บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

เอกสารและงานวิจัยที่เกี่ยวข้องกับงานวิจัยเรื่องการพัฒนาแบบทดสอบวัดทักษะการคิดคำนวณวิชาคณิตศาสตร์ โดยใช้ทฤษฎีการตอบสนองข้อสอบ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1 ผู้วิจัยได้ศึกษาเพื่อเป็นแนวทางในการดำเนินการวิจัย ดังต่อไปนี้

ทักษะการคิดคำนวณ

แบบทดสอบ

- ความหมายของแบบทดสอบ
- ลักษณะของแบบทดสอบ
- การสร้างแบบทดสอบ

ทฤษฎีการตอบสนองข้อสอบ

- หลักการของทฤษฎีการตอบสนองข้อสอบ
- ข้อตกลงของทฤษฎีการตอบสนองข้อสอบ
- พารามิเตอร์ของทฤษฎีการตอบสนองข้อสอบ
- รูปแบบของโมเดลโลจิสติก
- ฟังก์ชันสารสนเทศของข้อสอบและแบบทดสอบ
- การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบ
- สรุปแนวคิดทฤษฎีการตอบสนองข้อสอบ

เกณฑ์ปกติ

งานวิจัยที่เกี่ยวข้อง

ทักษะการคิดคำนวณ

โสภณ บำรุงสงค์ และสมหวัง ไตรต้นวงศ์ (2520 : 222) กล่าวไว้ว่า ความสามารถในการคิดคำนวณ เป็นความสามารถด้านการคิดเลขได้อย่างรวดเร็ว ถูกต้อง แม่นยำ และชำนาญในการบวก ลบคูณและหาร

สุรชัย ขวัญเมือง (2522 : 225 - 229) แบ่งทักษะการคิดคำนวณออกเป็น 3 ด้าน คือ ความรู้ที่เกี่ยวกับข้อเท็จจริง ความรู้เกี่ยวกับศัพท์ และความสามารถในการใช้กระบวนการในการแก้ปัญหา ลักษณะการทดสอบเป็นการวัดความคล่องแคล่วแม่นยำในการ บวก ลบ คูณ และหารตัวเลข (Compertational Skills) ไม่มีอิทธิพลของภาษาเข้ามาเกี่ยวข้องด้วยเลย

ไพศาล หวังพานิช (2526 : 128) กล่าวว่า ทักษะในการคิดคำนวณทางคณิตศาสตร์ (Operations) ได้แก่ ความคล่องแคล่วในการบวก ลบ คูณ หาร จำนวนต่าง ๆ ซึ่งในการทดสอบจะนิยมให้คำนวณเฉพาะตัวเลขเท่านั้น ไม่ต้องตั้งเป็นโจทย์ปัญหาให้คิดคำนวณ

พร้อมพรรณ อุดมสิน (2533 : 43) กล่าวว่า ทักษะในการคิดคำนวณได้แก่ การวัดความรู้ความจำแบบง่าย ๆ เกี่ยวกับสิ่งที่ผู้เรียนได้เรียนผ่านไปแล้ว เช่น ระลึกถึงแบบฝึกหัดที่ทำไปแล้วโดยไม่มีกระบวนการตัดสินใจ พฤติกรรมระดับนี้แบ่งเป็น 3 ขั้นคือ

1) ความรู้ความจำเกี่ยวกับข้อเท็จจริง (Knowledge of specific facts) หมายถึง การถามเพื่อจะวัดความรู้ความจำเกี่ยวกับเนื้อหาวิชาในรูปแบบหรือเกี่ยวกับที่ผู้เรียนได้รับจากการเรียนการสอนมาแล้ว นอกจากนี้ยังรวมถึงความรู้พื้นฐานซึ่งผู้เรียนต้องนำมาใช้

2) ความรู้ความจำเกี่ยวกับศัพท์และนิยาม (Knowledge of terminology) หมายถึง การถามให้ผู้เรียนบอกความหมายของคำศัพท์และนิยามต่าง ๆ ตามที่ได้เคยเรียนมาแล้ว โดยไม่ต้องอาศัยการคิดคำนวณแต่อย่างใดและไม่ต้องการหาความรู้อื่นมาช่วย

3) ความรู้ความจำเกี่ยวกับใช้กระบวนการคิดคำนวณ (Ability to carry outalgorithms) หมายถึง การที่ผู้เรียนสามารถนำสิ่งที่โจทย์กำหนดให้มาดำเนินการของกระบวนการคิดคำนวณในแบบที่ได้เรียนมาแล้ว ในขั้นนี้มิได้มุ่งหมายให้ผู้เรียนคิดหากระบวนการคิดคำนวณแบบใหม่ของตนเอง เช่น ลำดับขั้นตอนในการหารยาว ลำดับขั้นตอนในการหาตัวคูณร่วมน้อย หรือตัวหารร่วมมาก

ดวงเดือน อ่อนน่วม (2536 : 366) กล่าวว่า แบบทดสอบวัดทักษะการคิดคำนวณว่าเป็นแบบทดสอบวัดความรวดเร็วและแม่นยำในการคิดคำนวณ ลักษณะคำถามส่วนมากจึงเป็นตัวเลขโดด ๆ ถ้าจะใช้โจทย์ปัญหาด้วยก็ต้องเป็นโจทย์ปัญหาง่าย ๆ เพราะไม่ต้องการให้มีอิทธิพลของภาษาเข้ามาเกี่ยวข้อง

น้อมศรี เคท (2537 : 121 - A) ให้ความหมายว่า ทักษะการคิดคำนวณเป็นความสามารถในการคิดคำนวณเกี่ยวกับตัวเลขได้อย่างถูกต้องและรวดเร็ว ในการแก้โจทย์ปัญหาคณิตศาสตร์ นักเรียนจำเป็นต้องมีทักษะในการคำนวณ การมีทักษะในการคำนวณ คือ การที่นักเรียนสามารถบวก ลบ คูณ และหารได้อย่างถูกต้องแม่นยำและรวดเร็ว ดังนั้น เมื่อนักเรียนทราบประโยคสัญลักษณ์ของโจทย์ปัญหาแล้ว นักเรียนที่มีทักษะในการคำนวณ ครูควรหาทางช่วยนักเรียนจัดกิจกรรมหลาย ๆ อย่างที่จะส่งเสริมให้นักเรียนมีทักษะ เพื่อนักเรียนจะได้ใช้ทักษะในการคำนวณไปหาคำตอบในการแก้โจทย์ปัญหาได้อย่างถูกต้องแม่นยำและรวดเร็ว

สิริพร ทิพย์คง (2542 : 156) ทักษะมีบทบาทสำคัญในการเรียนวิชาคณิตศาสตร์ ถ้านักเรียนไม่มีทักษะในการบวก นักเรียนก็ยากที่จะเข้าใจเรื่องการคูณ ทักษะจะเกิดขึ้นได้เมื่อนักเรียนได้รับการฝึกฝน (drill) และการฝึกหัด (practice) อย่างสม่ำเสมอจนจำได้ ทำให้สามารถคิดคำนวณในเรื่องที่เรียนมาได้อย่างรวดเร็วและถูกต้อง ตัวอย่างของทักษะ ได้แก่ ทักษะการว่ายน้ำ ทักษะของการบวก ทักษะของการคูณ ทักษะการหาร ทักษะของการยกกำลัง เป็นต้น เมื่อนักเรียนได้ฝึกหัดหลาย ๆ ครั้งนักเรียนก็จะจำได้ มีความชำนาญ ทำได้รวดเร็วและถูกต้อง ในการฝึกทักษะนั้นต้องอาศัยเวลา และนักเรียนจะต้องมีความตั้งใจในการฝึกหัดทำ

การคิดคำนวณ ต้องอาศัยทักษะ การฝึกฝน ฝึกหัด จนมีความชำนาญ สามารถคิดคำนวณได้รวดเร็วและถูกต้อง (สิริพร ทิพย์คง, 2542 : 206)

สุวร กาญจนมยูร (2545 : 50) กล่าวว่า การที่นักเรียนจะสามารถหาคำตอบของโจทย์ปัญหานั้นได้โดยวิธีใด จะต้องอาศัยองค์ประกอบเกี่ยวกับการคำนวณ ขั้นนี้นักเรียนแต่ละคนต้องมีความสามารถในเรื่องต่อไปนี้

1. มีทักษะบวก ลบ คูณและหารจำนวนได้
2. มีทักษะการยกกำลังและการหารรากที่สอง รากที่สามของจำนวนได้
3. มีทักษะการแก้สมการ

สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ (2529 : 58) พฤติกรรมที่บ่งชี้ของทักษะการคิดคำนวณ คือ

1) คิดหาคำตอบจากประโยคสัญลักษณ์
2) คิดหาคำตอบจากโจทย์คำนวณ
3) คิดหาคำตอบจากการแทนค่าสูตรต่าง ๆ
4) คิดคำนวณหาพื้นที่ที่กำหนด
5) แสดงวิธีการคำนวณ

ฯลฯ

สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ กระทรวงศึกษาธิการ (2539 : 40) ได้กล่าวถึงความหมายของการวัดทักษะการคิดคำนวณว่า เป็นการตรวจสอบความสามารถของนักเรียนในการคิดหาคำตอบจากประโยคสัญลักษณ์เพื่อให้ได้คำตอบที่ถูกต้องและรวดเร็ว ข้อคำถามจะประกอบด้วยข้อคำถามจะประกอบด้วยตัวเลขและเครื่องหมาย บวก ลบ คูณ หาร เพราะไม่ต้องการให้มีอิทธิพลของภาษาเข้ามาทำให้ ผู้สอบต้องแปลความหรือแก้ปัญหา แต่ถ้ามีภาษาเข้ามาเกี่ยวข้องจะเป็นลักษณะของการบอกถึงจุดมุ่งหมายของข้อคำถาม

สำนักงานทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ (2540 : 8-9) ได้กล่าวถึงการวัดความสามารถทางการคิดคำนวณ (Numerical Ability) ไว้ว่า เป็นการวัดความเข้าใจในความคิดรวบยอดและหลักการทางคณิตศาสตร์ระดับเบื้องต้น ทักษะการคิดคำนวณระดับพื้นฐาน ความสามารถด้านเหตุผลเชิงปริมาณ (Quantitative Reasoning) การวิเคราะห์เปรียบเทียบและการประเมินข้อมูลเชิงปริมาณในรูปแบบต่าง ๆ เช่น ตาราง แผนภูมิ และประโยคสัญลักษณ์ทางคณิตศาสตร์ และความสามารถในการแก้โจทย์ปัญหาในรูปของจำนวนหรือปริมาณ การวัดความสามารถทางการคิดคำนวณกระทำใน 4 องค์ประกอบย่อยคือ

1) ความสามารถพื้นฐานทางคณิตศาสตร์ (Basic Mathermatical Ability) ข้อสอบวัดความรู้พื้นฐานทางคณิตศาสตร์จะครอบคลุมการใช้ Operation พื้นฐาน เช่น การบวก การลบ การคูณ การหาร การยกกำลัง และความสามารถในการตีความ และแก้ปัญหาโจทย์ที่ต้องอาศัยพื้นฐานความเข้าใจในความคิดรวบยอดและหลักการในคณิตศาสตร์ระดับเบื้องต้น เช่น เศษส่วน ทศนิยม ร้อยละ กำไร – ขาดทุน ระยะทาง พื้นทาง พื้นที่ ปริมาณ เวลา สมการ เรขาคณิตเบื้องต้น พีชคณิตเชิงเส้นขั้นพื้นฐานและกราฟ เป็นต้น

2) ความสามารถด้านการเปรียบเทียบเชิงปริมาณ (Quantitative Comparison) ข้อสอบด้านนี้วัดความสามารถของผู้เรียนในการคำนวณหาเหตุผล และเปรียบเทียบขนาดของปริมาณ 2 ปริมาณอย่างรวดเร็วและแม่นยำว่าปริมาณใดมีขนาดใหญ่กว่า เล็กกว่า เท่ากันหรือเปรียบเทียบกันไม่ได้เพราะข้อมูลที่กำหนดให้เพียงพอ การคำนวณและการหาเหตุผลจะต้องอาศัยความรู้ความเข้าใจเบื้องต้นในวิชาพีชคณิต เลขคณิต และเรขาคณิต

3) ความสามารถในการตีความข้อมูล (Data Interpretation) ข้อสอบด้านนี้วัดความสามารถของผู้เรียนด้านการอ่าน และวิเคราะห์ข้อมูลเชิงปริมาณที่เสนอในรูปแบบต่าง ๆ เช่น ตาราง แผนภูมิ กราฟ สมการหรือประโยคสัญลักษณ์ทางคณิตศาสตร์ การสังเคราะห์ข้อมูล และการเลือกใช้ข้อมูลที่เหมาะสมเพื่อตอบปัญหา

4) การประเมินความเพียงพอของข้อมูล (Evaluation of Data Sufficiency) เพื่อวัดความสามารถในการคิดวิเคราะห์และตีความปัญหาทางคณิตศาสตร์ การพิจารณาและจำแนกความเหมาะสมของข้อมูลในการแสวงหาคำตอบ และการประเมินความเพียงพอของข้อมูลสำหรับการแก้ปัญหา

สรุปความหมาย ทักษะการคิดคำนวณ หมายถึง ความสามารถในการคำนวณ การคิดหาคำตอบได้รวดเร็วภายในระยะเวลาจำกัด และถูกต้อง ในการบวก ลบ คูณ หาร

แบบทดสอบ

1. ความหมายของแบบทดสอบ

แบบทดสอบ (Tests) หมายถึง ชุดของข้อคำถาม (Items) หรือกลุ่มงานใด ๆ ที่สร้างขึ้นมาเพื่อใช้เป็นสิ่งเร้าในการวัดคุณสมบัติอย่างใดอย่างหนึ่งของผู้สอบ เมื่อผู้สอบได้รับการทดสอบหรือได้รับสิ่งเร้าดังกล่าวนี้ ก็จะแสดงพฤติกรรมตอบสนองออกมาให้อยู่ในรูปที่สามารถสังเกตได้และวัดได้ เช่น พูด เขียน แสดงท่าทาง หรือใช้สัญลักษณ์อย่างใดอย่างหนึ่ง เป็นต้น อันจะนำไปสู่การแปลความหมายผลการทดสอบได้ ตามความหมายที่กล่าวมานี้ แบบทดสอบจะหมายรวมถึงแบบทดสอบทั่ว ๆ ไป ได้แก่ แบบสอบถามและการสัมภาษณ์ด้วย เพราะมีข้อคำถามและมีการตอบสนองซึ่งสามารถสังเกตและวัดได้ แต่ในที่นี้ แบบทดสอบจะหมายถึงเฉพาะแบบทดสอบที่ใช้วัดพฤติกรรมด้านความรู้ และการคิด (Cognative domain) เท่านั้น ซึ่งอาจเรียกทั่ว ๆ ไปว่า “ข้อสอบ” ก็ได้ (พวงรัตน์ ทวีรัตน์ 2530 : 18) ในที่นี้จะกล่าวถึงเฉพาะแบบทดสอบปรนัย ดังนี้

พวงรัตน์ ทวีรัตน์ (2530 : 125) ข้อสอบปรนัย (Objective test) เป็นข้อสอบที่ใช้วัดพฤติกรรมด้านความรู้และความคิด (Cognitive domain) ได้โดยตรง และเป็นอย่างดี สามารถสร้างให้วัดเนื้อหาได้สอดคล้องกับหลักสูตร และถามได้อย่างกว้างขวาง ส่วนดีของข้อสอบปรนัยที่เห็นเด่นชัดประการหนึ่งคือ สามารถนำผลไปวิเคราะห์ทางสถิติได้มากมาย ทำให้การวัดผลมีลักษณะเป็นวิทยาศาสตร์ขึ้น ข้อสอบปรนัยที่เป็นที่รู้จักและใช้กันอย่างแพร่หลายมีอยู่ 4 รูปแบบคือ

1. แบบถูกผิด (True false item)

ลักษณะ เป็นข้อสอบที่มีคำตอบให้เลือกอยู่สองทาง เช่น ถูก – ผิด , ใช่ – ไม่ใช่ , จริง – ไม่จริง เป็นต้น ข้อสอบรูปแบบนี้จะมีค่าอำนาจจำแนกต่ำ เพราะผู้ตอบมีโอกาสเดาได้ถึง 50 % ข้อสอบแบบถูกผิดนี้ใช้ได้ดีกับการสอบวัดที่ต้องการตัวถูกจริง ๆ เท่านั้น

ข้อดี

1) สร้างง่าย และประหยัดเวลาในการสร้าง

2) ใช้วัดพฤติกรรมด้านความรู้ความจำเกี่ยวกับสูตร กฎ ทฤษฎี ข้อเท็จจริงต่าง ๆ รวมถึงความหมายของศัพท์ในวิชาต่าง ๆ ได้เป็นอย่างดี นอกจากนี้ถ้าผู้ออกข้อสอบมีทักษะหรือมีความชำนาญในการออกข้อสอบเป็นอย่างดี ก็สามารถสร้างข้อสอบวัดพฤติกรรมด้านความเข้าใจ และการนำไปใช้ได้ด้วย

3) สามารถออกข้อสอบได้มากข้อ และวัดเนื้อหาได้ครอบคลุมหลักสูตร ทำให้มีความเที่ยงตรงตามเนื้อหาดี

4) ตรวจง่ายและใช้เวลาในการตรวจน้อย

5) การตรวจคะแนนมีความยุติธรรม แน่นอน และเป็นปรนัย

ข้อเสีย

1) ข้อสอบแต่ละข้อมีโอกาสเดาถูก 50 % นับว่าเปอร์เซ็นต์การเดามีสูง ทำให้ข้อสอบขาดอำนาจจำแนก

2) ไม่สามารถวินิจฉัยหรือชี้ข้อบกพร่องของการเรียนแต่ละเรื่องได้ เพราะการตอบของนักเรียนมีการเดาเข้ามาเกี่ยวข้องอยู่มาก

3) ความเชื่อมั่นของคะแนนค่อนข้างต่ำ เนื่องจากการเดานั่นเองจึงควรออกข้อสอบให้มากข้อ

4) ไม่เหมาะสำหรับการถามในสิ่งที่ต้องการตีความ เพราะคำตอบอาจมีได้หลายกรณี

2. แบบจับคู่ (Matching item)

ลักษณะ เป็นข้อสอบที่กำหนดเป็นคำ วลี หรือประโยค ฯลฯ อยู่ 2 กลุ่ม โดยมากมักแยกเป็นแถวทางซ้ายมือกับขวามือ ที่มีความสัมพันธ์กัน ให้ผู้สอบพิจารณาจับคู่ให้ถูกต้อง โดยที่ถ้าข้อใดถูกเลือกจับคู่ไปแล้ว จะนำไปจับคู่กับข้ออื่นอีกไม่ได้

ข้อดี

1) สร้างง่าย ประหยัดเวลาในการสร้าง

2) ใช้วัดความรู้ความจำเกี่ยวกับการระลึกได้ถึงความสัมพันธ์ระหว่างข้อเท็จจริงสองเรื่องได้ดี

3) ขจัดโอกาสการเดาได้ดี (เมื่อกำหนดจำนวนข้อให้แตกต่างกันพอสมควร)

4) ตรวจง่าย ใช้เวลาในการตรวจน้อย

ข้อเสีย

1) วัดสมรรถภาพสมองขั้นสูงได้ไม่ดีนัก

2) ความยากแต่ละข้อไม่เท่ากัน ข้อแรก ๆ จะยากแต่ข้อท้าย ๆ จะง่ายขึ้น

3) คะแนนที่ได้ในการตอบแต่ละข้อไม่ยุติธรรม ข้อแรก ๆ ได้คะแนนยาก ส่วนข้อหลัง ๆ จะได้คะแนนง่ายขึ้น ทั้งนี้เนื่องจากเมื่อตัวเลือกใดถูกเลือกไปจับคู่กับข้อคำถามใดแล้ว จะไม่มีโอกาสจับคู่กับคำถามข้ออื่นได้อีก ทำให้โอกาสการเลือกจับคู่ข้อหลัง ๆ ได้ถูกมีมากขึ้น

3. แบบเติมคำหรือข้อความ (Completion item)

ลักษณะ เป็นข้อสอบที่เว้นช่องว่างต่อจากคำถามไว้ เพื่อให้ผู้สอบเขียนตอบแบบอธิบายสั้น ๆ หรือตอบเฉพาะคำ ข้อความ วลี เป็นต้น โดยให้ได้ความที่สมบูรณ์ที่สุด ตามนัยแห่งคำถามนั้น ๆ

ข้อดี

1) สร้างง่าย และประหยัดเวลาในการสร้าง

2) ใช้วัดพฤติกรรมด้านความรู้ความจำและความเข้าใจได้เป็นอย่างดี โดยเฉพาะด้านความรู้ความจำ ใช้วัดการระลึกได้เป็นอย่างดี เพราะผู้ตอบต้องนึกหาคำตอบเอง ไม่มีตัวเลือกเป็นสื่อเชื่อมโยงให้ระลึกได้ การที่มีตัวเลือกปรากฏให้เห็น ไม่เป็นการจำโดยแท้

3) ลดโอกาสการตอบถูกโดยการเดาได้มาก เพราะต้องเขียนคำตอบเอง

4) สร้างข้อคำถามได้มากข้อ วัดเนื้อหาได้กว้างขวางและครอบคลุมหลักสูตร ทำให้มีความเที่ยงตรงตามเนื้อหาพอดี

5) ช่วยให้ทราบแน่ชัดว่า ผู้ตอบมีความรู้ในเรื่องนั้น ๆ หรือไม่และมากน้อยเพียงใด

6) เหมาะสำหรับการถามในวิชาคณิตศาสตร์ และวิทยาศาสตร์

ข้อเสีย

1) ไม่เหมาะที่จะใช้วัดความสมรรถภาพสมองขั้นสูง เช่น การใช้เหตุผลการวิเคราะห์ และการสังเคราะห์

2) การตรวจให้คะแนน ตรวจยากและคะแนนอาจไม่ยุติธรรม เพราะคำตอบมีได้ต่าง ๆ ทำให้วินิจฉัยลำบาก ซึ่งเป็นผลทำให้คะแนนขาดความเป็นปรนัยและความเชื่อมั่นได้

3) ความบกพร่องทางภาษาของผู้ตอบ อาจทำให้คะแนนเกิดความคลาดเคลื่อนได้

4. แบบเลือกตอบ (Multiple choice item)

ลักษณะ ข้อสอบเลือกตอบประกอบด้วยส่วนที่สำคัญ 2 ส่วนคือ

ตอนนำ เป็นส่วนที่เป็นคำถามนำของเรื่องราว (Question or item or problem)

ตอนเลือก เป็นส่วนที่เป็นตัวเลือกที่สอดคล้องกับเรื่องราวในตอนนำ (Choice or option) ซึ่งส่วนที่เป็นตัวเลือกนี้ จะประกอบด้วยตัวเลือกที่เป็นคำตอบถูกหรือเรียกตัวถูก (Key response) กับตัวเลือกที่เป็นคำตอบผิด ซึ่งเรียกว่า “ตัวลวง” (Distractors)

