการใช้คำว่า "เกณฑ์" ของแบบสอบอิงเกณฑ์ ก็อาจทำให้แปลความหมายผิดไปได้ กล่าวคือ จะให้ความสำคัญกับเกณฑ์มาตรฐาน โดยไม่ต้องคำนึงถึงการกำหนดขอบเขตโดเมน และการสุ่มข้อสอบให้มีความเป็นตัวแทนของโดเมนดังกล่าว ด้วยเหตุผลต่าง ๆ ดังกล่าว ไฮฟ์ลีจึงได้เสนอแนะระบบการทดสอบแบบอิงโดเมนขึ้น ทำให้มีโครงการทดสอบอิงโดเมนอย่างกว้างขวางในมลรัฐมินเนโซต้า

ไฮฟ์ลี (Hively 1974 : 5-10) ได้กล่าวว่า ทฤษฎีเบื้องหลังของการทดสอบอิงโดเมน คือ ทฤษฎีการเรียนรู้ (Learning Theory) ผลการสอบอิงโดเมนมุ่งใช้เพื่อประเมินความก้าวหน้าในการเรียน และได้กล่าวถึงจุดเน้นของแบบสอบอิงโดเมนว่าการสร้างแบบสอบในแนวนี้ ต้องอยู่ภายใต้เงื่อนไขที่ว่า ขอบเขตเนื้อหาหรือภาระงานทั้งหมดที่ต้องการสอบวัด จะต้องได้รับการกำหนดอย่างชัดแจ้ง และข้อสอบในแบบสอบที่สร้างขึ้นตามโดเมนหนึ่ง ๆ จะต้องได้รับการสุ่มขึ้นมาจากขอบเขตเนื้อหา หรือภาระงานดังกล่าว เพื่อให้มีความเป็นตัวแทนอย่างแท้จริง และในการแปลความหมายของคะแนนสอบ จะใช้จำนวนข้อที่ตอบถูกในกลุ่มตัวอย่างของโดเมนนี้มาคำนวณหาค่าอัตราส่วนที่แท้จริงเมื่อเทียบกับพฤติกรรมทั้งหมดในโดเมน ซึ่งในประเด็นของการสร้างข้อสอบให้มีความเป็นตัวแทนตามโดเมนที่กำหนดขึ้นนั้น ไฮฟ์ลี ได้เสนอแนะเทคนิคกำหนดรูปแบบข้อสอบ (Item Form) อันเป็นเทคโนโลยีในการเขียนข้อสอบอย่างหนึ่งที่ไฮฟ์ลีได้เริ่มคิดค้นขึ้น ซึ่งจะได้เสนอในตอนต่อไป

นอกจากนี้ ก็มีนักวัดผลอีกหลายท่านที่ได้กล่าวถึงแบบสอบอิงโดเมน และนิยามคำว่า "โดเมน" ไว้ เช่น

เกลเซอร์ และ นิตโก้ (Glaser and Nitko 1971 : 653) ได้ให้ความหมายของแบบสอบอิงโดเมนว่า เป็นแบบสอบที่สร้างขึ้นเพื่อให้สามารถแปลผลของการวัดออกมาในรูปของมาตรฐานการกระทำที่เฉพาะเจาะจง ซึ่งทำได้โดยให้นิยามโดเมนของงานที่ผู้เรียนควรทำได้ จากนั้นนำกลุ่มตัวอย่างที่เป็นตัวแทนของงานในโดเมนนี้มาจัดเป็นแบบสอบผลจากการสอบจะแสดงถึงพฤติกรรมของผู้สอบแต่ละคน โดยอ้างอิงไปยังโดเมนที่ได้กำหนดไว้

เชคอฟท์ (Shaycoft 1979 : 4) ได้แบ่งแบบสอบอิงเกณฑ์ออกเป็น 2 ชนิดใหญ่ ๆ คือ แบบสอบอิงโดเมน และแบบสอบอิงจุดประสงค์ โดยได้กล่าวถึงแบบสอบอิงโดเมนว่า คะแนนทั้งหมดจากการสอบแบบสอบอิงโดเมน มีความหมายสมบูรณ์ในการแสดงถึงสัดส่วนของโดเมนที่ผู้สอบได้รอบรู้ การวัดแบบนี้เหมาะสมที่สุด เมื่อเนื้อหาที่จะวัดเป็นโดเมนที่ได้กำหนดไว้ชัดเจนแล้ว จำนวนเนื้อหามีขอบเขตจำกัด และกรอบของตัวอย่าง (sampling frame) มีครบทุกเนื้อหา ดังนั้นเมื่อจะทดสอบก็สามารถเอาตัวอย่างเนื้อหานั้นมาสอบได้

แฮมเบิลตัน และ สวามินาธาน (Hambleton and Swaminathan 1978 : 3) ได้ให้ข้อสังเกตว่าแบบสอบอิงเกณฑ์ที่มีใช้กันทั่วไปในปัจจุบันนั้น จัดเป็นประเภทแบบสอบอิงจุดประสงค์ ซึ่งเป็นแบบสอบที่ประกอบด้วยกลุ่มข้อสอบที่จับคู่กับวัตถุประสงค์ ไม่มีการกำหนดโดเมนของพฤติกรรมไว้ และไม่ได้พิจารณาถึงความเป็นตัวแทนของโดเมนพฤติกรรมนั้น ๆ ด้วย และแบบสอบอิงเกณฑ์ตามนิยามของโปแฟม (Popham 1978 : 13) มีลักษณะคล้ายคลึงกับแบบสอบอิงโดเมน

เบเคอร์ (Baker 1974 : 11) ได้อธิบายถึงคำว่า โดเมน ว่าประกอบด้วยกลุ่มย่อยของความรู้ ทักษะความเข้าใจ หรือทัศนคติ ซึ่งเป็นคุณลักษณะที่สำคัญตามเนื้อหาที่ซึ่งคาดหวังให้นักเรียนได้รับ

จอห์นสัน (Johnson, T.J. 1974 : 44-46) ได้จำแนกประเภทของ "โดเมน" ในการประเมินแบบอิงโดเมนไว้ 5 ประเภทคือ

1. โดเมนของวัตถุประสงค์ทางการศึกษา แบ่งเป็นวัตถุประสงค์ในด้านกลุ่มประสบการณ์ จิตพิสัย หรือ พุทธิพิสัย เป็นต้น และวัตถุประสงค์ในแง่ของทักษะ

2. โดเมนของสารสนเทศ เช่น เซตของความรู้ที่สัมพันธ์กับวัตถุประสงค์ของการเรียนการสอน เป็นต้น

3. โดเมนของเทคโนโลยี เช่น เซตของผลิตผลทางการศึกษา เป็นต้น

4. โดเมนของการสอน เช่น เซตของกิจกรรมการเรียนการสอน เป็นต้น

5. โดเมนของการปฏิบัติ (performance) เช่น เซตของข้อสอบ เป็นต้น

ซึ่งสามารถแสดงความสัมพันธ์ระหว่างโดเมนและสารสนเทศ (Information) ที่ได้ดังนี้

โดเมน	หน้าที่ของสารสนเทศที่ได้
1. วัตถุประสงค์ (Objective)	1. เราต้องการทำอะไร
2. ความรู้ (Knowledge)	2. เรารู้อะไร
3. ผลิตผล (Products)	3. สิ่งที่เราสามารถทำได้หรือสาธิตให้เห็นได้โดยใช้สื่อที่มีอยู่แล้ว
4. การสอน (Instruction)	4. สิ่งที่เราทำจริงด้วยสื่อที่มีอยู่
5. ข้อสอบ (Items)	5. สิ่งที่เราวัด

การกำหนดโดเมนทั้ง 5 ประเภทนี้ สัมพันธ์กันตามแนวคิดร่วมพื้นฐาน เป็นความสัมพันธ์เชิงเหตุผลต่อกัน โดยที่โดเมนวัตถุประสงค์ และโดเมนการเรียนการสอนมีความสัมพันธ์กันในระดับสูง ทัดเทียมกับความสัมพันธ์ของโดเมนการเรียนการสอน กับโดเมนของข้อสอบ โดเมนของความรู้เป็นตัวเชื่อมความสัมพันธ์ของโดเมนวัตถุประสงค์กับโดเมนการเรียนการสอน ส่วนโดเมนวัสดุอุปกรณ์หรือเทคโนโลยีนั้น เป็นวิธีการนำเอาความรู้มาถ่ายทอดให้กับผู้เรียน

สำหรับในประเทศไทย มีผู้กล่าวถึงแบบสอบอิงโดเมนและคำว่า โดเมนไว้หลายท่านในลักษณะที่คล้ายคลึงกัน เช่น

กมล ภู่ประเสริฐ (2520 : 15) กล่าวว่า แบบสอบอิงโดเมนใช้การกำหนดขอบเขตของพฤติกรรม หรือที่เรียกว่าการกำหนดโดเมน เป็นเกณฑ์สำหรับอ้างอิง ซึ่งการกำหนดโดเมนดังกล่าว เปรียบเสมือนการกำหนดขอบเขตของการสร้างคำถาม

โกวิท ประวาลพฤกษ์ (2523 : 18-19) ได้สรุปลักษณะของแบบสอบอิงเกณฑ์ไว้เป็น 2 ประเภท คือ แบบสอบอิงจุดประสงค์ ซึ่งยึดเอาจุดประสงค์เชิงพฤติกรรมเป็นเกณฑ์ในการชี้บ่งว่าผู้เรียนผ่านเกณฑ์ที่กำหนดไว้ในจุดประสงค์แล้วหรือไม่ และแบบสอบอิงโดเมน ซึ่งข้อสอบต่าง ๆ ได้จากการสุ่มตัวอย่างของพฤติกรรมที่กำหนดจากโดเมน และผลจากการสอบวัดสามารถอ้างอิงได้ว่า บุคคลมีความสามารถในระดับใดตามคะแนนมาตรฐานเนื้อหา

พวงรัตน์ ทวีรัตน์ (2529 : 31-38) กล่าวว่า แบบสอบที่นำมาใช้ในประเทศไทยขณะนี้ มี 2 ลักษณะ คือ แบบสอบอิงวัตถุประสงค์ และแบบสอบอิงโดเมน ซึ่งในการเขียนข้อสอบแบบสอบอิงวัตถุประสงค์จะยึดวัตถุประสงค์การสอนเป็นเกณฑ์ และอธิบายถึงคำว่า

"โดเมน" ว่า หมายถึง ประชากรพฤติกรรมหรือประชากรเนื้อหาของวิชาที่จะทำการสร้างข้อสอบ ในการเขียนข้อสอบโดยอิงพฤติกรรม จะยึดพฤติกรรมที่วิเคราะห์ได้จากเนื้อหาที่กำหนดไว้ในหลักสูตรเป็นเกณฑ์ในการเขียน

บุญเชิด ภิญโญอนันตพงษ์ (2526 : 19-29) ได้แบ่งแบบสอบอิงเกณฑ์เป็นประเภทแบบสอบอิงจุดประสงค์และแบบสอบอิงมวลความรู้ (หรืออิงโดเมน) เช่นกัน และกล่าวถึงแบบสอบอิงโดเมนว่า เป็นแบบสอบที่สร้างจากลักษณะเฉพาะของมวลความรู้หรือกลุ่มพฤติกรรมซึ่งสุ่มมาจากประชากรหรือจักรวาลของข้อสอบ การแปลความหมายของคะแนน โดยการนำเอาผลของการสอบไปเทียบกับเซตหรือกลุ่มของงานที่สุ่มมาจากจักรวาลของงานที่นิยามไว้อย่างดีแล้วว่านักเรียนมีความสามารถเท่าไร จากนั้นจะพยากรณ์ผลงานจากกลุ่มตัวอย่างหรือแบบสอบไปยังมวลความรู้หรือจักรวาลของงานทั้งหมดนั้น

สงบ ลักษณะ (2525 : 1-14) ได้เสนอข้อคิดเห็นว่า แบบสอบที่เหมาะสมที่จะใช้ประเมินผลสัมฤทธิ์ทางการเรียนของผู้เรียนได้ดีนั้น ควรเป็นแบบสอบอิงโดเมน และให้ความหมายของคำว่า "โดเมน" ว่าหมายถึง มวลพฤติกรรมหรือประชากรพฤติกรรมในขอบเขตเนื้อหาวิชาที่กำหนดหรือวิเคราะห์ไว้อย่างชัดเจน เพื่อที่จะสร้างข้อสอบ

จากความหมายและลักษณะของแบบสอบอิงโดเมนที่นักวัดผลการศึกษาได้กล่าวถึงแล้วนั้น อาจสรุปได้ว่า แบบสอบอิงโดเมนเป็นแบบสอบที่สร้างขึ้นจากโดเมนที่ได้กำหนดไว้อย่างชัดเจน แทนที่จะสร้างจากจุดประสงค์เชิงพฤติกรรมโดยตรงและประกอบด้วยกลุ่มตัวอย่างข้อสอบที่ได้รับการสุ่มมาจากประชากรข้อสอบทั้งหมดในโดเมนที่ได้กำหนดไว้นั้น และในการแปลความหมายของคะแนนสอบ จะใช้จำนวนข้อที่ตอบถูกในกลุ่มตัวอย่างของโดเมนนี้มาคำนวณหาค่าอัตราส่วนที่แท้จริง เมื่อเทียบกับประชากรข้อสอบทั้งหมดในโดเมน เช่น สุ่มข้อสอบมาได้ 10 ข้อ ถ้านักเรียนทำข้อสอบฉบับนี้ได้ 8 ข้อ แสดงว่านักเรียนมีความรู้ในเรื่องที่จะวัดทั้งหมดถึง 80% หรืออาจกล่าวได้ว่า ถ้านักเรียนทำข้อสอบทั้งหมดในโดเมนจะสามารถทำได้ในอัตราส่วน $\frac{8}{10}$ หรือ 80% ของข้อสอบ

1.2 <u>พัฒนาการของเทคโนโลยีการเขียนข้อสอบอิงโดเมน</u>

เป้าหมายของการสร้างและพัฒนาแบบสอบอิงโดเมน คือ กลุ่มข้อสอบจำนวนมากที่มีความครอบคลุมตามโครงสร้างเนื้อหาหรือโดเมนที่ต้องการวัด เพื่อเป็นตัวแทนที่แท้จริงในลักษณะย่อส่วน ตามสัดส่วนที่แท้จริงของโครงสร้างความรู้ โดยมุ่งใช้ผลการสอบในลักษณะของการตีความถึงระดับความรู้ความสามารถของผู้เรียนโดยพาดพิงไปถึงโดเมนนั้นได้ ดังนั้น กระบวนการในการสร้างข้อสอบจะเน้นการกำหนดขอบเขตโดเมนที่ต้องการวัดให้มีความชัดเจนและมีความเป็นตัวแทนของจักรวาลความรู้ให้มากที่สุด กล่าวคือ ในการเขียนข้อสอบอิงโดเมนจะเป็นการเขียนโดยยึดพฤติกรรมที่วิเคราะห์ได้จากเนื้อหาที่กำหนดไว้ในหลักสูตรเป็นเกณฑ์ โดยพิจารณาว่าหลักสูตรวิชานั้น ๆ ประกอบด้วยเนื้อหาที่สำคัญ ๆ อะไรบ้าง และเนื้อหานั้น ๆ ต้องการสอนให้ผู้เรียนเกิดพฤติกรรมอะไรบ้าง เมื่อวิเคราะห์ได้แล้วก็ออกข้อสอบให้ครอบคลุมพฤติกรรมนั้น ๆ นั่นเอง

นักวัดผลการศึกษาที่ได้ศึกษาวิธีการเขียนข้อสอบแบบนี้เป็นกลุ่มแรก คือ ไฮฟ์ลีและคณะ (Hively, et al. 1968 อ้างถึงใน ชูศักดิ์ 2529 : 133) แห่งมหาวิทยาลัยมินเนโซต้า ซึ่งได้พัฒนาแบบสอบคู่ขนานขึ้นหลายฉบับ เพื่อใช้วิจัยควบคู่ไปกับการพัฒนาหลักสูตรวิชาคณิตศาสตร์ และวิทยาศาสตร์ระดับประถมศึกษา โดยไฮฟ์ลีและคณะได้สร้างเครื่องทุ่นแรงในการเขียนข้อสอบชนิดหนึ่งเรียกว่า ฟอร์มข้อสอบ (Item Form) จึงถือว่า ฟอร์มข้อสอบเป็นเครื่องมือชนิดแรกของพัฒนาการของเทคโนโลยีการเขียนข้อสอบอิงโดเมน ในช่วงเวลาที่ใกล้เคียงกัน ที่มหาวิทยาลัยแคลิฟอร์เนีย เมืองลอสแองเจลีส ภายใต้การนำของโปแฟม (Popham 1978) ได้ตั้งสถาบัน IOX (Institute of Objectives Exchange) ซึ่งเดิมที่มีจุดมุ่งหมายที่จะเป็นผู้เชี่ยวชาญทางด้านจุดมุ่งหมายเชิงพฤติกรรม (BO) กล่าวคือ สถาบัน IOX นี้จัดตั้งขึ้นเพื่อเป็นศูนย์รวบรวม BO ของกระบวนการเรียนการสอนหลาย ๆ หมวดวิชา ให้ครอบคลุมมากที่สุด และสถาบัน IOX ได้จัดพิมพ์ BO ของแต่ละรายวิชาและตัวอย่างข้อสอบไว้บริการครู หรือผู้สนใจทั่วไป แต่ก็ไม่บรรลุเป้าหมาย เพราะ BO ของแต่ละข้อเป็นข้อความที่กำหนดไว้อย่างหลวม ๆ และไม่ได้กำหนดแนวทางเฉพาะให้กับผู้เขียนข้อสอบอย่างชัดเจน สถาบัน IOX จึงดัดแปลงฟอร์มข้อสอบของไฮฟ์ลีและคณะ โดยสร้างเครื่องทุ่นแรงขึ้นมาใหม่สองอย่าง เรียกว่า Amplified Objectives และ Test Specifications

ในปี 1970 บอร์มุท (Bormuth, 1970) ได้เสนอเทคนิคเก่าแก่ของการเขียนข้อสอบอิงโดเมนอีกวิธีหนึ่ง ซึ่งเรียกว่า "Items for prose learning" หรือเรียกว่า Linguistic-based Approach ซึ่งมีหลักและวิธีการแปลงประโยคสำคัญของเนื้อหาวิชาให้เป็นองค์ประกอบของมวลความรู้ (Domain) แล้วเขียนข้อสอบจากโดเมนดังกล่าว นอกจากนี้ยังมีเทคโนโลยีการเขียนข้อสอบอิงโดเมนอีกหลายแนวคิด ซึ่ง รอยด์ และ ฮาลาไดนา (Raid and Haladyna 1982) ได้รวบรวมเทคโนโลยีใหม่ของการกำหนดรายละเอียดของโดเมน ในการเขียนข้อสอบอิงโดเมนออกเป็น 6 ลักษณะ คือ

1. ฟอร์มข้อสอบ (Item Form หรือ Item Frame)
2. การแปลงประโยคของข้อความ (Linguistic Transformation Approach)
3. จุดประสงค์ขยายความ (Amplified Objective Approach)
4. การออกแบบแง่มุมคำถาม (The Facet design Approach หรือ Facet Theory)
5. การทดสอบอิงมโนภาพ (Concept-based Testing)
6. การใช้คอมพิวเตอร์ (Computer-based Method)

วิธีการเขียนข้อสอบในแต่ละแบบ กล่าวโดยสังเขปได้ดังนี้

1. ฟอร์มข้อสอบ

ในปี 1968 ไฮฟ์ลี และคณะ ได้ค้นคว้า วิธีการเขียนข้อสอบในวิชาคณิตศาสตร์และวิทยาศาสตร์ โดยการกำหนดฟอร์มข้อสอบ ซึ่งเป็นการกำหนดกฎเกณฑ์ในการเขียนข้อสอบให้ได้มาก ๆ ข้อ โดยที่กลุ่มข้อสอบเหล่านี้วัดโดเมนเดียวกันได้อย่างครอบคลุมและมีความเป็นเอกพันธ์

ฟอร์มข้อสอบมีส่วนประกอบที่สำคัญ 3 ส่วน ดังนี้

1) ส่วนที่คงที่ เป็นส่วนประกอบของโครงสร้างข้อความที่กำหนดไว้ตายตัวเป็นเสมือนตัวฟอร์มข้อสอบ หรือเปลือกหุ้มฟอร์มข้อสอบ (Item Form Shell) ที่ใช้สำหรับผลิตข้อสอบ