ข้อดี

1) สามารถวัดสมรรถภาพทางสมองด้านต่าง ๆ ได้ครบถ้วน ตั้งแต่พฤติกรรมขั้นต่ำคือความรู้ความจำไปจนถึงพฤติกรรมขั้นสูงสุดคือการประเมินค่า ทำให้สามารถใช้กับข้อสอบได้ทุกรายวิชาและมีความเที่ยงตรงตามโครงสร้างสูง

2) สามารถถามเนื้อหาได้กว้าง ครอบคลุมหลักสูตร ทำให้มีความเที่ยงตรงตามเนื้อหาสูง

3) ตรวจง่าย ใช้เวลาตรวจน้อย และสามารถตรวจให้คะแนนโดยใช้เครื่องจักรกลได้ ซึ่งช่วยลดค่าใช้จ่าย เวลา และแรงงานได้มาก

4) คะแนนมีความเชื่อมั่นสูง เพราะการที่มีตัวเลือกหลายตัวเป็นการช่วยลดการเดาลงได้มาก ซึ่งเป็นข้อได้เปรียบของข้อสอบเลือกตอบ เมื่อเทียบกับข้อสอบปรนัยแบบอื่น ๆ ทำให้เป็นที่นิยมใช้

5) มีลักษณะที่จะทำให้เกิดความเป็นปรนัยได้มากเมื่อเทียบกับข้อสอบรูปแบบอื่น ๆ

6) ครูสามารถนำผลที่ได้จากการตอบไปใช้ในการวินิจฉัยการเรียนของเด็กได้ว่า เขาบกพร่องในเนื้อหาตอนใด

7) สามารถตรวจสอบคุณภาพของข้อสอบได้ทั้งเป็นรายข้อและรายตัวเลือก ทำให้สามารถปรับปรุงแก้ไขข้อสอบให้มีคุณภาพได้อย่างตรงจุด

8) สามารถใช้ภาพประกอบการถามได้ดี เช่น ภาพ แผนภูมิ ตาราง กราฟ แผนที่ เป็นต้น ซึ่งจะช่วยลดคำอธิบายลงได้มาก และช่วยให้เข้าใจได้ชัดเจนและรวดเร็วขึ้นด้วย

9) มีรูปแบบการถามหลายอย่าง ซึ่งช่วยกระตุ้นให้ผู้ตอบต้องใช้ความคิด และความสามารถอย่างเต็มที่ เป็นการพัฒนาสมองของผู้ตอบเป็นอย่างดี ซึ่งสอดคล้องกับหลักการวัดผล

ข้อเสีย

1) สร้างยาก โดยเฉพาะการคิดตัวลวง ดังนั้นผู้ที่จะเขียนข้อสอบเลือกตอบจึงควรที่จะต้องศึกษาถึงเทคนิคการเขียนข้อคำถาม การเขียนตัวเลือกและฝึกฝนบ่อย ๆ เพื่อให้เกิดทักษะในการเขียน ก็จะทำให้สามารถสร้างข้อสอบปรนัยเลือกตอบที่มีคุณภาพได้

2) ใช้เวลาในการสร้างมาก

3) ค่าใช้จ่ายสูง และต้องใช้แรงงานสูง (ผู้ออกข้อสอบต้องใช้ความคิดและใช้สมองมาก)

4) การกำหนดตัวเลือกให้ผู้ตอบตอบเป็นการจำกัดความเป็นอิสระในการคิดของผู้ตอบ ข้อสอบเลือกตอบจึงไม่สามารถใช้วัดความคิดสร้างสรรค์ของผู้ตอบได้

2. ลักษณะของแบบทดสอบ

พร้อมพรรณ อุดมสิน (2544 : 109 -113) แบบทดสอบเป็นสิ่งจำเป็นอย่างยิ่งในกระบวนการวัดและประเมินผลการเรียนการสอน เพราะเป็นเครื่องมือในการวิจัย อันได้แก่ คะแนนจากแบบทดสอบซึ่งจะนำไปใช้ในการประเมินคุณภาพการเรียนการสอน และในการประเมินผลจะเป็นไปด้วยความถูกต้อง ยุติธรรม มีประสิทธิภาพ ก็ย่อมต้องอาศัยข้อมูลที่ดีมีความเที่ยงตรงสูง ซึ่งมาจากแบบทดสอบที่ดีมีคุณภาพนั่นเอง โดยทั่วไปลักษณะแบบทดสอบที่ดี มีดังนี้คือ

1. ความตรง (Validity)
2. ความเที่ยง (Reliability)
3. ความเป็นปรนัย (Objectivity)
4. ความยาก (Difficulty)

5. อำนาจจำแนก (Discrimination)

6. ความยุติธรรม (Fairness)

7. ยั่วยุ (Exemplary)

8. ถามลึก (Searching)

9. ประสิทธิภาพ (Effeciency)

10. ประโยชน์ใช้สอย (Usability)

ลักษณะ 10 ประการดังกล่าว จะเป็นเกณฑ์ที่ใช้ในการพิจารณาตัวแบบทดสอบที่จะใช้ว่าเป็นแบบทดสอบที่ดีมีคุณภาพ ซึ่งลักษณะต่าง ๆ นั้นมีรายละเอียดดังนี้

1. ความตรง หมายถึง ความสามารถของแบบทดสอบ ซึ่งจะวัดสิ่งที่ต้องการจะวัดให้ถูกต้องตามจุดประสงค์ เพราะจุดประสงค์สำคัญของการสอบนั้นเพื่อให้ได้ข้อมูลข่าวสารคะแนนที่เป็นตัวแทนสิ่งที่ต้องการวัด เช่น ต้องการวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ของนักเรียนก็ต้องสร้างแบบวัดผลสัมฤทธิ์ ถ้าต้องการวัดความถนัดทางการเรียนคณิตศาสตร์ก็ต้องสร้างแบบวัดความถนัด ถ้าสร้างเครื่องมือวัดความถนัดทางการเรียนคณิตศาสตร์แล้ว คะแนนที่ได้ออกมาสามารถวัดได้จริงว่า ผู้ที่ได้คะแนนสูงสุดคือผู้ที่มีความถนัดมากในการเรียนคณิตศาสตร์ ส่วนผู้ที่ได้คะแนนต่ำก็เป็นผู้ที่มีความถนัดน้อยในการเรียนคณิตศาสตร์ แบบทดสอบฉบับนี้ก็ได้ชื่อว่าวัดได้ตรงกับสิ่งที่ต้องการวัด

2. ความเที่ยง หมายถึง ความคงเส้นคงวาของคะแนนในการวัดแต่ละครั้ง หรือกล่าวได้ว่าใช้เครื่องมือนั้นวัดครั้งใด ๆ ก็ได้ค่าเท่าเดิมไม่เปลี่ยนแปลง เพราะถ้าหากข้อมูลไม่คงเส้นคงวา สับสนรวนเร เชื่อไม่ได้เพราะไม่เที่ยง กล่าวคือวัดซ้ำแล้วคะแนนเปลี่ยนแปลงไป สอบครั้งแรกได้คะแนนสูงครั้งต่อมาสอบในเงื่อนไขเดียวกันและกลุ่มเดียวกันอีก คะแนนที่ได้สัมพันธ์กับครั้งแรกน้อย เช่นนี้เชื่อว่าคะแนนไม่มีความเที่ยง เมื่อนำไปตัดสินหรือประเมินก็จะเกิดความคลาดเคลื่อนไม่แน่นอน

ค่าดัชนีที่ใช้วัดความเที่ยงใช้ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation : r) ระหว่างคะแนนสองชุดใช้แบบทดสอบฉบับเดียวกัน เมื่อสอบกับกลุ่มเดียวกัน ค่าความเที่ยงมีค่าอยู่ระหว่าง –1.00 กับ 1.00 ถ้าแบบสอบฉบับใดมีความเที่ยงใกล้ ๆ ศูนย์ ถือว่ามีความเที่ยงต่ำ คะแนนจากการสอบครั้งแรกกับครั้งหลังจะมีอันดับใกล้เคียงกัน แต่ถ้ามีความเที่ยงใกล้ๆ –1.00 จะมีความสัมพันธ์กลับกัน ใครเคยได้คะแนนสูงครั้งแรก สอบครั้งหลังจะได้คะแนนต่ำ ใครเคยได้คะแนนต่ำไว้ก่อนจะได้คะแนนสูงในครั้งหลัง

แบบทดสอบที่มีค่าความเที่ยงสูงจะให้ข้อมูลที่ดี ดังนั้นในการสอบจึงต้องใช้แบบที่มีความเที่ยงสูง ๆ ยิ่งสูงเท่าใดยิ่งดี ตามปกติแบบทดสอบที่ครูผู้สอนสร้างเองไม่ควร มีค่าความเที่ยงไม่ต่ำกว่า 0.60 จึงใช้ได้ผลดี

3. ความเป็นปรนัย เป็นคุณสมบัติที่ดีของแบบทดสอบ หมายถึง มีความแจ่มชัดในคำถามในการตรวจให้คะแนนและการแปลความหมายของคะแนน กล่าวคือ ข้อสอบที่มีลักษณะเป็นปรนัยจะมีคุณสมบัติ 3 ประการดังต่อไปนี้

3.1 แจ่มชัดในคำถาม หมายถึง ผู้สอบอ่านคำถามแล้วเข้าใจว่าต้องการถามอะไร และเข้าใจตรงกันทุกคน เข้าใจคำถามตรงกับผู้ออกข้อสอบ ถ้าข้อสอบข้อใดอ่านแล้วกำกวมจะมีความเป็นปรนัยน้อย

3.2 แจ่มชัดในการตรวจให้คะแนน คือ มีกฎเกณฑ์การให้คะแนนไว้ชัดเจน ไม่ว่าใครจะมาเป็นผู้ตรวจ จะตรวจเมื่อใดก็ให้คะแนนได้ตรงกัน ไม่มีความเห็นส่วนตัวเข้ามาปนกันในการให้คะแนน

3.3 แจ่มชัดในการแปลความหมายคะแนน หมายถึง บอกความหมายของคะแนนที่ได้ว่ามีความหมายอย่างไรได้อย่างมีความเที่ยงสูง เช่น สามารถเปรียบเทียบคะแนนที่ได้จริงว่าใครได้คะแนนสูง แสดงว่ามีความสามารถสูงกว่าคนที่ได้คะแนนต่ำจริง คุณสมบัติข้อนี้เกี่ยวข้องกับความตรงและความเที่ยงของข้อสอบอย่างมากด้วย

4. ความยาก หมายถึง ข้อสอบในแต่ละข้อของแบบทดสอบสามารถรวมถึงสัดส่วนหรือค่าร้อยละของคนที่ทำข้อสอบข้อนั้น ๆ ได้ถูก ซึ่งนักวัดผลใช้สัดส่วนหรือค่าร้อยละนี้เป็นดัชนีค่าความยาก ค่าความยากของแบบสอบจะมีค่าอยู่ระหว่าง 0.00 ถึง 1.00 ข้อสอบที่มีค่าความยากเป็น 1.00 เป็นข้อสอบที่ง่ายที่สุด เพราะทุกคนทำได้ถูกต้อง ข้อสอบที่มีค่าความยากปานกลางจึงมีค่าดัชนีความยาก 0.50 ซึ่งจะมีคนทำถูกครึ่งหนึ่งและทำผิดครึ่งหนึ่ง

ข้อสอบที่มีความยากปานกลางเป็นข้อสอบที่ดี เพราะจะช่วยให้แปลความหมายของคะแนนได้ดี ช่วยให้ข้อสอบมีค่าอำนาจจำแนกสูง ทำให้การกระจายของคะแนนมีความแปรปรวนสูง ซึ่งจะมีผลทำให้แบบสอบทั้งฉบับมีความเที่ยงสูงตามไปด้วย แต่แบบสอบฉบับหนึ่งยากที่จะมีข้อสอบที่มีความยากปานกลางทุกข้อ จึงนิยมคละกันระหว่างข้อที่ยากมาก ปานกลาง และยากน้อย เช่น แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนนิยมใช้ค่าความยากระหว่าง 0.20 ถึง 0.80

การเรียงข้อสอบในแบบทดสอบฉบับหนึ่ง ๆ สามารถเรียงได้หลายวิธี เช่น เรียงความยากของข้อสอบจากง่ายสุดไปหายากสุด หรือเรียงตามเนื้อหาโดยไม่คำนึงถึงความยาก หรืออาจจะเรียงตามเนื้อหาในแต่ละเนื้อหา และเรียงจากข้อสอบที่ยากน้อยไปหายากมากในแต่ละเนื้อหา ด้วยการเรียงตาม

ความยากจากง่ายสุดไปหายากสุด มีผลดีตรงที่ผู้เรียนจะพบข้อง่าย ๆ ก่อน ทำให้เกิดกำลังใจ แต่มีข้อเสียตรงที่ทำให้ผู้สอบเปลี่ยนความคิดตลอดเวลา มีความคิดไม่ต่อเนื่อง แต่การเรียงตามเนื้อหาโดยไม่คำนึงความยาก จะมีผลเสียที่ผู้สอบพบข้อยาก ๆ ก่อน จะเกิดความท้อใจที่จะทำต่อ ส่วนการเรียงตามเนื้อหาแล้วเรียงตามความยากจากง่ายไปหายากในเนื้อหาหนึ่ง ๆ ช่วยให้ผู้สอบใช้วงจรความคิดเดียวในเนื้อหาหนึ่ง ๆ และเกิดกำลังใจเมื่อพบข้อง่ายตอนต้น ๆ ของแต่ละเนื้อหา การเรียงลักษณะนี้ผู้สอบจะทำข้อสอบได้ดีกว่าแบบอื่น ๆ

5. อำนาจจำแนก คุณสมบัติของข้อสอบลักษณะนี้เป็นคุณสมบัติที่จะจำแนกคนออกเป็น เก่งมาก เก่งน้อย ตามความสามารถจริงของเขา แบบทดสอบที่มีอำนาจจำแนกสูงจะจำแนกผู้สอบออกตามความสามารถได้ดี และจำแนกได้ละเอียดถี่ถ้วนตั้งแต่เก่งสุดถึงอ่อนสุด โดยสามารถจำแนกได้ถูกต้องแทบทุกคน ซึ่งมีลักษณะที่สำคัญคือ มีพิสัยกว้าง ความแปรปรวนสูง ทั้งนี้เพราะโดยธรรมชาติแล้วไม่ควรมีใครเลยในกลุ่มที่จะมีความสามารถเท่ากันพอดี จะต้องแตกต่างกันไม่มากก็น้อย ถ้าแบบทดสอบมีอำนาจจำแนกสูงจริงแล้วคะแนนของแต่ละคนไม่ควรจะซ้ำกันเลย ซึ่งจะยังผลให้คะแนนมีการกระจายมาก

ค่าอำนาจจำแนกของข้อสอบจะหามาจาก สหสัมพันธ์แบบไบซีเรียล (Biserial Correlation) เขียนย่อว่า r_{bis} หรือ r ซึ่งจะมีค่าอยู่ระหว่าง –1.00 ถึง +1.00 ความหมายของค่าอำนาจจำแนกแต่ละลักษณะมีดังนี้

5.1 ข้อสอบใดที่คนเก่งทุกคนทำถูกหมด คนอ่อนทำผิดหมด ค่าอำนาจจำแนกจะมีค่า +1.00 ซึ่งจะเป็นข้อสอบที่ดีเยี่ยม

5.2 ข้อสอบข้อใดที่คนเก่งทุกคนทำผิดมาก คนอ่อนทำถูกหมด ค่าอำนาจจำแนกจะมีค่า –1.00 ข้อสอบนี้จัดเป็นข้อสอบที่ไม่ดี เพราะเป็นข้อสอบที่ลวงคนเก่งซึ่งจะผิดจุดประสงค์

5.3 ข้อสอบใดที่คนเก่งกับคนอ่อนทำได้ถูกเท่า ๆ กัน จะมีค่าอำนาจจำแนก 0.00 ข้อสอบนี้ไม่สามารถแยกคนเก่งออกจากคนอ่อนได้

จากดัชนีค่าอำนาจจำแนกตามที่กล่าวมา จะเห็นว่าแบบทดสอบที่ดีควรจะเลือกข้อสอบที่มีค่าอำนาจจำแนกสูงใกล้ +1.00 ในทางปฏิบัตินั้น ข้อสอบวัดผลสัมฤทธิ์ทางการเรียน ที่จะใช้ได้ควรเลือกข้อสอบที่มีค่าอำนาจจำแนกตั้งแต่ 0.20 ขึ้นไป

6. ความยุติธรรม เป็นแบบทดสอบที่ให้ความเสมอภาคเท่าเทียมกันที่ผู้สอบจะทำข้อสอบได้ตามความสามารถจริงของเขาในวิชานั้น ๆ ลักษณะที่สำคัญของแบบทดสอบที่มีความยุติธรรมมีลักษณะดังนี้

6.1 เป็นแบบสอบที่ไม่มีความลำเอียงเข้าข้างกลุ่มใด กล่าวคือ คะแนนที่ผู้สอบได้รับเป็นความสามารถจริง ๆ ของเขาในแขนงวิชานั้น ๆ เช่น วัดความสามารถทางคณิตศาสตร์ก็ไม่เปิดโอกาสให้ผู้ที่เก่งภาษาอังกฤษได้เปรียบ โดยใช้ศัพท์ภาษาอังกฤษหรือออกข้อสอบเป็นภาษาอังกฤษโดยผู้ที่มีความสามารถจริงทางคณิตศาสตร์เสียเปรียบเพราะจะต้องใช้ความรู้ทางภาษาเข้ามาเกี่ยวข้องด้วย ในกรณีที่สอนหลายกลุ่มไม่เหมือนกัน ก็ไม่ออกข้อสอบลำเอียงเข้ากับกลุ่มใดกลุ่มหนึ่ง

6.2 เป็นแบบทดสอบที่ไม่เปิดโอกาสให้คนเก่งหรืออ่อนเดาข้อสอบได้ ข้อสอบที่ยุติธรรมจะต้องวัดให้ครอบคลุม มีจำนวนมากข้อ และสอบหลาย ๆ ครั้งจึงจะดี

7. ยั่วยุ เป็นข้อสอบที่มีลักษณะท้าทาย ชวนให้คิดต่อ ใคร่อยากรู้เรื่องนั้นให้กว้างขวางลึกซึ้งยิ่งขึ้น ข้อสอบที่มีลักษณะเช่นนี้ต้องมีความยากง่ายพอเหมาะ ไม่ยากเกินไปและไม่ง่าย ไม่มีความซ้ำซาก ข้อสอบที่มีคุณสมบัติยั่วยุจะต้องมีคำถามที่เข้าใจวัดลึก ๆ จะช่วยปลุกสมองคิดต่อจากข้อสอบที่จะศึกษาต่อ อ่านต่อ ให้ผู้เรียนใคร่กระหายอยากรู้ต่อไป

8. ถามลึก ข้อสอบที่มีคุณค่าทางการศึกษาควรจะเป็นข้อสอบที่ให้ผู้สอบได้คิดค้นคำตอบด้วยความสามารถในระดับสติปัญญาที่อยู่ในขั้นสูง ข้อสอบที่ส่วนใหญ่วัดในด้านความจำจะมีคุณค่าน้อยกว่าการวัดที่ใช้คุณค่าสมองในระดับสูง เพราะข้อสอบที่ล้วงลึกจะทำให้ผู้สอบได้พัฒนาความสามารถที่กล้าแข็งต่อไป

9. ประสิทธิภาพ เป็นแบบทดสอบที่ให้คะแนนได้เที่ยงตรงมากที่สุด โดยใช้เวลา แรงงาน เงินทองน้อยที่สุด ข้อสอบที่มีคุณสมบัติเช่นนี้ ในแต่ละข้อจะต้องเป็นตัวแทนกลุ่มความรู้ได้กว้างวัดได้ตรง อำนาจจำแนกสูง มีความเที่ยงตรงสูง

10. ประโยชน์ใช้สอย หมายถึง การนำไปใช้สะดวก มีประโยชน์ใช้สอยได้สูง โดยมีลักษณะสำคัญดังนี้

10.1 สะดวกต่อการดำเนินการสอบ หมายถึง เป็นแบบทดสอบที่ง่ายต่อการดำเนินการสอบไม่ทำให้เกิดปัญหาหรือความยุ่งยากซับซ้อน ในการที่จะบรรลุเป้าหมายนี้ แบบทดสอบที่ดีจึงต้องมีคำชี้แจงสำหรับการดำเนินการสอบ เช่น ระบุว่าแบบทดสอบเป็นแบบใด มีกี่ข้อ ใช้เวลาทำเท่าไร ตอบอย่างไร มีตัวอย่างการตอบ การแก้คำตอบ การถามเมื่อมีข้อสงสัย การส่งกระดาษคำตอบและอื่น ๆ ที่จำเป็น ซึ่งมีลักษณะการดำเนินการสอบที่ไม่มีประสิทธิภาพหรือผิดพลาด จะมีผลต่อความตรงและความเที่ยงของแบบทดสอบ

10.2 เวลาที่ใช้ในการสอบพอเหมาะ การกำหนดเวลาให้เหมาะสมกับจำนวนข้อสอบมีผลต่อความเที่ยงตรงของข้อสอบ เช่น ให้เวลามากเกินไป ทำให้มีเวลาในการพิจารณามาก ยังผลให้การจำแนกคนได้น้องลง ถ้าให้เวลาน้อยเกินไป คะแนนจากการเดาจะเข้าไปปะปนมาก

10.3 สะดวกในการตรวจให้คะแนน แบบทดสอบที่มีประโยชน์ใช้สอยสูง จะมีการตรวจให้คะแนนสะดวก รวดเร็ว และถูกต้อง การให้คะแนนไม่ซับซ้อน ไม่ต้องใช้เวลาในการตรวจมาก

10.4 ง่ายต่อการแปลผลและการนำไปใช้ หมายถึง การแปลความหมายของคะแนนที่ได้มานั้นอย่างถูกต้อง ชัดเจน ซึ่งจะนำไปใช้ได้อย่างมีประสิทธิภาพ อันจะเป็นการช่วยเพิ่มคุณภาพของการสอบขึ้นอีกส่วนหนึ่ง

3. การสร้างแบบทดสอบ

พวงรัตน์ ทวีรัตน์(2530 : 46 – 56) กล่าวว่า การสร้างแบบทดสอบนั้นมี 6 ขั้นตอนที่สำคัญ คือ

1. ขั้นวางแผน
2. ขั้นเตรียมงาน
3. ขั้นลงมือปฏิบัติ
4. ขั้นประเมินหรือตรวจสอบคุณภาพ
5. ขั้นจัดพิมพ์
6. การจัดทำคู่มือการใช้

ขั้นตอน 6 ขั้นดังกล่าว ผู้สร้างได้ทราบขั้นตอนในการสร้างและปฏิบัติตามขั้นตอนแต่ละขั้น จะทำให้สามารถสร้างข้อสอบได้อย่างมีประสิทธิภาพ ซึ่งขั้นตอนแต่ละขั้นตอนนั้นมีรายละเอียดดังนี้

1. ขั้นวางแผน

โดยทั่วไป ในการสร้างข้อสอบควรจะกระทำกันในรูปของคณะกรรมการจะเหมาะสมกว่า เพราะจะทำให้ได้ข้อสอบที่มาจากหลายแนวความคิด และมีแง่มุมต่าง ๆ ในการถามกว้างขวางมากขึ้น แต่ในกรณีที่ครูผู้สอนจะต้องเป็นผู้ออกข้อสอบประจำวิชาผู้สอนแต่เพียงผู้เดียว ก็เป็นความจำเป็นที่ครูจะต้องออกข้อสอบวิชานั้น ๆ ทั้งฉบับเพียงผู้เดียว อย่างไรก็ตามไม่ว่าจะเป็นการออกข้อสอบในลักษณะของกลุ่มคณะกรรมการหรือครูผู้สอนเพียงคนเดียว สิ่งที่ควรปฏิบัติในการวางแผนการสร้างข้อสอบมีดังนี้

1) กำหนดจุดมุ่งหมาย ในการสร้างข้อสอบทุกครั้งจะต้องกำหนดจุดมุ่งหมายให้ชัดเจนและแน่นอนว่าเพื่อวัตถุประสงค์ใด ซึ่งมีจุดมุ่งหมายของการสอบในโรงเรียนมีหลายประการ เช่น

- เพื่อวัดความเจริญงอกงามในการเรียนรู้ของนักเรียนในช่วงเวลาต่าง ๆ เช่น เมื่อจบหน่วยการเรียนหนึ่ง ๆ เมื่อเรียนไปได้กลางภาคหรือเมื่อสิ้นสุดการเรียนในแต่ละภาค เป็นต้น
- เพื่อหาข้อบกพร่องในการเรียนคือการสอบเพื่อวินิจฉัย
- เพื่อนำผลมาใช้ในการลำดับความสามารถของนักเรียน
- เพื่อนำผลมาใช้ในการประเมินผลของหลักสูตร
- เพื่อการเลื่อนชั้นของผู้เรียน

ฯลฯ

นอกจากนี้ในการสอบทั่ว ๆ ไป เช่น การสอบเพื่อคัดเลือกคนเข้าทำงานหรือการสอบคัดเลือกนักเรียนเข้าเรียนในแต่ละระดับการศึกษา ก็ต้องกำหนดจุดมุ่งหมายให้ชัดเจนแน่นอนเช่นเดียวกัน การทราบจุดมุ่งหมายในการสอบวัดจะทำให้ทราบถึงลักษณะของข้อสอบที่จะนำไปใช้ เช่น

- ในกรณีที่ต้องการตรวจสอบความรู้เดิมหรือความรู้พื้นฐานของผู้เรียน ข้อสอบที่สร้างก็ควรออกเฉพาะเนื้อหาที่เป็นพื้นฐานจำเป็นของวิชานั้นเท่านั้นและไม่ควรเป็นข้อสอบที่ยากเกินไป เช่น อาจมีความยากอยู่ประมาณ 60 –70 % เป็นต้น
- ในกรณีที่ต้องการสอบเพื่อวัดความเจริญงอกงามของผู้เรียนในช่วงเวลาต่าง ๆ ข้อสอบที่ใช้ก็ควรเป็นข้อสอบย่อย ซึ่งวัดเฉพาะเนื้อหาตามที่ได้เรียนไปแล้ว เช่น เป็นข้อสอบเฉพาะหน่วยการเรียนหนึ่ง ๆ เป็นข้อสอบกลางภาค หรือ ปลายภาค เป็นต้น ซึ่งความยากง่ายของข้อสอบของเนื้อหาแต่ละเรื่อง สำหรับข้อสอบปลายภาคอาจมีความยากง่ายประมาณ 35 –70 % หรืออาจมีบางข้อที่ง่ายกว่านี้บ้าง และยากกว่านี้บ้างรวมเข้าไปด้วยก็ได้
- ในกรณีที่ต้องการค้นหาข้อบกพร่องของผู้เรียน ข้อสอบที่สร้างจะออกตามเนื้อหาเป็นตอน ๆ ซึ่งแต่ละตอนจะมีจำนวนข้อคำถามมากข้อเพื่อวัดความละเอียดและควรเป็นข้อสอบที่ค่อนข้างง่ายในกรณีที่เป็นข้อสอบคัดเลือก ก็ควรเป็นข้อสอบที่ค่อนข้างยาก เป็นต้น

2) กำหนดเนื้อหาและพฤติกรรมที่ต้องการวัด ขั้นนี้เป็นการกำหนดรายละเอียดของหัวข้อเนื้อหาและพฤติกรรมที่ต้องการวัด ซึ่งเป็นเรื่องที่มีความสำคัญมาก ถ้าสามารถกำหนดขอบข่ายของเนื้อหาและพฤติกรรมที่จะออกข้อสอบได้เหมาะสม ก็จะช่วยให้ได้ข้อสอบที่มีความเที่ยงตรง การที่จะสามารถกำหนดขอบเขตของเนื้อหาและพฤติกรรมที่ต้องการวัดได้เป็นอย่างดี ต้องอาศัยการศึกษาหลักสูตรและการสร้างตารางวิเคราะห์หลักสูตร

3) กำหนดชนิดและรูปแบบของข้อสอบ เนื่องจากในการสอบวัดเพื่อประเมินผลผู้เรียนนั้น สามารถใช้ข้อสอบได้หลายชนิดและหลายรูปแบบ ซึ่งข้อสอบแต่ละชนิดและแต่ละรูปแบบก็เหมาะที่จะใช้กับพฤติกรรมที่แตกต่างกัน ในการสอบวัดจึงต้องระมัดระวังในการเลือกใช้ชนิดและรูปแบบของข้อสอบให้เหมาะสม

4) กำหนดส่วนประกอบอื่น ๆ ที่จำเป็นในออกข้อสอบและในการสอบ ได้แก่ การกำหนดเวลาในการสร้างข้อสอบ บุคลากรในการสร้างข้อสอบ จำนวนข้อของข้อสอบ เวลาในการทำสอบ วิธีการตอบ กระดาษคำตอบ วิธีการตรวจให้คะแนน วิธีการตรวจสอบคุณภาพของข้อสอบ และวิธีการรายงานผลการสอบ เป็นต้น

2. ขั้นเตรียมงาน

ขั้นนี้เป็นการเตรียมสิ่งต่าง ๆ ที่จะต้องใช้และเอื้ออำนวยต่อการสร้างข้อสอบ ได้แก่

1) หลักสูตร หนังสือแบบเรียน แผนการสอน คู่มือครู

2) ทำการวิเคราะห์หลักสูตรรายวิชาจะออกข้อสอบ พร้อมกับบันทึกผลในตารางวิเคราะห์หลักสูตร

3) อุปกรณ์ต่าง ๆ สำหรับการพิมพ์ การอัดสำเนา

4) กระดาษคำตอบและอื่น ๆ

3. ขั้นลงมือปฏิบัติ

ขั้นนี้เป็นขั้นลงมือเขียนข้อสอบ ในกรณีที่การสร้างข้อสอบนั้นทำในรูปคณะกรรมการ คณะกรรมการก็จะตกลงแบ่งงานกันไปเขียนข้อสอบ โดยกำหนดเนื้อหาและจำนวนข้อให้แต่ละคนรับผิดชอบไป แล้วนัดหมายเวลาเพื่อประชุมวิจารณ์ข้อสอบที่สร้างขึ้น

ในขั้นนี้มีสิ่งที่ควรยึดถือเป็นหลักปฏิบัติ คือ

1) กรรมการทุกคนควรจะทราบถึงคุณลักษณะของข้อสอบที่ดี คุณลักษณะที่จำเป็นของผู้เขียนข้อสอบ และเทคนิคการเขียนข้อสอบชนิดต่าง ๆ เทคนิคการเขียนข้อสอบวัดพฤติกรรมด้านความรู้และความคิด

2) การออกข้อสอบควรยึดผลวิเคราะห์ในตารางวิเคราะห์เป็นหลัก

3) หากมีปัญหาด้านเนื้อหา หรือด้านเทคนิควิธีการออกข้อสอบก็ควรประชุมตกลงกัน

<u>หมายเหตุ</u> ในกรณีที่การสร้างข้อสอบ ครูผู้สอนประจำวิชาเป็นผู้ออกแต่เพียงคนเดียว ก็ควรจะปรึกษาผู้รู้และผู้เชี่ยวชาญในการออกข้อสอบหรือนักวัดผล

4. ขั้นประเมินหรือตรวจสอบคุณภาพ

ขั้นนี้มีวัตถุประสงค์เพื่อนำผลไปปรับปรุงข้อสอบ ขั้นนี้แยกเป็นขั้นย่อย ๆ ดังนี้

4.1 ขั้นประเมินเบื้องต้น ขั้นนี้ก็คือการวิจารณ์ข้อสอบนั่นเอง โดยผู้แทนกลุ่มวิชาซึ่งอาจได้แก่ ประธานหรือเลขากลุ่มหรือเลขากลุ่มทำการรวบรวมข้อสอบที่กรรมการแต่ละคนสร้างมาถ่ายเอกสาร พร้อมกับแจกจ่ายให้กรรมการแต่ละคนได้ใช้ในการวิจารณ์ การวิจารณ์ข้อสอบเป็นสิ่งดีและสิ่งจำเป็น โดยเฉพาะอย่างยิ่งสำหรับข้อสอบแบบปรนัยชนิดเลือกตอบ เป็นการร่วมกันพิจารณาปรับปรุง ข้อคำถามและตัวเลือกให้ดีขึ้น การวิจารณ์ข้อสอบเป็นการพิจารณาในเรื่องต่อไปนี้

1) ข้อคำถามวัดในสิ่งที่ต้องการวัดหรือไม่
2) ข้อคำถามชัดเจนเข้าใจตรงกันหรือไม่
3) ข้อคำถามมีคำตอบถูกที่แน่นอนเพียงคำตอบเดียวหรือไม่
4) ข้อคำถามใช้ภาษารัดกุม เหมาะสมกับระดับชั้นของนักเรียนหรือไม่
5) ในกรณีที่เป็นข้อสอบเลือกตอบ พิจารณาสิ่งต่อไปนี้ด้วย

- ตัวลวงเหมาะสมหรือไม่
- การเรียงข้อคำถามเรียงถูกต้องตามตามหลักหรือไม่ เช่น เรียงตามลำดับเนื้อหา เรียงจากง่ายไปหายาก เป็นต้น
- การเรียงตัวเลือกในแต่ละข้อเรียงเหมาะสมสวยงามหรือไม่

ข้อสอบที่ผ่านการวิจารณ์แล้ว ผู้ออกข้อสอบจะต้องนำไปปรับปรุงแก้ไขตามที่ได้รับคำแนะนำ ทั้งนี้เพื่อให้ได้ข้อสอบที่มีคุณภาพในเบื้องต้นก่อนจะนำไปทดลองจริง

4.2 ขั้นตรวจสอบคุณภาพหลังการทดลอง

ข้อสอบที่ได้ผ่านการวิจารณ์และได้ปรับปรุงแก้ไขจนเป็นที่เรียบร้อยแล้ว ก็จะนำไปจัดพิมพ์เพื่อนำไปทดลอง (Try out) และเมื่อนำไปทดลอง แล้วก็จะต้องนำมาตรวจคะแนน และทำการตรวจสอบคุณภาพอีก การตรวจสอบคุณภาพในขั้นนี้ เป็นการพิจารณาในเรื่องต่อไปนี้

1) ดูความยากง่ายของข้อสอบ (Difficulty)

2) ดูอำนาจจำแนกของข้อสอบ (Discrimination)

ซึ่งเป็นการตรวจสอบเป็นรายข้อและรายตัวเลือก วิธีการที่ใช้ในการตรวจสอบก็คือการวิเคราะห์ข้อสอบ (Item analysis) นั่นเอง

ผลการตรวจสอบคุณภาพสองประการข้างต้นนี้ จะทำให้ทราบว่าข้อสอบข้อใดยากเกินไปหรือง่ายเกินไป ข้อสอบข้อใดจำแนกเด็กเก่งกับเด็กอ่อนออกจากกันได้หรือไม่ และทำให้ทราบว่าข้อสอบข้อใดตัวเลือกดีหรือไม่ดีด้วย ซึ่งจะช่วยให้การปรับปรุงข้อสอบกระทำได้ตรงจุด

3) หาค่าสถิติพื้นฐานของข้อสอบ ได้แก่ ค่าเฉลี่ยและความแปรปรวน ค่าคะแนนเฉลี่ยจะชี้บอกให้ทราบว่า ข้อสอบนั้นโดยส่วนรวมทั้งฉบับมีความยากง่ายในระดับใด ส่วนค่าความแปรปรวนก็จะทำให้ทราบว่ากลุ่มที่ใช้ในการทดลองข้อสอบนั้นมีความสามารถแตกต่างกันหรือกระจายกันมากน้องเพียงใด ซึ่งจะเป็นประโยชน์ในการปรับปรุงข้อสอบด้วยประการหนึ่ง

<u>ข้อสังเกต</u> ในการตรวจสอบคุณภาพหลังการทดลอง อาจต้องมีการตรวจสอบมากกว่า 1 ครั้ง กล่าวคือ เมื่อนำไปทดลองครั้งแรก แล้วนำผลมาปรับปรุงแก้ไขก็จะต้องนำไปทดลองครั้งที่สอง แล้วทำการตรวจสอบอีก พิจารณาผลวิเคราะห์ที่ได้ ถ้ายังได้ผล ไม่เป็นไปตามเกณฑ์ที่กำหนดไว้ ก็จะต้องปรับปรุงแก้ไขแล้วนำไปทดลองอีก จนกว่าจะได้ข้อสอบที่มีผลวิเคราะห์รายข้อและรายตัวเลือกตามเกณฑ์ที่กำหนดไว้

<u>ข้อเสนอแนะสำหรับการทดลองข้อสอบ</u>

(1) กลุ่มตัวอย่างที่ใช้ในการทดลองข้อสอบ ควรประกอบด้วยนักเรียนที่ เก่ง อ่อน และปานกลางคละกันไป เพราะถ้านำไปทดลองกับกลุ่มตัวอย่างกลุ่มหนึ่งกลุ่มใดโดยเฉพาะ จะทำให้ไม่ทราบคุณภาพที่แท้จริงของข้อสอบ เช่น ถ้าสอบกับนักเรียนที่เรียนเก่งกว่าอย่างเดียว ข้อสอบนั้นก็จะง่าย ถ้าสอบกับนักเรียนที่เรียนอ่อนอย่างเดียว ข้อสอบนั้นก็จะยาก และในทั้ง 2 กรณี ข้อสอบนั้นจะไม่มีอำนาจจำแนก เพราะเด็กในแต่ละกลุ่มมีความสามารถพอ ๆ กัน เป็นต้น

(2) เวลาที่ใช้ในการทำสอบ ควรมีความเหมาะสมไม่มากเกินไปหรือน้อยเกินไป โดยทั่วไปข้อสอบที่ต้องการคิดคำนวณให้เวลาประมาณ 1 นาที ถ้าเป็นข้อสอบวิชาอื่น ๆ ก็ให้เวลาทำน้อยกว่าจำนวนเล็กน้อย เช่น สอบวิชาภาษาไทยชั้น ม.3 มีจำนวน 30 ข้อ อาจให้เวลา 25 นาที เป็นต้น การกำหนดเวลาอาจพิจารณาจากการทดลองครั้งแรกโดยใช้เกณฑ์ให้ผู้สอบ 80 – 90 % ทำเสร็จเป็นการหมดเวลาก็ได้

(3) ในการดำเนินการสอบต้องพยายามจูงใจให้นักเรียนมีความตั้งใจในการทำข้อสอบ โดยชี้แจงให้นักเรียนเข้าใจอย่างชัดเจนว่าการสอบนั้น ๆ ไม่มีผลต่อการสอบได้ตกของ

นักเรียน แต่กลับจะเป็นผลดีแก่นักเรียนเพราะจะทำให้เกิดทักษะในการทำข้อสอบ ทั้งนี้เพื่อจูงใจให้นักเรียนทำอย่างเต็มที่ ไม่ใช่ทำอย่างขอไปที

4.3 ขั้นตรวจสอบขั้นสุดท้าย

ข้อสอบที่ได้ผ่านการปรับปรุงแก้ไขจนได้ค่าคุณภาพของข้อสอบในเรื่องของความยากง่ายและอำนาจจำแนกตามเกณฑ์ที่กำหนดไว้เป็นที่เรียบร้อยแล้ว จะต้องนำข้อสอบนั้นไปสอบกับคนกลุ่มใหญ่ (ประมาณ 1,000 คนขึ้นไป) ที่ครอบคลุมความสามารถทั้งเก่ง อ่อน และปานกลางกระจายกันแล้วทำการประเมินคุณภาพครั้งสุดท้าย และสร้างเกณฑ์ปกติ

การประเมินคุณภาพขั้นสุดท้ายได้แก่ การหาค่าต่อไปนี้

- ความเชื่อมั่นและความคลาดเคลื่อนมาตรฐานในการวัดของแบบทดสอบทั้งฉบับ
- ค่าสถิติพื้นฐานทั้งฉบับ

<u>หมายเหตุ</u> ในการตรวจสอบคุณภาพของข้อสอบหลังการทดลอง (Try out) แต่ละครั้ง ควรหาความเชื่อมั่นและความคลาดเคลื่อนมาตรฐานในการวัดไว้ด้วยทุกครั้ง เพราะทำให้สามารถเปรียบเทียบผลของแต่ละครั้งได้

5. ขั้นจัดพิมพ์ การจัดพิมพ์ข้อสอบเป็นงานสำคัญอีกขั้นตอนหนึ่งกระทำภายหลังการประเมินคุณภาพของข้อสอบ แบ่งเป็นขั้นย่อย ๆ ดังนี้

5.1 การจัดพิมพ์ขั้นต้น กระทำภายหลังการวิจารณ์ข้อสอบและปรับปรุงแก้ไขในขั้นต้นเพื่อที่จะนำไปทดลองใช้ (Try Out) ครั้งแรก ซึ่งในการเตรียมต้นฉบับ เพื่อจัดพิมพ์นี้งานที่ต้องกระทำคือ

- การเรียงข้อสอบตามลำดับเนื้อหา หรือเรียงจากง่ายไปยาก
- การวางรูปแบบข้อสอบให้พอเหมาะกับหน้ากระดาษ
- การเขียนคำชี้แจงในการดำเนินการสอบ รวมถึงการระบุจำนวนข้อสอบ เวลาในการทำสอบ และการกรอกหัวกระดาษคำตอบด้วย

นอกจากนี้ในการจัดพิมพ์ข้อสอบ ยังต้องคำนึงสิ่งต่อไปนี้

- ความชัดเจนในการพิมพ์ และความสะอาดเรียบร้อย
- การพิสูจน์อักษรต้องกระทำอย่างรอบคอบอย่างไม่ให้มีที่ผิด
- การอัดสำเนาต้องชัดเจน และสะอาด

5.2 การจัดพิมพ์ภายหลังการทดลอง กระทำภายหลังการนำข้อสอบไปทดลองในแต่ละครั้ง หลังจากนำข้อสอบไปทดลอง แล้วทำการวิเคราะห์รายข้อ พร้อมกับนำผลมาพิจารณาปรับปรุงข้อสอบ เมื่อปรับปรุงจนเป็นที่เรียบร้อยแล้ว ก็จะต้องนำข้อสอบนั้นมาทำเป็นต้นฉบับเพื่อจัดพิมพ์ใหม่ ซึ่งการจัดพิมพ์ในขั้นนี้ ก็จะต้องพิจารณาถึงเรื่องต่อไปนี้

- จัดเรียงข้อสอบใหม่ โดยเรียงตามลำดับเนื้อหา และในแต่ละเนื้อหาเรียงตามผลการวิเคราะห์จากข้อง่ายไปหาข้อยาก (สำหรับข้อสอบปรนัยชนิดเลือกตอบ)
- เรียงลำดับตัวเลือกในแต่ละข้อคำถามโดยเรียงจากข้อความสั้นไปหาข้อความยาว เพื่อให้แลดูสวยงามเป็นระเบียบ (สำหรับข้อสอบปรนัยชนิดเลือกตอบ)
- ตรวจสอบสำนวนภาษาให้เหมาะสม
- ตรวจสอบความชัดเจนในการพิมพ์ และความสะอาดเรียบร้อย
- ทำการพิสูจน์อักษรอย่างรอบคอบ ให้มีความถูกต้อง
- การอัดสำเนาต้องชัดเจนและสะอาด

5.3 การจัดพิมพ์แบบทดสอบฉบับสมบูรณ์ หลังจากนำข้อสอบไปทดลองและทำการปรับปรุงแก้ไขจนได้ข้อสอบมีคุณภาพตามเกณฑ์ที่กำหนดแล้ว ก็จะต้องทำต้นฉบับเมื่อจัดพิมพ์แบบทดสอบฉบับสมบูรณ์ ซึ่งเป็นการพิมพ์ครั้งสุดท้าย การจัดพิมพ์ในขั้นนี้ควรปฏิบัติเพิ่มเติมจากสองขั้นแรกดังนี้

- ตรวจสอบจำนวนข้อสอบให้มีครบถ้วนตามที่กำหนด
- สำรวจการเรียงข้อสอบอีกครั้งหนึ่งว่าการเรียงถูกต้องตามหลักเกณฑ์และเหมาะสมสวยงามหรือไม่ โดยทั่วไปเรียงจากข้อความสั้นไปหาข้อความยาว (ตัวเลือก) และในกรณีเดียวกันก็พิจารณาตำแหน่งของตัวถูกด้วยโดยให้มีตัวถูกอยู่ในตำแหน่ง ก ข ค ง และ จ จำนวนพอ ๆ กัน (ถ้าเป็น 5 ตัวเลือก) สำหรับปรนัยชนิดเลือกตอบ
- ตรวจสอบการจัดวางข้อความให้เหมาะกับหน้ากระดาษพิมพ์ เพื่อให้แลดูสวยงามเป็นระเบียบ
- ตรวจสอบการจัดทำคำชี้แจงวิธีทำข้อสอบอีกครั้งหนึ่งให้มีหัวข้อต่าง ๆ ครบถ้วน ได้แก่ ชื่อวิชาที่สอบ วิธีการตอบ วิธีคิดข้อสอบ ซึ่งควรจะมีตัวอย่างประกอบด้วย การกรอกหัวกระดาษคำตอบ จำนวนข้อและเวลาที่ให้ทำ

6. การจัดทำคู่มือการใช้

กระบวนการสร้างแบบทดสอบผลสัมฤทธิ์ นอกจากจะประกอบด้วยขั้นตอน 5 ขั้นดังกล่าวข้างต้นแล้ว ยังมีอีกขั้นตอนหนึ่งที่มีความสำคัญ และเป็นขั้นตอนที่จะทำให้การสร้างแบบทดสอบ

ผลสัมฤทธิ์มีความสมบูรณ์ยิ่งขึ้น โดยเฉพาะอย่างยิ่งสำหรับการสร้างแบบทดสอบมาตรฐานถือว่าเป็นความจำเป็นอย่างยิ่ง ขั้นตอนนี้ก็คือ “ การจัดทำคู่มือการใช้ ”

การเขียนคู่มือการใช้ (Test Manual) ประกอบด้วยการให้รายละเอียดของหัวข้อสำคัญต่อไปนี้