2) ส่วนที่แปรเปลี่ยน เป็นส่วนประกอบของโครงสร้างข้อความที่เว้นว่างไว้อย่างน้อยที่สุดหนึ่งแห่ง สำหรับที่จะเอาเซตของคำหรือข้อความที่ต้องการถามใส่แทน

3) ส่วนที่จะนำไปเติมลงในส่วนที่แปรเปลี่ยน เป็นเซตของประโยคข้อความที่ได้รับการนิยามไว้อย่างชัดเจนสำหรับใช้เติมลงในส่วนที่แปรเปลี่ยน

แสดงว่า ฟอร์มข้อสอบต้องประกอบด้วย ประโยคที่คงที่ ส่วนที่เว้นว่างและการนิยามเซตขององค์ประกอบที่จะนำไปเติมลงในช่องว่าง ดังตัวอย่าง

ตัวอย่าง ฟอร์มข้อสอบที่วัดเนื้อหาวิชาคณิตศาสตร์

ประโยคข้อสอบ

จำนวนใดต่อไปนี้เป็นจำนวนเฉพาะ

ก ( ) ข ( ) ค ( ) ง ( )

ส่วนประกอบสำหรับเติมให้ข้อสอบสมบูรณ์

กำหนดจำนวนสี่จำนวนจาก ก ถึง ง ให้นักเรียนพิจารณาว่าจำนวนใดเป็นจำนวนเฉพาะ

จำนวนที่กำหนดให้ต้องเป็นจำนวนเต็มสองหรือสามหลัก และทุกจำนวนต้องเป็นเลขคี่มีจำนวนที่ไม่สามารถแยกตัวประกอบได้เพียงจำนวนเดียว นอกนั้นทั้งหมดสามารถแยกตัวประกอบได้

ข้อสอบตัวอย่าง

จำนวนใดต่อไปนี้เป็นจำนวนเฉพาะ

ก. 27 ข. 31 ค. 147 ง. 189

คำตอบถูก : 31

หมายเหตุ การกำหนดฟอร์มข้อสอบ อาจจะให้ส่วนที่แปรเปลี่ยนอยู่ที่ตัวเลือกดังตัวอย่างที่ให้ไปแล้ว หรือให้อยู่ที่ตัวประโยคข้อสอบหรือคำถามก็ได้ ดังตัวอย่างข้อสอบที่เขียนจากฟอร์มข้อสอบวิชาสถิติเบื้องต้น

สุ่มตัวอย่าง (รถบรรทุก) มา (100 คัน) ได้ (น้ำหนัก) เฉลี่ย (40,250 ปอนด์) และส่วนเบี่ยงเบนมาตรฐานของ (น้ำหนัก) เท่ากับ (2,500 ปอนด์) จงคำนวณขอบเขตของค่าเฉลี่ยจริงของ (น้ำหนัก) (รถบรรทุก) ด้วยความเชื่อมั่นระดับ (95%)

ข้อความที่อยู่ในวงเล็บเรียกว่า ส่วนที่แปรเปลี่ยน ถ้าเปลี่ยนข้อความของส่วนที่แปรเปลี่ยนเป็นเรื่องอื่นที่จะทำให้ได้ข้อสอบเพิ่มขึ้นมาอีก

ในการใช้ ฟอร์มข้อสอบนี้ โปแฟม (Popham 1980) ได้กล่าวว่า ทำให้เขียนข้อสอบได้ในวงจำกัด และวัดพฤติกรรมที่ค่อนข้างอยู่ในระดับที่ต่ำเป็นส่วนใหญ่

2. <u>การเปลี่ยนแปลงประโยคของข้อความ</u>

ในปี 1970 บอร์มุท เป็นคนแรกที่นำเทคโนโลยีการเขียนข้อสอบแบบการเปลี่ยนแปลงภาษาของข้อคำถามมาใช้ในการประเมินผลการเรียนรู้โดยวิธีการเปลี่ยนแปลงภาษาของข้อความเป็นข้อคำถาม ซึ่งต้องกำหนดกฎเกณฑ์เพื่อเป็นแนวทางให้ผู้เขียนข้อสอบและการเปลี่ยนแปลงภาษาของข้อคำถามมีอยู่ 2 วิธี คือ

(1) ข้อคำถามมาจากการเปลี่ยนแปลงประโยคที่กำหนดให้

(2) ข้อคำถามมาจากความสัมพันธ์ของประโยคเหล่านั้น แล้วใช้คำว่า ใคร อะไร หรือที่ไหน ฯลฯ มาแทนคำที่เป็นประธาน หรือกรรม ในประโยคแต่ละประโยคก็จะได้ประโยคข้อคำถามทันที เช่น

<u>ข้อคำถามมาจากการเปลี่ยนแปลงประโยคที่กำหนดให้</u> ลักษณะจะกำหนดข้อความมาให้ แล้วแปลงเป็นประโยคคำถาม

<u>ประโยคข้อความ</u>		<u>ประโยคคำถาม</u>
นักพัฒนาข้อสอบเป็นผู้คำนวณค่า		ใครเป็นผู้คำนวณค่าสัมประสิทธิ์
สัมประสิทธิ์ความเที่ยงตรง	แปลง →	ความเที่ยงตรง

การเปลี่ยนแปลงประโยคนี้คำถามอาจจะทำเป็นแบบให้ตอบสั้น ๆ หรือเป็นแบบข้อสอบเลือกตอบก็ได้

การเปลี่ยนแปลงประโยคข้อความมาเป็นประโยคคำถามนั้นนอกจากจะเป็นการวัดจำนวนแล้วยังสามารถที่จะเปลี่ยนเป็นคำถามที่วัดความเข้าใจได้โดยการเปลี่ยนจากประโยคข้อความเดิมมาเป็นประโยคใหม่ โดยที่ยังรักษาความหมายเดิมอยู่ ต่อจากนั้นจึงแปลงจากประโยคใหม่ให้เป็นข้อคำถามอีกทีหนึ่ง ซึ่งจะทำให้คำถามสามารถวัดพฤติกรรมสูงกว่าจำได้ แอนเดอร์สัน (Anderson, 1972) ได้เน้นถึงความสำคัญของการแปลงประโยคข้อความเป็นประโยคคำถาม อยู่ 2 ประการคือ

ก. คำนามหรือคำสรรพนามในประโยคต้องเปลี่ยนใหม่

ข. ประโยคข้อความเดิมแปลงมาเป็นประโยคใหม่ความหมายจะต้องเหมือนเดิม

ข้อคำถามจากความสัมพันธ์ของประโยคเหล่านั้น วิธีนี้จะต้องเขียนคำถามมาจากความสัมพันธ์ของประโยคดังตัวอย่าง

ประโยคที่สัมพันธ์กัน		ประโยคคำถาม
ก. มือของจิ๋มได้รับบาดเจ็บ		
ข. เมื่อเขากำลังล้างมีดของเขา	แปลง →	อะไรเป็นสาเหตุที่ทำให้มือของจิ๋ม
ค. เกิดอุบัติเหตุมีดบาดมือ		บาดเจ็บ

3. จุดประสงค์ขยายความ

ในปี 1978 โปแฟมและคณะได้เสนอเทคนิคการเขียนข้อสอบให้ครอบคลุมเนื้อหาโดยการอิงจุดประสงค์ขยายความ (Amplified Objectives) ซึ่งเป็นการผสมผสานระหว่างจุดประสงค์เชิงพฤติกรรม (BO) และ ฟอร์มข้อสอบเข้าด้วยกัน เพื่อเป็นการแก้ปัญหาการเขียนจุดประสงค์เชิงพฤติกรรมให้มีความชัดเจนยิ่งขึ้น จุดประสงค์ขยายความมีส่วนประกอบ 5 ส่วนดังนี้

1) คำอธิบายทั่วไป (General Description) ส่วนนี้จะระบุขอบเขตของความรู้ที่ต้องการวัด ซึ่งเป็นเสมือนจุดประสงค์หลักที่ต้องการประเมิน

2) จุดประสงค์ (Objective) ส่วนนี้เป็นการแปลงขอบเขตของความรู้หรือจุดประสงค์หลักให้เป็นจุดประสงค์เฉพาะ

3) ส่วนขยายความจุดประสงค์ (Amplified Objective) ส่วนนี้เป็นการขยายความจุดประสงค์ให้ละเอียดยิ่งขึ้น โดยการอธิบายฟอร์มข้อสอบ ทั้งภาคคำถามและภาคคำตอบ

4) เกณฑ์การตอบถูก (Criterion or Correctness) ส่วนนี้จะระบุถึงรายละเอียดของเงื่อนไขที่ตอบถูก

5) ข้อสอบตัวอย่าง ส่วนนี้เป็นการอธิบายตัวข้อสอบให้ชัดเจนยิ่งขึ้นโดยเขียนข้อสอบให้เป็นตัวอย่างไว้

ตัวอย่าง การใช้จุดประสงค์ขยายความเขียนข้อสอบ

1) คำอธิบายทั่วไป

ทักษะในการเรียบเรียงคำที่มีลักษณะชัดเจนและคลุมเครือ

2) จุดประสงค์

เมื่อกำหนดประโยคข้อความซึ่งละคำนามหรือคำกริยาไว้ นักเรียนสามารถเลือกคำที่กำหนดไว้ให้คำใดคำหนึ่งจากสองคำ มาเติมใส่ให้ประโยคข้อความที่กำหนดให้นั้นมีความสมบูรณ์ชัดเจน

3) ส่วนขยายความจุดประสงค์

ภาคคำถาม

1. ให้นักเรียนเลือกคำใดคำหนึ่ง จากคู่คำที่กำหนดไว้ เพื่อนำไปใส่ประโยคที่กำหนดให้มีความสมบูรณ์และชัดเจน โดยกา × ทับคำที่ต้องการนั้น

2. ประโยคที่กำหนดให้ต้องเป็นประโยคง่าย ๆ ที่ละคำนามหรือคำกริยา

3. แบบสอบแต่ละฉบับต้องละคำนามหรือคำกริยาในจำนวนที่เท่ากัน

4. คำศัพท์ที่เป็นคำนามหรือคำกริยาต้องเป็นศัพท์ที่เด็กนักเรียนระดับนั้นรู้จักดี

ภาคคำตอบ

1. กำหนดคำนามทั้งคู่หรือคำกริยาทั้งคู่ ซึ่งมีความหมายชัดเจนต่างกันให้นักเรียนเลือกตอบเพียงคำเดียว

2. ถ้าเป็นคำกริยา ต้องมีคำหนึ่งที่บ่งกริยาอาการเคลื่อนไหว เช่น ตะเกียกตะกาย กระโดด อีกคำหนึ่งจะต้องเป็นคำกริยาทั่ว ๆ ไปที่ไม่บ่งกิริยาอาการเคลื่อนไหวชัดเจน ซึ่งมักจะเป็นคำที่ใช้เชื่อมข้อความ เช่น เป็น ไป

3. ถ้าเป็นคำนาม ต้องมีคำหนึ่งที่เฉพาะเจาะจงและชัดเจนกว่า เช่น ช่างไม้ ช่างปูน ส่วนอีกคำหนึ่งเป็นคำทั่วไปไม่ชัดเจน มีความเป็นนามธรรม เช่น คน สิ่งของ

4) เกณฑ์การตอบถูก

คำตอบถูกต้องกา × ทับคำนามที่มีความหมายชัดเจนหรือคำกิริยาที่บ่งกริยาอาการเคลื่อนไหวชัดเจน

5) ข้อสอบตัวอย่าง

คำชี้แจง : ให้เลือกคำที่ต้องการที่อยู่ในวงเล็บ โดยกาทับตัวอักษรหน้าคำนั้น ซึ่งจะทำให้ข้อความของประโยคมีความหมายชัดเจน

ข้อสอบ : (ก. คน ข. ช่างปูน) กำลังซ่อมแซมส่วนที่ชำรุดของตัวตึก

คำตอบ : ช่างปูน

4. การออกแบบแง่มุมคำถาม

การเขียนข้อสอบตามแง่มุมของคำถามเป็นการเสนอแนะของกัตแมน (Guttman)ในปี 1969 ซึ่งเป็นฟอร์มข้อสอบชนิดหนึ่ง มีส่วนประกอบสำคัญอยู่ 3 ส่วน คือ ส่วนที่คงที่ ซึ่งเรียกว่า "ตัวฟอร์มข้อสอบ" หรือเปลือกหุ้มของข้อสอบ (Item form shell) ส่วนที่เว้นว่างไว้เรียกว่า "แง่มุมคำถาม" (Facet) ซึ่งต้องหาคำตอบมาเติมใส่ และส่วนที่เป็นคำตอบ เรียกว่า "ส่วนย่อยของแง่มุมคำถาม" (Facet Element) ซึ่งต้องนำไปเติมลงในที่ว่างไว้ ในฟอร์มข้อสอบให้เหมาะสม ซึ่งส่วนนี้ก็คือรายละเอียดของเนื้อหาวิชาที่ต้องการวัด ดังนั้นลักษณะของแง่มุมคำถามจะกำหนดประโยคมาแล้วเว้นช่องว่างไว้สำหรับให้นำชุดของคำตอบไปเติม จึงเรียกเป็นประโยคที่ใช้จับคู่ (Mapping Sentense)

ตัวอย่าง การเขียนข้อสอบตามแง่มุมของคำถาม

ประโยคที่ใช้จับคู่ : วิธี_A_ เป็นวิธีที่เหมาะสมที่สุดของการเขียนข้อสอบวิชา_B_ ในโดเมนที่เป็นเนื้อหา (content domain)

แง่มุมคำถาม A : กลวิธีกำหนดรายละเอียดของโดเมน

ส่วนย่อยของแง่มุมคำถาม :

(1) การแปลงประโยคของข้อความ

(2) ฟอร์มข้อสอบ

(3) ลักษณะเป็นปรนัย

(4) ประโยคที่ใช้จับคู่

แง่มุมคำถาม B : โดเมนที่เป็นเนื้อหา

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ส่วนย่อยของแง่มุมคำถาม

(1) การอ่าน (2) ภาษา

(3) คณิตศาสตร์ (4) วิทยาศาสตร์

(5) สังคมวิทยา

ส่วนประกอบของแง่มุมคำถามจะถูกนำมาจับคู่กันในความหมายหลาย ๆ ทาง ในตัวอย่างนี้จะมีส่วนย่อยในลักษณะจับคู่กันเป็น combination กันถึง 20 คู่ จากแง่มุมคำถาม A และแง่มุมคำถาม B ดังนั้นจะสามารถเขียนคำถามที่เป็นไปได้ถึง 20 ข้อ ในลักษณะคำถามถูก-ผิด

เช่น ตัวอย่างคำถาม : ในแง่มุมคำถาม A (1) กับแง่มุมคำถาม B (3) เป็นคำถามถูก-ผิด ดังนี้

"วิธี การแปลงประโยคข้อความเป็นวิธีที่เหมาะสมที่สุดของการเขียนข้อสอบวิชาคณิตศาสตร์ ในโดเมนที่เป็นเนื้อหา" ซึ่งข้อความนี้เป็นประโยคที่ผิดความหมาย ต้องกาเครื่องหมาย ✕ หน้าข้อความเป็นต้น ฯลฯ

5. การทดสอบอิงมโนภาพ

Markle และ Tieman ได้ใช้วิธีเขียนข้อสอบอิงโดเมนเพื่อวัดระดับมโนภาพของนักเรียน มโนภาพได้ถูกนิยามเช่นเดียวกับประเภทของจุดประสงค์เหตุการณ์หรือความสัมพันธ์ ซึ่งจะเปลี่ยนแปลงในตัวมันเอง แล้วจัดเข้าชุดเดียวกัน แล้วใช้ชื่อเดียวกัน การวัดมโนภาพทำได้โดยการตรวจสอบเพื่อสรุปอ้างอิงไปยังตัวใหม่ และแยกออกจากกลุ่มที่ไม่ใช่ตัวอย่าง กลุ่มที่เป็นตัวอย่างและไม่ใช่ตัวอย่างจะต้องแตกต่างไปจากที่สอนในห้อง เพื่อเป็นการทดสอบว่านักเรียนนี้มีความเข้าใจคำที่เป็นมโนภาพแล้วหรือยัง เช่น ให้หามโนภาพของคำที่มีความหมายตรงข้าม ครูอาจจะยกตัวอย่างคำที่มีความหมายตรงข้ามว่า "ดี-เลว" และคำที่ไม่ใช่คำตรงข้ามว่า หยิ่ง-ละโมบ หรืออาจจะใช้กลุ่มตัวอย่างคำ "ร้อน-หนาว"

ตัวอย่าง การเขียนข้อสอบตามวิธีอิงมโนภาพ

มโนภาพเกี่ยวกับคำศัพท์ : คำที่มีความหมายตรงข้ามกับคำต่อไปนี้

คุณลักษณะที่สำคัญ

(1) มีความหมายตรงข้ามกับคำที่ให้มา

(2) เป็นคำชนิดเดียวกับคำที่ให้มา

(3) เป็นคำใหม่ไม่ใช่เป็นคำที่ผันมาจากคำที่ให้มา

คุณลักษณะของตัวแปร

(4) อาจได้มาจากคำที่ต่างชนิดกัน เช่น

ก. คำนาม ข. คำกริยา ค. คำสรรพนาม

ง. คำวิเศษณ์ จ. คำคุณศัพท์ ฉ. คำบุพบท

(5) จำนวนพยางค์ของคำทั้งสองอาจจะมีลักษณะดังนี้

ก. เท่ากัน ข. ไม่เท่ากัน

(6) ความหมายที่ตรงข้ามกันจะต้องมีลักษณะดังนี้

ก. มีลักษณะต่อเนื่องกัน

ข. มีความหมายสองทาง

กลุ่มที่เป็นตัวอย่างที่ใช้ในการสอน			กลุ่มที่ไม่ใช่ตัวอย่างที่ใช้ในการสอน		
1. เลว : ดี	4จ, 5ก, 6ก		1. หยิ่ง : ละโมบ	ขาดคุณสมบัติ	1
2. อันตราย : ปลอดภัย	4ก, 5ก, 6ก		2. เหตุผล : อารมณ์	ขาดคุณสมบัติ	1
3. มีชีวิต : ตาย	4ง, 5ก, 6ข		3. เรา : เรา	ขาดคุณสมบัติ	1
4. เขา : หล่อน	4ค, 5ก, 6ข		4. เหนือ : บน	ขาดคุณสมบัติ	1
5. อย่างรวดเร็ว : อย่างเชื่องช้า	4ง, 5ข, 6ก		5. มีความสุข : เศร้า	ขาดคุณสมบัติ	2
6. ใน : นอก	4ค, 5ข, 6ก		6. มีความสุข : ไม่มีความสุข	ขาดคุณสมบัติ	3
			7. สามารถ : ไม่สามารถ	ขาดคุณสมบัติ	3
			8. โค้งแย้ง : ตกลง	ขาดคุณสมบัติ	2

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

กลุ่มที่เป็นตัวอย่างที่ใช้ในการสอน			กลุ่มที่ไม่ใช่ตัวอย่างที่ใช้ในการสอน		
1. ร้อน	: เย็น	4จ, 5ก, 6ก	1. จินตนาการ : ความคิดฟุ้งซ่าน		ขาดข้อ 1
2. สูญเสีย	: ได้รับ	4ก, 5ก, 6ก	2. เก้าอี้	: ม้านั่ง	ขาดข้อ 1
3. ทำให้สูง	: ทำให้ต่ำ	4ง, 5ข, 6ก	3. ข้างหลัง	: ต่อจาก	ขาดข้อ 1
4. เธอ	: ฉัน	4ค, 5ก, 6ข	4. มืดมน	: สว่าง	ขาดข้อ 2
5. อย่างรื่นเริง : อย่างโศกเศร้า		4ฉ, 5ก, 6ข	5. รุนแรง	: ไม่รุนแรง	ขาดข้อ 3
6. เหนือ	: ใต้	4ฉ, ข, 6ก	6. มีเหตุผล	: ไม่มีเหตุผล	ขาดข้อ 3
			7. อ่อนแอ	: รุนแรง	ขาดข้อ 2

ตัวอย่างข้อสอบ

คำคู่ใดต่อไปนี้ที่มีความหมายตรงข้ามกัน?