1) วัตถุประสงค์ของแบบทดสอบ
2) วิธีการใช้และวิธีดำเนินการสอบ
3) วิธีการตรวจให้คะแนน
4) การแปลความหมายคะแนน
5) เกณฑ์ปกติ

ทฤษฎีการตอบสนองข้อสอบ

1. หลักการของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการทดสอบแบบดั้งเดิม (CTT) เป็นที่รู้จักกันเป็นอย่างดีในวงการวัดผลทางการศึกษาและจิตวิทยามาเป็นเวลานาน แนวคิดจากทฤษฎีการทดสอบแบบดั้งเดิมถูกนำมาใช้อย่างกว้างขวางในการสร้าง วิเคราะห์ และตรวจสอบคุณภาพสำหรับการพัฒนาแบบทดสอบต่าง ๆ เช่น แบบทดสอบผลสัมฤทธิ์ แบบวัดความถนัด แบบวัดบุคลิกภาพ เป็นต้น ด้วยข้อจำกัดที่สำคัญของ CTT ได้แก่ การที่พารามิเตอร์ของข้อสอบผันแปรตามกลุ่มผู้สอบ และคะแนนที่สังเกตได้หรือค่าประมาณความสามารถของผู้ตอบไม่เป็นอิสระหรือขึ้นอยู่กับข้อสอบและแบบทดสอบที่นำมาใช้ นักทฤษฎีการทดสอบหลายท่านจึงได้พัฒนาระบบการวัดแนวใหม่ ทฤษฎีการทดสอบสำหรับการวัดแนวใหม่อันเป็นที่ยอมรับกันอย่างกว้างขวางในปัจจุบัน คือ ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) หรือเรียกโดยใช้ชื่อย่อว่า IRT

ทฤษฎีการตอบสนองข้อสอบ พยายามอธิบายความสัมพันธ์ระหว่างคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคล กับพฤติกรรมการตอบสนองข้อสอบของบุคคลนั้นว่ามีโอกาสตอบข้อสอบถูกมากน้อยเพียงไร ทฤษฎีมีพื้นฐานความเชื่อว่า พฤติกรรมการตอบสนองข้อสอบของผู้สอบซึ่งเป็นสิ่งที่สังเกตได้โดยตรงว่าถูกหรือผิด จะถูกกำหนดโดยลักษณะภายในหรือความสามารถที่อยู่ภายในตัวบุคคล ซึ่งเป็นสิ่งที่ไม่สามารถสังเกตได้โดยตรง ทฤษฎีนี้ได้อธิบายความสัมพันธ์ดังกล่าว

ในรูปของฟังก์ชันคณิตศาสตร์ หรือโมเดลที่แสดงความสัมพันธ์ระหว่างระดับความสามารถ คุณลักษณะของข้อสอบ และโอกาสการตอบข้อสอบได้ถูก

แนวคิดพื้นฐานของ IRT

ทฤษฎีการตอบสนองข้อสอบ เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างความสามารถที่มีอยู่ภายในบุคคลกับผลการตอบข้อสอบหรือข้อคำถามโดยใช้โค้งลักษณะข้อสอบ ซึ่งมีการกำหนดลักษณะของข้อสอบด้วยพารามิเตอร์ความยาก (b) อำนาจจำแนก (a) และโอกาสการเดาข้อสอบถูก (c) IRT จึงอยู่บนฐานความคิดที่สำคัญ 2 ประการ คือ

1) ผลการตอบข้อสอบหรือข้อคำถามของผู้ตอบ สามารถอธิบายได้ด้วยความสามารถที่มีอยู่ในตัวของผู้ตอบ

2) ความสัมพันธ์ระหว่างผลการตอบข้อสอบกับความสามารถที่มีอยู่ภายใน สามารถอธิบายได้ด้วยฟังก์ชันลักษณะข้อสอบ หรือโค้งลักษณะข้อสอบ

2. ข้อตกลงของทฤษฎีการตอบสนองข้อสอบ

ฟังก์ชันทางคณิตศาสตร์ที่ใช้ใน IRT กำหนดไว้ว่า ความน่าจะเป็นของการตอบข้อสอบต้องขึ้นอยู่กับความสามารถของผู้สอบและขึ้นอยู่กับคุณลักษณะของข้อสอบ โมเดลการตอบสนองข้อสอบจึงตั้งอยู่บนฐานความเชื่อหรือข้อตกลงเบื้องต้นหลายประการเกี่ยวกับลักษณะข้อมูลที่จะทำให้โมเดลสามารถนำไปใช้ได้อย่างเหมาะสม ถึงแม้ข้อตกลงเบื้องต้นบางประการจะตรวจสอบไม่ได้โดยตรง แต่เราสามารถเก็บรวบรวมข้อมูลหรือหลักฐานทางอ้อมมาช่วยสนับสนุนยืนยันได้ ข้อตกลงเบื้องต้นที่สำคัญมีดังนี้

2.1 ความเป็นเอกมิติ (Unidimensionality)

ข้อตกลงเบื้องต้นที่ใช้กันทั่วไปสำหรับ IRT คือ ข้อคำถาม/ข้อสอบทุกข้อในเครื่องมือ/แบบสอบมุ่งวัดเพียงคุณลักษณะเดียว หรือความสามารถเดียว (One abiltity) ซึ่งเรียกว่าความเป็นเอกมิติ (Unidimensionality)

2.2 ความเป็นอิสระ (Local Independence)

แนวคิดเกี่ยวกับ “ ความเป็นอิสระระหว่างข้อสอบและผู้สอบ ” มีความเกี่ยวข้องและโยงมาจาก “ความเป็นเอกมิติของแบบสอบ” ความเป็นอิสระของข้อสอบและผู้สอบ หมายถึง เมื่อมีการ

ควบคุมความสามารถ (θ) ที่ส่งผลต่อการตอบข้อสอบ หรือให้ θ คงที่แล้ว ผลการตอบข้อสอบแต่ละข้อจะต้องเป็นอิสระจากกัน หรือกล่าวอีกนัยหนึ่งได้ว่า เมื่อควบคุมอิทธิพลของ θ แล้วผลการตอบข้อสอบรายข้อ ไม่มีความสัมพันธ์กัน นั่นคือ โมเดลการตอบสนองมีเพียง θ ปัจจัยเดียวเท่านั้นที่มีผลต่อการตอบรายข้อ ความเป็นอิสระสามารถจำแนกการพิจารณาเป็นความอิสระระหว่างข้อสอบและความอิสระระหว่างผู้สอบดังต่อไปนี้

1) ความเป็นอิสระระหว่างข้อสอบ

เมื่อสุ่มผู้สอบ ซึ่งมีความสามารถ θ ขึ้นมา 1 คน ในการตอบข้อสอบ k ข้อ ให้ U_j เป็นผลการตอบหรือคะแนนข้อที่ j หลังจากควบคุม θ ของผู้สอบแล้ว คะแนนผลการตอบของผู้สอบนั้นในแต่ละข้อไม่สัมพันธ์กัน

ถ้าผลการตอบรายข้อของผู้สอบคนเดียวกันเป็นอิสระจากกัน ความน่าจะเป็นของแบบแผนการตอบข้อที่ k ของผู้สอบที่มีความสามารถ θ จะเท่ากับผลคูณระหว่าง ความน่าจะเป็นของผลการตอบข้อสอบแต่ละข้อ

2) ความเป็นอิสระระหว่างผู้สอบ

เมื่อสุ่มข้อสอบขึ้นมา 1 ข้อ ในการตอบข้อสอบของผู้สอบ n คน ให้ U_i เป็นผลการตอบหรือคะแนนข้อสอบของผู้สอบคนที่ i หลังจากควบคุม θ ของผู้สอบแต่ละคนแล้วคะแนนผลการตอบข้อนั้นของผู้สอบแต่ละคนไม่สัมพันธ์กัน

ถ้าผลการตอบข้อสอบข้อเดียวกันของผู้สอบแต่ละคนเป็นอิสระจากกัน ความน่าจะเป็นของแบบแผนการตอบข้อสอบของผู้สอบ n คน จะเท่ากับ ผลคูณระหว่างความน่าจะเป็นของผลการตอบข้อนั้นของผู้สอบแต่ละคน

2.3 โมเดลการตอบสนองข้อสอบ (Item Response Model)

IRT อยู่บนฐานความเชื่อว่า ฟังก์ชันลักษณะข้อสอบ หรือโค้งลักษณะข้อสอบ (ICC) สามารถสะท้อนความสัมพันธ์จริงระหว่างความสามารถของผู้สอบกับลักษณะของข้อสอบและผลการตอบข้อสอบ โมเดลการตอบสนองเสนอ ICC ซึ่งเป็นฟังก์ชันโลจิส ด้วยรูปลักษณ์ที่แตกต่างกัน ตามจำนวนพารามิเตอร์ที่ใช้บรรยายลักษณะของข้อสอบ โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนน 2 ค่า (Dichotomous Item Response Models) ที่ใช้กันแพร่หลาย ได้แก่ โมเดลการตอบสนองข้อสอบ 1-พารามิเตอร์ , 2-พารามิเตอร์และ3-พารามิเตอร์ การเลือกใช้จึงขึ้นอยู่กับจุดมุ่งหมายของงานและธรรมชาติของข้อมูล

โมเดลการตอบสนองข้อสอบแบบ 1-พารามิเตอร์ มีข้อตกลงเบื้องต้นว่า ข้อสอบแต่ละข้อมีพารามิเตอร์ c = 0 และพารามิเตอร์ a เท่ากัน แต่มีความแตกต่างกันเฉพาะ b เท่านั้น โมเดลนี้จึงเหมาะสำหรับใช้กับข้อสอบอิงเกณฑ์ที่ไม่สลับซับซ้อน ข้อสอบที่ค่อนข้างเรียบง่ายสำหรับพัฒนาคลังข้อสอบที่มีความเป็นเอกพันธ์

โมเดลการตอบสนองข้อสอบแบบ 2-พารามิเตอร์ มีข้อตกลงเบื้องต้นว่า ข้อสอบแต่ละข้อมีพารามิเตอร์ c = 0 มีความแตกต่างกัน a และ b เท่านั้น โมเดลนี้จึงเหมาะสำหรับใช้กับข้อสอบที่ต้องเติมคำตอบ หรือข้อสอบแบบเลือกตอบที่ไม่ยากมากนักและกลุ่มผู้สอบมีความพร้อมในการตอบ

โมเดลการตอบสนองข้อสอบแบบ 3-พารามิเตอร์ มีข้อตกลงเบื้องต้นว่า ข้อสอบแต่ละข้อมีความแตกต่างกันได้ทั้ง พารามิเตอร์ a , b และ c เท่านั้น โมเดลนี้จึงเหมาะสำหรับใช้กับข้อสอบแบบเลือกตอบทั่วไป ข้อสอบแบบหลายตัวเลือก เนื่องจากผู้สอบสามารถเดาคำตอบได้

2.4 การสอบที่ไม่แข่งขันด้านเวลา (Nonspeeded Test Administration)

IRT ถือว่าความสามารถ (θ) เป็นปัจจัยสำคัญต่อผลการสอบ ความเร็วในการตอบจะต้องไม่มีอิทธิพลต่อผลการตอบ การจัดการสอบจึงต้องไม่อยู่ในสถานการณ์ที่สอบแข่งขันกันด้วยเวลา การสอบจะต้องอยู่ในลักษณะที่ผู้สอบซึ่งมีความสามารถมีเวลาเพียงพอในการทำข้อสอบ

3. พารามิเตอร์ของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบประกอบด้วยค่าพารามิเตอร์และค่าคงที่ ดังนี้

1. พารามิเตอร์ของผู้สอบ

θ = ระดับความสามารถของผู้สอบ ซึ่งประมาณได้จากโมเดลตามทฤษฎีการตอบสนองข้อสอบ

- นิยมปรับให้เป็นคะแนนมาตรฐานที่มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1
- ค่า θ มีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ผลการวิเคราะห์ส่วนใหญ่มักให้ค่าอยู่ในช่วง –3 ถึง +3

โดย

$P_i(\theta)$ = ความน่าจะเป็นที่ผู้ตอบซึ่งมีความสามารถ θ จะตอบข้อสอบ i ได้ถูกต้อง

2. พารามิเตอร์ของข้อสอบ

b_i = ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i (Difficulty parameter)

- ตำแหน่งของโค้งบนสเกลของความสามารถ (θ) ที่ทำให้มีโอกาสตอบข้อสอบได้ถูกต้องเท่ากับ $\frac{1+c_i}{2}$
- สำหรับโมเดล 1- พารามิเตอร์ และ 2-พารามิเตอร์ $P_i(\theta) = 0.50$ ส่วนโมเดล 3-พารามิเตอร์ $P_i(\theta) = \frac{1+c_i}{2}$
- ในทางทฤษฎีมีค่าระหว่าง $(-\infty, +\infty)$ แต่ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่าระหว่าง –2.50 ถึง +2.50 ค่า b_i ที่อยู่ใกล้ –2.50 แสดงว่าเป็นข้อสอบที่ง่าย ส่วนค่า b_i ที่อยู่ใกล้ +2.50 แสดงว่าเป็นข้อสอบที่ยาก

a_i = ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i (Discrimination parameter)

- การจำแนกค่าความต่างของ $P_i(\theta)$ ระหว่างผู้สอบที่มีความสามารถ $\leq \theta$ กับ $> \theta$
- มีค่าเป็นสัดส่วนโดยตรงของค่าความชันของ ICC ที่ตำแหน่ง b_i ค่า a_i ที่สูงแสดงถึงการจำแนกผู้สอบที่มีความสามารถที่มีความสามารถแตกต่างกันได้ดี
- ในทางทฤษฎีมีค่าระหว่าง $(-\infty, +\infty)$ ควรมีค่าเป็น + ตามปกติมีค่าไม่เกิน +2.50 ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่า a_i อยู่ระหว่าง +0.50 ถึง +2.50

c_i = ค่าพารามิเตอร์โอกาสในการเดาข้อสอบได้ถูก (Guessing parameter)

- โอกาสในการตอบถูกของผู้สอบที่มีความสามารถต่ำ เป็นค่ากำกับต่ำสุด (lower assymtote) ของ ICC
- ในทางทฤษฎี มีค่าอยู่ระหว่าง 0 ถึง 1 โดยทั่วไปนิยมใช้ข้อสอบที่มีค่า c_i ไม่เกิน 0.30

- ตามปกติควรมีค่าต่ำกว่าโอกาสในการตอบถูกโดยการเดาตามทฤษฎี CTT

3. ค่าคงที่

e = 2.71828

- ค่าคงที่ของลอการิทึมธรรมชาติ (Natural log)

D = 1.70

- ค่าองค์ประกอบของการปรับเสกล (Scaling factor)เป็นค่าการปรับสเกลเพื่อทำให้ Logistic function กับ Normal ogive function ใกล้เคียงกันหรือมีค่าประมาณ θ ต่างกันไม่เกิน 0.01

4. รูปแบบของโมเดลโลจิสติก

จากแนวคิดตามทฤษฎีการตอบสนองข้อสอบได้มีการพัฒนารูปแบบขึ้นหลายรูปแบบซึ่งสามารถจำแนกได้ดังนี้

1. โมเดลการตอบสนองข้อสอบที่ใช้กับการตรวจให้คะแนนรายข้อแบบ 2 ค่า (Binary Item) เป็นข้อสอบที่ตรวจให้คะแนนแบบ 0 , 1 (ตอบผิดได้ 0 คะแนน ตอบถูกได้ 1 คะแนน) โมเดลประเภทนี้ถูกพัฒนาขึ้นในระยะเริ่มแรก เช่น กัตแมนสัมบูรณ์ (Guttman Perfect Scale) โมเดลระยะห่างแฝง (Latent Distance Model) โมเดลเชิงเส้นตรง (Linear Model) เป็นต้น ในระยะต่อมาได้มีการพัฒนาโมเดลประเภทนี้เพิ่มเข้ามา เช่น โมเดลปกติสะสม (Normal Ogive Model) แบบ 1, 2 , 3 พารามิเตอร์ และโมเดลโลจิสติก (Logistic Model) แบบ 1 , 2 , 3 , 4 พารามิเตอร์ เป็นต้น ซึ่งเป็นแบบจำลองที่ได้รับการปรับปรุงให้มีความสอดคล้องกับข้อมูลจากการสอบในสถานการณ์จริงมากขึ้น สามารถนำไปปฏิบัติได้ง่ายและมีผู้สนใจนำไปใช้อย่างกว้างขวาง

2. โมเดลการตอบสนองข้อสอบที่ใช้กับการตรวจให้คะแนนรายข้อแบบมากกว่า 2 ค่า (Multichotomous) เช่น Normal Response Model , Grade Response Model และ Partial Credit Model ซึ่งรูปแบบเหล่านี้มีจุดมุ่งหมายเพื่อเพิ่มความแม่นยำในการประมาณค่าความสามารถของผู้เข้าสอบโดยใช้ประโยชน์จากสารสนเทศทั้งจากการตอบข้อสอบถูกหรือผิด

3. โมเดลการตอบสนองข้อสอบที่ใช้กับการตรวจให้คะแนนรายข้อแบบต่อเนื่อง (Continuous) ได้แก่ Continuous Response Model ซึ่งจะเป็นประโยชน์ต่อการประมาณค่าความสามารถของผู้เข้าสอบทางจิตวิทยา หรือผู้ที่สนใจศึกษาทางด้านที่เกี่ยวกับทัศนคติ (Attitude) ซึ่งการตอบสนองข้อสอบเป็นมาตราที่ต่อเนื่อง (Continuous Scale)

ซึ่งในการวิจัยครั้งนี้ ผู้วิจัยจะกล่าวเฉพาะรูปแบบโลจิสติก (Logistic Model) เท่านั้น

4.1 รูปแบบโลจิสติกที่มี 1 พารามิเตอร์ (One – Parameter Logistic Model)

เป็นโมเดลที่อธิบายข้อสอบด้วยค่าพารามิเตอร์เพียงตัวเดียว คือค่าความยาก (b) โดยเชื่อว่าโอกาสที่ผู้สอบจะทำข้อสอบได้ถูกหรือไม่นั้นขึ้นอยู่กับระดับความสามารถ (θ) ของตนเองกับระดับความยากของข้อสอบ ดังนั้นจึงถือว่าการเดา (c) เป็นศูนย์ และค่าอำนาจจำแนก (a) ข้อสอบจะคงที่ทั้งฉบับ เขียนเป็นฟังก์ชันได้ดังนี้ (Hambleton, Swaminathan and Rogers, 1991 : 12)

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

เมื่อ $P_i(\theta)$ แทน โอกาสที่ผู้ตอบที่มีระดับความสามารถ θ จะทำข้อสอบข้อที่ i ได้ถูกต้อง

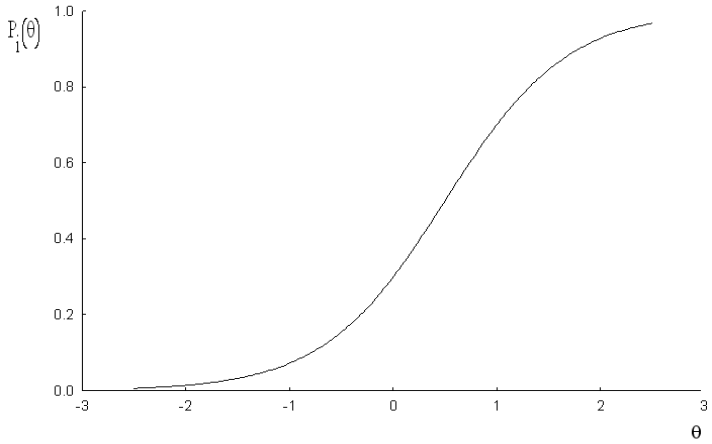
θ แทน ระดับความสามารถที่แท้จริงของผู้สอบ โดยมีพิสัยการกระจายอยู่ระหว่าง $-\infty$ ถึง $+\infty$ ซึ่งระดับความสามารถ –3 ถึง +3 ซึ่งระดับความสามารถ –3 หมายถึงผู้สอบมีระดับความสามารถต่ำมาก และความสามารถ +3 หมายถึงผู้สอบมีความสามารถสูงมาก

b_i แทน ความยากของข้อสอบข้อที่ i ค่าที่แสดงถึงระดับความสามารถของผู้สอบที่จุดโค้งคุณลักษณะข้อสอบที่มีความชันมากที่สุด ค่าความยากมีค่าตั้งแต่ $-\infty$ ถึง $+\infty$ โดยทั่วไปนิยมใช้ค่าที่อยู่ระหว่าง –2 ถึง +2 ซึ่งค่าที่เป็นลบแสดงว่าข้อสอบง่ายและค่าที่เป็นบวกแสดงว่าข้อสอบยาก

e แทน ค่าคงที่มีค่าเท่ากับ 2.718…

n แทน จำนวนข้อสอบในแบบสอบ

และมีโค้งลักษณะเฉพาะของข้อสอบเมื่อ ดังภาพประกอบ 1



ภาพประกอบ 1 โค้งลักษณะของข้อสอบที่มี 1 พารามิเตอร์ (b = 0.5)

4.2 รูปแบบโลจิสติกที่มี 2 พารามิเตอร์ (Two – Parameter Logistic Model)

โมเดลนี้เป็นรูปแบบที่เบอร์นบอม (Birnbaum, 1957 , 1958a , 1958b , 1968) ได้เสนอไว้แทนรูปแบบ Normal Ogive Model เนื่องมาจากรูปแบบโลจิสติก เป็นรูปแบบเชิงคณิตศาสตร์ ที่นำมาใช้ในทางปฏิบัติได้สะดวกกว่า รูปแบบนี้ถือว่าไม่มีการเดา มีฟังก์ชันการคำนวณคือ (Hambleton, Swaminathan and Rogers, 1991 : 15)

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

เมื่อ $P_i(\theta)$ แทน โอกาสที่ผู้ตอบที่มีระดับความสามารถ θ จะทำข้อสอบข้อที่ i ได้ถูกต้อง

θ แทน ระดับความสามารถที่แท้จริงของผู้สอบ โดยมีพิสัยการกระ

จายอยู่ระหว่าง $-\infty$ ถึง $+\infty$ ซึ่งระดับความสามารถ –3 ถึง +3 ซึ่งระดับความสามารถ –3 หมายถึงผู้สอบมีระดับความสามารถต่ำมาก และความสามารถ +3 หมายถึงผู้สอบมีความสามารถสูงมาก

b_i แทน ความยากของข้อสอบข้อที่ i ค่าที่แสดงถึงระดับความสามารถของผู้สอบที่จุดโค้งคุณลักษณะข้อสอบที่มีความชันมากที่สุด ค่าความยากมีค่าตั้งแต่ $-\infty$ ถึง $+\infty$ โดยทั่วไปนิยมใช้ค่าที่อยู่ระหว่าง –2 ถึง +2 ซึ่งค่าที่เป็นลบแสดงว่าข้อสอบง่ายและค่าที่เป็นบวกแสดงว่าข้อสอบยาก

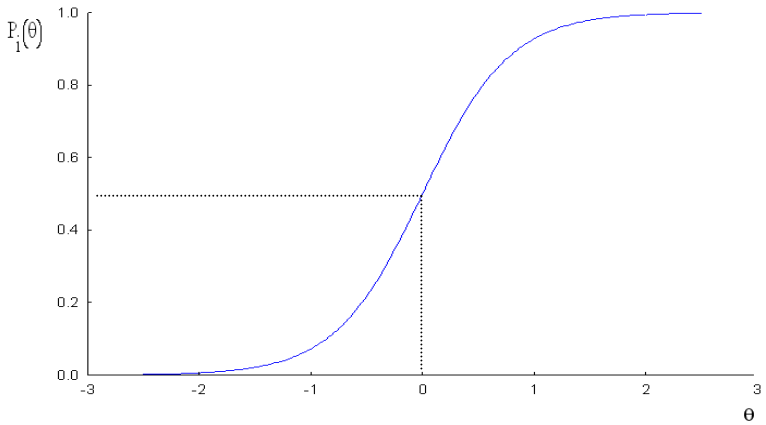
a_I แทน อำนาจจำแนกของข้อสอบข้อที่ i ค่าที่บอกความสามารถของข้อสอบที่จำแนกกลุ่มบุคคลได้ระหว่างผู้ที่มีความสามารถสูงกว่าค่าความยากและผู้ที่มีความสามารถต่ำกว่าค่าความยากซึ่งมีสัดส่วนโดยตรงกับความชันของโค้งคุณลักษณะของข้อสอบ ณ จุดเปลี่ยนโค้ง หรือจุดที่ชันที่สุดของโค้งนั่นเอง ค่าอำนาจจำแนกมีค่าตั้งแต่ $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัตินิยมใช้ค่าตั้งแต่ 0.50 ถึง 2.50 ค่า a_I สูงแสดงว่าข้อสอบนั้นมีโค้งที่ชัน จำแนกผู้สอบที่มีความสามารถแตกต่างกันได้ดี

e แทน ค่าคงที่มีค่าเท่ากับ 2.718…

D แทน Scaling Factor มีค่าเท่ากับ 1.7

n แทน จำนวนข้อสอบในแบบสอบ

และมีโค้งลักษณะเฉพาะของข้อสอบ เมื่อ ดังภาพประกอบ 2



ภาพประกอบ 2 โค้งลักษณะของข้อสอบที่มี 2 พารามิเตอร์ (a = 1.5 , b = 0)

4.3 รูปแบบโลจิสติกที่มี 3 พารามิเตอร์ (Three – Parameter Logistic Model)

โมเดลนี้ได้พัฒนามาจาก เพื่อให้เหมาะสมกับแบบสอบที่มีอิทธิพลจากการเดาเข้ามาแฝงอยู่ด้วย ซึ่งมีฟังก์ชันในการคำนวณคือ (Hambleton, Swaminathan and Roger, 1991 : 18)

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

เมื่อ $P_i(\theta)$ แทน โอกาสที่ผู้ตอบที่มีระดับความสามารถ θ จะทำข้อสอบข้อที่ i ได้ถูกต้อง

θ แทน ระดับความสามารถที่แท้จริงของผู้สอบ โดยมีพิสัยการกระจายอยู่ระหว่าง $-\infty$ ถึง $+\infty$ ซึ่งระดับความสามารถ –3 ถึง +3 ซึ่งระดับความสามารถ –3 หมายถึงผู้สอบมีระดับความสามารถต่ำมาก และความสามารถ +3 หมายถึงผู้สอบมีความสามารถสูงมาก

b_i แทน ความยากของข้อสอบข้อที่ i ค่าที่แสดงถึงระดับความสามารถของผู้สอบที่จุดโค้งคุณลักษณะข้อสอบที่มีความชันมากที่สุด ค่าความยากมีค่าตั้งแต่ $-\infty$ ถึง $+\infty$ โดยทั่วไปนิยมใช้ค่าที่อยู่ระหว่าง –2 ถึง +2 ซึ่งค่าที่เป็นลบแสดงว่าข้อสอบง่ายและค่าที่เป็นบวกแสดงว่าข้อสอบยาก

a_I แทน อำนาจจำแนกของข้อสอบข้อที่ i ค่าที่บอกความสามารถของข้อสอบที่จำแนกกลุ่มบุคคลได้ระหว่างผู้ที่มีความสามารถสูงกว่าค่าความยากและผู้ที่มีความสามารถต่ำกว่าค่าความยากซึ่งมีสัดส่วนโดยตรงกับความชันของโค้งคุณลักษณะของข้อสอบ ณ จุดเปลี่ยนโค้ง หรือจุดที่ชันที่สุดของโค้งนั่นเอง ค่าอำนาจจำแนกมีค่าตั้งแต่ $-\infty$ ถึง $+\infty$

แต่ในทางปฏิบัตินิยมใช้ค่าตั้งแต่ 0.50 ถึง 2.50 ค่า a_I สูงแสดงว่าข้อสอบนั้นมีโค้งที่ชัน จำแนกผู้สอบที่มีความสามารถแตกต่างกันได้ดี

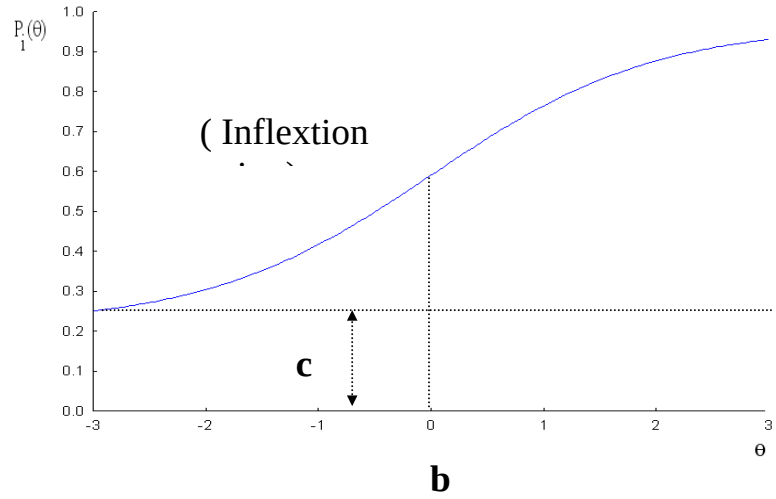
c_i แทน ค่าการเดาของข้อสอบข้อที่ i ซึ่งเป็นค่าความน่าจะเป็นที่ผู้สอบมีความสามารถต่ำมากจะมีความน่าจะเป็นในการตอบข้อสอบข้อที่ i ได้ถูกหรือเป็นค่ากำกับโค้งที่ต่ำที่สุด (Lower Asymptote) ของโค้งลักษณะข้อสอบมีพิสัยอยู่ระหว่าง 0 ถึง 1 โดยทั่วไปนิยมใช้ข้อสอบที่มีค่า c_i อยู่ระหว่าง 0.00 ถึง 0.30

e แทน ค่าคงที่มีค่าเท่ากับ 2.718…

D แทน Scaling Factor มีค่าเท่ากับ 1.7

n แทน จำนวนข้อสอบในแบบสอบ

และมีโค้งลักษณะเฉพาะของข้อสอบเมื่อ ดังภาพประกอบ 3



ภาพประกอบ 3 โค้งลักษณะของข้อสอบที่มี 3 พารามิเตอร์ (a = 0.59 , b = 0 , c = 0.25)

4.4 รูปแบบโลจิสติกที่มี 4 พารามิเตอร์ (Four – Parameter Logistic Model)

แมคโดแนล บาร์ตันและลอร์ด (Mcdonald, Barton and Lord) คิดใช้ขึ้นปี ค.ศ. 1967 และ 1981 (Hableton and Swaminathan, 1985 : 48) เพื่อนำค่าพารามิเตอร์เกี่ยวกับความคลาดเคลื่อนอื่น ๆ

ในการทดสอบมาคำนวณในรูปแบบด้วย เช่น ความสะเพร่า ความรอบรู้มากกว่าเนื้อหาที่ทดสอบ (จนคิดว่าข้อสอบที่ถูกต้องยังไม่ถูกต้องมากเพียงพอ) เป็นต้น รูปแบบนี้เขียนฟังก์ชันได้ดังนี้

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

เมื่อ γ หมายถึง ความสะเพร่าและความคลาดเคลื่อนอื่น ๆ มีค่าต่ำกว่า 1 เล็กน้อยปัจจุบันนี้แนวคิดของรูปแบบนี้ ยังไม่มีการนำมาใช้ในทางปฏิบัติ

ลักษณะเฉพาะของโค้งข้อสอบเมื่อระดับความสามารถของผู้สอบ (θ) เปลี่ยนแปลงไปจะมีรูปแบบแตกต่างกัน รูปแบบต่าง ๆ เหล่านี้เปลี่ยนแปลงไปตามฟังก์ชันการตอบสนองรายข้อ (Item Response Function) หรือที่นิยมเรียกกันทั่วไปว่า โค้งคุณลักษณะของข้อสอบ รูปแบบนี้มีลักษณะเฉพาะต่อไปนี้ (Warm, 1979 : 42)

1. เส้นกำกับบน (Upper Assymplote) จะมีค่าสูงสุดได้เกือบเท่ากับ 1.0 แต่ไม่เท่ากับ 1.0 เพราะความน่าจะเป็นในการตอบข้อสอบได้ถูกมีค่าไม่เท่ากับ 1.0 หรือ 100 % เพราะไม่มีอะไรแน่นอนที่สุด

2. เมื่อผู้สอบมีความสามารถมากขึ้น ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องมีมากขึ้น เส้นโค้งลักษณะเฉพาะของข้อสอบจะสูงขึ้นเรื่อย ๆ เมื่อเคลื่อนจากซ้ายไปขวา และจุดที่โค้งมีความชันมากที่สุด คือจุดที่ดีที่สุดในการแบ่งผู้สอบที่มีความสามารถต่างกัน (จุดนี้เรียกว่าจุด a ซึ่งเป็นค่าอำนาจจำแนก)

3. เส้นกำกับล่าง (Lower Assymplote) มีค่าได้น้อยที่สุดเกือบ 0.0 แต่ไม่เท่ากับ 0.0 เพราะการสอบย่อมมีการเดาบ้างไม่มากก็น้อย จะไม่มีการเดาเลยย่อมเป็นไปไม่ได้

5. ฟังก์ชันสารสนเทศของข้อสอบและแบบทดสอบ

การวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบจะใช้แบบแผนการตอบสนองแบบสอบเป็นรายข้อในการประมาณค่าความสามารถของผู้สอบ ดังนั้นการประเมินคุณภาพของแบบสอบ จึงสามารถพิจารณาจากความถูกต้องแม่นยำในการประมาณความสามารถของผู้สอบ ซึ่งมีดัชนีชี้ถึงความถูกต้องแม่นยำดังกล่าว เรียกว่าสารสนเทศแบบสอบ ซึ่งเป็นค่าฟังก์ชันสารสนเทศของแบบสอบเกิดจากผลรวมเชิงพีชคณิตของค่าฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อรวมเข้าด้วยกัน ดังนั้นผลของข้อสอบแต่ละข้อจะมีผลต่อแบบสอบทั้งฉบับ คุณสมบัตินี้ไม่สามารถจะนำเสนอได้ในการวัดแบบดั้งเดิม โดยค่าฟังก์ชันสารสนเทศของข้อสอบเป็นดัชนีคุณลักษณะของข้อสอบหลายลักษณะ ได้แก่

ค่าความยาก ค่าอำนาจจำแนก และค่าความแปรปรวนของคะแนนรายข้อ เพื่อบ่งชี้คุณภาพของข้อสอบ เนื่องจากค่าสารสนเทศของแบบสอบมีค่าสูงในช่วง θ ใด ก็จะมีความถูกต้องแม่นยำสูงในการประมาณค่าความสามารถของผู้สอบในช่วง θ นั้น ๆ โดยมีความคลาดเคลื่อนมาตรฐานของการประมาณค่าต่ำ ด้วยคุณสมบัติความไม่แปรเปลี่ยนตามกลุ่มตัวอย่างของค่าพารามิเตอร์ของข้อสอบจากการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ จึงทำให้ค่าฟังก์ชันสารสนเทศเหมาะสมที่จะใช้เป็นดัชนีบ่งบอกคุณภาพของข้อสอบและแบบสอบ แทนการหาค่าความเที่ยงและความคลาดเคลื่อนมาตรฐานของการวัดตามทฤษฎีการสอบแบบดั้งเดิม (Hambleton , 1977 อ้างถึงใน เพ็ญพรรณ กรึงไกร, 2539 : 23)

5.1 ฟังก์ชันสารสนเทศของข้อสอบ (Item and Test Information)

ฟังก์ชันสารสนเทศของข้อสอบเป็นดัชนีผสมที่สร้างจากดัชนีคุณลักษณะของข้อสอบหลายลักษณะ ประกอบด้วยค่าพารามิเตอร์ความยาก ค่าพารามิเตอร์อำนาจจำแนก และค่าความแปรปรวนของคะแนนรายข้อ เพื่อใช้บ่งชี้คุณภาพของข้อสอบ (Birnbaum , 1968 อ้างถึงใน ศิริชัย กาญจนวาสี, 2534 (ค))

$$I_i(\theta) = \frac{\left[P_i'(\theta)\right]^2}{P_i(\theta)Q_i(\theta)} \quad , i = 1, 2, \ldots, k$$

$I_i(\theta)$ แทน ค่าฟังก์ชันสารสนเทศหรือค่าสารสนเทศที่ได้รับจากข้อสอบข้อที่ i สำหรับผู้ตอบที่มีความสามารถ θ

$P_i'(\theta)$ แทน P_i' = ความชันของฟังก์ชันการตอบสนองข้อสอบข้อที่ i ณ ตำแหน่งความสามารถ θ

$P_i(\theta)$ แทน P_i = ความน่าจะเป็นที่ผู้ตอบซึ่งมีความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

$Q_i(\theta)$ แทน $Q_i = 1 - P_i(\theta)$

สูตรการคำนวณฟังก์ชันสารสนเทศของข้อสอบ สำหรับโมเดลการตอบสนองข้อสอบโลจิสแบบ 1 - พารามิเตอร์ 2 - พารามิเตอร์ และ 3 - พารามิเตอร์ ตลอดจนค่าความชันของฟังก์ชัน

การตอบสนอง ณ ตำแหน่ง θ (P_i') ค่าสารสนเทศสูงสุดของข้อสอบ $[I_i(\theta)_{max}]$ และตำแหน่งของ θ ที่มีสารสนเทศของข้อสอบสูงสุด (θ_{max}) ดังตาราง 2

ตาราง 2 การคำนวณค่าฟังก์ชันสารสนเทศ ค่าสารสนเทศสูงสุดของข้อสอบและตำแหน่ง θ ที่มีสารสนเทศของข้อสอบสูงสุด สำหรับโมเดลการตอบสนองข้อสอบโลจิสแบบ 1, 2 และ 3 พารามิเตอร์

ค่าประมาณ	1 - Parameter	2 - Parameter	3 – Parameter
$I_i(\theta)$	$D^2P_iQ_i$	$D^2a_i^2P_iQ_i$	$D^2a_i^2Q_i(P_i-c_i)^2/(1-c_i)^2$
P_i'	DP_iQ_i	$Da_iP_iQ_i$	$Da_iQ_i(P_i-c_i)/(1-c_i)$
$I_i(\theta)_{max}$	$\frac{1}{4}D^2$	$\frac{1}{4}D^2a_i^2$	$D^2a_i^2\frac{\left[1-20c_i-8c_i^2+(1+8c_i)^{3/2}\right]}{8(1-c_i^2)}$
θ_{max}	b_i	b_i	$b_i+\frac{1}{Da_i}\left[In1+\frac{(1+c_i)^{1/2}}{2}\right]$

จากสูตรของฟังก์ชันสารสนเทศของข้อสอบ เมื่อพิจารณาเทียบกับค่าพารามิเตอร์ และของข้อสอบ พอสรุปเป็นแนวทางโดยทั่วไปดังนี้

1) ค่าสารสนเทศของข้อสอบจะสูงขึ้น สำหรับผู้สอบมีความสามารถ θ ใกล้กับค่าพารามิเตอร์ b ของข้อสอบ และค่าสารสนเทศของข้อสอบจะลดลง สำหรับผู้สอบที่มีความสามารถ θ ไกลจากค่าพารามิเตอร์ b ของข้อสอบ

2) ค่าสารสนเทศของข้อสอบโดยทั่วไปจะมีค่าสูงขึ้น a ถ้าค่าพารามิเตอร์ของข้อสอบ มีค่ามากขึ้น

3) ค่าสารสนเทศของข้อสอบมีค่าสูงขึ้น เมื่อค่าพารามิเตอร์ c ของข้อสอบเข้าใกล้ 0

4) $I_i(\theta)$ จะมีค่าสูงสุด ณ ตำแหน่ง θ_{max} ถ้า $c_i = 0$, $\theta_{max} = b$ แต่ถ้า $c_i > 0$, $\theta_{max} > b$

5.2 ฟังก์ชันสารสนเทศของแบบสอบ (Test Information)

การวิเคราะห์ตามทฤษฎี IRT จะใช้แบบแผนการตอบสนองแบบสอบเป็นรายข้อในการประมาณค่าความสามารถของผู้สอบ ดังนั้น การประเมินคุณภาพของแบบสอบ จึงสามารถพิจารณาจากความถูกต้องแม่นยำในการประมาณความสามารถของผู้ตอบ โดยใช้ดัชนีตัวหนึ่งเรียกว่า สารสนเทศของแบบสอบ [Test Information ; $I(\theta)$] ซึ่งเป็นค่าฟังก์ชันสารสนเทศของแบบสอบอันเกิดจากผลรวมเชิงพีชคณิตของค่าฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อรวมเข้าด้วยกันทั้งฉบับ ณ ตำแหน่ง θ เดียวกัน ดังสูตร

$$I(\theta) = \sum_{i=1}^{k} I_i(\theta) \qquad , i = 1, 2, \ldots, k$$

เมื่อ $I(\theta)$ แทน ค่าฟังก์ชันสารสนเทศ หรือค่าสารสนเทศที่ได้รับจากแบบสอบสำหรับผู้ตอบที่มีความสามารถ θ

ค่าฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อ จึงมีส่วนอย่างเป็นอิสระจากกัน ต่อค่าสารสนเทศของแบบสอบ ลักษณะเช่นนี้ไม่ได้เกิดกับทฤษฎีการทดสอบแบบดั้งเดิม เนื่องจากค่าความยากและค่าอำนาจจำแนกของข้อสอบแต่ละข้อต่างก็ส่งผลต่อความเที่ยงของแบบสอบทั้งฉบับ แต่ไม่สามารถคำนวณค่าของแต่ละข้อต่างได้อย่างเป็นอิสระจากกัน ดังนั้นคะแนนที่ได้จึงขึ้นอยู่กับลักษณะเฉพาะของกลุ่มข้อสอบ และแบบเฉพาะฉบับที่เลือกมาใช้

เนื่องจากค่าสารสนเทศมีความสัมพันธ์ผกผันกับความคลาดเคลื่อนมาตรฐานการประมาณค่า ดังนั้นถ้าค่าสารสนเทศของแบบสอบมีค่าสูงในช่วง θ ใดก็จะมีความถูกต้องแม่นยำสูง ในการประมาณค่าความสามารถของผู้ตอบในช่วง θ นั้น ๆ โดยมีความคลาดเคลื่อนมาตรฐานของการประมาณค่าต่ำ

5.3 ประสิทธิภาพสัมพันธ์ (Relative Efficiency)

1) การเปรียบเทียบค่าฟังก์ชันสารสนเทศระหว่างแบบสอบที่วัดคุณลักษณะเดียวกัน สามารถทำได้โดยคำนวณประสิทธิภาพสัมพันธ์ (Relative Efficiency ; $RE(\theta)$) ซึ่งเป็นอัตราส่วนระหว่างค่าฟังก์ชันสารสนเทศของแบบสอบต่างฉบับ ณ ตำแหน่ง θ เดียวกันดังนี้

$$RE(\theta) = \frac{I_A(\theta)}{I_B(\theta)}$$

เมื่อ $I_A(\theta)$ แทน ค่าฟังก์ชันสารสนเทศของแบบสอบฉบับ A ณ ตำแหน่งบนสเกลความสามารถร่วมกันระดับ θ

$I_B(\theta)$ แทน ค่าฟังก์ชันสารสนเทศของแบบสอบฉบับ B ณ ตำแหน่งบนสเกลความสามารถร่วมกันระดับ θ

การแปลความหมายของค่าดัชนีประสิทธิภาพสัมพัทธ์ สามารถนำไปช่วยคัดเลือกแบบสอบได้ดังนี้

(1) $RE(\theta) = 1$ แสดงว่าแบบสอบทั้งสองฉบับมีประสิทธิภาพเท่ากัน สำหรับผู้ตอบที่มีระดับความสามารถ θ

(2) $RE(\theta) > 1$ แสดงว่าแบบสอบฉบับ A มีประสิทธิภาพสูงกว่าแบบสอบฉบับ B สำหรับผู้ตอบที่มีความสามารถ θ

(3) $RE(\theta) < 1$ แสดงว่าแบบสอบฉบับ A มีประสิทธิภาพต่ำกว่าแบบสอบฉบับ B สำหรับผู้ตอบที่มีความสามารถ θ

สมมติว่า $I_A(\theta) = 25.0$ และ $I_B(\theta) = 20.0$

$$RE(\theta) = \frac{25}{20} = 1.25$$

แสดงว่า แบบสอบฉบับ A มีประสิทธิภาพสูงกว่าแบบสอบฉบับ B สำหรับใช้สอบกับผู้สอบที่มีระดับความสามารถ θ อาจกล่าวได้ว่าแบบสอบฉบับ A ทำหน้าที่เสมือนเป็นแบบสอบที่มีความยาวกว่าแบบสอบฉบับ B 25% หรือแบบสอบฉบับ A สามารถตัดข้อออกได้อีก 20% ก็ยังคงมีความแม่นยำในการประมาณค่าความสามารถ ณ ตำแหน่ง θ ได้ดีเท่ากับแบบสอบฉบับ B

(2) การเปรียบเทียบค่าฟังก์ชันสารสนเทศเฉลี่ยระหว่างแบบสอบที่วัดคุณลักษณะเดียวกันสามารถทำได้โดยคำนวณประสิทธิภาพสัมพัทธ์เฉลี่ย (RAI) ซึ่งเป็นอัตราส่วนระหว่างค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบสอบต่างฉบับ ณ ทุกตำแหน่ง แสดงว่าแบบสอบฉบับ มีประสิทธิภาพสูงกว่าแบบสอบฉบับ สำหรับผู้ตอบที่มีความสามารถ θ ดังนี้

$$RAI(\theta; X, Y) = \frac{AI(\theta, X)}{AI(\theta, Y)}$$

เมื่อ $AI(\theta, X)$ แทน ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบสอบฉบับ X ณ ทุกตำแหน่ง

$AI(\theta, X)$ แทน ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบสอบฉบับ Y ณ ทุกตำแหน่ง