ก. จินตนา : ความคิดฟุ้งซ่าน

ข. ทำให้สูง : ทำให้ต่ำ

ค. มีเหตุผล : ไม่มีเหตุผล

ง. อ่อนแอ : รุนแรง

คำตอบที่ถูกต้อง ข้อ ข.

6. การเขียนข้อสอบโดยใช้คอมพิวเตอร์

การใช้คอมพิวเตอร์สร้างข้อสอบนับว่าเป็นเทคโนโลยีการเขียนข้อสอบขั้นสูงสุด ซึ่งพัฒนามาจากการเก็บรวบรวมข้อสอบไว้ในคลังข้อสอบ แล้วป้อนข้อสอบเหล่านี้เข้าไปในเครื่องจำของคอมพิวเตอร์ ต่อมาจึงมีการพัฒนาให้คอมพิวเตอร์เขียนข้อสอบขึ้นมา โดยทำเป็นฟอร์มข้อสอบแล้วเขียนเป็นโปรแกรม สั่งให้เครื่องผสมประโยค ประโยคข้อสอบออกมาจำนวนมากมาย ความจริงแล้วการใช้คอมพิวเตอร์เขียนข้อสอบก็คือ การใช้ฟอร์มข้อสอบสร้างข้อสอบนั่นเองเพียงแต่ใช้เครื่องแทนคนเท่านั้น

สำหรับในประเทศไทย ก็ได้มีการกล่าวถึงการสร้างแบบสอบอิงโดเมน และได้นำแนวคิดของเทคโนโลยีการเขียนข้อสอบมาดัดแปลงใช้ในประเทศไทย เช่น

สมหวัง พิธิยานุวัฒน์ (2529 : 76-77) ได้กล่าวว่าโดยทั่วไปในการพัฒนาแบบสอบอิงปริเขตหรืออิงโดเมน มีขั้นตอนที่สำคัญ คือ

1. การกำหนดวัตถุประสงค์ของการสอน

2. การกำหนดโดเมนที่ชัดเจน ถ้าโดเมนกว้างเกินไปก็ควรกำหนดเป็นโดเมนย่อยเพื่อให้โดเมนมีความเป็นตัวแทนของจักรวาลความรู้

3. พัฒนาข้อสอบโดยกำหนดรูปแบบข้อสอบ

4. บรรณาธิกรข้อสอบ ทั้งในเชิงหลักเหตุผลและหลักประจักษ์

5. พัฒนาแบบสอบทั้งฉบับ และนำไปทดสอบหาคุณภาพ และจัดทำให้อยู่ในรูปมาตรฐานเพื่อนำไปใช้ต่อไป

โกวิท ประวาลพฤกษ์ (2523 : 22) ได้เสนอบทความเกี่ยวกับการสร้างแบบสอบอิงโดเมน ซึ่งสรุปเป็นขั้นตอนได้ดังนี้

1. การกำหนดขอบเขตประชากรของความรู้หรือความสามารถ เช่น โดเมน คือ ประชากรของการคูณเลขหลักเดียว ซึ่งประชากรนี้จะเล็กใหญ่ตามแต่เนื้อหาที่จะกำหนดได้

2. กำหนดรูปแบบข้อสอบ และสร้างข้อสอบตามรูปแบบที่ได้กำหนดไว้

3. เลือกข้อสอบมาใช้โดยวิธีการทางสถิติ ในกรณีที่ข้อสอบไปซ้ำกับที่เคยใช้สอนหรือใช้ฝึกหัด ก็ถือว่าข้อสอบนั้นวัดความรู้ความจำ แต่ในกรณีที่ไม่ตรงก็ถือว่า วัดการนำไปใช้ หรืออาจใช้วิธีการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) ก็ได้

4. ในการแปลผล สามารถนำคะแนนการตอบถูกมาคำนวณเพื่อสรุปได้ว่าเขามีความสามารถในระดับใด หรือมีความรอบรู้ในระดับใด ได้ด้วยวิธีการทางสถิติเกี่ยวกับการคาดหมายแบบต่าง ๆ ตามวิธีการสุ่มข้อสอบที่ใช้

บุญเชิด ภิญโญอนันตพงษ์ (2526 : 24-25) กล่าวถึงขั้นตอนในการสร้างแบบสอบอิงมวลความรู้หรืออิงโดเมน ไว้ดังนี้

1. กำหนดขอบข่ายเนื้อหาหรือกฎเกณฑ์สำหรับใช้เป็นเกณฑ์พิจารณาตัดทอนหรือนับรวมข้อสอบ

2. สร้างฟอร์มข้อสอบหรือกฎเกณฑ์ต่าง ๆ สำหรับใช้ในการผลิตเชตของข้อสอบที่วัดพฤติกรรมเดียวกัน

3. เขียนข้อสอบตามเกณฑ์ที่สร้างขึ้นในแต่ละฟอร์มข้อสอบ แล้วให้ผู้เชี่ยวชาญตรวจสอบกฎเกณฑ์สำหรับผลิตข้อสอบและตรวจสอบดูว่าข้อสอบกับมวลความรู้สอดคล้องกันหรือไม่ด้วย

4. เตรียมตัวอย่างข้อสอบให้เป็นตัวแทนที่ดีของแต่ละฟอร์มข้อสอบในแต่ละมวลความรู้ เพื่อให้การเขียนข้อสอบแต่ละข้อ เขียนได้ชัดเจนไม่กำกวมและไม่เกิดข้อบกพร่องทางโครงสร้าง

สงบ ลักษณะ (2525 : 16) ได้พัฒนาเทคนิคการเขียนข้อสอบอิงโดเมนขึ้นโดยอาศัยแนวคิดที่คล้ายคลึงกับจุดประสงค์ขยายความของโปแฟม และรูปแบบข้อสอบของไฮฟลีเรียกว่า"ลักษณะเฉพาะของข้อสอบ" (Item Specification) โดยได้ให้นิยามไว้ว่า

"การกำหนดลักษณะเฉพาะของข้อสอบ เป็นการจัดระเบียบหรือสร้างกฎเกณฑ์สำหรับการเขียนข้อสอบให้รัดกุม รอบคอบ เด่นชัด สมบูรณ์ด้วยคุณภาพต่าง ๆ เช่น มีเหตุผลเป็นปรนัย วัดในสิ่งที่ต้องการวัดได้"

ลักษณะเฉพาะของข้อสอบ เป็นแบบฟอร์มที่กำหนดขึ้น เพื่ออำนวยความสะดวกในการเขียนหรือผลิตข้อสอบ ซึ่งมีส่วนประกอบที่สำคัญ 5 ส่วน ดังนี้

1. พฤติกรรมที่ต้องการวัด ส่วนนี้จะระบุพฤติกรรมหลักที่วิเคราะห์ได้จากเนื้อหาและจุดประสงค์ที่ต้องการวัด

2. พฤติกรรมย่อย ส่วนนี้จะระบุพฤติกรรมเฉพาะที่แตกย่อยมาจากพฤติกรรมหลัก

3. คำอธิบาย ส่วนนี้จะเขียนขยายพฤติกรรมย่อย โดยเขียนในรูปจุดประสงค์เชิงพฤติกรรมที่ครอบคลุมสถานการณ์เงื่อนไขอย่างละเอียด และมีพฤติกรรมที่คาดหวังที่สังเกตเห็นได้

4. ลักษณะคำถามและคำตอบ ส่วนนี้จะกำหนดฟอร์มข้อสอบทั้งส่วนที่เป็นคำถามและคำตอบ

ส่วนที่เป็นคำถาม จะกำหนดสิ่งเร้า สถานการณ์ เงื่อนไข และคำสั่งที่จะให้ผู้สอบทำ

ส่วนที่เป็นคำตอบ จะกำหนดฟอร์มของการตอบที่อาจเป็นแบบเลือกตอบ หรือเติมคำ และต้องระบุเกณฑ์ของการกำหนดคำตอบถูก วิธีเขียนตัวลวง และอาจต้องระบุวิธีการให้คะแนนด้วย

5. ข้อสอบตัวอย่าง ส่วนนี้จะช่วยขยายคำอธิบายให้ชัดเจนยิ่งขึ้น

<u>หมายเหตุ</u> การกำหนดลักษณะเฉพาะข้อสอบจะกำหนดพฤติกรรมย่อยหนึ่งพฤิตกรรมต่อหนึ่งแบบฟอร์ม ดังนั้น พฤติกรรมหลักที่ประกอบด้วยหลายพฤติกรรมย่อยจึงต้องเขียนแบบฟอร์มหลายแผ่น

ในการวิจัยครั้งนี้ ผู้วิจัยได้ใช้เทคนิคการกำหนดลักษณะเฉพาะของข้อสอบตามแบบฟอร์มของ สงบ ลักษณะ ช่วยในการเขียนข้อสอบอิงโดเมนขึ้น ดังนั้นจะกล่าวถึงรายละเอียดของการเขียนข้อสอบตามวิธีนี้ ซึ่งมีขั้นตอนในการสร้างข้อสอบดังนี้

1. วิเคราะห์เนื้อหาวิชาที่จะออกข้อสอบเป็นหัวเรื่องใหญ่ ๆ ที่สำคัญ ๆ

2. กำหนดประเภทพฤติกรรมใหญ่ ๆ ที่จะออกข้อสอบ

3. กำหนดพฤติกรรมที่ต้องการจะวัดและพฤติกรรมย่อย ในขั้นนี้ผู้วางแผนสร้างข้อสอบจะนำเนื้อหาวิชาและพฤติกรรมหลักที่วิเคราะห์มาแล้ว มาสร้างเป็นรายการพฤติกรรมที่ต้องการจะวัด และในแต่ละพฤติกรรมที่ต้องการจะวัดก็นำมาแตกออกเป็นพฤติกรรมย่อย ๆ ที่สำคัญ ภายใต้พฤติกรรมที่จะวัดนั้น

ในกรณีนี้พฤติกรรมที่ต้องการจะวัดหมายถึง พฤติกรรมใหญ่ หรือกลุ่มหรือหมวดหมู่ของพฤติกรรมย่อย การกำหนดพฤติกรรมที่ต้องการวัด ซึ่งเป็นพฤติกรรมใหญ่นี้ อาจใช้จุดประสงค์ของหลักสูตร และจุดประสงค์ของรายวิชาเป็นเครื่องช่วยได้

ส่วนพฤติกรรมย่อยนั้น มีความคล้ายคลึงกับจุดประสงค์เชิงพฤติกรรมที่แตกออกมาจากพฤติกรรมใหญ่ การเขียนพฤติกรรมย่อย อาจได้รับความช่วยเหลือจากจุดประสงค์ของการเรียนรู้ในแต่ละวิชาได้

ตัวอย่างของขั้นตอนที่ 1-3 แสดงได้ตามตารางดังนี้

ตารางการวิเคราะห์เนื้อหาและพฤติกรรมที่ต้องการวัดวิชาคณิตศาสตร์

หัวข้อเนื้อหา	พฤติกรรม		
	ความเข้าใจ	การคำนวณ	การแก้ปัญหา
1. การบวก ลบ คูณ หาร	-	12	-
2. เศษส่วน (อัตราส่วน) ร้อยละ	4	2	-
3. แผนภูมิ, ตาราง, กราฟ	2	-	-
4. การวัด	-	-	6
5. การแก้ปัญหาในชีวิตจริง	-	-	9

ประเภทพฤติกรรม มี 3 ประเภท

จำนวนพฤติกรรมใหญ่ที่ต้องการวัด มี 6 พฤติกรรม

จำนวนพฤติกรรมย่อยที่ต้องการวัด มี 35 พฤติกรรม

ตัวอย่าง

พฤติกรรมที่ต้องการจะวัด ความสามารถ (คำนวณได้) ในการบวก-ลบ คูณ หาร

พฤติกรรมย่อย
1. สามารถบวกเลขจำนวนเต็มได้
2. สามารถบวกเลขเศษส่วนได้
3. สามารถบวกเลขทศนิยมได้
ฯลฯ
12. สามารถหารเลขทศนิยมได้

4. นำพฤติกรรมย่อยแต่ละตัวมาสร้างลักษณะเฉพาะของข้อสอบ ดังตัวอย่าง

ตัวอย่าง ลักษณะเฉพาะของข้อสอบ

พฤติกรรมที่ต้องการวัด

1. ความสามารถในการคิดคำนวณตัวเลขชนิดต่าง ๆ

พฤติกรรมย่อย

1.1 ความสามารถในการบวกเลขจำนวนเต็ม

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

คำอธิบาย

เมื่อกำหนดเลขจำนวนเต็มไม่เกิน 4 หลัก ให้ 3 จำนวน นักเรียนสามารถบวกเลขจำนวนดังกล่าวเข้าด้วยกันอย่างถูกต้อง

ลักษณะคำถาม	ลักษณะของคำตอบ
1. กำหนดจำนวนเลขให้ 3 จำนวน ประกอบด้วย 2 หลัก, 3 หลัก และ 4 หลัก	1. ใช้คำตอบชนิด 4 ตัวเลือก ก. การจัดจำนวนเลขในแต่ละคำตอบจะต้องมีเลขหลักหน่วยตรงกัน ข. ใช้จุลภาคแบ่งจำนวนเลขหลักพัน ค. เรียงลำดับจากมากไปหาน้อย หรือจากน้อยไปหามาก
2. โจทย์คำถาม ให้เขียนเป็นประโยค มีคำว่า "ผลบวก"	
3. เลขจำนวน 3 จำนวน ไม่เรียงกันตามจำนวนหลัก (ไม่ใช่ 2, 3, 4 หรือ 4, 3, 2 หลัก)	
4. การบวกต้องมีการทดอย่างน้อย 2 ครั้ง	2. การสร้างตัวเลือก ก. ตัวเลือกที่บวกถูกต้อง ข. ตัวลวงตัวหนึ่งจะเกิดจากการทดเลขผิดหรือลืมทดเลขอย่างน้อย 1 ครั้ง ค. ตัวเลขที่ลวงตัวหนึ่งเกิดจากการจัดเลขในการบวกไม่ตรงกันในบางหลัก ง. ตัวลวงตัวหนึ่งเกิดจากผลบวกที่มีหลักใดหลักหนึ่งมีค่ามากหรือน้อยความจริงอยู่ 1 จ. อาจใช้ตัวเลือก "ไม่มีคำตอบถูก" ได้ แต่ต้องเอาไว้ที่ตัวเลือกสุดท้าย
5. ในการเลือกจำนวนเลข 3 จำนวนในโจทย์จะไม่มีเลขใดปรากฏมากเกิน 2 ครั้ง	

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างข้อสอบ

ผลบวกของเลขจำนวนต่อไปนี้มีค่าตรงกับคำตอบใด

3649 + 27 + 458

ก. 4,074

ข. 4,174

ค. 4,184

ง. 4,354

ข้อดีและข้อจำกัดของลักษณะเฉพาะของข้อสอบ (บุญเชิด ภิญโญอนันตพงษ์ 2526 : 84)

ข้อดี

1. ระบุสิ่งที่ต้องการวัดได้ละเอียดชัดเจน
2. มีกฎเกณฑ์การเขียนข้อสอบที่เป็นระเบียบรัดกุมและเป็นวิทยาศาสตร์
3. สามารถแปลความหมายคะแนนข้อสอบในลักษณะภาพรวมได้

ข้อจำกัด

1. มีความยุ่งยากและซับซ้อนในการเขียนลักษณะเฉพาะของข้อสอบ
2. การกำหนดกฎเกณฑ์การเขียนตัวเลือกอาจใช้ได้เฉพาะบางเนื้อหา

1.3 ประโยชน์ของการทดสอบอิงโดเมน

ประโยชน์ที่สำคัญของการทดสอบอิงโดเมน คือ การที่สามารถพัฒนากลุ่มตัวอย่างของข้อสอบ เพื่อวัดโดเมนเดียวกันได้เป็นจำนวนมาก คะแนนที่ได้จากการสอบแต่ละกลุ่มโดเมน จะเป็นการประมาณคะแนนที่เป็นอิสระต่อกันของคะแนนโดเมนของผู้เรียน และข้อสอบแต่ละโดเมนของแบบสอบอิงโดเมนนี้ สามารถใช้วัดความก้าวหน้าเป็นระยะ ๆ และสามารถวัดพฤติกรรมที่ซับซ้อนได้

มิลแมน (Millman 1974 : 315) ได้กล่าวสรุปว่า ประโยชน์ที่สำคัญของแบบสอบอิงโดเมน คือ การประมาณคะแนนโดเมนหรือระดับของการปฏิบัติของนักเรียนเป็นเปอร์เซนต์ของประชากรข้อสอบที่นักเรียนตอบได้ถูกต้องหรือในทิศทางที่กำหนดให้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

I 10301999

เซ็นซัน และ ราเบล (Sension and Rabehl 1974 : 22-28) ได้กล่าวถึงประโยชน์ของการทดสอบอิงโดเมน ซึ่งสรุปได้ดังนี้

1. ช่วยในการกำหนดตารางการวิเคราะห์ และการประเมินวัตถุประสงค์และเป้าหมายของการเรียนการสอน

2. ช่วยระบุกิจกรรมการเรียนการสอนตามขั้นตอนที่สัมพันธ์กัน

3. ใช้ในการประเมินผลกระบวนการเรียนการสอน โดยใช้วัดผลสัมฤทธิ์จากการสุ่มเนื้อหา หรือทักษะของแต่ละโดเมนอย่างมีระบบ จึงทำให้ได้ข้อสอบที่เป็นตัวแทนของโดเมนอย่างแท้จริง มีความตรงสูง

ซึ่งข้อมูลต่าง ๆ ที่ได้จากการทดสอบอิงโดเมน จะให้ประโยชน์ทั้งในระดับรายบุคคลและระดับกลุ่ม ดังนี้

<u>ระดับรายบุคคล</u> จากการเอาข้อสอบในแบบสอบอิงโดเมนที่สุ่มจากโดเมนหนึ่งไปสอบเด็ก ถ้าคนใดไม่ผ่านก็อาจใช้แบบสอบคู่ขนาน ซึ่งได้ข้อสอบจากโดเมนเดียวกันไปทดสอบใหม่ หลังจากการสอนซ่อมเสริมแล้ว

<u>ระดับกลุ่ม</u> การทดสอบอิงโดเมน จะให้ข้อมูลที่เฉพาะเจาะจงเกี่ยวกับผลสัมฤทธิ์ของนักเรียนในระดับกลุ่ม ซึ่งจะเป็นการช่วยคณะครูในโรงเรียนในการประเมินและการวางแผนโปรแกรมการเรียนการสอน

มิลเลอร์ (Miller 1974 : 29-35) กล่าวว่า คะแนนโดเมนจะบ่งถึงความมีประสิทธิภาพและประสิทธิผลของโปรแกรม และการสุ่มข้อสอบจากโดเมนของเนื้อหาทั้งหมดตามวิธีการที่ถูกต้อง จะประหยัดเวลาและแรงงานทั้งของผู้เรียน ผู้ตรวจ ตลอดจนผู้ที่จะนำผลการสอบมาประเมินผลโปรแกรมด้วย

สำหรับนักการศึกษาของไทย ก็ได้กล่าวถึงประโยชน์ของการใช้แบบสอบอิงโดเมนไว้ดังนี้ (สงบ ลักษณะ 2525 ก : 1-14)

1. ช่วยให้การวางแผนจัดการเรียนการสอนของแต่ละวิชามีเหตุผล โดยที่เน้นกลุ่มย่อย ๆ ของเนื้อหาและพฤติกรรมสำคัญที่ควรแก่การสอน และการประเมิน แทนที่จะหลงทางด้วยชุดเล็กชุดน้อยของจุดประสงค์เชิงพฤติกรรม

2. การประเมินผลแบบนี้ช่วยให้ได้ข้อมูลหลายระดับ ตั้งแต่ระดับโดเมนที่เราสามารถตรวจสอบคะแนนของโดเมนได้ว่า ผู้เรียนสามารถกระทำอะไรได้เพียงใด ระดับวิชาที่ช่วยให้ทราบระดับความสำเร็จของวิชา ถ้าบกพร่องก็ทราบทันทีว่าบกพร่องในโดเมนใด เพราะมีขอบข่ายเนื้อหาที่ชัดเจนเพื่อจัดกิจกรรมการเรียนการสอน นอกจากนี้ช่วยให้ทราบความสำเร็จระดับหมวดวิชาและระดับหลักสูตรรวมทุกวิชาด้วย