6 การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบ

คุณสมบัติที่สำคัญของทฤษฎีการตอบสนองข้อสอบ ในเรื่องค่าพารามิเตอร์ของข้อสอบไม่ขึ้นอยู่กับความสามารถของผู้สอบ ความสามารถของผู้สอบก็ไม่ขึ้นอยู่กับความยากหรือง่ายของข้อสอบ และค่าพารามิเตอร์ไม่เปลี่ยนแปลงไปตามตำแหน่งของข้อสอบ คุณลักษณะดังกล่าว คือ ความเป็นปรนัยของการวัดผล (Objectivity of Measurement) ที่มีความสำคัญมากในการวัดผล (อ้างถึงในสุพัฒน์ สุกมลสันต์ , 2528 : 68) จึงสามารถนำเอาคุณสมบัติเด่นเหล่านี้ มาประยุกต์ใช้ในการวัดผลการศึกษาได้อย่างมีประสิทธิภาพในหลาย ๆ เรื่อง พอสรุปได้ดังนี้

6.1 ใช้วิเคราะห์ข้อสอบ (Item Analysis) การวิเคราะห์ข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ จะสามารถแก้ปัญหาที่ทฤษฎีการทดสอบดั้งเดิม (Classiical Test Theory) แก้ไม่ได้ (Hambleton ,1979 : 14-15) อันได้แก่

- ค่าสถิติและค่าพารามิเตอร์ต่างขึ้นอยู่กับสภาพกลุ่มตัวอย่างที่ทำการสอบ
- การเปรียบเทียบความสามารถหรือคุณลักษณะ (Trait) ใด ๆ ในแต่ละบุคคลจะเทียบกันไม่ได้ ก็ต่อเมื่อต้องสอบด้วยแบบสอบฉบับเดียวกัน
- ความคลาดเคลื่อนมาตรฐานในการวัด ของแต่ละคนเท่ากัน ซึ่งลอร์ด และโนวิก (Lord and Novick , 1968)ได้แสดงให้เห็นว่า ความคลาดเคลื่อนมาตรฐานในการวัดสำหรับคะแนนปานกลางจะมีขนาดเล็กกว่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับสูงหรือคะแนนต่ำ

6.2 ใช้ในการสร้างคลังข้อสอบ (Item Bank) เนื่องจากข้อสอบที่วิเคราะห์แล้วค่าพารามิเตอร์ที่ได้มีค่าไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง ซึ่งถ้าได้บันทึกค่าฟังก์สารสนเทศของข้อสอบ (Item Information Function : IIF) ของแต่ละข้อเก็บไว้ ข้อสอบเหล่านี้ ก็จะสามารถนำมาสร้างเป็นแบบทดสอบใหม่ ที่มีค่าพารามิเตอร์ต่าง ๆ สอดคล้องกับเกณฑ์ หรือจุดมุ่งหมายของแบบทดสอบได้

6.3 ใช้วินิจฉัยความผิดปกติของผู้สอบ (Individual Diagnosis) ในกรณีที่โค้งลักษณะเฉพาะของข้อสอบ (Item Characteristics Curve : ICC) ของข้อสอบไม่เหมาะสมกับรูปแบบ แสดงว่าอาจมีบางสิ่งบางอย่างผิดปกติในตัวผู้สอบ เช่น อาจมีความสามารถอื่นแฝงเข้ามาในความสามารถที่เราต้องการวัด

6.4 ใช้ตรวจสอบความเป็นอคติของข้อสอบ (Item Bias) จากการพิจารณาโค้งลักษณะเฉพาะของข้อสอบ ของผู้สอบแต่ละกลุ่ม ถ้าโค้งลักษณะเฉพาะของข้อสอบนั้น ๆ แตกต่างกัน ตามกลุ่มที่นำมาทดสอบ เช่น คนในเมืองกับคนนอกเมือง หรือชายกับหญิง เป็นต้น แสดงว่าข้อสอบนั้นมีอคติเกิดขึ้น (Wright , 1977 : 12)

6.5 ใช้ในการวัดระดับความสามารถของแต่ละบุคคล (Self Tailoring) ในบรรดาข้อสอบที่วิเคราะห์แล้วเก็บไว้ในคลังข้อสอบ สามารถสุ่มข้อสอบเหล่านั้นที่มีระดับความยากเรียงกันตามลำดับเพียงจำนวนเล็กน้อยมาทดสอบกับแต่ละบุคคล ก็จะสามารถทราบความสามารถของผู้สอบ

6.6 ใช้ในการจัดชั้นเรียน (Grade Placement Tailoring) จากค่าความสามารถของผู้สอบที่ได้จากการวิเคราะห์แล้ว สามารถนำมาจัดชั้นเรียนให้เหมาะสมกับระดับความสามารถของผู้เรียนเป็นกลุ่มตามแผนการจัดการเรียนการสอน หรือตามนโยบายทางวิชาการได้

6.7 ใช้ในการเปรียบเทียบคะแนนต่างชุด (Equating Score) ในแบบสอบที่วิเคราะห์แล้ว 2 ชุด ที่ ต่างกันแต่วัดในสิ่งเดียวกัน จะสามารถนำคะแนนของผู้สอบในแบบสอบฉบับหนึ่งเปรียบเทียบกับคะแนนของผู้สอบในแบบสอบอีกฉบับหนึ่งได้ ทั้งนี้เพราะคะแนนแต่ละชุดของทฤษฎีการตอบสนองข้อสอบนั้น ถูกแปลงเป็นคะแนนมาตรฐานที่สามารถเปรียบเทียบกันได้

6.8 ใช้หาค่าการเดา (Guessing) ในทฤษฎีการทดสอบแบบดั้งเดิม ได้มีการแก้สูตรการเดา แต่ในทางปฏิบัติจริงแล้วถ้าผู้สอบมีเวลาเพียงพอ มีตัวเลือกที่มีประสิทธิภาพ และผู้สอบได้แสดงความสามารถกันอย่างเต็มที่ การเดาแบบสุ่มน่าจะเกิดขึ้นได้น้อยมาก ถ้าเป็นเช่นนี้สูตรแก้การเดาจึงเสมือนเป็นการลงโทษผู้สอบผู้นั้น สำหรับในทฤษฎีการตอบสนองข้อสอบจะไม่ใช้สูตรแก้การเดา ซึ่งหาได้จากการนำรูปแบบทางคณิตศาสตร์เข้ามาช่วยสนับสนุน ทำให้หลักการพิจารณาค่าการเดามีความสมเหตุสมผลยิ่งขึ้น

7 สรุปแนวคิดทฤษฎีการตอบสนองข้อสอบ

จากแนวคิดทฤษฎีการตอบสนองข้อสอบมีผู้สรุปไว้ดังนี้ (ชัยพจน์ รักงาม, 2538 : 12 - 13) แนวคิดทฤษฎีการตอบสนองข้อสอบมีความเชื่อว่าความสามารถของผู้สอบและโอกาสในการตอบข้อสอบถูก สามารถเขียนเป็นสมการได้ และมีความสัมพันธ์เป็นโค้งโลจิสติก (Logistic Curve) นั่นคือมีลักษณะเป็นฟังก์ชันที่เพิ่มขึ้นทางเดียว

จากแนวคิดทฤษฎีการตอบสนองข้อสอบสามารถเขียนเป็นรูปแบบทางคณิตศาสตร์ได้ทำให้เห็นความสัมพันธ์ระหว่างความสามารถของผู้สอบและโอกาสที่จะตอบข้อสอบถูกชัดเจนเรียกว่า ทฤษฎีแกร่ง (Strenght Theory หรือ Hard Theory)

สาระสำคัญของทฤษฎีนี้ ค่าคุณลักษณะของข้อสอบ (a ,b และ c) เป็นค่าพารามิเตอร์ประจำข้อสอบ นั่นคือค่าคุณลักษณะข้อสอบจะเป็นค่าคงที่ ไม่ว่าจะนำไปสอบกับผู้สอบกลุ่มใด ต่างกับทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) ซึ่งความยากของข้อสอบขึ้นอยู่กับกลุ่มนักเรียนที่เรานำข้อสอบนั้นไปสอบ ถ้ากลุ่มนักเรียนนั้นเก่ง ข้อสอบก็จะถูกวิเคราะห์ว่าข้อสอบนั้นง่าย เพราะคนส่วนใหญ่ในกลุ่มทำได้ ในทำนองเดียวกันถ้านำข้อสอบเดียวกันมาสอบกลุ่มนักเรียนที่อ่อน ข้อสอบนั้นก็จะถูกวิเคราะห์ออกมาว่ายาก เพราะคนส่วนใหญ่ในกลุ่มทำไม่ได้ แต่สำหรับทฤษฎีการตอบสนองข้อสอบนั้น ถ้านำไปสอบกับกลุ่มอ่อนก็จะได้ความสัมพันธ์ของสมการในตำแหน่งความสามารถต่ำ แต่ถ้าไปสอบกับนักเรียนเก่งก็จะได้ความสัมพันธ์ของสมการในตำแหน่งความสามารถสูง ซึ่งเป็นโค้งความสัมพันธ์อันเดียวกัน

ค่าพารามิเตอร์ของข้อสอบในทฤษฎีการตอบสนองข้อสอบแต่ละตัวมีความหมาย ดังนี้

θ : คือ ระดับความสามารถของผู้สอบในสเกลคะแนนมาตรฐาน อยู่ระหว่าง –3 ถึง +3 โดยประมาณ ผู้ที่มีความสามารถปานกลางจะมีค่า θ อยู่ระหว่าง –1.0 ถึง +1.0 ถ้าค่า θ มากกว่า 1.0 แสดงว่าค่าความสามารถค่อนข้างสูง ถ้าน้อยกว่า –1.0 แสดงว่าค่าความสามารถค่อนข้างต่ำ

a : คือ ความชัน (Slope) หรือค่าอำนาจจำแนก (Item Discriminating) ของสมการที่จุดบนโค้งที่ตรงกับค่าโอกาสที่จะทำข้องสอบถูกเท่ากับ 0.5 (ค่าแกนตั้งที่ 0.5) เป็นค่าพารามิเตอร์ที่แสดงถึงความสามารถของข้อสอบในการจำแนกบุคคลที่มีความสามารถต่างไปออกได้มากน้อยเพียงใด ถ้าค่าความชันสูงคนที่มีความสามารถต่างกันเพียงเล็กน้อยก็ส่งผลให้โอกาสที่ทำข้อสอบถูกต่างกันมาก ถ้าข้อสอบใดมีค่า a มากกว่า 0.5 หมายความว่า ข้อสอบข้อนั้นมีการจำแนกอยู่ในระดับที่ดี ถ้าค่า a มีค่าตั้งแต่ 0.3 ถึง 0.5 หมายความว่า ข้อสอบนั้นมีการจำแนกได้ในระดับปานกลาง ถ้าค่า a มีค่าน้อยกว่า 0.3 หมายความว่า ข้อสอบข้อนั้นมีการจำแนกไม่ดี

b : คือ ค่าความยาก (Item Difficulty) ก็คือตำแหน่งบนสเกลของค่าความสามารถ (แกนนอน) ที่ตรงกับ ค่าโอกาสที่จะทำข้อสอบถูกเท่ากับ 0.5 นั้น ค่าความยากก็คือ ค่าความสามารถของผู้สอบที่มีโอกาสทำข้อสอบข้อนั้นถูกร้อยละ 50 ค่าความยากเป็นค่าที่บอกถึงความเหมาะสมในการที่จะเอาข้อสอบข้อนี้ไปใช้ ถ้าค่าความยากสูงกว่า 1.0 ก็ควรนำไปใช้กับผู้มีความสามารถสูง (กลุ่มเก่ง) เช่น นำไปใช้ในการสอบแข่งขัน ถ้าค่าความยากต่ำกว่า –1.0 ก็ควรนำไปใช้กับผู้เรียนที่มีความสามารถต่ำ

(กลุ่มอ่อน) เช่น นำไปใช้กับนักเรียนอ่อนเพื่อวินิจฉัยข้อบกพร่องในการเรียนวิชานั้น ส่วนข้อที่มีค่า b อยู่ระหว่าง –1.0 ถึง +1.0 เหมาะสำหรับไปใช้กับนักเรียนที่มีความสามารถปานกลาง เช่น ใช้ในการเรียนการสอนในชั้นปกติ หรือเรียกว่า วัดผลสัมฤทธิ์นั่นเอง

c : คือ ค่าจุดตัดบนแกนตั้ง (ค่าโอกาสที่จะทำข้อสอบข้อนั้นถูก) เรียกว่า ค่าโอกาสการเดา (Item Guessing) เพราะเป็นโอกาสที่จะทำข้อสอบนั้นถูกโดยไม่มีความสามารถ มีค่าที่เป็นไปได้อยู่ระหว่าง 0 ถึง 1 ถ้าข้อสอบข้อนั้นไม่สามารถที่จะตอบถูกด้วยการเดาแล้ว $c_i = 0$ และปกติควรจะมีค่าน้อยกว่า 0.3 ซึ่งความหมายของค่าการเดามีดังนี้ ถ้า $c < 0.2$ แสดงว่าข้อสอบนั้นมีตัวเลือกดีมาก ถ้า c มีค่าในช่วงตั้งแต่ 0.2 ถึง 0.3 แสดงว่าข้อสอบนั้นมีตัวเลือกดี ถ้าค่า $c > 0.3$ แสดงว่าข้อสอบนั้นมีตัวเลือกที่เดาถูกได้ง่าย

เกณฑ์ปกติ

1. ความหมายและลักษณะของเกณฑ์ปกติ

เกณฑ์ปกติ (Norms) มีผู้ให้ความหมายไว้หลายท่านด้วยกัน ดังนี้

อาดัมส์ (Adams , 1964 : 634) ได้ให้ความหมายว่า เกณฑ์ปกติ หมายถึง การอธิบายผลของการกระทำ (Performance) ที่เป็นส่วนเฉลี่ย หรือลักษณะปานกลาง และไม่ใช่สิ่งที่ยึดถือเป็นมาตรฐาน (Standard)

ชวาล แพรัตกุล (2519 : 275) ได้อธิบายเกณฑ์ปกติว่า เป็นปริมาณคุณภาพปานกลางของคุณลักษณะต่าง ๆ และเป็นสถานภาพตามความจริงในปัจจุบัน

สมบูรณ์ ชิตพงษ์ และสำเริง บุญเรืองรัตน์ (2524 : 56) ได้ให้ความหมายของเกณฑ์ปกติ หมายถึง ตัวเลขที่มีไว้เพื่อเปรียบเทียบความสามารถของบุคคล เป็นคะแนนเฉลี่ย หรือจุดกึ่งกลางของคะแนนกลุ่มตัวอย่างที่เรียกว่า กลุ่มอ้างอิง (Norm Group หรือ Reference Group) ที่ทำการสุ่มเลือกมา และมีตารางคะแนน (Norms Table) แสดงความสัมพันธ์ระหว่างคะแนนดิบ (Raw Score) กับคะแนนแปลงรูป (Derived Score) ซึ่งจะบอกให้ทราบว่าบุคคลอยู่ในตำแหน่งใด หรือคะแนนของบุคคลนั้น ๆ ใกล้เคียงคะแนนเฉลี่ยของกลุ่มอ้างอิงหรือไม่

สมบูรณ์ ภู่นวล (2525 : 122) ให้ความหมายว่า เกณฑ์ปกติ คือ ปริมาณคุณภาพกลาง ๆ ของคุณลักษณะต่าง ๆ เช่นในการสอบวัดก็คือ คะแนนเฉลี่ยของวิชาใด เรื่องใด ของเด็กกลุ่มใดก็ได้ ก็เป็นเกณฑ์ปกติของเด็กกลุ่มนั้น

ล้วน สายยศ และอังคณา สายยศ (2539 : 313 - 314) เกณฑ์ปกติ หมายถึง ข้อเท็จจริงทางสถิติที่บรรยายการแจกแจงของคะแนนจากประชากรที่นิยามไว้อย่างดีแล้ว และเป็นคะแนนตัวที่จะบอกระดับความสามารถของผู้สอบว่า อยู่ระดับใดของกลุ่มประชากร

จากความหมายข้างต้นสรุปได้ว่า เกณฑ์ปกติ หมายถึง คะแนนที่เทียบคะแนนดิบให้เป็นคะแนนมาตรฐาน ซึ่งเป็นคะแนนมีไว้เพื่อบอกและเปรียบเทียบระดับความสามารถของบุคคลในเรื่องใดเรื่องหนึ่ง ว่าอยู่ในช่วงใดของกลุ่มประชากร โดยบอกความสัมพันธ์ระหว่างคะแนนดิบ (Raw score) กับคะแนนทีปกติ (Normalized T-score) ที่แปลงมาจากคะแนนดิบ

2. หลักการสร้างเกณฑ์ปกติ

หลัก 3 ประการที่ต้องคำนึงถึงในการสร้างเกณฑ์ปกติ มีดังนี้ (ล้วน สายยศ และอังคณา สายยศ, 2539 : 314 - 315)

2.1 ความเป็นตัวแทนที่ดี ก่อนการสร้างเกณฑ์ปกติต้องวางแผนการสุ่มให้ดี เพื่อให้เกณฑ์ปกติเชื่อมั่นได้ เนื่องจากการสุ่มกลุ่มตัวอย่างของประชากร ต้องอาศัยความน่าจะเป็น สามารถทำได้หลายวิธี ทั้งนี้ต้องเลือกสุ่มตามความเหมาะสมโดยการพิจารณาประชากรเป็นสำคัญ เช่น ถ้าประชากรมีลักษณะเป็นอันหนึ่งอันเดียวกันไม่มีคุณสมบัติที่แตกต่างกัน ก็ใช้วิธีการสุ่มแบบธรรมดา (Simple Random Sampling) ถ้าระหว่างประชากรกลุ่มย่อยมีลักษณะแตกต่างกัน เช่น ขนาดโรงเรียนต่างกัน จะต้องใช้วิธีสุ่มแบบแบ่งชั้น (Stratified Random Sampling) คือ การสุ่มมาจากประชากรทุกกลุ่มย่อย ในทางตรงกันข้าม ถ้าระหว่างประชากรกลุ่มย่อยมีลักษณะเหมือนกัน เช่น นักเรียนในแต่ละห้องเรียน มีการคละกันระหว่างเด็กเก่ง ปานกลาง และอ่อน ก็ให้ใช้วิธีการสุ่มแบบแบ่งกลุ่ม (Cluster Random Sampling) คือ สุ่มเพียงบางกลุ่มจากประชากรกลุ่มย่อย ซึ่งในการสร้างเกณฑ์ปกติซึ่งมักใช้การสุ่มทั้ง 3 วิธีข้างต้นนี้มากที่สุด

2.2 มีความเที่ยงตรง เกณฑ์ปกติที่สร้างจะต้องสามารถแปลความหมายได้ตรงกับความเป็นจริง นั่นคือ เมื่อนำคะแนนดิบไปเทียบกับเกณฑ์ปกติที่ทำไว้แล้ว สามารถแปลความหมายได้ตรงกับความเป็นจริง เช่น คะแนนสอบวิชาหนึ่งของผู้สอบได้ 20 คะแนน ตรงกับเปอร์เซนต์ไทล์ที่ 50 และตรงกับ (T) 50 แปลความหมายได้ว่า ผู้สอบนั้นมีความสามารถปานกลางของกลุ่มซึ่งตรงกับสภาพความเป็นจริงของผู้สอบ

2.3 มีความทันสมัย เนื่องจากเกณฑ์ปกติขึ้นอยู่กับความสามารถของประชากรกลุ่มนั้น ๆ และมีปัจจัยที่ช่วยให้ความสามารถของประชากรเปลี่ยนแปลงได้ เช่น การพัฒนาคนที่มีอยู่ตลอดเวลา

เทคโนโลยี สภาพแวดล้อม หรืออาหารการกิน ดังนั้น เกณฑ์ปกติที่เคยศึกษาไว้นานแล้วหลายปีอาจมีความคลาดเคลื่อนหรืออาจผิดพลาดจากความเป็นจริง จำเป็นต้องสร้างขึ้นมาใหม่ให้ทันสมัยโดยทั่ว ๆ ไปแล้วเกณฑ์ปกติควรเปลี่ยนทุก ๆ 5 ปี

3. ชนิดของเกณฑ์ปกติ

ชนิดของเกณฑ์ปกติแบ่งได้ตามลักษณะของประชากร และตามลักษณะของการใช้สถิติการเปรียบเทียบ ดังนี้ (ล้วน สายยศและอังคณา สายยศ, 2539 : 315 - 317)

3.1 การแบ่งชนิดของเกณฑ์ปกติตามลักษณะประชากร

3.1.1 เกณฑ์ปกติระดับชาติ (National Norms) เป็นเกณฑ์ปกติที่ต้องใช้ประชากรทั่วประเทศดังนั้นต้องสุ่มตัวอย่างให้ครอบคลุมทั่วประเทศ จำนวนตัวอย่างที่เข้าสอบจึงมีจำนวนมาก

3.1.2 เกณฑ์ปกติระดับท้องถิ่น (Local Norms) เป็นเกณฑ์ปกติที่ระดับประชากรเล็กลงมา เช่น ระดับจังหวัด หรือระดับอำเภอ ซึ่งมีค่าใช้จ่ายน้อยลง และเป็นประโยชน์ในการเปรียบเทียบคะแนนของผู้สอบกับคนทั้งจังหวัด หรืออำเภอ

3.1.3 เกณฑ์ปกติของโรงเรียน (School Norms) เป็นเกณฑ์ปกติของโรงเรียนเดียวหรือกลุ่มโรงเรียนในเครือ ใช้ประเมินนักเรียนแต่ละคนกับนักเรียนส่วนรวมของโรงเรียน และใช้ประเมินการพัฒนาของโรงเรียนได้ด้วย โดยพิจารณาจากการศึกษาแต่ละปีว่า เด่น หรือ ด้อยกว่าปีที่สร้างเกณฑ์ปกติเอาไว้

3.2 การแบ่งชนิดของเกณฑ์ปกติตามลักษณะของการใช้สถิติการเปรียบเทียบ

3.2.1 เกณฑ์ปกติ เปอร์เซนต์ไทล์ (Percentile norms) เป็นเกณฑ์ปกติที่เป็นคะแนนจัดอันดับเท่านั้น จะนำไปบวกลบกันไม่ได้ แต่สามารถเทียบและแปลความหมายได้ เพราะเกณฑ์ปกติแบบนี้สร้างจากคะแนนดิบที่มาจากประชากร หรือกลุ่มตัวอย่างที่เป็นตัวแทนที่ดีที่นำมาเทียบกับตำแหน่งเปอร์เซ็นไทล์ ซึ่งแปลความหมายในรูปของร้อยละของตัวคะแนนที่จุดใต้จุดคะแนนดิบนั้น ๆ เช่น ผู้สอบคนหนึ่งสอบได้คะแนน 25 คะแนนเมื่อนำไปเทียบกับเกณฑ์ปกติตรงกับตำแหน่งเปอร์เซ็นต์ไทล์ที่ 80 ก็หมายความว่ามีคนเข้าสอบ 100 คน ผู้สอบคนนั้นมีความสามารถเหนือกว่าคนอื่น 80 คน

3.2.2 เกณฑ์ปกติคะแนนที (T – score norms) นิยมใช้กันมากเพราะเป็นคะแนนมาตรฐานสามารถนำมาบวก ลบและเฉลี่ยได้ มีค่าเหมาะสมในการแปลความหมาย คือมีค่าตั้งแต่

0 ถึง 100 มีคะแนนเฉลี่ย 50 ความเบี่ยงเบนมาตรฐาน 10 เรียกคะแนนชนิดนี้ว่า คะแนนทีปกติ (Normalized T - Score)

3.2.3 เกณฑ์ปกติสเตไนน์ (Stanine norms) เป็นคะแนนมาตรฐานชนิดหนึ่ง แต่มีค่าเพียง 9 ตัว (Standard nine points) โดยมีค่าตั้งแต่ 1 ถึง 9 คะแนนเฉลี่ยอยู่ที่ 5 คะแนน มีความเบี่ยงเบนมาตรฐานประมาณ 2 คะแนน วิธีการหามักจะเทียบจากเปอร์เซนต์ของความถี่ที่คะแนนเรียงตามค่าจะสะดวกกว่า ดังตัวอย่างในตารางต่อไปนี้

คะแนนสเตไนน์ที่	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
ร้อยละของจำนวนคนที่อยู่ในสเตไนน์	4	7	12	17	20	17	12	7	4

3.2.4 เกณฑ์ปกติตามระดับชั้น (Grade Norms) เป็นการหาเกณฑ์ปกติตามระดับชั้นว่าคะแนนเท่าไรควรอยู่ชั้นไหนจึงจะเหมาะสม แบบทดสอบที่จะทำเกณฑ์ปกติชนิดนี้จะต้องเป็นเนื้อหาเดียวกัน เพราะจะสามารถอธิบายการแปลผลเปรียบเทียบได้ ดังนั้น วิชาที่นิยมสร้างเกณฑ์ปกติชนิดนี้จะเป็นวิชาพื้นฐาน เช่น คำศัพท์ คณิตศาสตร์เบื้องต้น นอกจากนี้ยังนิยมใช้กับแบบทดสอบมาตรฐานวัดผลสัมฤทธิ์ทางการเรียน (Standardized Achievement Test) เพราะหลักสูตรค่อนข้างจะเทียบเคียงได้และนิยมใช้กันมากในระดับประถมศึกษา

3.2.5 เกณฑ์ปกติตามอายุ (Age Norms) เกณฑ์ปกติแบบนี้คล้ายคลึงกับแบบตามระดับชั้นต่างกันที่แบบนี้จำแนกตามอายุ นิยมใช้กับแบบทดสอบที่ไม่ได้วัดผลสัมฤทธิ์ตามหลักสูตร เช่น ใช้กับแบบทดสอบวัดความถนัด สติปัญญา อารมณ์ และอื่น ๆ ซึ่งคะแนนปรับเปลี่ยนของเกณฑ์แบบนี้เป็นไปตามอายุ

ในการวิจัยครั้งนี้ผู้วิจัยสร้างเกณฑ์ปกติในรูปคะแนนทีปกติ (Normalized T - Score) และเปอร์เซนต์ไทล์ (Percentile Norms)

4. ความจำเป็นที่ต้องมีเกณฑ์ปกติ

การสอบวัดใด ๆ จะมีความหมายก็ต่อเมื่อผลการสอบนั้นสามารถบอกถึงสถานภาพของผู้สอบว่า มีระดับความสามารถหรือสมรรถภาพเด่นด้อยกว่าบุคคลอื่น ๆ ภายในกลุ่มนั้นเพียงใด เพื่อครูอาจารย์จะได้นำผลนั้นไปใช้ในการประเมินค่า หรือแยกคัดจัดประเภทนักเรียนเหล่านั้นได้ถูกต้องว่าใครควรจะสอบได้สอบตก หรือเป็นเด็กประเภทเก่ง – อ่อน หรือควรจะรับไว้เข้าเรียนหรือไม่รับไว้

เข้าเรียน นั่นก็คือการสอบที่ดี จะต้องประกอบด้วยเครื่องมือหรือแบบทดสอบที่ดี ร่วมกับวิธีการวัดและการแปลผลที่ถูกต้อง รวมทั้ง 3 ประการ จึงจะถือว่าการสอบนั้นบรรลุผลตามความปรารถนา

วิธีประเมินค่าคะแนนและการแปลผลต่าง ๆ เท่าที่ผ่านมาและที่ยังมีอยู่บ้างในปัจจุบันยังไม่สมบูรณ์ เช่น การกำหนดไว้เป็นเกณฑ์ตายตัวว่านักเรียนจะต้องสอบให้ได้อย่างน้อย 50 % จึงนับว่าเป็นผู้ที่สอบผ่านหรือมีความรู้เข้าเกณฑ์ตามที่ต้องการ ยังไม่เหมาะสม เป็นต้น ความผิดพลาดของการประเมินค่าแบบนี้อยู่ตรงความหลงผิด 3 ประการด้วยกันดังนี้ (ชวาล แพรัตกุล 2514 : 32–35)

<u>ประการแรก</u> จากที่นึกเอาเองว่าตัวคำถามหรือข้อสอบต่าง ๆ ที่ออกมานั้น มีคุณภาพดีเหมาะสมกับเด็กนักเรียนระดับนั้นแล้ว หรือบางทีถึงกับเชื่อว่า มีความยากง่ายเป็นมาตรฐานสำหรับชั้นนั้นแล้วก็มี โดยไม่ได้พิจารณาว่า ข้อสอบนั้นมีคุณสมบัติดีด้านใด หรือมีความยากง่ายเป็นเท่าใด

<u>ประการที่สอง</u> ได้แก่การหลงผิดที่ไปยึดเอาตัวเลขคะแนนมาเป็นเครื่องชี้บอกปริมาณความรู้ของนักเรียน เช่น มีความคิดว่า เด็กที่สอบได้ 20 คะแนน จะมีความรู้มากเป็น 2 เท่าของเด็กที่สอบได้ 10 คะแนน ทำนองเดียวกันก้อนหิน 20 กิโลกรัม ก็หนักเป็น 2 เท่าของหิน 10 กิโลกรัม หรือเงิน 20 บาท ก็มีค่ามากเป็น 2 เท่าของเงิน 10 บาท เป็นต้น ความเข้าใจประเภทนี้เองที่ทำให้เกิดเกณฑ์ตัดสินได้ตกที่ 50 % กันขึ้น เพราะคิดไปว่า เด็กที่สอบได้ 50 % เป็นผู้ที่มีความรู้ขนาดปานกลาง สมควรยอมให้เลื่อนชั้นได้ แต่ถ้าถามว่า 50 % นี้หมายถึงอะไร จะหมายถึงผู้ที่มีความรู้ครึ่งหนึ่งของหลักสูตร หรือครึ่งหนึ่งของครูที่สอน ก็ไม่สามารถชี้แจงได้อีกเช่นกัน นอกจากจะตอบว่า 50 % ของคะแนนเต็มของข้อสอบฉบับนั้นเท่านั้นเอง

การที่เห็นแต่เพียงตัวเลขคะแนนของวิชาหนึ่งมากกว่าอีกวิชาหนึ่ง แล้วจะไปลงสรุปว่าเด็กนั้นเก่งวิชาแรกมากกว่าวิชาหลัง เช่นนี้ ย่อมไม่ถูกต้องด้วยเหตุผลอันใดเลย เริ่มตั้งแต่คะแนนเต็มของแต่ละวิชานั้นไม่เท่ากัน ซึ่งจะมีผลให้นักเรียนมีโอกาสได้คะแนนสูง ๆ ในวิชาที่มีคะแนนเต็มมากกว่าในวิชาที่มีคะแนนเต็มน้อย และถึงแม้ว่าจะนำคะแนนดิบเหล่านั้นมาทำเป็นเปอร์เซ็นต์เสียให้หมด โดยหวังจะให้มีฐานหรือคะแนนเต็มของทุกวิชาให้เป็น 100 หน่วยเท่า ๆ กันก็ตาม แต่ตัวเลขเปอร์เซ็นต์ที่ได้นั้น ก็ยังยังคงมีสภาพเป็นเพียงการอนุมานที่เลื่อนลอยอยู่มากเช่นเดียวกันเพราะไปหวังเอาเองว่าเด็กที่สอบเลขในใจได้คะแนน 50 จาก 100 คะแนน ถ้าเราขยายคะแนนเต็มของวิชานี้ให้มากถึง 100 คะแนนเรื่องนี้ก็คล้าย ๆ กับว่าถ้ามีคำถามอยู่ 10 ข้อ เด็กก็จะต้องตอบถูกถึง 50 ข้อด้วย ซึ่งมีโอกาสที่จะ

เป็นจริงน้อยมาก จึงเห็นได้ว่าการคิดคะแนนเป็นแบบเปอร์เซ็นต์ก็ไม่ทำให้เกิดอะไรดีขึ้นกว่าคะแนนเดิมหรือคะแนนดิบ

<u>ประการที่สาม</u> ได้แก่ ความไม่รู้เรื่องธรรมชาติของตัวเลขคะแนน โดยนึกไปว่าคะแนนของแต่ละวิชามีหน่วยเป็นชนิดเดียวกันและเท่ากัน เพราะนับเพิ่มขึ้นทีละหนึ่ง แบบเดียวกับตัวเลขในคณิตศาสตร์ จึงสามารถนำมารวมกันได้โดยตรง ความเข้าใจผิดในเรื่องนี้ก็เทียบได้กับการที่เอาเงิน 1 ดอลลาร์ไปรวมกับ 1 บาท แล้วได้เป็น 2 ซึ่งเป็นวิธีการรวมที่ไม่มีความหมายอะไร

สาเหตุที่ไม่สามารถนำคะแนนดิบของวิชาต่าง ๆ มาเทียบและรวมกันได้โดยตรงนั้น ก็เนื่องจากความแตกต่างในธรรมชาติและเนื้อหาของแต่ละวิชา เช่น คำถามในวิชาภาษาไทย เลขคณิต และศิลปศึกษา ต่างต้องการความรู้ความสามารถในการตอบแตกต่างกันไปคนละชนิดทั้งสิ้น ฉะนั้น 1 คะแนนของแต่ละวิชาเหล่านี้จึงมีคุณค่าไม่เหมือนกันหรือไม่เท่ากัน และไม่สามารถมารวมกันได้โดยตรงด้วย ถึงแม้ในวิชาเดียวกันก็ตามข้อสอบที่ถามเมื่อตอนต้นภาคเรียน ก็ยังไม่สามารถนำมารวมกันได้อยู่นั่นเอง เพราะฉบับหลัง ๆ ก็ย่อมจะยากกว่าฉบับแรกอยู่เสมอทั้ง ๆ ที่ทั้งสองฉบับนั้นก็ต่างก็ยัง มุ่งหวังสมรรถภาพชนิดเดียวกันอยู่

ความผิดพลาดในการสอบวัดทั้ง 3 ประการที่กล่าวมานี้โดยเฉพาะอย่างยิ่งในเรื่องการตัดสินได้ – ตกที่ 50% นั้น นับว่าเป็นความบกพร่องที่มีผลต่อชีวิตมนุษย์เป็นอันมาก ความผิดพลาดนี้อาจชี้ให้เห็นได้ง่าย ๆ ว่าคะแนนต่าง ๆ ที่เด็กสอบได้ในแต่ละครั้ง ๆ นั้นจะขึ้นอยู่กับความยากง่ายของข้อสอบเป็นประการแรก และเป็นประการที่สำคัญที่สุด นั้นคือ ถ้าคราวใดกรรมการออกข้อสอบยากเกินกว่าภูมิปัญญาของนักเรียนแล้ว คะแนนจากการสอบครั้งนั้นก็จะลดลงต่ำทันที นักเรียนทั่วไปจะไม่สามารถทำคะแนนได้ถึง 50% กลายเป็นสอบตกกันหมด และโดยนัยตรงข้าม ถ้าข้อสอบครั้งใดถามง่ายกว่าธรรมดาคะแนนของผู้สอบในครั้งนั้นก็จะสูงกว่าปรกติทำให้นักเรียนสอบไล่ได้กันมากขึ้น

ด้วยประการนี้ เพื่อให้เกิดความยุติธรรมแก่ทุกฝ่าย การวัดผลที่เป็นวิทยาศาสตร์ตามแบบประชาธิปไตย จึงนิยมใช้คะแนนเฉลี่ยจากการสอบเหล่านั้นมาเป็นหลักเริ่มต้นในการวัด แต่ทั้งนี้มิได้หมายความว่า ให้เอาคะแนนเฉลี่ยมาใช้เป็นจุดเริ่มต้นของการวัดเพื่อนำไปใช้ในการตัดสินได้ – ตก แต่หมายถึงการเอาคะแนนเฉลี่ยมาใช้เป็นจุดเริ่มต้นของการวัดเพื่อนำไปตัดสินได้ – ตกอีกทีหนึ่ง หลักการใหม่นี้มีประโยชน์ตรงที่สามารถช่วยขจัดความบกพร่องของการสอบได้คะแนนหลายประการ เช่น ในบางโอกาสที่ข้อสอบยากเกินไปจริง ๆ ผู้สอบต่างก็ได้คะแนนน้อย ๆ ไปตามกัน คะแนนเฉลี่ยของข้อสอบนั้นก็จะลดตามลงด้วย ดังนั้นจุดหลักของการวัดก็จะเลื่อนต่ำลงไปเองโดยอัตโนมัติ หรือบังเอิญว่าข้อสอบนั้นง่ายเกินไป เด็ก ๆ ก็จะได้คะแนนมากขึ้น ซึ่งก็เท่ากับเลื่อนคะแนนเฉลี่ยและ

จุดหลักให้สูงตามขึ้นด้วย ผู้ที่สอบได้ 60 % ก็อาจเป็นคนรั้งท้ายหรือสอบตกไปเลยก็ได้ ดังนั้นจะเห็นได้ว่าการใช้คะแนนเฉลี่ยเป็นจุดหลัก ความไม่ยุติธรรมอันเนื่องมาจากข้อสอบยากหรือง่ายทั้งสองกรณีนี้ก็จะสิ้นไปเองโดยปริยาย

วิธีการวัดผลที่ใช้คะแนนเฉลี่ยเป็นจุดหลัก แล้วจำแนกออกเป็นชั้น ๆ ไปตามสภาพความเป็นจริงของนักเรียนเหล่านั้น เรียกว่า การสร้างมาตรฐานเกณฑ์ปกติ (Standardized norm) ของผลการสอบเหล่านั้น

การสร้างเกณฑ์ปกติ มีความสำคัญตรงที่จะต้องแบ่งคะแนนออกจากจุดหลักเป็นช่วง ๆ ให้เท่ากันอย่างมีความหมาย ทั้งตอนบนตอนล่าง และในการนี้จะใช้คะแนนดิบหรือเปอร์เซ็นต์มาแบ่งไม่ได้

การสร้างเกณฑ์ปกติสามารถทำได้โดยเปลี่ยนคะแนนดิบของทุกวิชาให้เป็นคะแนน T ปกติ (Normalized T-score) หมดทุกฉบับ การปฏิบัติเช่นนี้ก็ด้วยวัตถุประสงค์ สามารถเปรียบเทียบกันได้โดยตรงและสะดวกต่อการใช้ยิ่งขึ้น และอีกประการหนึ่งก็คือนั้นก็เพราะมีความเห็นว่าคะแนน T ปกตินี้มีหน่วยเท่ากัน และมีหลักการสมบูรณ์มากกว่าคะแนนแปลงรูปอื่น ๆ ด้วย

การสร้างเกณฑ์ปกติของแบบทดสอบต่าง ๆ ให้เป็นคะแนน ปกติร่วมกันนี้ก็เสมือนกับที่เราเปลี่ยนมาตรเงินปอนด์ ดอลลาร์ และเงินกีบ ให้เป็นมาตรเงินบาทให้หมด นั่นเอง ฉะนั้นคะแนนใหม่ที่ได้จึงมีความหมายเป็นอย่างเดียวกัน และสามารถนำมารวมกันได้โดยตรงอย่างถูกต้องตามหลักวิชา ยิ่งกว่านั้น มาตรคะแนน T นี้ก็จำง่ายใช้สะดวก มีตั้งแต่ 0 ถึง 100 คล้าย ๆ กับสตางค์ของเงินบาทหรือเซนติเมตรของไม้เมตร โดยมีรายเฉลี่ยอยู่กึ่งกลางตรงที่ T 50 พอดีและในการใช้คะแนนชนิดนี้ก็มิต้องไปกังวลว่า วิชาใดจะมีคะแนนเต็ม คะแนนเฉลี่ย หรือความเบี่ยงเบนมาตรฐานเป็นเท่าใดเลย เพราะไม่ว่าวิชานั้นจะมีค่าเฉลี่ยเป็นเท่าใดก็ตาม เมื่อนำมาแปลงเป็นคะแนน T แล้ว ต่างก็จะมีคะแนนเฉลี่ยใหม่เท่ากับ 50 หน่วย และความเบี่ยงเบนมาตรฐานก็จะเท่ากับ 10 หน่วย เหมือนกันหมดทุกวิชากลายเป็นตัวเลขมาตรเดียวกันตลอด (สำเริง บุญเรืองรัตน์ : 175-177)

5. การใช้เกณฑ์ปกติ

5.1 วิธีอ่านเกณฑ์ปกติ

เมื่อครูตรวจกระดาษคำตอบว่า นักเรียนคนใดตอบถูกกี่ข้อ หรือได้คะแนนดิบเท่าใดแล้วก็ไปเปิดบัญชีอ่านให้เป็นคะแนน T ปกติได้เลยทันที การอ่านเกณฑ์ปกติจะต้องระวังให้ถูกแถว

ของวิชานั้น ๆ เสมอ วิธีที่ดีควรอ่านไปทีละวิชาจนหมดจำนวนนักเรียนที่เข้าสอบและเมื่อจะอ่านเกณฑ์ปกติของวิชาใดก็ควรใช้กระดาษปิดตัวเลขของเกณฑ์อื่น ๆ เสียก่อนเพื่อกันการอ่านผิดช่อง

5.2 วิธีบันทึกคะแนน T ปกติ

ควรบันทึกด้วยหมึกหรือดินสอคนละสีกับตัวเลขคะแนนดิบและควรตกลงกันว่าช่องที่ 1 สำหรับเขียนคะแนนดิบ ส่วนช่องที่ 2 สำหรับเขียนคะแนน T ปกติ เป็นต้น

5.3 การแปลคะแนน T ปกติที่อ่านได้จากเกณฑ์ปกติ

คะแนนชนิดนี้นิยมแปลความหมายเป็นในรูปของเปอร์เซ็นต์ที่อยู่เหนือกว่าผู้อื่นเปอร์เซ็นต์อยู่เหนือ มีความหมายว่า นักเรียนที่สอบได้คะแนน T 30 ก็อ่านจากตารางว่าเขาเก่งกว่าหรือมีความรู้เหนือกว่าผู้อื่นอยู่ 2.28 % คือใน 100 คน เขาจะมีความรู้สูงกว่าเด็กคนอื่น ๆ อยู่ 2.28 หรือ 2 ซึ่งก็เท่ากับเป็นเด็กจวนได้ที่โหล่นั่นเอง

ถ้าจะเทียบให้เห็นง่ายขึ้นอีกก็อาจสมมติได้ว่าเอาความรู้ของเด็ก 100 คนมายืนเรียงกันเป็นแถวบนขั้นบันไดความรู้ ตั้งแต่ 0 ขั้นที่พื้นที่จนถึงยอด 100 ขั้นแล้ว นักเรียนคนที่สอบได้ 30 ของวิชาใด ๆ ก็ตามเสมือนกับเขายืนอยู่บนบันไดขั้นที่สูงกว่าคนอื่น ๆ อยู่เพียง 2 คน ซึ่ง ณ ตรงนั้นจะมีเด็กเป็นรองอยู่ใต้เขา 2 คน พร้อม ๆ กับอีก 98 คน อยู่สูงกว่าเขาในทำนองเดียวกัน ผู้ที่สอบได้คะแนน T 50 ในวิชาใด ๆ ก็ตาม ก็อ่านจากบัญชีตรงกับ T ที่อยู่เหนือผู้อื่น 50.00 ฉะนั้นแปลว่า เด็กคนนั้นมีความรู้ระดับปานกลางพอดีคือ เขายืนอยู่ตรงกึ่งกลางของบันไดความรู้ มีผู้อื่นอยู่ใต้เขา 50 คน พร้อม ๆ กับอีก 50 คน อยู่เหนือเขา

5.4 การประเมินคะแนน T ปกติ

เมื่อสามารถวัดว่า นักเรียนคนใดได้คะแนน T เท่าใดและเสมือนกับเขายืนอยู่บนบันไดความรู้ขั้นที่เท่าใดแล้ว ต่อไปก็เป็นการประเมินราคาของคะแนนเหล่านั้นว่ามีคุณภาพสูง – ต่ำ หรือดีเลวปานใด อันเป็นเรื่องของการตัดสินชี้ขาดหรือการตีราคาโดยสรุปอย่างมีหลักเกณฑ์ให้กับนักเรียนผู้นั้น ว่าควรจะยกย่องให้เขาเป็นผู้มีความรู้ความสามารถเทียบเท่ากับระดับใด ตามหลักสากลนิยมแบ่งคุณภาพเหล่านี้ออกเป็น 5 ระดับ ดังเช่นในขั้นอุดมศึกษาก็อาจตั้งชื่อเป็นระดับA , B , C , D , E หรืออาจเป็นระดับ ก , ข , ค , ง และ จ ก็ได้ ซึ่งชื่อต่าง ๆ เหล่านี้ก็พออาจเทียบได้กับระดับดีมาก – ดี – พอใช้ – ยังไม่พอใช้ – อ่อน นั่นเอง

การกำหนดวิธีการประเมินคุณภาพผลการทดสอบอาจแบ่งคะแนน T ออกเป็น 5 ระดับ ดังนี้

ตั้งแต่ T 65 และสูงกว่า แปลว่า ดีมาก

ตั้งแต่ T 55 – 65	แปลว่า	ดี
ตั้งแต่ T 45 – 55	แปลว่า	พอใช้
เฉพาะตรง T 50	แปลว่า	มีความสามารถปานกลางพอดี
ตั้งแต่ T 35 – 45	แปลว่า	ยังไม่พอใช้
ตั้งแต่ T 35 และต่ำกว่า	แปลว่า	อ่อน

การแบ่งระดับข้างต้นนี้ จะมีคะแนน T บางตัวซ้ำกันที่ตรงหัวและตรงท้ายของช่องคะแนน ดังเช่นที่ T 55 เป็นต้น การที่เป็นเช่นนี้ ก็เฉพาะตรง T 55 นั้นเป็นจุดแบ่งเขตระหว่างกลุ่ม ดังนั้นถ้านักเรียนคนใดได้คะแนน T ตรงจุดแบ่งเขตเหล่านั้นพอดี คือที่ 35 , 45 , 55 และ 65 แล้วก็อาจเกิดลังเลไม่แน่ใจว่าควรให้เขาอยู่กลุ่มใด วิธีแก้ไขในเรื่องนี้ก็ให้ถือหลักไว้ว่า ให้เลื่อนนักเรียนที่คาบเส้นผู้นั้นขึ้นไปอยู่ในกลุ่มสูงที่ถัดไปเสมอ เพื่อผลทางจิตวิทยา เพราะโอกาสที่นักเรียนคนเดียวกันนั้นจะได้คะแนนตรงนั้นซ้ำ ๆ กัน มีอยู่น้อยมาก (พวงรัตน์ ทวีรัตน์ 2536 :358)

6. ประโยชน์ของเกณฑ์ปกติ

ประโยชน์ของเกณฑ์ปกติมีความสำคัญ 3 ประการ

6.1 ใช้สำหรับเปลี่ยนคะแนนดิบ ให้เป็นหน่วยที่มีความหมายยิ่งขึ้นด้วยเหตุผลที่ว่าตัวเลขคะแนนดิบที่ได้จากการสอบต่าง ๆ นั้น ยังมีหน่วยไม่เท่ากันเป็นคนละชนิด และยังแปลความหมายไม่ได้แน่ชัดว่าหมายถึงอะไรกันแน่ ดังนั้นในการวัดความสามารถใด ๆ จึงมีความจำเป็นที่ต้องปรับคะแนนของแต่ละวิชาเหล่านั้น ให้เป็นหน่วยเดียวกันเสียก่อน ดังเช่นเกณฑ์ปกตินี้ จึงจะสามารถนำมาเปรียบเทียบ และอธิบายความหมายของตัวเลขเหล่านั้นได้ชัดเจนและมีหลักเกณฑ์ยิ่งขึ้น

คุณประโยชน์เหล่านี้เป็นผลพลอยได้จากการใช้เกณฑ์ปกตินี้ ก็ได้แก่การที่สามารถนำตัวเลขคะแนนเหล่านั้นมารวมหรือหารเฉลี่ยกันได้โดยตรงอย่างตัวเลขธรรมดา เช่น นักเรียนคนหนึ่งสอบวิชาวิทยาศาสตร์ฉบับเนื้อหาได้ T 55 วิทยาศาสตร์ทักษะได้ T 60 และวิทยาศาสตร์ฉบับทัศนคติได้ T 40 ดังนั้นคะแนนเฉลี่ยทั้ง 3 ฉบับของนักเรียนคนนี้เท่ากับ $\frac{55+60+40}{3} = 51.66$ หรือเท่ากับ T 52 ซึ่งอาจประเมินได้ว่าเขามีความสามารถด้านวิทยาศาสตร์อยู่ในระดับปานกลางหรือพอใช้

การที่เราสามารถนำคะแนน Tปกติของวิชาต่าง ๆ มารวมกันได้โดยตรงนี้ก็เพราะเราได้แปลงคะแนนดิบของแต่ละวิชาเหล่านั้นด้วยเกณฑ์ปกติให้เป็นมาตราเดียวร่วมกันหมด จึงสามารถนำมาบวก ลบ คูณ หารกันได้อย่างถูกหลักวิชา ซึ่งคุณสมบัติเช่นนี้คะแนนดิบเป็นเปอร์เซ็นต์ไม่มี

6.2 ใช้ในการประเมินผลการศึกษา กล่าวคือ เมื่อครู – อาจารย์ได้ทำการสอนนักเรียนไประยะหนึ่ง ก็คงอยากทราบถึงประสิทธิภาพในการสอนของตนและคุณภาพในการเรียนของนักเรียนว่ามีผลสัมฤทธิ์ถึงระดับใด ดังความปรารถนา เป็นต้น ก็สามารถกระทำได้โดยนำแบบทดสอบมาสอบวัดแล้วนำผลการสอบในแต่ละวิชา (ซึ่งอ่านจากเกณฑ์ปกติ) มาเปรียบเทียบกับเกณฑ์ที่กำหนดไว้ ก็จะทราบว่าผลการเรียนของนักเรียนในแต่ละวิชาอยู่ในระดับใด วิชาใดถึงเกณฑ์ ที่กำหนดไว้และวิชาใดยังไม่ถึงเกณฑ์ นอกจากนี้ก็สามารถเปรียบเทียบได้ว่า นักเรียนมีผลการเรียนในวิชาได้ดีกว่ากันด้วย แต่ทั้งนี้แบบทดสอบที่ใช้ในการสอบวัดจะต้องมีความเที่ยงตรงและเชื่อถือได้สูง การประเมินผลนั้นจึงจะถูกต้องตรงกับความเป็นจริง นี่คือประโยชน์ที่เด่นชัดของเกณฑ์ปกติ

6.3 ใช้ในการแนะแนว นักเรียนในแต่ละชั้นและในแต่ละห้องย่อมมีระดับความรู้ความสามารถแตกต่างกันไปเป็นรายบุคคล ครู อาจารย์ ผู้สอน ย่อมมีความปรารถนาที่จะชักจูงส่งเสริมให้นักเรียนแต่ละคนไคร่เรียนตามความถนัด ประสบความสำเร็จในการเรียนและการงานต่อไป ซึ่งเป็นเรื่องของการแนะแนวนั่นเอง ซึ่งในการนี้ ย่อมต้องเริ่มด้วยการวินิจฉัยสมรรถภาพของนักเรียนแต่ละบุคคลก่อนว่ามีความเด่นด้อยในทางใด จึงจะสามารถให้คำแนะนำแก่เขาได้อย่างมีหลักเกณฑ์ ซึ่งการวินิจฉัยนี้ก็กระทำโดยเปรียบเทียบความสามารถของเขาในด้านต่าง ๆ เพื่อดูว่าเขามีความสามารถหรือมีสมรรถภาพด้านใดสูงที่สุด ซึ่งต้องอาศัยการพิจารณาจากคะแนนปกติของเขาแต่ละด้าน ซึ่งอ่านมาจากเกณฑ์ปกตินั่นเอง

งานวิจัยที่เกี่ยวข้อง

พูลศิริ แกล้วกลางศึก (2514 : 64 – 72) ได้ดัดแปลงแบบสอบวัดความถนัด 3 ชุด ซึ่งเป็นส่วนหนึ่งของแบบสอบความถนัด ดี เอ ที (DAT) มาเพื่อใช้กับนักเรียนชั้นมัธยมศึกษาปีที่ 4 ได้แก่ แบบสอบเหตุผลเชิงภาษา ความสามารถเชิงตัวเลข และเหตุผลเชิงนามธรรม โดยนำแบบสอบความสามารถเชิงตัวเลขไปสอบกับนักเรียนชายจำนวน 44 คน นักเรียนหญิง 139 คน ที่เรียนแผนกวิทยาศาสตร์ จากโรงเรียนรัฐบาลและโรงเรียนราษฎร์ ในเขตพระนครธนบุรี จำนวน 4 โรงเรียน โดยมุ่งวัดความเข้าใจด้านตัวเลข และความคล่องแคล่วในการจัดกระทำกับตัวเลข ลักษณะของข้อสอบเป็นการคำนวณทางเลขคณิต (Arithmetic Computation) ข้อสอบเป็นแบบเลือกตอบ 5 ตัวเลือก คือ ประกอบด้วยข้อสอบ 40 ข้อ ใช้เวลาในการสอบ 30 นาที ดัดแปลงโดยการแปลและเรียบเรียงคำสั่งชี้แจงวิธีทำแบบสอบจากภาษาอังกฤษเป็นภาษาไทย หาความเที่ยงโดยวิธีแบ่งครึ่ง (Split Half) ของ

เพศชายและหญิง และหาความตรงจากโดยใช้คะแนนจากแบบสอบถามสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนในหมวดวิชาต่าง ๆ ประจำภาคกลาง โดยใช้สูตรเพียร์สันโพรดัคโมเมนต์ (Pearson Product Moment) ได้ผลดังนี้ ความเที่ยงของแบบสอบความสามารถเชิงตัวเลข แยกเพศชายและเพศหญิงเท่ากับ .83 และ .86 ตามลำดับ ค่าสัมประสิทธิ์สหสัมพันธ์ของความตรงเชิงทำนายระหว่างแบบสอบความสามารถเชิงตัวเลขสำหรับนักเรียนชายกับวิชาวิทยาศาสตร์ให้ค่าสูงสุดเท่ากับ .44 ส่วนนักเรียนหญิงคะแนนจากแบบสอบความสามารถเชิงตัวเลขกับคะแนนรวมให้ค่าสูงสุดเป็น .50 รองลงมาเป็นวิชาคณิตศาสตร์และวิทยาศาสตร์ ตามลำดับ

ปราณี เศาภายน (2518 : 53 – 54) ศึกษาเกี่ยวกับคุณภาพของแบบทดสอบด้านตัวเลข ซึ่งได้วิเคราะห์ข้อสอบความถนัดทางการเรียนของมหาวิทยาลัยศรีนครินทรวิโรฒประสานมิตร เพื่อคำนวณหาระดับอำนาจจำแนก ระดับความยากง่าย สัมประสิทธิ์ความเชื่อมั่น สัมประสิทธิ์ความเที่ยงตรงและปกติวิสัย (Norms) ของข้อสอบ โดยใช้คะแนนจากผู้ที่สมัครสอบคัดเลือกเข้าศึกษาต่อในระดับ ปริญญาตรี ปีการศึกษา 2517 จำนวน 2,178 คน ผลการศึกษาพบว่า แบบทดสอบคณิตศาสตร์มีอำนาจจำแนกระหว่าง -.12 - .62 และระดับความยากอยู่ระหว่าง .10 - .64 สัมประสิทธิ์ความเชื่อมั่นเท่ากับ .65 และสัมประสิทธิ์ความเที่ยงตรงเท่ากับ .68

ชลลดา ชินะศิริกุล (2521 : 53 – 65) ได้พัฒนาแบบสอบชุดความถนัดจำแนกด้านจำนวนโดยใช้นักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 1,448 คน และนักเรียนชั้นมัธยมศึกษาปีที่ 5 จำนวน 1,094 คน สังกัดกองการมัธยมศึกษา จำนวน 84 โรงเรียน จาก 35 จังหวัด ซึ่งเป็นแบบสอบปรนัยชนิดเลือกตอบ 5 ตัวเลือก จำนวน 40 ข้อ คำถามเป็นการคิดคำนวณ 2 แบบ คือ การคำนวณตัวเลข (Arithmetic Computation) และการคำนวณโจทย์ปัญหา (Problem Solving) ให้เวลาในการตอบ 40 นาที พบว่าค่าความยากของข้อสอบสำหรับชั้นมัธยมศึกษาปีที่ 3 อยู่ในช่วง .244 ถึง .645 ค่าความเที่ยงตรงของแบบสอบสำหรับชั้นมัธยมศึกษาปีที่ 3 และชั้นมัธยมศึกษาปีที่ 5 มีค่า .703 และ .828 ตามลำดับ ค่าความตรงร่วมสมัยสำหรับชั้นมัธยมศึกษาปีที่ 3 และชั้นมัธยมศึกษาปีที่ 5 อยู่ในช่วง 0.45 ถึง .61 และ .28 ถึง .51 ตามลำดับ

สุขุม มูลเมือง (2522 : 16 - 25) ได้สร้างแบบทดสอบวินิจฉัยข้อบกพร่องในการเรียนเลขทศนิยม ชั้นมัธยมศึกษาปีที่ 1 แบบทดสอบที่สร้างขึ้นมี 4 ฉบับ คือ

แบบทดสอบฉบับที่ 1 แบบทดสอบพื้นฐานความเข้าใจทศนิยม

แบบทดสอบฉบับที่ 2 การบวกการลบทศนิยม

แบบทดสอบฉบับที่ 3 การคูณการหารทศนิยม

แบบทดสอบฉบับที่ 4 โจทย์ปัญหาทศนิยม

ผลการศึกษาพบว่า แบบทดสอบชุดนี้มีค่าความยากของข้อสอบรายข้อตั้งแต่ 0.50 – 0.93 ค่าอำนาจจำแนกของข้อสอบตั้งแต่ 0.02 – 0.64 ค่าความเที่ยงของแบบทดสอบทั้ง 4 ฉบับมีค่าตั้งแต่ 0.9163 – 0.9683 สำหรับค่าความตรงเชิงเนื้อหาของแบบทดสอบ ให้ผู้เชี่ยวชาญเป็นผู้พิจารณา ปรากฏว่า แบบทดสอบทั้ง 4 ฉบับ วัดเรื่องทศนิยมจริง

สุนันทา จันพลา (2524 : 45 - 129) ได้สร้างแบบทดสอบวินิจฉัยในการเรียนเรื่องการบวก การลบ การคูณ และการหาร สำหรับชั้นประถมศึกษาปีที่ 3 แบบทดสอบที่สร้างขึ้นมี 4 ฉบับคือ

แบบทดสอบฉบับที่ 1 การบวก

แบบทดสอบฉบับที่ 2 การลบ

แบบทดสอบฉบับที่ 3 การคูณ

แบบทดสอบฉบับที่ 4 การหาร

ผลการศึกษาพบว่า แบบทดสอบชุดนี้มีมีค่าความยากตั้งแต่ 0.53 – 0.95 ค่าอำนาจจำแนกตั้งแต่ 0.01 – 0.94 ค่าความเที่ยงมีค่าตั้งแต่ 0.87 – 0.92 สำหรับค่าความตรงทางเนื้อหาของแบบทดสอบ โดยใช้ผู้เชี่ยวชาญพิจารณา ผลปรากฏว่า แบบทดสอบทั้ง 4 ฉบับ วัดเรื่อง การบวก การลบ การคูณ การหาร จริง

จิตรเมธี สายสุ่ม (2534 : 70 – 71) ได้ศึกษาทักษะทางคณิตศาสตร์ที่มีอิทธิพลต่อความสำเร็จในการแก้โจทย์ปัญหาคณิตศาสตร์ของนักเรียนมัธยมศึกษาปีที่ 3 ในจังหวัดเพชรบูรณ์ พบว่า ทักษะทางคณิตศาสตร์ 7 ทักษะ คือ ทักษะการอ่านเพื่อวิเคราะห์โจทย์ปัญหาคณิตศาสตร์ ทักษะการเขียนรูปหรือแผนภาพ ทักษะหาความสัมพันธ์ในปริภูมิที่กำหนด ทักษะการนำหลักการหรือทฤษฎีที่เรียนรู้แล้วมาใช้ ทักษะการแปลภาษาจากโจทย์ให้เป็นภาษาคณิตศาสตร์ ทักษะการคิดคำนวณ ทักษะการตรวจสอบผลลัพธ์ มีความสัมพันธ์ทางบวกกับผลสำเร็จในการแก้โจทย์ปัญหาคณิตศาสตร์ อย่างมีนัยสำคัญทางสถิติระดับ .01 โดยมีค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณเท่ากับ .86178 และสามารเป็นตัวพยากรณ์ได้ร้อยละ 74.27 และทักษะทางคณิตศาสตร์มีอิทธิพลต่อความสำเร็จในการแก้โจทย์คณิตศาสตร์ โดยทักษะการอ่านเพื่อวิเคราะห์โจทย์คณิตศาสตร์ ทักษะการนำหลักการหรือทฤษฎีที่เรียนรู้แล้วมาใช้ ทักษะการแปลภาษาคณิตศาสตร์ ทักษะการคิดคำนวณ และทักษะการตรวจสอบผลลัพธ์ มีอิทธิพลทั้งทางตรงและทางอ้อมต่อผลสำเร็จในการแก้โจทย์ปัญหาคณิตศาสตร์ ส่วนทักษะการเขียนรูปหรือแผนภาพและทักษะการหาความสัมพันธ์ในปริภูมิที่กำหนด มีอิทธิพลทางอ้อมต่อผลสำเร็จในการแก้โจทย์คณิตศาสตร์

นิตยา โอบอ้อม (2540 : 46 – 47) ศึกษาการพัฒนาโปรแกรมคอมพิวเตอร์สำหรับตรวจสอบความรอบรู้ตามจุดประสงค์ในคิดคำนวณพื้นฐานในวิชาคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 3 และตรวจสอบความตรงและความเที่ยงวิธีการตรวจสอบความรอบรู้ตามจุดประสงค์โดยใช้คอมพิวเตอร์ที่พัฒนาขึ้น โปรแกรมคอมพิวเตอร์สำหรับตรวจสอบความรอบรู้ตามจุดประสงค์ที่พัฒนาดังกล่าว ใช้รูปแบบของการตรวจสอบความรู้ตามลำดับขั้นเนื้อหา เครื่องมือที่ใช้ในการวิจัยคือโปรแกรมคอมพิวเตอร์สำหรับการตรวจสอบความรอบรู้ตามจุดประสงค์ในการคิดคำนวณพื้นฐาน 4 ทักษะ ได้แก่ ทักษะการบวก ทักษะการลบ ทักษะการคูณและทักษะการหาร กลุ่มตัวอย่างในการวิจัยคือนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 ของโรงเรียนวัดพลับพลาชัย สังกัดสำนักงานการประถมศึกษากรุงเทพมหานคร จำนวน 136 คน

ผลการวิจัยพบว่า วิธีการตรวจสอบความรอบรู้โดยใช้โปรแกรมคอมพิวเตอร์สำหรับตรวจสอบความรอบรู้ตามจุดประสงค์ในการคิดคำนวณพื้นฐานที่พัฒนาขึ้น มีความคงที่ในการตรวจสอบความรอบรู้และมีความตรงเชิงเกณฑ์สัมพันธ์เมื่อใช้ผลการตรวจสอบความรอบรู้โดยครูเป็นเกณฑ์

สุกัญญา โพธิสุวรรณ (2540 : บทคัดย่อ) ได้ศึกษาการพัฒนาแบบฝึดหัดทักษะการวิเคราะห์โจทย์ปัญหาคณิตศาสตร์สำหรับนักเรียนชั้นประถมศึกษาปีที่ 5 เพื่อเปรียบเทียบผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นประถมศึกษาปีที่ 5 กลุ่มที่เรียนโดยใช้แบบฝึกหัดทักษะการวิเคราะห์โจทย์ปัญหาคณิตศาสตร์กับกลุ่มที่เรียนโดยใช้แบบฝึกหัดหนังสือแบบเรียนคณิตศาสตร์และเพื่อศึกษากระบวนการวิเคราะห์โจทย์ปัญหาของนักเรียนทั้งสองกลุ่ม ผลการวิจัยพบว่า ผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ของนักเรียนที่เรียนโดยใช้แบบฝึกทักษะการวิเคราะห์โจทย์ปัญหาสูงกว่านักเรียนที่เรียนโดยใช้แบบฝึกหัดในหนังสือแบบเรียนคณิตศาสตร์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กระบวนการแก้โจทย์ปัญหาของนักเรียนที่เรียนโดยใช้แบบฝึกทักษะการวิเคราะห์โจทย์ปัญหา มีการอ่านวิเคราะห์ข้อมูลส่วนประกอบของโจทย์ปัญหา การเลือกวิธีคำนวณ การเขียนประโยคสัญลักษณ์ การแสดงวิธีทำ และการตรวจคำตอบ ส่วนนักเรียนที่เรียนโดยใช้แบบฝึกหัดในหนังสือแบบเรียนคณิตศาสตร์ส่วนใหญ่จะไม่มีกระบวนการวิเคราะห์ข้อมูลแต่คิดแก้ปัญหาโดยแสดงวิธีทำตามตัวอย่างในแบบเรียน

เพลินพิศ เสือชาวนา(2540 : บทคัดย่อ) ได้ศึกษาความสัมพันธ์ระหว่างทักษะการวิเคราะห์ปัญหา การแปลภาษาโจทย์ การคิดคำนวณกับความสามารถในการแก้โจทย์ปัญหาคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2540 ของโรงเรียนสังกัดสำนักงานการประถมศึกษาจังหวัดพัทลุง จำนวน 381 คน ผลการวิจัยพบว่า

1. ทักษะการวิเคราะห์ปัญหามีความสัมพันธ์เชิงบวกกับความสามารถในการแก้โจทย์ปัญหาคณิตศาสตร์ อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .1902

2. ทักษะการแปลภาษาโจทย์มีความสัมพันธ์เชิงบวกกับความสามารถในการแก้โจทย์ปัญหาคณิตศาสตร์ อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .4263

3. ทักษะการคิดคำนวณความสัมพันธ์เชิงบวกกับความสามารถในการแก้โจทย์ปัญหาคณิตศาสตร์ อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .2172

ปัญญลักษณ์ สุวรรณรัตน์ (2545 : บทคัดย่อ) ได้สร้างและพัฒนาแบบทดสอบวัดความสามารถทางคณิตศาสตร์สำหรับนักเรียนชั้นประถมศึกษาปีที่ 5 รวมทั้งสร้างเกณฑ์ปกติของแบบทดสอบในรูปคะแนนทีปกติ กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ภาคเรียนที่ 1 ปีการศึกษา 2544 ของโรงเรียนสังกัดสำนักงานการประถมศึกษาจังหวัดสงขลา โดยใช้แบบทดสอบวัดความสามารถทางคณิตศาสตร์สำหรับนักเรียนชั้นประถมศึกษาปีที่ 5 จำนวน 4 ด้าน คือ แบบทดสอบวัดความสามารถด้านความรู้ความเข้าใจ จำนวน 35 ข้อ แบบทดสอบวัดความสามารถด้านทักษะการคิดคำนวณ จำนวน 30 ข้อ แบบทดสอบวัดความสามารถด้านกระบวนการทางคณิตศาสตร์ จำนวน 35 ข้อ และแบบทดสอบวัดความสามารถด้านการแก้โจทย์ปัญหา จำนวน 35 ข้อ พบว่า

- แบบทดสอบวัดความสามารถด้านความรู้ความเข้าใจ มีค่าความยากตั้งแต่ .33 - .73 ค่าอำนาจจำแนกตั้งแต่ .26 - .64 ค่าความเชื่อมั่นเท่ากับ .858 และค่าคะแนนทีปกติ T18 – T76

- แบบทดสอบวัดความสามารถด้านทักษะการคิดคำนวณ มีค่าความยากตั้งแต่ .26 -.77 ค่าอำนาจจำแนกตั้งแต่ .26 - .75 ค่าความเชื่อมั่นเท่ากับ .889 และค่าคะแนนทีปกติ T26 – T76

- แบบทดสอบวัดความสามารถด้านกระบวนการทางคณิตศาสตร์ มีค่าความยากตั้งแต่ .32 - .76 ค่าอำนาจจำแนกตั้งแต่ .26 - .72 ค่าความเชื่อมั่นเท่ากับ .824 และค่าคะแนนทีปกติ T18 – T82

- แบบทดสอบวัดความสามารถด้านการแก้โจทย์ปัญหา มีค่าความยากตั้งแต่ .25 - .75 ค่าอำนาจจำแนกตั้งแต่ .25 - .62 ค่าความเชื่อมั่นเท่ากับ .808 และค่าคะแนนทีปกติ T18 – T78

จากผลการศึกษางานวิจัยดังกล่าวจึงทำให้ผู้วิจัยมีความสนใจที่จะศึกษาเกี่ยวกับทักษะการคิดคำนวณวิชาคณิตศาสตร์เพราะทักษะการคิดคำนวณเป็นพื้นฐานสำคัญต่าง ๆ ต่อไป นักเรียนจึงต้องมีความแม่นยำในทักษะการคิดคำนวณเพื่อจะนำไปสู่คำตอบที่ถูกต้องอีกทั้งยังส่งผลในการเรียนรู้วิชาคณิตศาสตร์ในเนื้อหาต่อไปและในการเรียนระดับที่สูงขึ้น