3. เป็นแนวที่เหมาะกับการประเมินหลักสูตร เพราะทราบทั้งระดับความสำเร็จของตัวผู้เรียน ระดับโรงเรียน ระดับจังหวัด เขตและประเทศ ซึ่งสามารถชี้จุดเฉพาะของความสำเร็จหรือความล้มเหลว ทั้งในแง่เทียบกับเกณฑ์สมบูรณ์และ เกณฑ์สัมพันธ์

4. ช่วยให้สามารถปรับปรุงหลักสูตรและการจัดการเรียนการสอนได้ตรงจุดตรงประเด็น

2. มโนทัศน์เกี่ยวกับโมเดลโลจีสติก 3 พารามิเตอร์

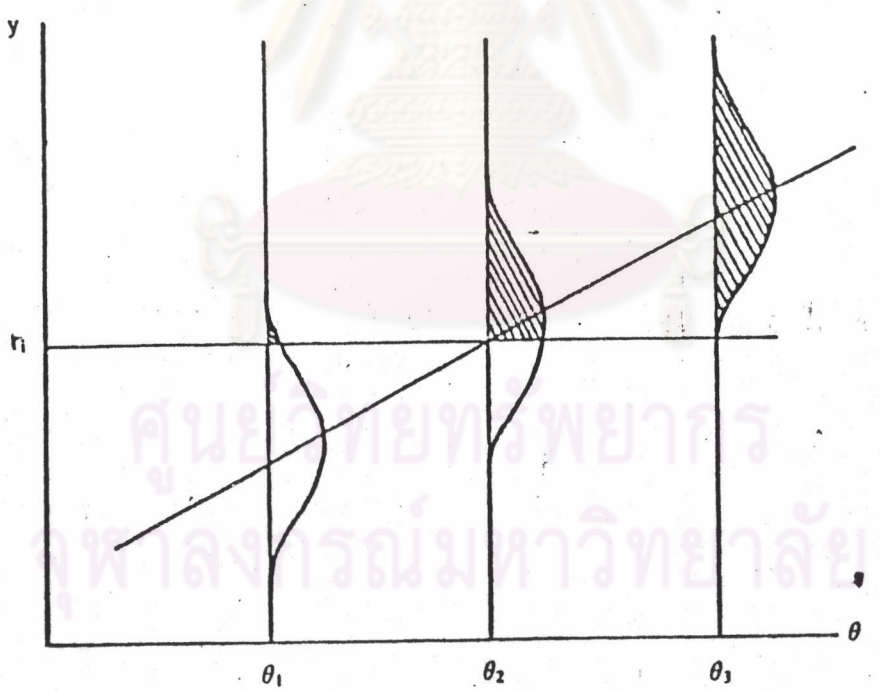
2.1 ความเป็นมาของโมเดลโลจีสติก 3 พารามิเตอร์

โมเดลโลจีสติก 3 พารามิเตอร์ เป็นโมเดลเฉพาะโมเดลหนึ่งของทฤษฎีการตอบสนองข้อสอบ ซึ่งแนวคิดและหลักการเบื้องต้นของทฤษฎีนี้ เฟอร์กูสัน และ ลอเลย์ (Ferguson and Lawley) ได้ริเริ่มขึ้นในปี 1942 และ 1943 (Warm 1979 : 19) ต่อมาในปี 1952 ลอร์ด (Lord) ได้เสนอทฤษฎีในรูปโค้งลักษณะข้อสอบ (Item Characteristic Curve : ICC) ของแต่ละข้อ โดยกล่าวว่าพฤติกรรมการตอบสนองข้อสอบแต่ละข้อของผู้สอบสามารถอธิบายได้ด้วยโค้งลักษณะข้อสอบซึ่งมีลักษณะเป็นโค้งปกติสะสม หรือที่เรียกว่าโมเดลโค้งปกติสะสม (Normal Ogive Model) ซึ่งโมเดลนี้จะกล่าวถึงพารามิเตอร์ 2 ตัว คือ ค่าความยากและอำนาจจำแนก แต่มีปัญหาในการนำแนวคิดของทฤษฎีไปใช้ เนื่องจากมีการคำนวณยุ่งยากและขาดแคลนคอมพิวเตอร์ที่จำเป็นต้องใช้วิเคราะห์ข้อมูลตามทฤษฎี ลอร์ดจึงหยุดความสนใจทฤษฎีนี้ไประยะหนึ่ง

ต่อมาในปี 1960 ราส์ช (Rasch) ได้เสนอแนวคิดหนึ่งของ IRT ในรูปที่มีพารามิเตอร์ตัวเดียว คือ ค่าความยาก ซึ่งรูปแบบนี้เรียกกันทั่วไปว่า ราส์ชโมเดล (Rasch Model) แนวคิดนี้ทำให้มีผู้สนใจ IRT กันมากขึ้น จนในปี 1968 เบิร์นบอม (Birnbaum)

ได้เสนอแนวคิดเกี่ยวกับโมเดลโลจีสติก (Logistic Model) ที่ใช้พารามิเตอร์ 2 ตัว คือ ค่าความยาก และอำนาจจำแนก ซึ่งเป็นโมเดลที่ง่ายกว่าของลอร์ด จึงทำให้โมเดลโลจีสติกเป็นที่นิยมแพร่หลายและมีการพัฒนาขึ้นเรื่อย ๆ จนกระทั่งสามารถใช้กับพารามิเตอร์ตัวเดียว สามตัว และสี่ตัวได้

สำหรับแนวคิดของโมเดลโลจีสติกนั้น มาจากแนวคิดเดียวกับโมเดลโค้งปกติสะสม ซึ่งมุ่งที่จะอธิบายความสัมพันธ์ระหว่างลักษณะภายในตัวบุคคลกับพฤติกรรมที่บุคคลตอบสนองต่อข้อสอบ โดยให้คะแนนที่ได้รับจากการตอบข้อสอบ (Y) แทนพฤติกรรมการตอบสนองต่อข้อสอบ ให้ θ แทนลักษณะหรือความสามารถในตัวบุคคล และ r_i เป็นเกณฑ์ในการทำข้อสอบข้อนั้นว่าถูกหรือผิด ดังนั้นถ้า $Y > r_i$ แสดงว่าผู้สอบทำข้อสอบข้อ i ได้ถูก และถ้า $Y < r_i$ แสดงว่าผู้สอบทำข้อสอบข้อ i ผิด ความสัมพันธ์ระหว่างความสามารถ (θ) กับพฤติกรรมการตอบสนองข้อสอบ (Y) แสดงได้ดังแผนภาพที่ 1



แผนภาพที่ 1 ความสัมพันธ์ระหว่างความสามารถกับพฤติกรรมการตอบสนองข้อสอบ

จากภาพจะเห็นว่า ถ้านำโอกาสการตอบถูก (พื้นที่ส่วนที่แรเงา) ในระดับความสามารถต่าง ๆ มาเขียนกราฟใหม่ จะได้โค้งลักษณะข้อสอบ (ICC) เป็นรูปต่าง ๆ ทั้งนี้ขึ้นอยู่กับฟังก์ชั่นทางคณิตศาสตร์ และจำนวนพารามิเตอร์ที่จะใช้อธิบายแต่ในการอธิบายโค้งคุณลักษณะข้อสอบ โดยใช้ฟังก์ชั่นโลจีสติก จะมีวิธีการคำนวณทางคณิตศาสตร์ได้ง่ายและสะดวกกว่าการใช้ฟังก์ชั่นโค้งปกติสะสม นอกจากนี้แล้วในสถานการณ์สอบจริงอาจจะมีผู้ที่มีความสามารถสูงตอบข้อสอบผิดด้วยความเลินเล่อ กรณีเช่นนี้ โมเดลโลจีสติกจะมีความแกร่งต่อข้อมูลแบบนี้มากกว่าโมเดลปกติสะสม จึงทำให้โมเดลโลจีสติกเป็นที่นิยมกันมากในการปฏิบัติงานจริง (Lord 1980 : 14)

2.2 <u>ข้อตกลงเบื้องต้นของโมเดลโลจีสติก 3 พารามิเตอร์</u>

เนื่องจาก โมเดลโลจีสติก 3 พารามิเตอร์ เป็นรูปแบบหนึ่งของโมเดลโลจีสติก ซึ่งเป็นโมเดลเฉพาะของทฤษฎี IRT ดังนั้นข้อตกลงเบื้องต้นที่สำคัญจึงเหมือนกับข้อตกลงเบื้องต้นของทฤษฎี IRT ดังนี้

1. Unidimensionality เป็นการสมมุติว่าข้อสอบในแบบสอบมีลักษณะเป็นเอกพันธ์ นั่นคือ แบบสอบนั้นจะต้องมุ่งวัดความสามารถอย่างใดอย่างหนึ่งเพียงความสามารถเดียว ส่วนการตรวจสอบว่าข้อสอบมีคุณสมบัติดังกล่าวหรือไม่ อาจทำได้โดยการวิเคราะห์องค์ประกอบ (Factor Analysis)

2. Local Independence เป็นข้อตกลงเกี่ยวกับความเป็นอิสระในการตอบสนองต่อข้อสอบ ซึ่งมี 2 ลักษณะ คือ

2.1 มีความเป็นอิสระทางสถิติ (Statistically independence) กล่าวคือ ข้อสอบแต่ละข้อเป็นอิสระไม่เกี่ยวข้องกัน ดังนั้นคำตอบในแต่ละข้อของแต่ละคนเป็นอิสระต่อกัน แต่รวมกันแล้วจะวัดคุณลักษณะเดียวเท่านั้น

2.2 มีความเป็นอิสระจากตำแหน่ง (uncorrelated independence) กล่าวคือ ข้อสอบแต่ละข้อจะปรากฏอยู่ในตำแหน่งใดของข้อสอบก็ได้ จะไม่มีผลต่อการตอบของผู้สอบ

ลอร์ด (Lord 1980 : 19) ได้กล่าวว่า เมื่อแบบสอบมีคุณสมบัติของ Unidimensionality แล้ว จะมีคุณสมบัติ Local Independence ด้วย ดังนั้นไม่จำเป็นต้องกล่าวแยกกันก็ได้

3. Dichotomous กล่าวคือ การให้คะแนนการตอบข้อสอบ เป็น 0 หรือ 1 เท่านั้น

ในการนำทฤษฎี IRT มาใช้ให้สอดคล้องกับข้อตกลงเบื้องต้นดังกล่าว จะมีความเหนือกว่าและช่วยแก้จุดอ่อนของทฤษฎีการวัดแบบคลาสสิคอล (Classical Test Theory) ได้ดังนี้

1. ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ (Invariant of Item Parameter) กล่าวคือ ค่าพารามิเตอร์ของข้อสอบ ได้แก่ ค่าอำนาจจำแนก ค่าความยากและค่าการเดา จะไม่แปรเปลี่ยนไปตามกลุ่มของผู้สอบ

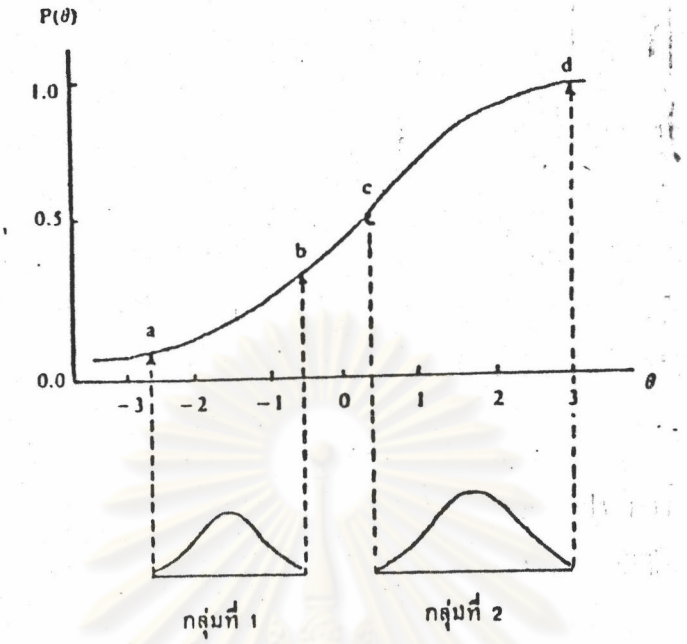
2. จะใช้ข้อสอบกับใครก็ได้ (Person Free) กล่าวคือ ไม่ว่าจะนำข้อสอบไปใช้สอบกับกลุ่มบุคคลใด โค้งลักษณะข้อสอบก็จะคงเดิม

3. จะใช้ข้อสอบข้อใดก็ได้ (Item Free) ในการประมาณค่าความสามารถแท้ (θ) ของผู้สอบ จะใช้ข้อสอบชุดใด จำนวนเท่าใดก็ได้ ไม่จำเป็นต้องใช้แบบสอบชุดเดียวกันหรือแบบสอบคู่ขนาน บางครั้งอาจใช้ข้อสอบเพียง 3-5 ข้อ ก็สามารถประมาณค่าความสามารถแท้ของผู้สอบได้แล้ว

4. ค่าความคลาดเคลื่อนมาตรฐานในการวัดของผู้สอบแต่ละคน สามารถประมาณได้โดยอิสระจากกัน ทำให้การประมาณค่าความสามารถของผู้เข้าสอบแต่ละคนมีความแม่นยำสูง

เกี่ยวกับความไม่แปรเปลี่ยนของค่าพารามิเตอร์นั้นสามารถอธิบายได้ด้วยแผนภาพที่ 2

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



แผนภาพที่ 2 ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบที่ประมาณค่าจากกลุ่มที่แตกต่างกัน

จากภาพแสดงให้เห็นว่า เส้นโค้งลักษณะข้อสอบในช่วง a-b จะอธิบายถึงโอกาสในการตอบข้อสอบข้อนี้ได้ถูกของผู้สอบที่มีความสามารถต่ำ (กลุ่มที่ 1) และเส้นโค้งลักษณะข้อสอบในช่วง c-d ก็แสดงถึงโอกาสในการตอบข้อสอบได้ถูกของผู้สอบที่มีความสามารถสูง (กลุ่มที่ 2) ถ้าข้อสอบข้อนี้วัดความสามารถอันเดียวกันของกลุ่มตัวอย่างทั้งสองกลุ่มแล้ว (Unidimensionality) เส้นโค้งลักษณะข้อสอบย่อมมีเส้นเดียว เส้นโค้งในช่วง a-b และช่วง c-d ย่อมเป็นส่วนหนึ่งของเส้นโค้งลักษณะข้อสอบข้อนี้ ดังนั้นตัวพารามิเตอร์ ซึ่งเป็นตัวกำหนดรูปร่างลักษณะของเส้นโค้งย่อมมีได้ค่าเดียว

2.3 <u>รูปแบบของโมเดลโลจีสติก</u>

โมเดลโลจีสติก เป็นโมเดลทางคณิตศาสตร์ ที่สำคัญโมเดลหนึ่งของทฤษฎี IRT ซึ่งได้รับการพัฒนาและมีการนำไปใช้มากที่สุดในสถานการณ์ของการสอบ แบ่งออกเป็น 4 โมเดลย่อย ตามจำนวนพารามิเตอร์ข้อสอบดังนี้ (Hambleton 1985 : 35-49)

1. <u>โมเดลโลจิสติกที่มีพารามิเตอร์ 3 ตัว</u> (Three-Parameter Logistic Model)

เป็นโมเดลที่มีพารามิเตอร์อธิบายโค้งลักษณะข้อสอบ 3 ตัว ซึ่งมีรูปสมการดังนี้

$$P_i(\Theta) = c_i + (1-c_i)\frac{e^{Da_i(\Theta-b_i)}}{1+e^{Da_i(\Theta-b_i)}}$$

หรือ $P_i(\Theta) = c_i + (1-c_i)\left[1+e^{-Da_i(\Theta-b_i)}\right]^{-1}$, $(i = 1,2..,n)$

เมื่อ $P_i(\Theta)$ คือ ความน่าจะเป็นที่ผู้สอบที่มีระดับความสามารถ Θ ตอบข้อสอบข้อ i ถูกต้อง

Θ คือ ระดับความสามารถของผู้สอบที่ประมาณได้จากคะแนนรวมของการตอบแบบสอบ โดยการปรับให้เป็นคะแนนที่มีค่าเฉลี่ยเท่ากับ 0 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 โดยมีพิสัยการกระจายอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัติระดับความสามารถจะมีค่าอยู่ระหว่าง -3 ถึง +3 ซึ่งมีระดับความสามารถ -3 หมายถึง ผู้สอบมีระดับความสามารถต่ำมาก และระดับความสามารถ +3 หมายถึงผู้สอบมีระดับความสามารถสูงมาก

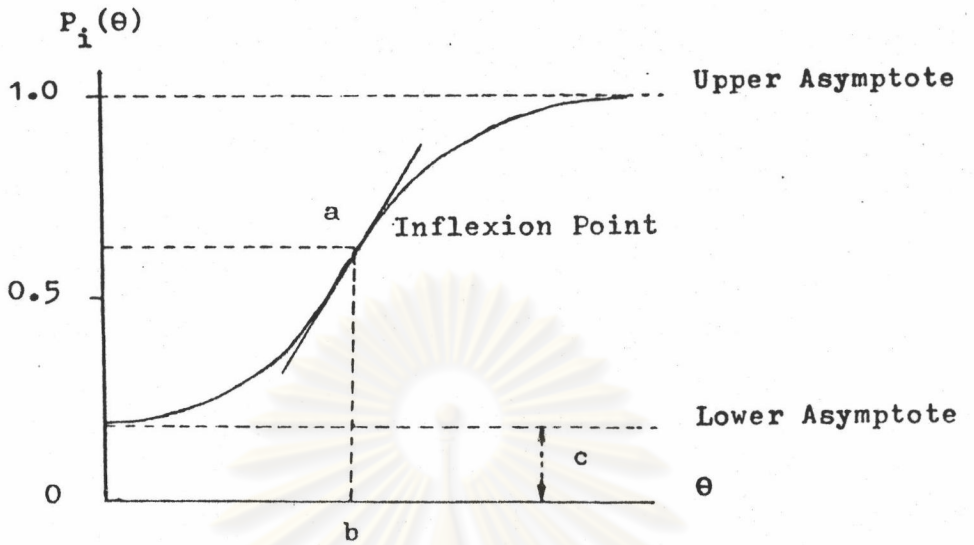
a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อ i ที่เป็นสัดส่วนโดยตรงกับความชันของโค้งลักษณะข้อสอบ ณ จุดเปลี่ยนโค้ง หรือที่จุด $\Theta = b_i$ โดยทั่วไปพิสัยของค่าอำนาจจำแนกของข้อสอบอยู่ระหว่าง 0 ถึง 2

b_i คือ ค่าความยากของข้อสอบข้อ i ที่แสดงถึงระดับความสามารถ ที่จุดโค้งลักษณะข้อสอบมีความชันมากที่สุด โดยทั่วไปมีพิสัยของค่าความยากของข้อสอบอยู่ระหว่าง -2 ถึง +2 ซึ่งค่าที่ใกล้ -2 หมายถึง ข้อสอบที่ง่ายมาก และค่าที่ใกล้ +2 หมายถึง ข้อสอบที่ยากมาก

c_i คือ ค่าการเดาของข้อสอบข้อ i ซึ่งเป็นค่าความน่าจะเป็นที่ผู้สอบมีความสามารถต่ำมาก จะมีโอกาสตอบข้อสอบข้อ i ถูก หรือเป็นค่ากำกับโค้งที่ต่ำสุด (Lower asymptote) ของโค้งลักษณะข้อสอบมีค่า 0 ถึง 1

D คือ a scaling factor มีค่าเท่ากับ 1.7

e คือ ค่าคงที่ มีค่าเท่ากับ 2.71828...



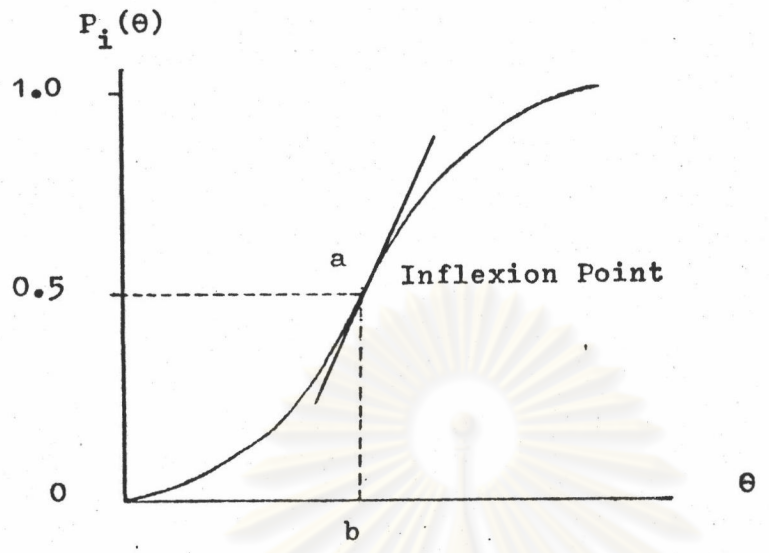
แผนภาพที่ 3 ความหมายของค่าพารามิเตอร์ของข้อสอบ 3 พารามิเตอร์

2. <u>โมเดลโลจิ่สติกที่มีพารามิเตอร์ 2 ตัว</u> (Two-Parameter Logistic Model)

เป็นโมเดลที่กำหนดให้ค่าการเดาของข้อสอบเป็น 0 รูปสมการก็จะเหลือค่าพารามิเตอร์เพียง 2 ตัว คือ ค่าอำนาจจำแนกของข้อสอบ และค่าความยากของข้อสอบดังนี้

$$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}} \quad \text{หรือ}$$

$$P_i(\theta) = \left[1 + e^{-Da_i(\theta-b_i)}\right]^{-1} \quad ; (i = 1, 2,..,n)$$



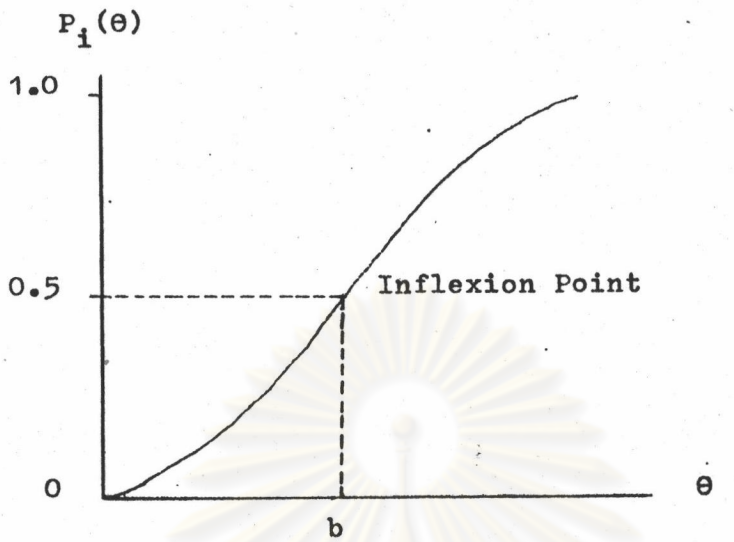
แผนภาพที่ 4 ความหมายของค่าพารามิเตอร์ของข้อสอบ 2 พารามิเตอร์

3. <u>โมเดลโลจีสติกที่มีพารามิเตอร์ 1 ตัว</u> (One-Parameter Logistic Model)

เป็นโมเดลที่มีข้อตกลงเบื้องต้น คือ ค่าอำนาจจำแนกของข้อสอบทุกข้อมีค่าเท่ากัน และค่าการเดาของข้อสอบเท่ากับ 0 รูปสมการก็จะเหลือค่าพารามิเตอร์ความยากเพียง 1 ตัว มีที่มีค่าแปรเปลี่ยนดังนี้

$$P_i(\theta) = \frac{e^{D(\theta-b_i)}}{1+e^{D(\theta-b_i)}} \quad \text{หรือ}$$

$$P_i(\theta) = \left[1 + e^{-D(\theta-b_i)}\right]^{-1} \quad ; (i = 1, 2,..,n)$$



แผนภาพที่ 5 ความหมายของค่าพารามิเตอร์ของข้อสอบ 1 พารามิเตอร์

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

4. โมเดลโลจิสติกที่มีพารามิเตอร์ 4 ตัว (Four-Parameter Logistic Model)

เป็นโมเดลที่แตกต่างจากโมเดลโลจีสติกที่มีพารามิเตอร์ 3 ตัว โดยเพิ่มพารามิเตอร์อีก 1 ตัว คือ Y_i ซึ่งมีค่าต่ำกว่า 1 เล็กน้อย ซึ่งหมายถึง ค่าที่ผู้สอบที่มีความสามารถสูงมีความสะเพร่าตอบข้อสอบไม่ถูก แต่เป็นโมเดลที่ยังไม่สามารถหาวิธีปฏิบัติได้ โดยมีรูปสมการดังนี้

$$P_i(\Theta) = c_i + (y_i - c_i) \frac{e^{Da_i(\Theta - b_i)}}{1+e^{Da_i(\Theta - b_i)}} \quad \text{หรือ}$$

$$P_i(\Theta) = c_i + (Y_i - c_i)\left[1 + e^{-Da_i(\Theta - b_i)}\right]^{-1} \quad ; (i = 1,2,..,n)$$

3. การใช้โมเดลโลจีสติก 3 พารามิเตอร์ ในการพัฒนาแบบสอบอิงโดเมน

3.1 การประมาณค่าความสามารถและค่าพารามิเตอร์ของข้อสอบ

ในการประมาณค่าความสามารถ (Θ) และค่าพารามิเตอร์ของข้อสอบนั้น ทำได้โดยการวิเคราะห์ข้อสอบด้วยคอมพิวเตอร์ โปรแกรม LOGIST Version 1.0 ซึ่งพัฒนาขึ้นโดย Wood, Wingersky และ Lord ในปี 1976 โดยโปรแกรมจะใช้วิธีประมาณค่าแบบความเป็นไปได้สูงสุด (Maximum Likelihood Estimation) ซึ่งวิธีการประมาณค่าแบบนี้มีคุณสมบัติที่สำคัญและมีประโยชน์ภายใต้สภาพการณ์ทั่ว ๆ ไปดังนี้ คือ (Hambleton and Swaminathan 1985 : 88-89)

1. มีความคงที่ (Consistent) ได้แก่ เมื่อขนาดของกลุ่มตัวอย่างและจำนวนข้อสอบเพิ่มขึ้น ตัวประมาณค่าจะเป็นค่าที่ถูกต้อง

2. ทำหน้าที่ของค่าสถิติที่เพียงพอ เมื่อมีค่าสถิติที่เพียงพอ (Functions of sufficient statistics when sufficient statistics exist) กล่าวคือ ค่าสถิติที่เพียงพอจะมีสารสนเทศทั้งหมดเกี่ยวกับค่าพารามิเตอร์ ไม่จำเป็นต้องใช้ข้อมูลใด ๆ เพิ่มเติม

3. มีประสิทธิภาพ (Efficient) นั่นคือ ตัวประมาณค่าแบบความเป็นไปได้สูงสุด มีความแปรปรวนน้อยที่สุด

4. มีลักษณะของการกระจายเข้าใกล้โค้งปกติ (Asymptotically normally distributed)

สำหรับโมเดลโลจีสติก 3 พารามิเตอร์ จะทำการประมาณค่าพารามิเตอร์ของข้อสอบแต่ละข้อ 3 ตัว คือ a_i, b_i และ $c_i (i = 1, 2, .., n)$ และค่าความสามารถของผู้สอบแต่ละคน (θ_a) เมื่อ $a = 1, 2, .., N$ ซึ่งมีฟังก์ชั่นในการประมาณค่าพารามิเตอร์ดังกล่าว เรียกว่า ฟังก์ชั่นความน่าจะเป็น (Likelihood Function) แสดงได้ดังสมการ

$$L(U/\theta, b, a, c) = \prod_{a=1}^{N} \prod_{i=1}^{n} P_{ia}^{U_{ia}} Q_{ia}^{1-U_{ia}} \quad \ldots\ldots(1)$$

เมื่อ P_{ia} คือ โอกาสที่ผู้สอบคนที่ a จะตอบข้อสอบข้อที่ i ได้ถูก หรือคือ $P_i(\theta)$ นั่นเอง

Q_{ia} คือ $1 - P_{ia}$

n คือ จำนวนข้อสอบ

N คือ จำนวนผู้สอบ

U_{ia} คือ คะแนนในข้อที่ i ของผู้สอบคนที่ a โดยถ้าตอบถูกจะมีค่าเท่ากับ 1 ตอบผิดมีค่าเท่ากับ 0 และถ้าเว้นไม่ตอบจะมีค่าเท่ากับ 1/A เมื่อ A คือจำนวนตัวเลือกของข้อสอบข้อที่ i

การพิจารณาการประมาณความน่าจะเป็นสูงสุดของค่าพารามิเตอร์ θ, a, b และ c โดยการทำให้ค่าพารามิเตอร์เหล่านี้มีค่าสูงสุดในลอการิทึม (Logarithm) ของฟังก์ชั่นความน่าจะเป็นดังนี้ คือ

$$\ln L(U/\theta, b, a, c) = \sum_{a=1}^{N} \sum_{i=1}^{n} (U_{ia} \ln P_{ia} + (1-U_{ia}) \ln Q_{ia}) \ldots(2)$$

แต่เนื่องจาก (2) ไม่เป็นสมการเส้นตรง ในการคำนวณค่าพารามิเตอร์ดังกล่าว จึงต้องใช้วิธีการประมาณค่าซ้ำกันหลาย ๆ ครั้ง จนกระทั่งค่าประมาณจะเป็นอันหนึ่งอันเดียวกัน ในการประมาณค่าซ้ำ ๆ กันนี้ จะใช้วิธีของนิวตัน (Newton-Raphson Procedure) ดังสมการ

$$\theta_{m+1} = \theta_m - \left[\frac{d}{d\theta} \ln L(U/\theta,b,a,c)\right]_m / \left[\frac{d^2}{d\theta^2} \ln L(U/\theta,b,a,c)\right]_m \quad ...(3)$$

เมื่อ θ_m และ θ_{m+1} คือ ผลลัพธ์การประมาณค่าขั้นที่ m และ m+1 ตามลำดับ ซึ่งเมื่อผลการประมาณค่าใน 2 ขั้นนี้แตกต่างกันเพียงเล็กน้อย ก็จะหยุดประมาณถือว่ามีค่าคงที่

กล่าวโดยสรุปการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบมี 2 ขั้นตอน คือ การประมาณค่าความสามารถของผู้สอบ N คน โดยการปรับค่าความสามารถในการทำซ้ำแต่ละครั้ง จนมีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 และการนำค่าความสามารถนี้ไปประมาณค่าพารามิเตอร์ของข้อสอบ n ข้อ ซึ่งในโปรแกรมโลจีสต์ มีการทำงาน 4 ขั้นตอนดังนี้

1. กำหนดให้ค่าอำนาจจำแนกของข้อสอบทุกข้อเท่ากับ 1 และค่าการเดาของข้อสอบเท่ากับ 1/A-0.05 เมื่อ A แทน จำนวนตัวเลือกของข้อสอบ แล้วทำการประมาณค่าความสามารถของผู้สอบแต่ละคน และค่าความยากของข้อสอบแต่ละข้อ

2. นำค่าความสามารถของผู้สอบแต่ละคนจากขั้นที่ 1 ไปประมาณค่าอำนาจจำแนก ค่าความยาก และค่าการเดาของข้อสอบแต่ละข้อ

3. นำค่าอำนาจจำแนกและค่าการเดาของข้อสอบจากขั้นที่ 2 ไปประมาณค่าความสามารถของผู้สอบแต่ละคน และค่าความยากของข้อสอบแต่ละข้อ

4. นำค่าความสามารถของผู้สอบแต่ละคนจากขั้นที่ 3 ไปประมาณค่าอำนาจจำแนก ค่าความยาก และค่าการเดาของข้อสอบแต่ละข้อ

ในการประมาณค่าความสามารถแบบเป็นไปได้สูงสุด เมื่อทราบค่าพารามิเตอร์ของข้อสอบแล้ว จะใช้การประมาณค่าความสามารถอย่างมีเงื่อนไข (Conditional Maximum Likelihood Estimation of ability) โดยกำหนดค่าความสามารถเริ่มต้นสำหรับผู้สอบคนที่ a เป็น (Hambleton and Swaminathan 1985 : 84-86)

$$\theta_{oa} = \ln\left[r_a/(n-r_a)\right] \quad ...(4)$$

เมื่อ r_a = จำนวนข้อสอบที่ผู้สอบคนที่ a ตอบถูก

n = จำนวนข้อสอบที่ผู้สอบคนที่ a ทำทั้งหมด

ln = natural logarithm

และหาค่า Θ ครั้งที่ m+1 จากสมการ

$$\Theta_{m+1} = \Theta_m - h_m \quad \ldots\ldots(5)$$

โดยที่ $$h_m = \left[\frac{\partial LnL}{\partial \Theta_a}\right] \Big/ \left[\frac{\partial^2 LnL}{\partial \Theta_a^2}\right] \quad \ldots\ldots(6)$$

เมื่อ $h_m = .001$ จึงสิ้นสุดกระบวนการทำซ้ำ

และสำหรับรูปแบบการตอบข้อสอบที่มีจำนวนพารามิเตอร์ 3 ตัว สามารถประมาณค่า h_m ได้จากสมการ

$$h_m = D \sum_{i=1}^{n} \frac{a_i(P_{ia}-c_i)(U_{ia}-P_{ia})}{(1-c_i)P_{ia}} \Big/ D^2 \sum_{i=1}^{n} \frac{a_i^2 (P_{ia}-c_i)(U_{ia}c_i - P_{ia}^2) Q_{ia}}{P_{ia}^2 (1-c_i)^2} \quad \ldots(7)$$

เมื่อพิจารณาค่าอนุพันธ์ครั้งที่ 1 $(\frac{\partial \ln L}{\partial \Theta_a})$ และครั้งที่ 2 $(\frac{\partial^2 \ln L}{\partial \Theta_a^2})$ ของฟังก์ชั่นความน่าจะเป็นสูงสุด ของโมเดลโลจีสติก 3 พารามิเตอร์ จะได้

$$\frac{\partial \ln L}{\partial \Theta_a} = \sum_{i=1}^{n} k_i u_{ia} - \sum_{i=1}^{n} k_i P_{ia} = 0 \quad (a = 1,\ldots,N) \ldots(8)$$

โดยที่ $k_i = Da_i(P_{ia} - c_i) / P_{ia}(1 - c_i)$

เนื่องจากค่าอำนาจจำแนกของข้อสอบ (a_i) มีค่าเป็นบวก และในโมเดล 3 พารามิเตอร์นี้ $P_{ia} - c_i$ ไม่เป็นลบ (เพราะ c_i เป็นขีดจำกัดล่าง) K_i เป็นจำนวนเชิงบวก ซึ่งถ้าผู้สอบตอบข้อสอบผิดทุกข้อ $u_i = 0$ สำหรับ $i = 1,\ldots,n$ ดังนั้นสมการความเป็นไปได้ในสมการที่ 8 จึงลดลงเป็น

$$\sum_{i=1}^{n} k_i P_{ia} = 0 \quad \ldots\ldots(9)$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

และเนื่องจาก k_i มีค่าเป็นบวก และ P_{ia} คือ ความน่าจะเป็นของการตอบถูก สมการนี้จึงใช้ได้เมื่อ $\theta_a = -\infty$ เท่านั้น ในทำนองเดียวกัน เมื่อผู้สอบตอบข้อสอบทุกข้อ $u_{ia} = 1$ สำหรับข้อสอบทุกข้อ และสมการความเป็นไปได้จะลดลงเป็น

$$\sum_{i=1}^{n} k_i = \sum_{i=1}^{n} k_i P_{ia}$$

และใช้ได้เมื่อ $\theta_a = +\infty$ จึงปรากฏว่า ตัวประมาณค่าแบบความเป็นไปได้สูงสุดจะใช้ไม่ได้ในกรณีเหล่านี้ ผู้สอบที่ได้คะแนนเต็มหรือศูนย์ จะถูกกำจัดออกจากกระบวนการประมาณค่าแบบนี้ ซึ่ง ลอร์ด (Lord 1980 อ้างถึงใน Hambleton and Swaminathan 1985 : 309) ได้เสนอว่า ค่าประมาณของความสามารถ ($\hat{\theta}$) อาจถูกถ่ายโยง (transform) ไปยังมาตราคะแนน (score metric) โดยทางโค้งคุณลักษณะแบบสอบ (Test Characteristic Curve) โดยคะแนนจะอยู่ในช่วง 0 ถึง n เมื่อ n คือ จำนวนข้อสอบ ซึ่งการถ่ายโยงนี้จะเป็นการเลี่ยงปัญหาของการที่ไม่สามารถประมาณค่าของผู้สอบที่ได้คะแนนเต็ม หรือศูนย์ได้

3.2 การประมาณค่าคะแนนโดเมน และความสัมพันธ์กับค่าความสามารถ

จุดมุ่งหมายที่สำคัญประการหนึ่งของแบบสอบอิงโดเมน ก็คือ มุ่งประมาณค่าคะแนนโดเมน (Domain Score : π) ได้อย่างน่าเชื่อถือ ซึ่งจากแนวคิดของทฤษฎี IRT นั้น คะแนนโดเมน หมายถึง ค่าคาดหวัง (Expected Value) ของคะแนนจากข้อสอบทุกข้อในโดเมน (Hambleton and De Gruijter 1983 : 358) และจะมีความสัมพันธ์กับคะแนนความสามารถ (θ) ซึ่งแสดงได้ด้วยฟังก์ชั่นคุณลักษณะแบบสอบ (Test Characteristic Function) ในรูปที่ไม่ใช่การถ่ายโยงเชิงเส้น (non-linear transformation) โดยสามารถอธิบายได้ดังนี้ (Hambleton and Swaminathan 1985 : 61-62)

ให้ $\hat{\pi}$ คือ Observed Domain Score หรือสัดส่วนของคะแนนการตอบถูกที่สังเกตได้ และใช้เป็นค่าประมาณของคะแนนโดเมน (π) ซึ่งกำหนดได้เป็น

$$\hat{\pi} = \frac{1}{n} \sum_{i=1}^{n} u_i$$

และจากนิยามของคะแนนโดเมน หรือ คะแนนจริง (True score)

$$E(\hat{\pi}) = \pi$$

ดังนั้น $$E(\hat{\pi}) = \pi = \frac{1}{n} \sum_{i=1}^{n} E(u_i)$$

และสำหรับผู้สอบที่มีระดับความสามารถ Θ

$$\hat{\pi}/\Theta = \frac{1}{n} \sum_{i=1}^{n} (u_i/\Theta) \quad \ldots\ldots\ldots\ldots\ldots\ldots (13)$$

และ $$E(\hat{\pi}/\Theta) = \pi/\Theta$$

$$= \frac{1}{n} \sum_{i=1}^{n} E(u_i/\Theta)$$

เมื่อ u_i คือตัวแปรเชิงสุ่ม ที่มีค่า 0 หรือ 1

ดังนั้น $$E(u_i/\Theta) = (u_i = 1)\ P_i(u_i = 1/\Theta) +$$

$$(u_i = 0)\ P_i(u_i = 0/\Theta)$$

$$= 1.\ P_i(u_i = 1/\Theta) + 0.P_i(u_i = 0/\Theta)$$

$$= P_i(u_i = 1/\Theta)$$

$$= P_i(\Theta)$$

จะได้ $$\pi/\Theta = \frac{1}{n} \sum_{i=1}^{n} P_i(\Theta) \quad \ldots\ldots\ldots\ldots\ldots\ldots (14)$$

เมื่อ n คือ จำนวนข้อสอบในแบบสอบ

$P_i(\Theta)$ คือ โอกาสที่ผู้สอบซึ่งมีความสามารถระดับ Θ จะทำข้อสอบข้อได้ถูกต้อง

สมการนี้เรียกว่า ฟังก์ชั่นคุณลักษณะแบบสอบ ซึ่งสำหรับโมเดลโลจีสติกที่มี 3 พารามิเตอร์

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

จะสามารถคำนวณค่า $P_i(\Theta)$ ได้จากสูตร ดังนี้

$$P_i(\Theta) = c_i + (1-c_i)\frac{e^{Da_i(\Theta-b_i)}}{1+e^{Da_i(\Theta-b_i)}} \text{ หรือ} \quad \ldots\ldots(15)$$

$$P_i(\Theta) = c_i + (1-c_i)\left[1+e^{-Da_i(\Theta-b_i)}\right]^{-1} \quad \ldots\ldots(16)$$

โดยทั่วไป ถ้าแบบสอบมีค่าอำนาจจำแนกไม่สูงมาก โค้งคุณลักษณะแบบสอบจะเกือบเป็นเส้นตรง ในช่วงค่าความสามารถระดับกลาง ๆ ถ้าระดับความสามารถของผู้สอบตกในช่วงกลางของสเกลความสามารถ จะมีการทำให้ผิดส่วนเล็กน้อย อย่างไรก็ดี ถ้ามีการกระจายของความสามารถในช่วงกว้าง ๆ ค่าคะแนนโดเมนที่ประมาณในช่วงความสามารถกลาง ๆ จะถี่ และเพื่อให้คะแนนโดเมนมีการกระจายให้มากเท่าที่เป็นไปได้ ควรใช้โค้งคุณลักษณะแบบสอบที่ชันมาก ๆ เท่าที่จะเป็นไปได้ในช่วงความสามารถกลาง ๆ ซึ่งสามารถสร้างแบบสอบในลักษณะเช่นนี้ได้ด้วยข้อสอบที่มีค่าอำนาจจำแนกสูง และค่าความยากในระดับกลาง ๆ และถ้าต้องการให้คะแนนโดเมนมีการกระจายที่ช่วงความสามารถสูง ๆ แบบสอบต้องประกอบด้วยข้อสอบที่มีค่าความยากสูงและค่าอำนาจจำแนกสูง ๆ

ลักษณะดังกล่าวจะมีความตรง เมื่อทราบคะแนนโดเมนที่แท้จริง (True domain score : π) ซึ่งเราจะไม่สามารถหา π ได้ แต่เราสามารถกำหนดค่าประมาณของคะแนนโดเมน ซึ่งเป็นคะแนนโดเมนที่สังเกตได้ (Observed Domain Score : $\hat{\pi}$) จากสมการ

$$\hat{\pi}/\hat{\Theta} = \frac{1}{n}\sum_{i=1}^{n} P_i(\hat{\Theta}) \quad \ldots\ldots(17)$$

ให้ระลึกว่า $\hat{\pi}/\Theta$ ที่กำหนดไว้ในสมการ 13 เป็นค่าประมาณ π ที่ไม่คลาดเคลื่อน (Unbias) ของคะแนนที่สังเกตได้ (Observe score) คะแนนโดเมนของประชากรกำหนดไว้ในสมการ 14 ส่วนค่า $\hat{\pi}/\hat{\Theta}$ ที่กำหนดไว้ในสมการ 17 ซึ่งอยู่ในเทอมของโค้งคุณลักษณะแบบสอบ ไม่ใช่ค่าประมาณคะแนนโดเมนที่ไม่คลาดเคลื่อน เพราะ $P_i(\hat{\Theta})$ เป็นการถ่ายโยงที่ไม่เป็นเส้นตรง (non-linear transformation) ของค่าความสามารถ อย่างไรก็ดี ถ้าแบบสอบมีความยาว $\hat{\Theta} \longrightarrow \Theta$ $\hat{\pi}$ สามารถใช้เป็น π และจะมีความคล้ายคลึงกันมาก ระหว่างการกระจายของคะแนนโดเมนที่แท้จริง (π) และคะแนนโดเมนที่สังเกตได้ ($\hat{\pi}$)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ดังนั้น ตามมโนทัศน์ของโมเดลโลจีสติก 3 พารามิเตอร์ เมื่อทราบค่าพารามิเตอร์ของข้อสอบหลายข้อที่อยู่ในโดเมนเดียวกัน หรือทราบค่าพารามิเตอร์ของข้อสอบชุดหนึ่งที่เป็นตัวแทนที่ดีของโดเมน จะสามารถประมาณคะแนนโดเมนได้จากค่าเฉลี่ยของความน่าจะเป็นในการตอบข้อสอบได้ถูกของข้อสอบทุกข้อในแบบสอบ ที่ระดับความสามารถต่าง ๆ ได้จากสมการที่ 14 และ 15 หรือ 16 หรือสมการที่ 17 ดังกล่าวแล้ว และระดับของเกณฑ์มาตรฐานก็จะมีความสัมพันธ์กันในลักษณะเดียวกัน คือ

$$\pi_0 = \frac{1}{m} \sum_{i=1}^{m} P_i(\theta_0) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (18)$$

เมื่อ $P_i(\theta_0)$ คือ โอกาสที่ผู้สอบซึ่งมีความสามารถตรงกับระดับเกณฑ์มาตรฐานจะทำข้อสอบข้อ i ได้ถูก

สำหรับคะแนนโดเมนจะเป็นค่าสถิติที่ใช้คาดคะเนระดับผลสัมฤทธิ์ทางการเรียนที่ไม่คลาดเคลื่อนและพอเพียงได้ ก็ต่อเมื่อมีการกำหนดกลุ่มของข้อสอบที่เป็นไปได้ในแต่ละโดเมนไว้ล่วงหน้า และข้อสอบในแบบสอบอิงโดเมน ต้องเป็นกลุ่มตัวอย่างแบบสุ่มของข้อสอบในโดเมนนั้น และเมื่อข้อสอบที่รวมกันเป็นแบบสอบมีความเป็นตัวแทนของกลุ่มข้อสอบจากโดเมนข้อสอบที่มุ่งวัดคุณลักษณะของผู้สอบ ฟังก์ชั่นคุณลักษณะแบบสอบจะถ่ายโยง (transform) ค่าประมาณคะแนนความสามารถ ($\hat{\theta}$) ไปในความหมายของค่าประมาณคะแนนโดเมน ($\hat{\pi}$) ดังนั้นคะแนนความสามารถจะให้พื้นฐานสำหรับการแปลความหมายอิงเนื้อหาของคะแนนแบบสอบของผู้สอบ เช่น บอกถึงระดับปฏิบัติ (สัดส่วนของการรอบรู้เนื้อหา ของผู้สอบ ที่สัมพันธ์กับจุดประสงค์ที่มุ่งวัดจากแบบสอบ และการแปลความหมายนี้จะให้ความหมายที่ไม่พาดพิงถึงการปฏิบัติ (performance) ของผู้สอบคนอื่น ๆ สเกลของคะแนนโดเมนจะกำหนดไว้ในช่วง $[0, 1]$ และจะขึ้นอยู่กับกลุ่มข้อสอบที่ใช้ ส่วนคะแนนความสามารถจะอยู่ในช่วง $(-\infty, \infty)$ และเป็นอิสระจากข้อสอบที่ใช้

3.3 การตรวจสอบคุณภาพแบบสอบ

การตรวจสอบคุณภาพของแบบสอบตามแนวคิดของทฤษฎี IRT จะพิจารณาจากค่าฟังก์ชั่นสารสนเทศแบบสอบ (Test Information Function) ซึ่งได้จากผลรวมของฟังก์ชั่นสารสนเทศข้อสอบ (Item Information Function) โดยที่ค่าฟังก์ชั่นสารสนเทศข้อสอบหรือแบบสอบเป็นดัชนีผสม (Composite index) ที่สร้างจากดัชนีบอกคุณลักษณะของข้อสอบหลายลักษณะ เช่น ค่าอำนาจจำแนก ค่าความยาก เป็นต้น รวมเป็นดัชนีเพียงตัวเดียวเพื่อชี้ถึงคุณภาพของข้อสอบหรือแบบสอบได้ จะเห็นว่าในทฤษฎี IRT จะใช้ค่าฟังก์ชั่นสารสนเทศแบบสอบ แทนการหาค่าความเที่ยงของคะแนนและความคลาดเคลื่อนมาตรฐานของการวัด ตามแนวคิดของทฤษฎีคลาสสิคอล (Hambleton 1977 : 64)

สำหรับในโมเดลโลจีสติก 3 พารามิเตอร์ จะสามารถหาค่าฟังก์ชั่นสารสนเทศข้อสอบ ในแต่ละระดับความสามารถ (θ) ได้จากสูตร (Lord 1980 : 72-74)

$$I\{\theta, u_i\} = \frac{(1.7a_i)^2(1-c_i)}{\left[c_i + e^{1.7a_i(\theta-b_i)}\right]\left[1+e^{-1.7a_i(\theta-b_i)}\right]^2} \quad ..(19)$$

และสามารถกำหนดโค้งสารสนเทศข้อสอบได้จากสมการ

$$I\{\theta, u_i\} = (P_i')^2/P_iQ_i \quad .....(20)$$

$I\{\theta, u_i\}$ คือ ค่าฟังก์ชั่นสารสนเทศข้อสอบ

P_i' คือ ความชันของ ICC ที่ระดับความสามารถ θ

P_i คือ ความน่าจะเป็นที่ผู้สอบที่มีความสามารถ θ จะตอบข้อสอบข้อ i ได้ถูกต้อง

Q_i คือ $1 - P_i$

และจะสามารถหาค่าฟังก์ชั่นสารสนเทศแบบสอบได้จากผลรวมของฟังก์ชั่นสารสนเทศข้อสอบทั้งหมดในแบบสอบ เขียนเป็นสมการได้ดังนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$I\{\theta\} = \sum_{i=1}^{n} I\{\theta, u_i\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (21)$$

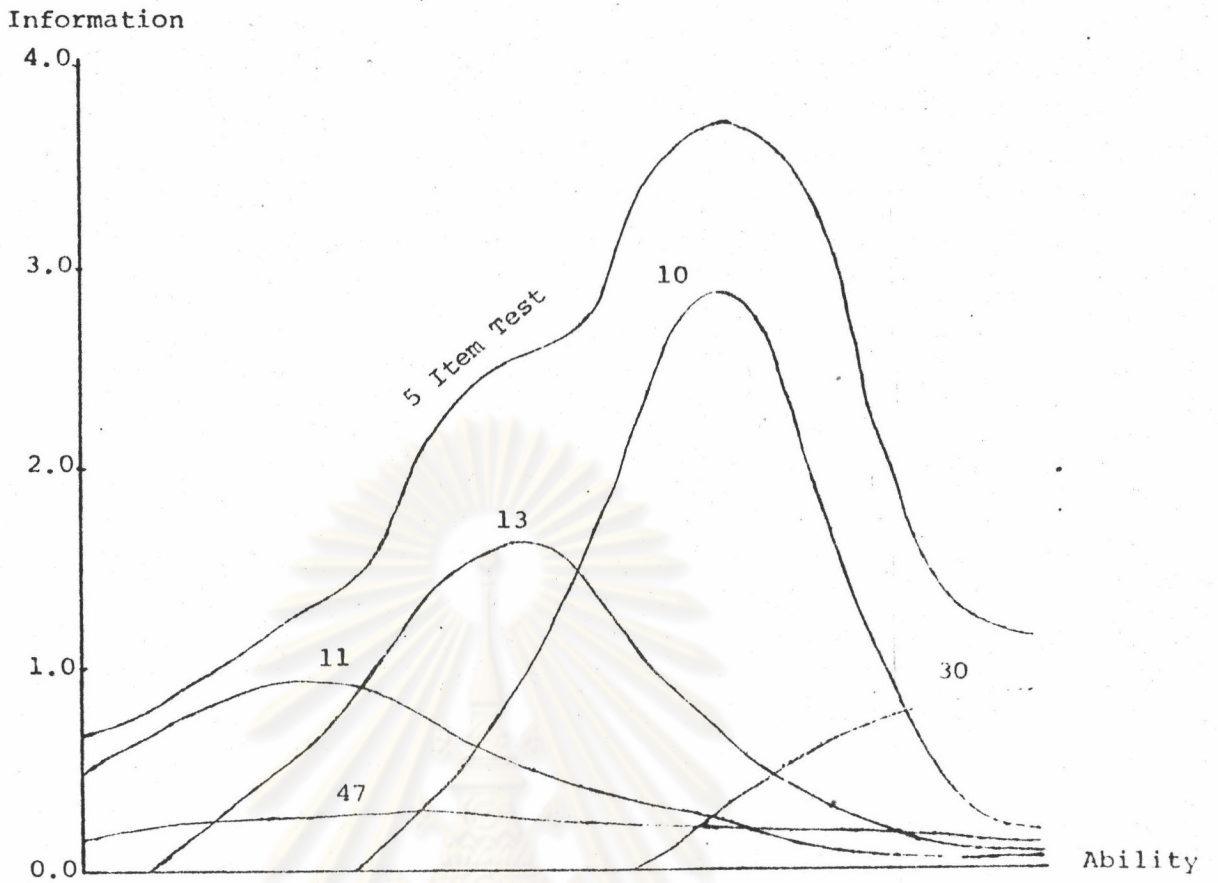
เมื่อ $I\{\theta\}$ คือ ค่าฟังก์ชั่นสารสนเทศแบบสอบ

จากสมการ (20) จะเห็นว่าค่าฟังก์ชั่นสารสนเทศแต่ละข้อขึ้นอยู่กับความชันของโค้งลักษณะข้อสอบ และความแปรปรวนของการตอบข้อสอบถูกของแต่ละข้อในแต่ละระดับความสามารถ θ และยิ่งความชันของ ICC มีค่ามาก ประกอบกับค่าความแปรปรวนมีค่าน้อย ๆ โค้งสารสนเทศข้อสอบที่ระดับความสามารถนั้นจะยิ่งสูงขึ้น สำหรับโค้งสารสนเทศข้อสอบที่มีค่าสูงสุด ณ ระดับของความสามารถใดก็จะจำแนกระดับความสามารถของผู้สอบได้ดี ณ ระดับความสามารถนั้น (Hambleton 1977 : 66 อ้างถึงใน ประดิษฐ์ เรืองตระกูล 2529 : 23-24)

สำหรับค่าฟังก์ชั่นสารสนเทศของแบบสอบฉบับหนึ่งจะมีได้หลายค่าแตกต่างกันไปตามระดับของ θ และ ณ ระดับความสามารถใดที่ค่าฟังก์ชั่นสารสนเทศแบบสอบที่มีค่าสูงแสดงว่าแบบสอบมีคุณภาพดีมากในการประมาณค่าความสามารถที่ระดับนั้น แฮมเบิลตันและสวามินาธาน (Hambleton and Swaminathan 1985 : 107) กล่าวว่า เมื่อแบบสอบมีข้อสอบจำนวนมากขึ้น ค่าฟังก์ชั่นสารสนเทศแบบสอบก็จะเพิ่มมากขึ้น

ดังนั้นถ้าเรามีกลุ่มข้อสอบที่ทราบโค้งสารสนเทศ เราจะสามารถสร้างแบบสอบให้มีโค้งสารสนเทศแบบสอบ ณ ระดับหนึ่งของความสามารถที่เราต้องการได้ เช่น เพื่อคัดเลือกนักเรียนให้ได้รับทุน ก็ต้องใช้ข้อสอบที่มีประสิทธิภาพสูงสุดที่ระดับความสามารถสูง ๆ นั่นคือให้มีโค้งสารสนเทศแบบสอบสูง ณ ระดับความสามารถสูง ๆ นั่นเอง

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



แผนภาพที่ 6 แสดงโค้งสารสนเทศข้อสอบและโค้งสารสนเทศแบบสอบของข้อสอบ 5 ข้อ (Lord 1980 : 22)

ในการสอบทุกครั้ง จะมีความคลาดเคลื่อนในการวัดของแบบสอบ และสำหรับค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ (The Standard Error of Estimate Ability : S.E.E.) หมายถึง ส่วนเบี่ยงเบนมาตรฐานในการวัดของการประมาณค่าความสามารถ (θ) ในกลุ่มผู้สอบที่มีความสามารถแตกต่างกันไป ซึ่งประมาณค่าได้จากสูตร (Warm 1978 : 77)

$$S.E.E. = \frac{1}{\sqrt{I(\theta)}} \quad \text{.....(21)}$$

ความคลาดเคลื่อนมาตรฐานในการประมาณค่านี้ขึ้นอยู่กับและผันแปรไปตามฟังก์ชั่นของข้อสอบและกลุ่มผู้สอบที่ใช้ ยิ่งความยากของข้อสอบเหมาะกันกับความสามารถของผู้สอบ และจำนวนข้อและจำนวนผู้สอบยิ่งมาก ความคลาดเคลื่อนมาตรฐานจะยิ่งต่ำลง

ค่าฟังก์ชั่นสารสนเทศแบบสอบจะมีค่าแปรเปลี่ยนไปตามสเกลของความสามารถ และถ้าฟังก์ชั่นสารสนเทศแบบสอบมีค่ามาก จะทำให้ค่า S.E.E. มีค่าน้อย ซึ่งเป็นค่าที่เราต้องการมากในการสอบ

3.4 การกำหนดคะแนนเกณฑ์

แวน เดอ ลินเดน (Van der Linden 1982 : 295-307) ได้เสนอว่าโมเดลในทฤษฎี IRT อาจนำมาใช้ประโยชน์ในการวิเคราะห์ผลจากการพิจารณากำหนดคะแนนเกณฑ์ของข้อสอบ โดยผู้เชี่ยวชาญได้ เนื่องจากในแนว IRT เชื่อว่าข้อสอบแต่ละข้อสามารถกำหนดให้มีฟังก์ชั่นคุณลักษณะ (Characteristic) เป็นตัวแสดงความน่าจะเป็นในการตอบข้อสอบได้ถูก ซึ่งขึ้นกับระดับความสามารถของนักเรียน (θ) และคุณสมบัติของข้อสอบ ได้แก่ ค่า a,b,c และ เคน (Michael T. Kane 1987 : 333-345) ได้เสนอวิธีการแปลงคะแนนเกณฑ์ที่ได้จากการตัดสินของผู้เชี่ยวชาญ ซึ่งอยู่บนสเกลของคะแนนจริงหรือ คะแนนโดเมนให้อยู่บนสเกลของระดับความสามารถ โดยอาศัยโค้งลักษณะแบบสอบ (Test Characteristic Curve) ซึ่งมีวิธีการดังนี้ คือ

1. คำนวณหาค่าเฉลี่ยของระดับเกณฑ์ผ่านต่ำสุด (Minimum Pass levels) ในการตอบข้อสอบข้อที่ i ที่ได้จากการพิจารณาของผู้เชี่ยวชาญ (MiR)

2. หาคะแนนจุดตัดหรือคะแนนเกณฑ์ที่อยู่บนสเกลของคะแนนจริง ซึ่งคำนวณได้จากผลรวมของ MiR ของข้อสอบทุกข้อ

และถ้าต้องการกำหนดคะแนนเกณฑ์บนสเกลของระดับความสามารถ (θ) ก็ทำในขั้นตอนที่ 3 ต่อไป

3. ใช้โค้งคุณลักษณะแบบสอบ ซึ่งได้จากผลรวมของโค้งคุณลักษณะข้อสอบ (Item Characteristic Curve) เพื่อหาคะแนนเกณฑ์บนสเกลของระดับความสามารถ ซึ่งสัมพันธ์กับคะแนนจริงในขั้นที่ 2 ดังสมการ

$$\sum_i M_{iR} = \sum_i P_i(\theta^*) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (22)$$

เมื่อ θ^* คือ คะแนนเกณฑ์บนสเกลของระดับความสามารถ

M_{iR} คือ ค่าเฉลี่ยของเกณฑ์ผ่านระดับต่ำสุดของผู้เชี่ยวชาญ k คน บนข้อสอบแต่ละข้อ

$$\sum_i \left[M_{iR} - P_i(\hat{\Theta}^*) \right] = 0 \quad \ldots\ldots(23)$$

และเนื่องจาก $M_{iR} = P_i(\hat{\Theta}^*_{iR})$ ดังนั้นสามารถเขียน (23) ได้ใหม่

$$\sum \left[P_i(\hat{\Theta}^*_{iR}) - P_i(\hat{\Theta}^*) \right] = 0 \quad \ldots\ldots(24)$$

กำหนดให้ $\hat{\Theta}^*_{iR} \longrightarrow \Theta^*$ ดังนั้นความสัมพันธ์นี้สามารถประมาณได้ด้วยฟังก์ชั่นเชิงเส้น (Linear function)

$$\sum_i P'_i(\hat{\Theta}^*_{iR})(\hat{\Theta}^*_{iR} - \hat{\Theta}^*) \simeq 0 \quad \ldots\ldots(25)$$

$$\hat{\Theta}^* \simeq \frac{1}{\sum_i P'_i(\hat{\Theta}^*_{iR})} \sum_i P'_i(\hat{\Theta}^*_{iR})\, \hat{Q}^*_{iR} \quad \ldots\ldots(26)$$

และเนื่องจาก $\dfrac{P'_i(\hat{\Theta}^*_{iR})}{\sigma_i(M_{iR})} \simeq \dfrac{1}{\sigma_i(\hat{\Theta}^*_{iR})}$ $\ldots\ldots(27)$

ดังนั้นสมการ 26 เขียนได้เป็น

$$\hat{\Theta}^* \frac{1}{\sum_i \sigma_i(M_{iR}) / \sigma_i(\hat{\Theta}^*_{iR})} \sum_i \frac{\sigma_i(M_{iR})}{\sigma_i(\hat{\Theta}^*_{iR})} \hat{\Theta}^*_{iR} \quad \ldots\ldots(28)$$

ค่าประมาณของ Θ^* ในสมการ (26) และ (28) ยังไม่ให้การตีความที่ชัดเจน ซึ่งสามารถปรับปรุงได้ โดยการถ่วงด้วยน้ำหนักผลรวมของ M_{iR} กับน้ำหนักผลรวมของโค้งคุณลักษณะข้อสอบ ได้แก่ $1/\sigma_i(M_{iR})$ ในสมการ (23) และ (24) ดังนั้นจะได้

$$\hat{\Theta}^*_w \simeq \frac{1}{\sum_i \left[P'_i(\hat{\Theta}^*_{iR}) / \sigma_i(M_{iR}) \right]} \sum_i \frac{P'_i(\hat{\theta}^*_{iR})}{\sigma_i(M_{iR})} \hat{\Theta}^*_{iR} \quad \ldots\ldots(29)$$

และจาก (27) ดังนั้น

$$\hat{\Theta}^*_w \simeq \frac{1}{\sum_i \left[(1/\sigma_i(\hat{\Theta}^*_{iR}) \right]} \sum_i \frac{1}{\sigma_i(\hat{\Theta}^*_{iR})} \hat{\Theta}^*_{iR} \quad \ldots\ldots(30)$$

3.5 การนำโมเดลโลจีสติก 3 พารามิเตอร์ไปใช้โดยทั่วไป

โมเดลโลจีสติก 3 พารามิเตอร์ สามารถนำไปใช้ประโยชน์ในการทดสอบได้ละเอียดและถูกต้องมากกว่าโมเดลอื่น ๆ และ ยูริ (Urry 1970 อ้างถึงใน สุพัฒน์ สุกมลสันต์ 2530 : 57) ได้ทำการศึกษาเปรียบเทียบผลการใช้โมเดลโลจีสติก ทั้งแบบ 1, 2 และ 3 พารามิเตอร์ ได้ข้อสรุปว่า ชนิด 3 พารามิเตอร์ เหมาะกับแบบสอบเลือกตอบมากที่สุด เพราะพิจารณารวมถึงค่าการเดา ค่าความยาก และค่าอำนาจจำแนก ซึ่งสัมพันธ์กับคุณลักษณะของการตอบแบบสอบเลือกตอบเป็นอย่างมาก ซึ่งประโยชน์ในการนำโมเดลโลจีสติก 3 พารามิเตอร์ไปใช้ในการทดสอบ สามารถสรุปได้ดังนี้

1. ใช้ในการวิเคราะห์ข้อสอบ (Item Analysis) เนื่องจากการวิเคราะห์ข้อสอบตามรูปแบบคลาสสิคอลนั้นมีข้อจำกัดหลายประการ เช่น (1) ค่าสถิติและค่าพารามิเตอร์ต่าง ๆ ขึ้นอยู่กับสภาพกลุ่มตัวอย่างที่ทำการสอบ (2) การเปรียบเทียบความสามารถหรือคุณลักษณะ (Trait) ใด ๆ ในแต่ละบุคคลจะเปรียบเทียบกันได้ก็ต่อเมื่อสอบด้วยแบบสอบฉบับเดียวกัน (3) มีข้อตกลงที่ว่าความคลาดเคลื่อนมาตรฐานในการวัดของแต่ละคนเท่ากัน แต่ในสภาพความเป็นจริง ค่าความคลาดเคลื่อนในการวัดนี้จะแปรเปลี่ยนไปตามความสามารถของแต่ละบุคคล ซึ่งปัญหาต่าง ๆ เหล่านี้ จะสามารถแก้ไขได้ด้วยการวิเคราะห์ด้วยโมเดลโลจีสติก 3 พารามิเตอร์

2. ใช้ในการสร้างธนาคารข้อสอบ (Item bank) เนื่องจากข้อสอบที่วิเคราะห์แล้ว ค่าพารามิเตอร์ที่ได้มีลักษณะไม่แปรเปลี่ยน (invariant) ดังนั้น จากค่าฟังก์ชั่นสารสนเทศของข้อสอบ (Item Characteristic Function) แต่ละข้อจะช่วยในการนำข้อสอบเหล่านี้มาสร้างเป็นแบบสอบฉบับใหม่ ที่มีค่าพารามิเตอร์ต่าง ๆ สอดคล้องกับเกณฑ์หรือจุดมุ่งหมายของแบบสอบได้

3. ใช้ในการกำหนดเกณฑ์ของระดับการรอบรู้ (master level) ของแบบสอบอิงเกณฑ์ จากการวิเคราะห์ข้อสอบด้วยโมเดลโลจีสติก คะแนนของผู้สอบจะถูกแปลงให้เป็นคะแนนความสามารถ (ability score) ซึ่งสามารถเปรียบเทียบกับคะแนนความสามารถที่เป็นเกณฑ์คงที่ได้ ทำให้เราทราบได้ว่า ระดับการรอบรู้ขั้นต่ำ (minimum master level) ของข้อสอบแต่ละข้อควรเป็นเท่าใด

4. ใช้ในการวัดระดับความสามารถของแต่ละบุคคล (Self-Tailoring Test) เราอาจสุ่มข้อสอบที่วิเคราะห์แล้ว ที่มีระดับความยากเรียงกันตามลำดับ เพียงจำนวนเล็กน้อยมาทดสอบหาความสามารถของบุคคลได้

5. ใช้วินิจฉัยความสามารถของผู้สอบ (Individual Diagnosis) ในกรณีที่โค้งคุณลักษณะของข้อสอบไม่เหมาะสม (fit) กับโมเดล แสดงว่าอาจมีบางอย่างที่ผิดปกติในตัวผู้สอน เช่น อาจมีความสามารถอื่นแฝงเข้ามาในความสามารถที่เราต้องการวัด

6. ใช้ตรวจสอบความเป็นอคติของข้อสอบ (Item Bias) ซึ่งทำได้โดยการศึกษาจากโค้งคุณลักษณะข้อสอบ (Item Characteristic Curve) ของผู้สอบแต่ละกลุ่ม ถ้าโค้งคุณลักษณะข้อสอบนั้น ๆ แตกต่างกันตามกลุ่มที่นำมาทดสอบ แสดงว่าข้อสอบข้อนั้นมีอคติเกิดขึ้น

7. ใช้ในการเปรียบเทียบคะแนนต่างชุด (Equating Score) เราสามารถนำคะแนนของผู้สอบจากแบบสอบ 2 ฉบับที่ต่างกัน แต่วัดในสิ่งเดียวกัน มาเปรียบเทียบกันได้ ทั้งนี้เพราะจากการวิเคราะห์คะแนนการตอบข้อสอบของผู้สอบด้วยโมเดลโลจีสติก คะแนนจะถูกแปลงเป็นคะแนนมาตรฐาน ซึ่งสามารถเปรียบเทียบกันได้

8. ใช้ในการจัดชั้นเรียน (Grade-Placement Tailoring) สามารถนำค่าความสามารถของผู้สอบที่ได้จากการวิเคราะห์ มาจัดชั้นเรียนให้เหมาะสมกับระดับความสามารถของผู้เรียนเป็นกลุ่ม ตามแผนการจัดการเรียนการสอนได้

IRT เป็นทฤษฎีที่มีรูปแบบคณิตศาสตร์ที่ซับซ้อน ซึ่งอยู่บนพื้นฐานของข้อตกลงเบื้องต้นที่แกร่ง แต่ในปัจจุบันก็ได้มีการพัฒนาโปรแกรมคอมพิวเตอร์ที่จะนำมาใช้กับทฤษฎีอย่างแพร่หลาย สำหรับโอกาสที่ผู้ปฏิบัติจะนำ IRT มาประยุกต์ใช้ในการพัฒนาแบบสอบ ควรจะพิจารณาในประเด็นต่อไปนี้ (Hambleton and Swaminathan 1985 : 305)

1. วัตถุประสงค์เพื่อพัฒนาแบบสอบ หรือการวิเคราะห์ข้อมูลแบบสอบที่มี

2. แบบสอบมีความเป็นเอกพันธ์ (Unidimensional) หรือไม่

3. โมเดลไหนที่เหมาะสม (fit) กับข้อมูลมากที่สุด

4. ควรปรับปรุงข้อมูลให้มีความเหมาะสมกับโมเดลหรือโมเดลที่เลือกใช้ต้องเหมาะสมกับข้อมูล

5. ขนาดกลุ่มตัวอย่างเพียงพอหรือไม่

6. โปรแกรมคอมพิวเตอร์ที่ใช้เหมาะสมหรือไม่

7. แหล่งข้อมูลที่มีเพียงพอหรือไม่

8. วิธีประมาณค่าวิธีใดที่เหมาะสม

9. ควรจะรายงานคะแนนจากแบบสอบ (test score) อย่างไร

10. การประยุกต์ใช้วิธี IRT จะให้คำตอบที่ต้องการหรือไม่

คำตอบของคำถามเหล่านี้ ในบางกรณีอาจจำเป็นทุกประเด็น อย่างไรก็ดี ตัววัตถุประสงค์ที่ชัดแจ้งจะช่วยกำหนดทิศทางว่าควรจะดำเนินการอย่างไร

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

4. งานวิจัยที่เกี่ยวข้อง

4.1 งานวิจัยที่เกี่ยวข้องกับการทดสอบอิงโดเมน

สำหรับในต่างประเทศ ประเด็นเกี่ยวกับแบบสอบอิงโดเมน ส่วนใหญ่จะพบในรูปของการเสนอบทความแสดงแนวคิดเกี่ยวกับการให้ความหมายจัดประเภท ตลอดจนการสร้างและพัฒนาแบบสอบ สำหรับโครงการและงานวิจัยที่เกี่ยวข้องเท่าที่ค้นคว้าได้จะเสนอได้ดังนี้

ในปี 1966 ไฮฟลีและคณะ ซึ่งเป็นคณะอาจารย์ในมหาวิทยาลัยมินเนโซตา ได้เริ่มโครงการ MINNEMAST ซึ่งเป็นโครงการการทดสอบอิงโดเมนในวิชาวิทยาศาสตร์และคณิตศาสตร์ในระดับชั้นอนุบาลจนถึงชั้นประถมปีที่ 6 ในมลรัฐมินเนโซตา ซึ่งจุดมุ่งหมายโครงการเป็นไปเพื่อการประเมินผลโครงการโดยเฉพาะที่มีต่อผู้เรียน และพัฒนาแบบสอบอิงโดเมนเพื่อวัดผลสัมฤทธิ์ขึ้นในการกำหนดโดเมน จะกำหนดเป็นโดเมนย่อย ๆ ที่ประกอบกันขึ้นเป็นโดเมนใหญ่ โดยใช้วิธีอุปนัย (induction) ซึ่งมีหลักการว่า พฤติกรรมที่แสดงออกถึงการเรียนรู้ตามที่นักวิชาการของสาขาวิชาระบุไว้อย่างแน่นอนและชัดเจนแล้ว สามารถนำมาจัดเข้าเป็นกลุ่มหรือโดเมนของข้อสอบได้ และในการกำหนดโดเมนให้มีความเป็นตัวแทนที่ดีและวิธีการพัฒนาข้อสอบด้วยนั้น ไฮฟลีได้ใช้เทคนิครูปแบบข้อสอบ (Item Form) ซึ่งระบบการทดสอบที่ใช้อิงโดเมนเหล่านี้โดยการนำกฎของการสุ่มข้อสอบมาจากโดเมนและนำข้อสอบเหล่านี้มาใช้วัดความรู้ผู้เรียนแต่ละคน เพื่อนำผลการวัดมาประมาณความรู้ความสามารถของผู้เรียนตามโดเมนเป็นรายบุคคล และในการทดสอบไม่มีการกำหนดเกณฑ์ผ่าน ทั้งนี้เพราะโปรแกรมอยู่ในระยะทดลองเพื่อการปรับปรุงให้ดีขึ้น

ในปี 1983 ฮาลาไดนาและรอยด์ (Haladyna and Roid 1983) ได้ศึกษาเปรียบเทียบผลของความคลาดเคลื่อนในการประมาณค่าความสามารถของบุคคล จากการใช้แบบสอบที่มีวิธีการสร้างต่างกัน 2 แบบ คือ แบบแรกสร้างโดยสุ่มข้อสอบขึ้นมาจากประชากรข้อสอบที่มีอยู่ทั้งหมด แบบที่สองเป็นการเลือกให้เหมาะกับระดับความสามารถของผู้สอบ ข้อสอบเหล่านี้เป็นประชากรข้อสอบทั้งหมดในโดเมนซึ่งผ่านการวิเคราะห์ด้วยราส์ชไมเดลแล้ว จากนั้นได้ทดลองสุ่มสร้างแบบสอบให้มีจำนวนข้อเป็น 10, 20, 30 และ 40 ข้อ ตามลำดับ โดยในการสร้างแบบสอบย่อย จะมีวิธีการสร้าง 4 วิธี คือ 1) การสุ่มข้อสอบขึ้นมาจากโดเมน 2) การเลือกข้อสอบให้มีความยากตรงกับระดับความสามารถของผู้สอบ 3) การเลือกข้อสอบ

ให้มีค่าความยากกระจายมากกว่าระดับความสามารถของผู้สอบเล็กน้อย 4) การเลือกข้อสอบให้มีค่าความยากกระจายมากกว่าระดับความสามารถของผู้สอบ จากนั้นศึกษาเปรียบเทียบอัตราความแตกต่างเฉลี่ยระหว่างคะแนนที่สอบได้กับคะแนนโดเมนที่ประมาณได้ (AAD) จากแบบสอบที่มีความยาวระดับต่าง ๆ ที่กำหนดหารด้วยส่วนเบี่ยงเบนมาตรฐาน (เรียกค่าที่ได้ว่า อัตราส่วน AAD/SD) จากการศึกษาพบว่า วิธีที่ 2 จะได้ค่าอัตราส่วน AAD/SD ต่ำสุด วิธีที่ 1 ซึ่งแบบสอบได้รับการสุ่มข้อสอบขึ้นมามีค่าดัชนีดังกล่าวต่ำเป็นอันดับที่สอง ส่วนแบบสอบที่มีค่าความยากกระจายมาก จะยิ่งมีค่าดัชนีดังกล่าวสูงมากยิ่งขึ้น ผู้เขียนได้เสนอว่า วิธีที่ 2 จะใช้ได้ดีในกรณีที่ต้องการวัดคุณลักษณะความสามารถที่ต่อเนื่องกัน ส่วนวิธีที่ 1 ก็ยังคงให้ผลดีและสะดวกในการปฏิบัติดังนั้นน่าจะใช้เทคนิคทั้งสองในการสร้างแบบสอบอิงเกณฑ์ การที่จะตัดสินว่าวิธีการใดดีกว่าจะต้องรอดูผลจากประโยชน์ที่ได้จากการสอบจริง

ในปี 1984 เพิร์ฟและคณะ (A.C. Purves et al. 1984) ได้ทำการวิจัยเกี่ยวกับการให้คำสั่งในการเขียนเรียงความที่เป็นแบบอิงโดเมนในหลายประเทศ ทั้งในยุโรป สหรัฐอเมริกา และเอเชีย คณะผู้วิจัยได้พัฒนาระบบการกำหนดหัวข้อการเขียนเรียงความแบบอิงโดเมน ซึ่งประกอบด้วย 15 มิติ ได้แก่ คำสั่ง สิ่งเร้า การกำหนดแนวคิดในการเขียน (Cognitive demand) เป้าหมาย บทบาท กลุ่มผู้รับสาร (audience) เนื้อหา การกำหนดโวหาร สีลาภาษา การตระเตรียมหาข้อมูลล่วงหน้า ความยาว รูปแบบการเขียน การกำหนดเวลาการร่างข้อความก่อนลงมือเขียนจริง และเกณฑ์ การวิจัยทดลองใช้ระบบอิงโดเมนตามมิติทั้ง 15 มิติข้างต้น พบว่า มีประโยชน์และสามารถใช้เป็นเครื่องมือในการกำหนดการเขียนเรียงความได้ โดยทำให้ผู้เขียนตระหนักถึงรายละเอียดและความซับซ้อนต่าง ๆ ของงานเขียน ตลอดจนสามารถใช้ในการประเมินงานเขียนและงานวิจัยได้อย่างดี เนื่องจากเป็นเครื่องมือที่ครอบคลุมโดเมน ทำให้เปรียบเทียบการกำหนดงานเขียนต่าง ๆ ได้

สำหรับในประเทศไทยการศึกษาวิจัยเกี่ยวกับแบบสอบอิงโดเมน ยังคงมีไม่มากนัก ส่วนใหญ่จะเป็นการสร้างและพัฒนาแบบสอบอิงเกณฑ์ในรูปของการยึดวัตถุประสงค์เฉพาะในการสร้าง ซึ่งยังไม่มีการนิยามขอบเขตโดเมนที่ชัดแจ้ง และไม่มีการใช้เทคโนโลยีการเขียนข้อสอบเข้ามาช่วยเป็นหลักในการสร้างข้อสอบที่เป็นตัวแทนของเนื้อหาภายใต้ขอบเขตที่กำหนดด้วย อีกทั้งวิธีการวิเคราะห์และการตรวจสอบคุณภาพแบบสอบก็ยังคงใช้แนวคิดแบบคลาสสิคอลทั้งหมด ส่วนงานวิจัยที่เกี่ยวกับแบบสอบอิงโดเมน จะพบในรูปของวิทยานิพนธ์นิสิตนักศึกษาระดับมหาบัณฑิต โดยจะเน้นไปในทางการสร้างแบบสอบเพื่อใช้ประกอบการเรียนการสอนในระดับชั้น

ต่าง ๆ และเน้นในด้านการค้นคว้าเพื่อความก้าวหน้าทางวิชาการ ซึ่งจะเสนอตามลำดับปีการศึกษาที่ได้ดำเนินการดังนี้

งานวิจัยชิ้นแรกที่สร้างแบบสอบตามแนวคิดของการอิงโดเมน เริ่มในปี พ.ศ. 2521 โดยกาญจนา วัธนสุนทร (2522) ได้สร้างแบบสอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 เรื่องสมการ โดยใช้เทคนิครูปแบบข้อสอบ (Item Form) แบบสอบแบ่งเป็น 4 ฉบับย่อย ฉบับละ 10 ข้อ รวมเป็น 40 ข้อ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนจำนวน 60 คน ทำการวิเคราะห์คุณภาพของแบบสอบโดยใช้ดัชนี S คำนวณค่าอำนาจจำแนก ส่วนค่าความยากคำนวณจากสัดส่วนของผู้ตอบข้อสอบข้อนั้น ๆ ถูก ในการหาเกณฑ์ที่เหมาะสมได้ทดลองกำหนดเกณฑ์การตัดสินเป็น .6, .7 และ .8 แล้วคำนวณหาค่าความเที่ยงและความตรงโดยใช้สูตรของลิฟวิงสตันและสูตรของคาร์เวอร์ ตามลำดับ พบว่าเมื่อเกณฑ์เป็น .6 จะให้ค่าความเที่ยงและความตรงสูงสุด และถ้าเกณฑ์สูงขึ้น ค่าความเที่ยงและความตรงจะลดลง

ในปี พ.ศ. 2522 ชมพู จันทรอมรพร (2523) ได้ปรับปรุงแบบสอบที่กาญจนา วัธนสุนทร สร้างไว้ แล้วคำนวณหาจุดตัดของแบบสอบโดยใช้กระบวนการเชิงทฤษฎีการตัดสินใจของเบส์ ใช้กลุ่มตัวอย่างจำนวน 80 คน พบว่าเกณฑ์ของแบบสอบฉบับที่ 1 เป็น 60 เปอร์เซนต์ ส่วนฉบับที่ 2, 3 และ 4 เป็น 50 เปอร์เซนต์

ในปี พ.ศ. 2523 ประภา แก่นเพิ่ม (2524) ได้ศึกษาเกี่ยวกับความถูกต้องในการกำหนดจุดตัดด้วยวิธีการกำหนดเกณฑ์ระดับผ่านต่ำสุด โดยได้นำแบบสอบของกาญจนา วัธนสุนทร ที่ ชมพู จันทรอมรพร ได้ปรับปรุงแล้วไปปรับแก้อีกเล็กน้อย ไปศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 1 ที่มีผลการเรียนคาบเส้นตามที่ครูระบุ และนักเรียนที่มีผลการเรียนคาบเส้นในอดีต จำนวน 82 คน และ 67 คน ตามลำดับ เพื่อศึกษาเปรียบเทียบระหว่างการกำหนดเกณฑ์มาตรฐานขั้นต่ำจากการพิจารณาของครูกับผลการเรียนของนักเรียนที่ใช้เป็นตัวตรวจสอบวิธีการกำหนดจุดตัดขั้นต่ำใช้ 2 วิธี คือ วิธีของกิลเบิร์ท และวิธีของมหาวิทยาลัยอิลลินอยส์ ผลการศึกษา พบว่า คะแนนจุดตัดที่ได้จากการกำหนดเกณฑ์ระดับผ่านต่ำสุดมีความถูกต้อง โดยพิจารณาจากจุดตัดที่ได้จากการทดสอบนักเรียนคาบเส้นที่ใช้เป็นเกณฑ์ กับจุดตัดที่ได้จากการกำหนดเกณฑ์ระดับผ่านต่ำสุดมีความแตกต่างกันน้อยมากไม่เกินร้อยละ 10

ในปี พ.ศ. 2526 ประเทือง ทาสีแสง (2527) ได้สร้างแบบสอบอิงโดเมนวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 เรื่องคู่ลำดับและกราฟ โดยสร้างขึ้นตามลักษณะเฉพาะ

(Item Specification) ที่กำหนดขึ้นตามกลุ่มพฤติกรรมย่อย และจัดแบ่งเป็นแบบสอบคู่ขนาน จำนวน 3 ชุด ได้นำไปศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 1 จำนวน 167 คน ทำการวิเคราะห์คุณภาพของแบบสอบ โดยใช้ดัชนี S คำนวณค่าอำนาจจำแนก ส่วนค่าความยากคำนวณจากสัดส่วนของผู้ตอบข้อสอบข้อนั้น ๆ ถูก การหาความเที่ยงใช้วิธีของ สวามินาธาน แฮมเบิลตัน และอัลจินา ค่าความตรงใช้วิธีการให้ผู้เชี่ยวชาญด้านเนื้อหาพิจารณาตามวิธีของ โรวิเนลลีและแฮมเลิลตัน และสำหรับการหาเกณฑ์ใช้วิธีของ ฮวิน และการให้ผู้เชี่ยวชาญเนื้อหาวิชาพิจารณาจากเกณฑ์ที่ลดจาก 100% ตามเทคนิคของกลาส แล้วหาคะแนนเฉลี่ยเป็นคะแนนเกณฑ์ที่เหมาะสม ซึ่งได้คะแนนเกณฑ์ของแบบสอบคู่ขนาน ชุดที่ 1, 2 และ 3 คิดเป็นร้อยละ 80, 87 และ 70 ของคะแนนเต็มตามลำดับ

ในปี พ.ศ. 2527 กำจัด เกตุสุวรรณ (2528) ได้สร้างแบบสอบอิงโดเมนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 5 เรื่องลำดับ โดยใช้วิธีการกำหนดลักษณะเฉพาะของข้อสอบตามกลุ่มพฤติกรรมย่อย และจัดเป็นแบบสอบคู่ขนาน จำนวน 4 ชุดย่อย ได้นำไปศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 5 จำนวน 160 คน ทำการวิเคราะห์คุณภาพแบบสอบ โดยใช้ดัชนี S คำนวณค่าอำนาจจำแนก และคำนวณค่าความยากจากสัดส่วนของผู้ตอบข้อสอบข้อนั้น ๆ ถูก การหาค่าความเที่ยงใช้สูตรของ คาร์เวอร์ ตรวจหาความตรงตามเนื้อหาโดยอาศัยดุลยพินิจของผู้เชี่ยวชาญตามวิธีของ โรวิเนลลี และแฮมเบิลตัน สำหรับการกำหนดจุดตัดหรือเกณฑ์ของแบบสอบ ใช้วิธี Pmax ของฮวิน ซึ่งได้คะแนนเกณฑ์ของแบบสอบคู่ขนาน ชุดที่ 1, 2, 3 และ 4 คิดเป็นร้อยละ 70, 45, 44 และ 50 ตามลำดับ

ในปี พ.ศ. 2529 สุนันทา พากเพียร (2530) ได้ทำการวิจัยเกี่ยวกับการสำรวจความต้องการในการเรียนภาษาอังกฤษของนักเรียน ผู้ปกครอง และครูผู้สอน ระดับชั้นมัธยมศึกษาตอนต้น ในโรงเรียนรัฐบาล เขตกรุงเทพมหานคร เครื่องมือที่ใช้สร้างโดยการกำหนดวัตถุประสงค์ขยายความ (Amplified Objective) ซึ่งพัฒนาจากโดเมนหลักสูตรและเนื้อหาจากแบบเรียนต่าง ๆ ในระดับมัธยมศึกษาตอนต้น

จะเห็นว่างานวิจัยทั้งหมดเกี่ยวกับแบบสอบอิงโดเมนดังที่ได้กล่าวแล้ว ยังคงใช้แนวคิดแบบคลาสสิคอลอยู่

4.2 งานวิจัยที่เกี่ยวข้องกับโมเดลโลจีสติก 3 พารามิเตอร์ และการนำไปใช้พัฒนาแบบสอบอิงโดเมน

ในปี 1981 ฮัทเทน (Hutten 1981 : 4799-A) ได้ศึกษาเกี่ยวกับความเหมาะสม (fit) ของข้อมูลจากการตอบแบบสอบกับราส์ชโมเดล และโมเดลโลจีสติก 3 พารามิเตอร์ โดยใช้แบบสอบ 4 ชุด ๆ ละ 40 ข้อ สอบกับนักเรียน 1,000 คน พบว่าข้อมูลของแบบสอบเหมาะสมกับโมเดลทั้งสอง คิดเป็นร้อยละ 80 โดยร้อยละ 65 ของแบบสอบเหมาะที่จะใช้กับโมเดลโลจีสติก 3 พารามิเตอร์ มากกว่าจะใช้กับราส์ชโมเดล และยังพบว่าความเหมาะสมกับโมเดลและคุณสมบัติการวัดเพียงคุณลักษณะเดียว (Unidimensionality) มีความสัมพันธ์กันอย่างมีนัยสำคัญ และในการประมาณค่าความสามารถของผู้สอบเมื่อจำนวนข้อสอบของแบบสอบแต่ละชุดลดลงมาจาก 40 ข้อ เป็น 20 ข้อ ด้วยวิธีวิเคราะห์ของราส์ชโมเดล พบว่ามีความสัมพันธ์กัน คือมีค่า .923 ส่วนผลจากการวิเคราะห์ด้วยโมเดลโลจีสติก 3 พารามิเตอร์ จะมีความสัมพันธ์กัน คือมีค่า .866 นอกจากนี้ยังพบด้วยว่า ในกรณีการวิเคราะห์กับกลุ่มตัวอย่างขนาดเล็ก 250 คน สำหรับโมเดลโลจีสติก 3 พารามิเตอร์ จะพบว่าค่าความยากได้ แต่ค่าพารามิเตอร์อื่น ๆ จะไม่สามารถเชื่อถือได้ ดังนั้นในการใช้โมเดลโลจีสติก 3 พารามิเตอร์ จึงควรใช้กลุ่มตัวอย่างไม่ต่ำกว่า 1,000 คน

ในปี 1982 แวน เดอ ลินเดน (Van der Linden 1982 : 295-307) ได้ศึกษาเกี่ยวกับการใช้ IRT ในการเปรียบเทียบความถูกต้องในการตัดสินคะแนนเกณฑ์โดยวิธีของแองกอฟและวิธีนีเดลสกี้ โดยใช้ค่าความน่าจะเป็นในการตอบข้อสอบถูก ตามแนวคิดของ IRT ซึ่งคำนวณจากค่าฟังก์ชันคุณลักษณะข้อสอบเป็นเกณฑ์ วิธีการศึกษาโดยการให้ผู้เชี่ยวชาญ 8 และ 9 คน กำหนดความน่าจะเป็นในการตอบข้อสอบได้ถูกของนักเรียนคาบเส้นตามวิธีของแองกอฟ และนีเดลสกี้ ตามลำดับ ซึ่งจะได้เป็นคะแนนเกณฑ์ที่อยู่บนสเกลของคะแนนจริง แล้วแปลงให้เป็นคะแนนเกณฑ์ที่อยู่บนสเกลของระดับความสามารถ (θ) โดยอาศัยโค้งคุณลักษณะแบบสอบ จากนั้นได้เปรียบเทียบความน่าจะเป็นในการตอบข้อสอบถูกระหว่างการตัดสินโดยผู้เชี่ยวชาญและจากการคำนวณหาค่าฟังก์ชันคุณลักษณะข้อสอบ ($P_i(\theta)$) ของข้อสอบแต่ละข้อจากการคำนวณค่าดัชนีความสอดคล้อง (Index of Consistency) ซึ่งผลการศึกษาพบว่าดัชนีความสอดคล้องของวิธีแองกอฟกับค่า $P_i(\theta)$ มีค่า 0.77 ซึ่งสูงกว่าวิธีของนีเดลสกี้ ซึ่งเท่ากับ 0.68 และค่าความเคลื่อนเฉลี่ยจากวิธีแองกอฟ มีค่า 0.18 ส่วนวิธีนีเดลสกี้ มีค่าเป็น 0.25

ในปี 1983 แฮมเบิลตันและกรุจเตอร์ (Hambleton and Gruijter 1983 : 355-366)ได้ประยุกต์ใช้ทฤษฎี IRT ในการคัดเลือกข้อสอบแบบสอบอิงเกณฑ์ โดยพิจารณาจากค่าอำนาจจำแนก ค่าความยาก พบว่าแบบสอบที่มีจำนวนข้อ มีค่าความยากเท่ากันและตรงกับระดับความสามารถของผู้สอบ แบบสอบที่ประกอบด้วยข้อสอบที่มีค่าอำนาจจำแนกสูงกว่า จะได้ค่าความคลาดเคลื่อนต่ำกว่า และได้เสนอไว้ด้วยว่าถ้ากลุ่มตัวอย่างที่ใช้วิเคราะห์มีขนาดเล็กควรวิเคราะห์ด้วยโมเดลที่มี 1 พารามิเตอร์ แต่ถ้ากลุ่มตัวอย่างมีขนาดใหญ่จะเลือกใช้โมเดลไหนมาวิเคราะห์ให้ขึ้นอยู่กับความต้องการว่าจะพัฒนาแบบสอบถึงขั้นไหนและเป็นแบบสอบที่มีความสำคัญเพียงใด

และในปีเดียวกันนี้ แฮมเบิลตัน (Hambleton 1983 : 34-43)ได้ศึกษาเกี่ยวกับการเปรียบเทียบโมเดลโลจีสติก 1-2-3 พารามิเตอร์ในการประมาณคะแนนโดเมนและตัดสินจำแนกผู้รอบรู้/ไม่รอบรู้ เพื่อเปรียบเทียบว่ารูปแบบใดจะให้คะแนนโดเมนที่มีความคงที่มากกว่า โดยใช้แบบสอบที่มีความยาว 10, 15, 20 และ 40 ข้อ มีจุดตัดหลายค่า ได้แก่ -1.50, 0.00, 1.50 และเปรียบเทียบที่ระดับ θ หลายค่า (ในช่วงจากต่ำมากไปสูง) กลุ่มตัวอย่างที่ใช้จำนวน 2,000 คน เกณฑ์ที่ใช้ในการตัดสิน มี 3 เกณฑ์ คือ เกณฑ์ค่าเฉลี่ยของความเบี่ยงเบนสัมบูรณ์ระหว่างคะแนนโดเมนที่เป็นคะแนนจริงของผู้สอบกับคะแนนโดเมนที่ประมาณได้ด้วย แต่ละโมเดลโลจีสติก เกณฑ์เกี่ยวกับความลำเอียงของคะแนนโดเมนที่ประมาณได้ กับโมเดลที่ใช้ และเกณฑ์จุดตัด ผลการศึกษาพบว่าในการเปรียบเทียบความถูกต้องแม่นยำของการประมาณคะแนนโดเมนที่ระดับความสามารถจาก -2.00 ถึง 1.00 โมเดลชนิด 3 พารามิเตอร์จะให้ค่าการประมาณคะแนนโดเมนได้ดีกว่าชนิด 1, 2 พารามิเตอร์ ส่วนค่าความลำเอียงในการประมาณคะแนนโดเมนด้วยโมเดลโลจีสติกทั้ง 3 รูปแบบ มีค่าน้อยมาก และในแง่ความถูกต้องแม่นยำของการตัดสินจำแนกผู้รอบรู้/ไม่รอบรู้ พบว่าโมเดลโลจีสติกชนิด 1 และ 3 พารามิเตอร์ จะให้ค่าใกล้เคียงกัน ยกเว้นที่ระดับความสามารถต่ำ ๆ (-2.0 ถึง -.5) ที่ชนิด 3 พารามิเตอร์ จะให้ค่าที่ถูกต้องแม่นยำกว่า ส่วนชนิด 2 พารามิเตอร์ จะให้ค่าไม่แน่นอน และแฮมเบิลตันเสนอว่า เมื่อขนาดกลุ่มตัวอย่างมากพอ และต้องการค่าประมาณที่ถูกต้อง หรือเมื่อสนใจพิเศษเกี่ยวกับผู้สอบที่ระดับความสามารถต่ำ เช่น ต้องการจำแนกผู้สอบที่จำเป็นต้องได้รับการปรับปรุงแก้ไขหรือเมื่อต้องการประเมินความสามารถของผู้สอบก่อนสอน ควรใช้โมเดลโลจีสติกชนิด 3 พารามิเตอร์

สำหรับประเทศไทย ในปัจจุบันได้เริ่มตื่นตัวเกี่ยวกับการนำโมเดลโลจีสติกมาประยุกต์ใช้ในการสร้างและพัฒนาแบบสอบบ้างแล้ว โดยจะปรากฏในรูปของบทความของนักการศึกษาและวิทยานิพนธ์ระดับมหาบัณฑิตและดุษฎีบัณฑิต ในสาขาด้านการวัดและประเมินผลการศึกษา สำหรับในงานวิจัยฉบับนี้ จะเสนอเฉพาะงานวิจัยที่เกี่ยวข้องกับโมเดลโลจีสติกและการใช้พัฒนาแบบสอบอิงโดเมน ดังนี้

ในปี พ.ศ. 2527 ปนัดดา วัยวัฒนะ (2528) ได้ศึกษาเกี่ยวกับผลการวิเคราะห์ข้อสอบโดยวิธีโลจีสติกโมเดลกับวิธีแบบเดิม (คลาสสิคอลโมเดล) โดยมีจุดมุ่งหมายเพื่อเปรียบเทียบและทดสอบความแตกต่างของผลที่ได้จากการวิเคราะห์ข้อสอบโดยวิธีโมเดลโลจีสติก ชนิด 1, 2 และ 3 พารามิเตอร์ กับวิธีแบบเดิม กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2527 ในเขตอำเภอเมือง จังหวัดเชียงใหม่ จำนวน 1,667 คน เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลเป็นแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ (ค 101) วัดพฤติกรรมด้านทักษะการคิดคำนวณตามแนวคิดของวิลสัน คำนวณหาความเที่ยงของข้อสอบที่ได้รับการคัดเลือกไว้จากการวิเคราะห์ข้อสอบแต่ละวิธีจากสูตรแอลฟา (Coefficient Alpha) ใช้การทดสอบซี (Z-Test) หาความสอดคล้องในการคัดเลือกข้อสอบเป็นรายคู่ โดยใช้สหสัมพันธ์ฟาย (Phi-Coefficient) และทดสอบความมีนัยสำคัญของความสอดคล้องในการคัดเลือกข้อสอบโดยใช้ไคสแควร์ (Y^2-Test) ซึ่งจากผลการวิจัยพบว่า จำนวนข้อของข้อสอบที่ได้รับการคัดเลือกไว้จากการวิเคราะห์ข้อสอบแบบเดิมกับแบบ 1 พารามิเตอร์ แบบเดิมกับแบบ 2 พารามิเตอร์ และแบบเดิมกับแบบ 3 พารามิเตอร์ มีความแตกต่างอย่างมีนัยสำคัญที่ระดับ .05 โดยจำนวนข้อที่ได้รับการคัดเลือกจากวิธีแบบเดิม จะมีจำนวนมากกว่าวิธีแบบ 1, 2 และ 3 พารามิเตอร์ ส่วนค่าความเที่ยงของข้อสอบที่ได้รับการคัดเลือกไว้จากการวิเคราะห์ข้อสอบในแต่ละวิธี เมื่อปรับจำนวนข้อให้เท่ากันจะมีความแตกต่างกันอย่างไม่มีนัยสำคัญที่ระดับ .05 โดยค่าความเที่ยงของข้อสอบที่ได้รับการคัดเลือกไว้จากการวิเคราะห์ด้วยโมเดลโลจีสติก จะสูงกว่าโดยวิธีแบบเดิม

ในปี พ.ศ. 2527 ประดิษฐ์ เรืองตระกูล (2528) ได้ศึกษาเกี่ยวกับการประยุกต์ใช้โมเดลโลจีสติก 3 พารามิเตอร์ ในการวิเคราะห์แบบสอบวินิจฉัยวิชาคณิตศาสตร์ เรื่องสมการ โดยแบ่งเนื้อหาย่อยได้ 6 กลุ่ม ซึ่งแยกย่อยได้เป็น 13 จุดประสงค์การเรียนรู้ ในขั้นแรกสร้างเป็นแบบสอบเชิงสำรวจ ประเภทอัตนัย 120 ข้อ เพื่อค้นหาความบกพร่อง แล้ว

พัฒนาเป็นแบบสอบวินิจฉัย ประเภทเลือกตอบ 4 ตัวเลือก จำนวน 60 ข้อ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2527 จำนวน 1,913 คน วิเคราะห์ข้อสอบด้วยโปรแกรมไลจีส 5 ตรวจสอบคุณภาพแบบสอบด้านความเที่ยงด้วยสูตรแบบไบโนเมียล (Binomial) และตรวจสอบฟังก์ชันสารสนเทศแบบสอบ ในด้านความตรงได้ตรวจสอบความตรงเชิงเนื้อหาและความตรงเชิงวินิจฉัย ตามวิธีของโรวิเนลลีและแฮมเบิลตัน ตรวจสอบความตรงร่วมสมัยและความตรงเชิงพยากรณ์ โดยหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนของแบบสอบกับเกรดในภาคเรียนที่ 1 และภาคเรียนที่ 2 ตามลำดับ ซึ่งจากการศึกษาได้แบบสอบวินิจฉัยที่มีคุณภาพและมีประสิทธิภาพสูงสุด เมื่อวัดกับผู้สอบที่มีระดับความสามารถปานกลาง คือมีค่า θ ในช่วง -1.5 ถึง 1.5 ข้อสอบที่คัดเลือกไว้มีค่าอำนาจจำแนกในช่วงตั้งแต่ 0.19036 ถึง 2.00000 ค่าความยากตั้งแต่ -3.75288 ถึง 1.90366 และมีค่าการเดาตั้งแต่ 0.0 ถึง 0.5 ค่าความเที่ยงจากสูตรแบบไบโนเมียลของแบบสอบทั้งฉบับเป็น .9126 สำหรับค่าความตรงเชิงเนื้อหา ความตรงเชิงวินิจฉัย มีคะแนนเฉลี่ยผ่านเกณฑ์ที่ยอมรับ คือ 2.5 และค่าความตรงร่วมสมัย ความตรงเชิงพยากรณ์ เป็น .578 และ .665 ที่ระดับนัยสำคัญ .001 ตามลำดับ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย