บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้ผู้วิจัยได้เสนอรายละเอียดของเอกสารและงานวิจัยที่เกี่ยวข้องดังนี้

ทฤษฎีและความเชื่อเกี่ยวกับสมอง

ทฤษฎีองค์ประกอบเดียว (Uni Factor Theory)

ทฤษฎีสององค์ประกอบ (The Two Factor Theory)

ทฤษฎีหลายองค์ประกอบ (Multiple Factor Theory)

ทฤษฎีโครงสร้างสามมิติของสมอง (Three Faces of Intellect Model)

ทฤษฎีไฮเอราชิคอล (Hierachical Theory)

ทฤษฎีพัฒนาการทางสติปัญญาของเพียเจท์กับการคิดหาเหตุผล

ความหมายของความถนัด

ความถนัดด้านเหตุผลและการทดสอบความถนัดด้านเหตุผล

ทฤษฎีการทดสอบดั้งเดิม

โมเดลของทฤษฎีการทดสอบดั้งเดิม

ข้อตกลงเบื้องต้น

ข้อสรุปที่ได้จากทฤษฎีการทดสอบดั้งเดิม

การวิเคราะห์คุณภาพของแบบทดสอบ

การวิเคราะห์คุณภาพของข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ

หลักการของทฤษฎีการตอบสนองข้อสอบ

ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

พารามิเตอร์ของทฤษฎีการตอบสนองข้อสอบ

โมเดลในทฤษฎีการตอบสนองข้อสอบ

ฟังก์ชันสารสนเทศของแบบทดสอบ

การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบ

สรุปแนวคิดทฤษฎีการตอบสนองข้อสอบ

งานวิจัยที่เกี่ยวข้อง

ทฤษฎีและความเชื่อเกี่ยวกับสมอง

ทฤษฎีองค์ประกอบเดียว (Uni Factor Theory)

ทฤษฎีนี้มีความเชื่อว่าสมรรถภาพทางสมองของมนุษย์ ประกอบด้วยภาคเดียว ส่วนเดียวหรือที่เรียกว่าองค์ประกอบทั่ว ๆ ไปหรือเป็นความสามารถทั่ว ๆ ไป (General Ability) จะเป็นผู้สั่งงานในการประกอบกิจกรรมทุกอย่าง ผู้สร้างทฤษฎีนี้คือ บีเนท์และซิมอน (Binet and Simon, 1905. อ้างอิงถึงใน ล้วนและอังคณา สายยศ, 2541 : 27) ซึ่งบีเนท์และซิมอนได้เน้นว่าองค์ประกอบทั่ว ๆ ไปของสมองนั้น เป็นองค์ประกอบย่อย ๆ หลายองค์ประกอบที่สลับซับซ้อน ดังนั้นในการสร้างแบบทดสอบ บีเนท์และซิมอน จึงวัดความสามารถที่ได้จากการวัดเป็นคะแนนความสามารถโดยรวม (ทองห่อ วิภาวัน, 2523 : 20)

ทฤษฎีสององค์ประกอบ (The Two Factor Theory)

ทฤษฎีนี้เสนอโดยนักจิตวิทยาชาวอังกฤษชื่อ ชาร์ล สเปียร์แมน (Charles Spearman) เป็นทฤษฎีที่เกิดจากการวิเคราะห์คุณลักษณะโดยกระบวนการทางสถิติ พบว่ากิจกรรมทางสมองมีองค์ประกอบอยู่ 2 ส่วน คือองค์ประกอบร่วมเรียกว่าองค์ประกอบทั่วไป (General Factor หรือ G – Factor) เป็นความสามารถพื้นฐานของสมองและความสามารถทั่ว ๆ ไปในการประกอบกิจกรรม แต่ละคนจะมีความสามารถพื้นฐานมากน้อยแตกต่างกันและองค์ประกอบเฉพาะ (Specific Factor หรือ S – Factor) เป็นความสามารถเฉพาะลงไปในกิจกรรมอย่างเดียวที่ทำให้มนุษย์แตกต่างกัน เป็นความสามารถพิเศษที่มีอยู่ในแต่ละบุคคล ทฤษฎีนี้มองความสำคัญที่องค์ประกอบทั่วไปเป็นสำคัญ ไม่แตกต่างกับทฤษฎีองค์ประกอบเดียวมากนักเพียงแต่มีองค์ประกอบย่อยเพิ่มขึ้นมาอีก (ล้วน และอังคณา สายยศ, 2541 : 28)

ทฤษฎีหลายองค์ประกอบ (Multiple Factor Theory)

ทฤษฎีนี้เป็นที่ยอมรับกันอย่างกว้างขวางของนักจิตวิทยาชาวอเมริกัน ผู้นำในการสร้างทฤษฎีนี้ คือ เทอร์สโตน (Thurstone) ซึ่งเทอร์สโตนได้ใช้หลักวิธีการวิเคราะห์องค์ประกอบ (Factor Analysis) ทำให้สามารถแยกแยะความสามารถของสมองเป็นส่วนย่อย ๆ ประกอบกันเป็นกลุ่มหลายกลุ่ม แต่ละกลุ่มมีหน้าที่เฉพาะอย่างไปหรือทำงานร่วมกันบ้างก็ได้ องค์ประกอบย่อย ๆ นี้ เทอร์สโตน ให้ชื่อว่าสมรรถภาพพื้นฐานทางสมอง (Primary Mental Abilities) แยกองค์ประกอบย่อย ๆ โดยยึดน้ำหนักขององค์ประกอบเป็น 7 สมรรถภาพ คือ (สมบูรณ์ ชิตพงศ์ และสำเริง บุญเรืองรัตน์, 2541 : 29)

1. สมรรถภาพทางด้านภาษา (Verbal Factor) หรือ V – Factor เป็นสมรรถภาพในการเข้าใจคำศัพท์ ข้อความ บทกวี หรือเรื่องราวต่าง ๆ ในด้านภาษาและการเลือกใช้ภาษาได้อย่างเหมาะสม

2. สมรรถภาพทางด้านตัวเลข หรือคณิตศาสตร์ (Number Factor) หรือ N – Factor เป็นสมรรถภาพในการเข้าใจและคิดคำนวณตัวเลขทางคณิตศาสตร์

3. สมรรถภาพทางด้านความจำ (Memory Factor) หรือ M – Factor เป็นสมรรถภาพในการระลึกหรือจดจำเหตุการณ์หรือเรื่องราวต่าง ๆ ได้อย่างถูกต้องแม่นยำ

4. สมรรถภาพทางด้านความคล่องแคล่วในการใช้คำ (Word Fluency) หรือ W –Factor เป็นสมรรถภาพในการใช้ถ้อยคำต่าง ๆ ได้อย่างรวดเร็วถูกต้อง

5. สมรรถภาพทางด้านเหตุผล (Reasoning Factor) หรือ R – Factor เป็นสมรรถภาพในการจัดประเภทอุปมาอุปไมย และสรุปความได้อย่างมีเหตุผล

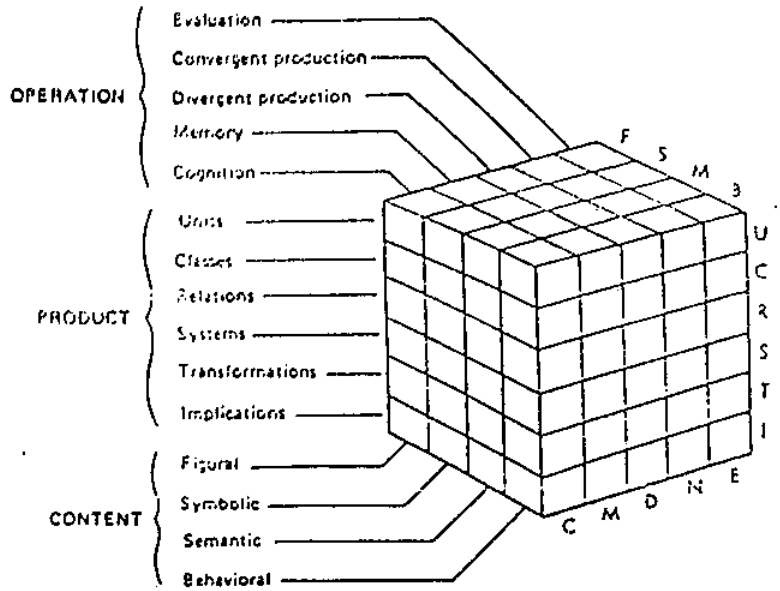
6. สมรรถภาพทางด้านมิติสัมพันธ์ (Spatial Factor) หรือ S – Factor เป็นสมรรถภาพในการที่จะมองเห็นความสัมพันธ์ด้านมิติ (Space) ต่าง ๆ ได้

7. สมรรถภาพทางด้านการรับรู้ (Perceptual Factor) หรือ P – Factor เป็นสมรรถภาพในการที่จะรับรู้ได้อย่างรวดเร็วและถูกต้อง สามารถมองเห็นรายละเอียดต่าง ๆ ได้

สมรรถภาพสมองทั้ง 7 ประการนี้ ถือว่าเป็นสมรรถภาพพื้นฐานทางสมองที่มีอยู่ในมนุษย์ทุกคน แต่มากน้อยแตกต่างกันออกไปในแต่ละด้านตามแต่ละบุคคล แม้แต่ในบุคคลเดียวกันก็ยังมีสมรรถภาพแต่ละด้านไม่เท่ากัน การที่แต่ละคนมีสมรรถภาพที่แตกต่างกันเช่นนี้จะเป็นดัชนีชี้ให้เห็นถึงความแตกต่างกันแต่ละบุคคล (Individual Differences) ซึ่งมีความสำคัญต่อการเรียนและการประกอบอาชีพให้สอดคล้องกับความสามารถของแต่ละบุคคล

ทฤษฎีโครงสร้างตามมิติของสมอง (Three Faces of Intellect Model)

ทฤษฎีนี้สร้างขึ้นโดย กิลฟอร์ด (Guildford) นักจิตวิทยาชาวอเมริกันได้ทำวิจัยเกี่ยวกับความถนัด แล้วเสนอทฤษฎีเกี่ยวกับสมรรถภาพทางสมองว่า โครงสร้างและองค์ประกอบของสมรรถภาพทางสมอง สามารถมองได้ในรูป 3 มิติ (Three Faces of Human Intellect) ดังภาพประกอบ 1 ตามลักษณะโครงสร้างของสมองตามภาพนี้ประกอบด้วย 3 มิติดังนี้ (สมบูรณ์ ชิตพงศ์และสำเริง บุญเรืองรัตน์, 2524 : 8-9)



ภาพประกอบ 1 โครงสร้างทางสมองตามทฤษฎีของกิลฟอร์ด

1. มิติแรก วิธีการคิดหรือการปฏิบัติ (Operation) ประกอบด้วย 4 ลักษณะดังนี้

1.1 การรู้ (Cognition) หมายถึง ความสามารถทางสมองของบุคคลที่จะรู้จักหรือค้นพบ และเข้าใจสิ่งต่าง ๆ

1.2 การจำ (Memory) หมายถึง ความสามารถที่จะคิดหรือตอบสนองต่อสิ่งต่าง ๆ ได้

1.3 การคิดอเนกนัย (Divergent Thinking) หมายถึง ความสามารถที่จะคิดหรือกระทำตอบต่อสิ่งต่าง ๆ ได้หลายทาง และเป็นความสามารถที่คิดกระทำตอบนั้นในทางที่ดีที่สุด

1.4 การคิดเอกนัย (Convergent Thinking) หมายถึง ความสามารถของมนุษย์ที่คิดกระทำตอบต่อสิ่งต่าง ๆ ในทางที่ดีและถูกต้องสมเหตุสมผลที่สุด

1.5 การประเมินผล (Evaluation) หมายถึง ความสามารถของมนุษย์ในการพิจารณาตัดสินหรือประเมินคุณค่าสิ่งต่าง ๆ ได้ถูกต้อง ดี สมเหตุผลและเป็นไปตามกฎเกณฑ์ต่าง ๆ

2. มิติ 2 เนื้อหา (Content) หมายถึง สิ่งต่าง ๆ ที่ก่อให้เกิดความคิดหรือความรู้สึกอาจเป็นสิ่งของ เรื่องราว สัญลักษณ์หรือเหตุการณ์ต่าง ๆ มิติเนื้อหาแบ่งได้เป็น 4 ประเภท ดังนี้

2.1 ภาพ (Figural) หมายถึง สิ่งที่เป็นรูปธรรมต่าง ๆ มีโครงสร้างที่สามารถมองเห็นหรือให้ความรู้สึกได้ เช่น บ้าน รถ นก ภาพ แสง เสียง ความร้อน เป็นต้น

2.2 สัญลักษณ์ (Symbolic) หมายถึง เครื่องหมายหรือสัญญาณต่าง ๆ เช่น ตัวเลข ตัวอักษร สัญลักษณ์ ชาตุ โน้ตดนตรี สัญญาณจราจร เป็นต้น

2.3 ภาษา (Semantic) หมายถึง ถ้อยคำต่าง ๆ

2.4 พฤติกรรม (Behavioral) หมายถึง การกระทำหรือการแสดงออกต่าง ๆ ของบุคคล เช่น ความต้องการ ทัศนคติ อารมณ์ เป็นต้น

3. มิติ 3 ผลของการคิด (Product) หมายถึง ผลที่เกิดจากการที่มนุษย์คิดในสิ่งต่าง ๆ แล้วสามารถจัดเป็นระบบเป็นเหล่า เป็นพวก หรือสามารถดัดแปลงปรับปรุงนำไปใช้หรือสร้างสรรค์สิ่งต่าง ๆ ออกมาใช้ได้ มิตินี้แบ่งออกเป็น 6 ประเภท ดังนี้

3.1 หน่วย (Unit) หมายถึง สิ่งย่อยที่สุดของสิ่งต่าง ๆ ที่มีคุณสมบัติเฉพาะตัวไม่เหมือนกับสิ่งอื่น ๆ เช่น นกฮูก นกเอี้ยง นกพิราบ เป็นต้น

3.2 จำพวก (Classes) หมายถึง กลุ่มต่าง ๆ ที่มีคุณสมบัติร่วมกัน เช่น นกฮูก นกเอี้ยง นกพิราบ ก็จัดได้เป็นสัตว์จำพวกนก

3.3 ความสัมพันธ์ (Relation) หมายถึง ความเชื่อมโยงของสิ่งต่าง ๆ เช่น พ่อคู่กับแม่ ชายคู่ กับหญิง ความสัมพันธ์ของสิ่งเหล่านี้คือ การมีเพศต่างกัน

3.4 ระบบ (System) หมายถึง กลุ่มของสิ่งต่าง ๆ ที่เชื่อมโยงกันโดยมีหลักเกณฑ์รวมกันอยู่อย่างหนึ่ง เช่น เลขชุด 1 3 5 7 9 จัดเป็นระบบเลขคี่ 2 6 8 10 จัดเป็นระบบเลขคู่ เป็นต้น

3.5 การแปลงรูป (Transformation) หมายถึง การเปลี่ยนแปลง แก้ไข ปรับปรุง ดัดแปลงสิ่งต่าง ๆ ออกมาในรูปแบบใหม่ เช่น การแปลความ ย่อความ เป็นต้น

3.6 การประยุกต์ (Implication) หมายถึง การคาดคะเน การคาดหวัง การทำนายจากสิ่งหรือเหตุการณ์ต่าง ๆ ว่าจะมีสิ่งใดเกิดขึ้นตามมา

ตามลักษณะโครงสร้างทางสมอง 3 มิตินี้ ปรากฏว่าสมองคนเรามีอยู่ 120 ช่อง แต่ละช่องจะเป็นตัวแทนของสมรรถภาพสมอง 1 สมรรถภาพ แต่ละสมรรถภาพนั้นสามารถอธิบายด้วยมิติทั้ง 3 ได้เสมอ เช่น

D M R หมายถึง ความสามารถด้านการคิดอเนกนัยโดยใช้ภาษาแบบความสัมพันธ์

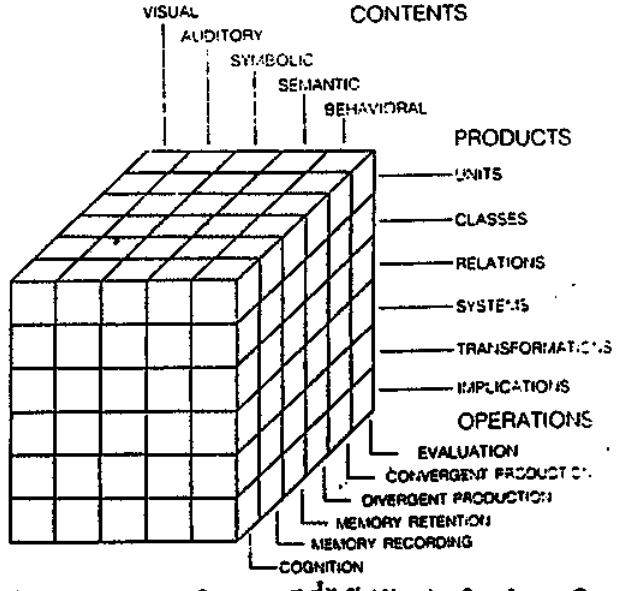
C F U หมายถึง ความสามารถด้านการรู้โดยใช้ภาพแบบเป็นหน่วย เป็นต้น

Operation ใช้ตัวย่อแรก	Content ตัวที่ 2	Product ตัวที่ 3
C- Cognition	F- Figural	U-Unit
M- Memory	S- Symbolic	C- Class
D – Divergent Product	M- Sembolic	R- Relation
N- Convergent Product	B- Behavioral	S- System
E- Evaluation		T-Transformation
		I- Implication

ภาพประกอบ 2 ตารางโครงสร้างทางสมองสัญลักษณ์ย่อของมิติ 3 มิติ

ในปี 1988 กิลฟอร์ด(Guild,1988 อ้างถึงในถ้วนและอังคณา สายยศ,2541:51)ได้เสนอบทความ Some change in the structur-of-intellect Model โดยเพิ่มเนื้อหาเป็น5 อย่างโดยมี Figural แล้วแตกเป็น Visual กับ Auditory Visual เป็นความสามารถในการมองเห็น ส่วน Auditory เป็นความสามารถในการรับรู้ทางการได้ยิน

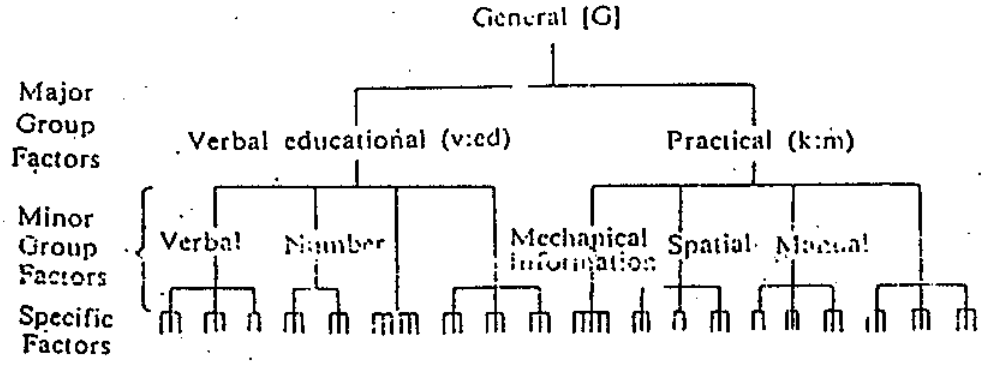
ด้าน Operations เดิมมี 5อย่างเพิ่มเป็น 6อย่าง โดยแยกความจำ(Memory)ออกเป็น 2 อย่างคือ Memory Recording ซึ่งหมายถึงความจำในช่วงสั้น (Short-term memory)นั่นเอง ส่วนความจำอีกอย่างหนึ่งคือ Memory Retention เป็นความจำที่ทิ้งช่วงนั่นคือเป็นเวลาในการจำนานๆ นั่นเอง ดังนั้น Micromodel ของทฤษฎีกิลฟอร์ดอันใหม่ก็จะมีจำนวน 5 x 6 x 6 =180 หน่วย จะวัดเชาวน์ปัญญาให้ครอบคลุมจะต้องสร้างเครื่องมือวัดให้ครอบคลุมทั้ง 180 องค์ประกอบซึ่งในทางปฏิบัติไม่สามารถสอบได้หมด



ภาพประกอบ 3 แนวคิดทฤษฎีที่ได้ปรับปรุงใหม่ของ กิลฟอร์ด

ทฤษฎีไฮเอราซิคอล (Hierachical Theory)

มีนักจิตวิทยาชาวอังกฤษกลุ่มหนึ่ง ได้จัดรูปแบบการประกอบกันขององค์ประกอบอีกรูปแบบหนึ่ง กลุ่มนี้คือ เบิร์ท (Burt) เวอร์นอน (Vernon) และฮัมเฟรย์ (Humphreys) โดยเฉพาะเวอร์นอน (Vernon) ได้เสนอโครงสร้างของเชาว์ปัญญาในปี ค.ศ. 1960 โดยเริ่มอธิบายตามแบบของสเปียร์แมน (ล้วน และอังคณา สายยศ, 2527 : 30-31) นั่นคือ เวอร์นอนเริ่มจุดแรกด้วยความสามารถทางสมองทั่วไป ที่เรียกว่า Major Group Factor (G-Factor) ขั้นต่อไปแบ่งออกเป็น 2 องค์ประกอบใหญ่ ๆ (Major Group Factor) คือองค์ประกอบด้านความถนัดทางการเรียน เรียกว่า Verbal Education (V : Ed) และองค์ประกอบด้านความถนัดทางอาชีพ เรียกว่า Practical Mechanical (K : m) ซึ่งองค์ประกอบใหญ่ 2 องค์ประกอบนี้ยังแบ่งเป็นองค์ประกอบย่อย (Minor Group Factor) ลงไปอีก เช่นองค์ประกอบด้าน Verbal Education (V : Ed) ประกอบด้วยความสามารถด้านภาษา (Verbal) ความสามารถด้านตัวเลข (Numerical) และอื่น ๆ ในทำนองเดียวกัน องค์ประกอบด้าน Practical Mechanical แบ่งย่อยออกเป็น Mechanical Information, Spatial และ Manual เป็นต้น ระดับที่ต่ำสุดขององค์ประกอบในรูปแบบนี้ ยังมีองค์ประกอบย่อย ๆ ไปอีก เรียกว่าองค์ประกอบเฉพาะ (Specific Factor) ถ้าพิจารณาดูโครงสร้างอันนี้แล้วไม่ต่างอะไรกับลักษณะของต้นไม้ที่แผ่กิ่งก้านใหญ่เล็กลงไปตามลำดับ ดังภาพประกอบ 4



ภาพประกอบ 4 โครงสร้างทฤษฎีไฮเอราซิคอล (Hierachical Theory)

6. ทฤษฎีพัฒนาการทางสติปัญญาของเพียเจท์กับการคิดหาเหตุผล

เพียเจท์ (De Cecco, 1967 : 264 – 265 Citing Piaget) แบ่งลำดับขั้นของการพัฒนาการทางสติปัญญาของเด็ก ซึ่งเกี่ยวข้องกับการคิดหาเหตุผลไว้สี่ลำดับขั้น คือ

1. ขั้นการคิดก่อนสร้างสังกัป พบในเด็กอายุ 2 – 4 ปี เด็กในขั้นนี้จะเรียนรู้โดยการกระทำ

ที่ซ้ำ ๆ กัน โดยจะใช้ประสาทสัมผัส แต่ก็ยังไม่สามารถสร้างสังกัปได้ จึงไม่รู้จักใช้สัญลักษณ์เด็กจะสร้างแบบแผนการรับรู้ขึ้นเอง เช่น เด็กอายุ 3 ขวบ ยังไม่มีความเข้าใจในเรื่องการเท่ากัน การจัดลำดับสิ่งของที่มีค่ากลาง ๆ การคิดหาเหตุผลจะมีลักษณะที่คิดจากส่วนย่อยสู่ส่วนย่อย ดังตัวอย่าง ก ชอบ ข แล้ว ข ต้องชอบ ก ด้วย

2. ขั้นการคิดโดยใช้ญาณ พบในเด็กอายุระหว่าง 4 – 7 ปี เด็กเริ่มมีการคิดหาเหตุผลแต่มีอย่างจำกัด และยังไม่อิสระ เด็กเริ่มเข้าใจสัญลักษณ์ และสามารถสร้างสังกัปได้บ้างแล้ว

3. ขั้นการคิดด้วยรูปธรรม พบในเด็กอายุ 7 – 11 ปี ในขั้นนี้อย่างน้อยที่สุดจะมีการคิดหาเหตุผลที่สอดคล้องตามหลักตรรกศาสตร์ที่เพียเจท์ เรียกว่า การใช้ตัวปฏิบัติการ เด็กจะรวบรวมสังกัปและนำมาใช้ สามารถคิดหาเหตุผลถูกต้องตามหลักตรรกศาสตร์เชิงคณิตศาสตร์ได้

4. ขั้นการคิดตามหลักตรรกศาสตร์ พบในเด็กอายุ 11 – 15 ปี เด็กจะนำความคิดที่ถูกต้องตามหลักตรรกศาสตร์ไปใช้แก้ปัญหาในสถานการณ์ที่เป็นรูปธรรม มีการคิดที่อิสระขึ้น และที่สำคัญก็คือจะมีการคิดย้อนกลับอย่างสมบูรณ์ และยังสามารถคิดในสิ่งที่เป็นนามธรรมได้ สามารถสรุปและอนุมานจากสมมติฐานได้ และสามารถหาความรู้ในลักษณะที่เป็นวิทยาศาสตร์ โดยสอดคล้องตามหลักตรรกศาสตร์ได้ จึงนับได้ว่า เด็กในขั้นนี้มีการคิดหาเหตุผลสูงสุด

ดี เซคโค (De Cecco, 1967 : 265 – 266) กล่าวถึง พัฒนาการทางสติปัญญาตามทฤษฎีของเพียเจท์ว่าในขั้นการคิดด้วยรูปธรรม เด็กมีการคิดหาเหตุผลแล้ว และสิ่งที่ปรากฏมาขั้นนี้ ได้แก่

1. การจำแนกประเภท เป็นความสามารถที่จะจำแนกวัตถุสิ่งของเป็นหมวดหมู่ได้

2. การหาความสัมพันธ์ เป็นความสามารถที่คิดหาความสัมพันธ์ที่ไม่สมมาตรได้ เช่น ก ยาวกว่า ข ดำเป็นพ่อแดง เป็นต้น

3. การเข้าใจเกี่ยวกับจำนวน เป็นผลร่วมของความสามารถในการจำแนกประเภทและการหาความสัมพันธ์

ด้านความสามารถในการจัดเป็นหมวดหมู่ในขั้นการคิดด้วยรูปธรรม เด็กต้องอาศัยตัวปฏิบัติการดังนี้

1. ตัวปฏิบัติการเกี่ยวกับการจำแนกประเภท เป็นความสามารถในการใช้ตัวปฏิบัติการเกี่ยวกับความสัมพันธ์ระหว่างส่วนรวมและส่วนย่อย

2. ตัวปฏิบัติการเกี่ยวกับการจัดเรียงลำดับ เป็นความสามารถในการสรุปเกี่ยวกับการจัดเรียงวัตถุสิ่งของที่ตั้งอยู่บนรากฐานของความสัมพันธ์เชิงตรรกศาสตร์

3. การเข้าใจโครงสร้างเกี่ยวกับรูปธรรม เป็นความสามารถที่ตั้งอยู่บนพื้นฐานของตรรกศาสตร์ เกี่ยวกับการจำแนกประเภทและความสัมพันธ์ ซึ่งกลุ่มวัยรุ่นจะมีความสามารถเหล่านี้มาก (Inhelder and Piaget, 1958 : 8 – 16)

ความถนัดและการวัดความถนัด

ความหมายของความถนัด

คำว่า Aptitude รากศัพท์มาจาก Aptos ซึ่งเป็นภาษากรีกแปลว่า ความเหมาะสมกับ (Fitted For) (ทองห่อ วิภาวิน, 2523 : 14)

สมศักดิ์ สินธุระเวชญ์ (2526 : 1) ได้ให้ความหมายว่าความถนัด หมายถึงสมรรถวิสัยและทิศทางแห่งความงอกงามของสมอง หรือขีดระดับความสามารถขั้นสูงสุดของบุคคลที่เขาอาจมีได้ต่อการเรียนรู้และการฝึกฝนในวิทยาการและทักษะต่าง ๆ ถ้าหากเขาได้รับการสอน การฝึกและประสบการณ์ที่เหมาะสม และยังได้ให้แนวคิดเกี่ยวกับความหมายของความถนัดดังนี้

1. ความถนัดไม่หมายถึงความรู้ "ความรู้" ในที่นี้หมายถึง การรักษาไว้ซึ่งเรื่องราวต่างๆ และประสบการณ์ทั้งปวงที่บุคคลได้เคยผ่านมาในอดีต และความถนัดมิใช่หมายถึงความจำและมิได้วัดด้วยการให้ระลึกออกมาตรง ๆ แต่ต้องเป็นพลังงานหรือสัมฤทธิ์ผลทางสมองที่สามารถเอาการฟัง และการเห็นมาผสมเคล้ากันจนกลายเป็นสมรรถภาพถูกใหม่ที่มีสรรพคุณใหม่

2. ความถนัดไม่ได้หมายถึงความเร็วหรือเอาความเร็วเป็นปัจจัยสำคัญ แต่เอาระดับความยากของงานนั้นเป็นหลักในการพิจารณา คือถ้าใครสามารถทำงานที่ยากและซับซ้อนได้ถูกต้องมาก เรียกว่าเป็นคนฉลาดมาก มีปัญญามาก

3. ความถนัดไม่ได้หมายถึงกรรมพันธุ์ วิทยาการปัจจุบันสามารถพิสูจน์ว่ากรรมพันธุ์สามารถส่งผลต่อโครงสร้างและเนื้อเยื่อของร่างกายได้จริง แต่ไม่มีการรับรองในเรื่องการสืบเชื้อสายทางสติปัญญาและความเฉลียวฉลาด การที่เชื่อว่าเด็กในครอบครัวหนึ่งมีหัวดีหรือปัญญาดีเหมือนพ่อแม่นั้นยังไม่ถูกต้องนัก เป็นเพราะคตินี้ลืมเรื่องการฝึกฝนกับโอกาสที่จะได้รับการฝึกฝนและสิ่งแวดล้อม ทั้ง ๆ ที่สามประการนี้มีอิทธิพลในการหล่อหลอมชีวิตมนุษย์เป็นอันมาก

4. ความถนัดไม่หมายถึงสมรรถภาพชนิดเดียว แต่เดิมมีความเชื่อว่าสมองมนุษย์เป็นสิ่งเดียวกันและสามารถแสดงถึงพฤติกรรมออกมาได้หลายด้าน แต่ต่อมาเชื่อว่าสมองมนุษย์มีการทำงานที่สัมพันธ์สืบเนื่องกันตลอดเวลา ปัจจุบันวิทยาการด้านวัดผล เช่น วิธีวิเคราะห์องค์ประกอบ (Factor Analysis) สามารถพิสูจน์จากความจริงได้ว่าสมองมนุษย์ประกอบด้วยสมรรถภาพเฉพาะ เรื่องราวหลายชนิดโดยแต่ละชนิดมีสรรพคุณแตกต่างกันไปและคนที่มีสมรรถภาพด้านใดด้านหนึ่งเป็นพิเศษนั้นจะมีความสามารถอื่น ๆ มาประกอบด้วยเสมอ ดังนั้นการใช้แบบทดสอบเพื่อวัดความถนัดจึงควรมีการทดสอบหลาย ๆ ด้านของสมอง

5. ความถนัดไม่หมายถึงพรหมลิขิต หมายถึงว่าความไง่หรือฉลาด ไม่ได้เป็นบุญวาสนาที่ติดมาแต่กำเนิด แต่เชื่อว่าพฤติกรรมของสิ่งมีชีวิตจะแก่กล้าขึ้นก็ด้วยอาศัยสิ่งภายนอกมารบเร้าและตัวเองทำปฏิกิริยาโต้ตอบออกไปเท่านั้น พวกเชื่อความถนัดมองเห็นว่ามนุษย์ไม่ใช่ขอนไม้แต่สามารถที่จะหล่อหลอมปรับปรุงแก้ไขตัวเองได้

ล้วน และอังคณา สายยศ (2527 : 16) ได้ให้ความหมายของความถนัดว่า หมายถึงความสามารถที่บุคคลได้รับประสบการณ์ ฝึกฝนตนเองและมีการสั่งสมไว้มาก จนเกิดเป็นทักษะพิเศษเด่นชัดด้านใดด้านหนึ่งพร้อมที่จะปฏิบัติกิจกรรมนั้นได้อย่างดี

สมบูรณ์ ชิตพงศ์ และสำเริง บุญเรืองรัตน์ (2524 : 16) ได้ให้ความเห็นเกี่ยวกับความถนัดว่า ปราชญ์ทางการวัดผลทั้งหลายต่างมีความเชื่ออยู่ว่า อินทรีย์จะแก่กล้าขึ้นต่อเมื่ออินทรีย์นั้นได้รับการเร้าจากสิ่งแวดล้อมทั้งหลาย อันหมายถึงประสบการณ์ต่าง ๆ ที่ผ่านเข้ามาในชีวิตของบุคคลและบุคคลนั้นจะทำการตอบสนองต่อสิ่งแวดล้อมนั้น ๆ ยิ่งได้มีการตอบสนองอยู่เสมอ ๆ เชาว์ภายในสมองก็ได้รับการยั่วยุให้แก้ปัญหาบ่อย ๆ ผลการตอบสนองทำให้บุคคลนั้นเป็นผู้มีความสามารถที่มีสมรรถภาพที่แก่กล้าขึ้น เมื่อเป็นเช่นนี้พวกที่เชื่อทฤษฎีความถนัดจึงไม่เชื่อว่าสมรรถภาพของมนุษย์มีขีดจำกัดตายตัว แต่เชื่อว่าสมรรถภาพมนุษย์สามารถที่จะพัฒนาขึ้นได้เรื่อย ๆ

บุญส่ง นิลแก้ว (2519 : 177) ได้ให้ความหมายว่า ความถนัดเป็นสมรรถภาพหรือศักยภาพที่มีอยู่ในตัวบุคคลอันจะก่อให้เกิดความสำเร็จในการทำกิจกรรมใด ๆ

วอเร็น (Warren, อ้างถึงในล้วน และอังคณา สายยศ (2527 : 25) ให้ความหมายความถนัดว่า หมายถึงสภาวะหรือคุณลักษณะที่แสดงถึงความสามารถของแต่ละบุคคลที่จะเรียนรู้ เกิดทักษะหรือมีปฏิกิริยาตอบสนองเฉพาะอย่างเมื่อได้รับการฝึกหัด

ฟรีแมน (Freeman, อ้างถึงใน ล้วน และอังคณา สายยศ (2527 : 25) ได้ให้นิยามเชาว์ปัญญาในรูปความถนัดไว้ว่า ความถนัดเป็นวิสัยสามารถที่ได้เรียนรู้อย่างชำนาญจากการฝึกอบรมทั้งในระบบและนอกระบบมาจำนวนหนึ่ง และนิยามความถนัดทางการเรียน (Scholastic Aptitude) ว่าเป็นคุณลักษณะพิเศษส่วนบุคคลที่ทำให้ผู้นั้นมีความสำเร็จทางวิชาการ

ส่วนครอนบัช (Cronbach อ้างถึงใน ล้วน และอังคณา สายยศ (2527 : 25) ได้ให้ความหมายของความถนัดทางการเรียนว่า เป็นกลุ่มความสามารถทางสมองที่ร่วมกันทำงานเพื่อเพิ่มพูนความสำเร็จในกิจกรรมทางปัญญา

นอกจากนี้ยังมีผู้ให้ความหมายของความถนัดอีกมากมาย แต่โดยสรุปความถนัด หมายถึง ความสามารถและความพร้อมที่เกิดขึ้นจากการพัฒนา ฝึกฝน การได้รับประสบการณ์ที่บุคคลนั้นได้รับหรือเกิดขึ้น ทำให้บุคคลมีระดับความสามารถและศักยภาพในด้านการเรียนรู้และการปฏิบัติที่แตกต่างกัน

ความถนัดด้านเหตุผลและการทดสอบความถนัดด้านเหตุผล

ความถนัดด้านเหตุผล หมายถึง การคิดอย่างมีวิจารณญาณแล้ววินิจฉัยลงสรุปอย่างถูกต้องในแบบทดสอบมาตรฐานนั้นการวัดเหตุผลมักจะอยู่ในกลุ่มความสามารถใหญ่ ๆ 3 ด้าน คือด้านความสามารถในการจำแนกประเภท(Classification) ด้านความสามารถในการหาความสัมพันธ์หรืออุปมาอุปไมย (Analogy)และความสามารถ ในการสรุปอ้างอิงโดยหลักการตรรกะ (Inference) (ล้วน สายยศ และ อังคณา สายยศ,2541:30)

ความถนัดด้านเหตุผล หรือ ความสามารถในการใช้วิจารณญาณ เป็นความสามารถที่สำคัญของการเรียนทุกชนิด และทุกระดับชั้น เป็นการแสดงออกโดยการหาคำตอบหรือข้อยุติที่เหมาะสมที่สุดในการเปรียบเทียบหลาย ๆสิ่งพร้อมกันเช่นความสามารถในการจัดประเภท การอุปมาอุปไมย และการสรุปความเป็นต้น (ทองห่อ วิภาวิน,2523:50-69)

องค์ประกอบด้านเหตุผล (Reasoning Factor)เรียกย่อๆ ว่า R-Factor เป็นองค์ประกอบด้านหนึ่งที่นิยมใช้วัดความถนัดทางการเรียน ในการคิดหาเหตุผลความสัมพันธ์ในรูปแบบต่างๆกัน การมองความสำคัญ ความสามารถในการจัดประเภท แยกประเภทในอุปมาอุปไมยได้ถูกต้อง สามารถสรุปความจากสถานการณ์ที่กำหนดให้ได้อย่างสมเหตุสมผลและถูกต้อง(ชาญวิทย์ เทียมบุญประเสริฐ,2534:60)ซึ่งมีผู้กล่าวถึงรูปแบบของแบบทดสอบที่ใช้ไว้ดังนี้

ล้วน สายยศ และ อังคณา สายยศ(2541:107)แบ่งลักษณะแบบทดสอบความถนัดด้านเหตุผลไว้ 6 แบบ

1. การจำแนกประเภท (Classification)
2. การอุปมาอุปไมย (Analogy)
3. แบบอนุกรมภาพหรืออนุกรมมิติ
4. แบบทดสอบสรุปความ
5. ตัวร่วมตัวต่าง
6. ความสามารถด้านวิเคราะห์

JOHN F. KENNEDY LIBRARY
PRINCE OF SONGKLA UNIVERSITY
PATTANI THAILAND

วัญญา วิศาลาภรณ์(2525:99)แบ่งลักษณะแบบทดสอบความถนัดด้านเหตุผลไว้ 7 แบบ

1. ข้อสอบการหาเหตุผลแบบต่างประเภท
2. ข้อสอบการหาเหตุผลแบบจัดเข้าพวก
3. ข้อสอบการหาเหตุผลแบบไม่เข้าพวก
4. ข้อสอบการหาเหตุผลแบบอุปมาอุปไมย
5. ข้อสอบการหาเหตุผลแบบสรุปความ
6. ข้อสอบการหาเหตุผลแบบเชิงนามธรรม
7. ข้อสอบการหาเหตุผลแบบอนุกรมมิติ

วิเชียร เกตุสิงห์ (2517:57) แบ่งลักษณะแบบทดสอบความถนัดด้านเหตุผลไว้ 4 แบบ

1. แบบอุปมาอุปไมย
2. แบบจัดประเภท
3. แบบสรุปความ
4. แบบเรียงลำดับ

สมบูรณ์ ชิตพงษ์และสำเริง บุญเรืองรัตน์ (2524:31-45) แบ่งลักษณะแบบทดสอบความถนัดด้านเหตุผลไว้ 5 แบบ

1. แบบทดสอบด้านการจัดเข้าพวก
2. แบบทดสอบด้านอุปมาอุปไมย
3. แบบทดสอบสรุปความ
4. แบบทดสอบอักษรเรียงลำดับ
5. แบบทดสอบเรียงลำดับภาพ แบ่งเป็นความสัมพันธ์ทางเดียวและแบบความสัมพันธ์หลายๆทาง

สมศักดิ์ สินธุระเวชญ์ (2526:103) แบ่งลักษณะแบบทดสอบความถนัดด้านเหตุผลไว้ 6 แบบ

1. แบบอุปมาอุปไมย
2. แบบจัดเข้าพวก
3. แบบไม่เข้าพวก
4. แบบความสัมพันธ์ทางเดียว

5. แบบอนุกรมมิติ

6. แบบสรุปความ

แบบทดสอบความถนัดตามองค์ประกอบด้านเหตุผลนี้ มีหลายรูปแบบแต่รูปแบบที่นิยมใช้กันมี 4 รูปแบบคือ (ชาญวิทย์ เทียมบุญประเสริฐ,2534:60)

1. การจัดประเภท (Classification)

1.1 แบบเข้าพวกภาษา

1.2 แบบเข้าพวกรูปภาพ

1.3 แบบไม่เข้าพวกภาษา

1.4 แบบไม่เข้าพวกรูปภาพ

2. อุปมาอุปไมย (Analogy)

2.1 อุปมาอุปไมยที่เป็นภาษา

2.2 อุปมาอุปไมยที่เป็นภาพ

3. สรุปความ (Inference)

4. อนุกรมภาพ

4.1 อนุกรมภาพทางเดียว

4.2 อนุกรมภาพสองทาง

จะเห็นได้ว่า การวัดสมรรถภาพสมองด้านเหตุผลนั้นมีผู้กล่าวถึงรูปแบบของแบบทดสอบความถนัดด้านเหตุผล ที่ใช้แตกต่างกัน ดังนั้นในงานวิจัยนี้ได้เลือกรูปแบบของแบบทดสอบความถนัดด้านเหตุผลที่นิยม
ใช้มา 9 รูปแบบคือ

1. แบบจัดประเภท ชนิดเข้าพวกภาษา

2. แบบจัดประเภท ชนิดเข้าพวกรูปภาพ

3. แบบจัดประเภท ชนิดไม่เข้าพวกภาษา

4. แบบจัดประเภท ชนิดไม่เข้าพวกรูปภาพ

5. แบบอุปมาอุปไมยภาษา

6. แบบอุปมาอุปไมยภาพ

7. แบบสรุปความ

8. แบบอนุกรมภาพทางเดียว

9. แบบอนุกรมภาพสองทาง

ทฤษฎีการทดสอบแบบดั้งเดิม

ทฤษฎีการทดสอบแบบดั้งเดิม เป็นองค์ความรู้ทั่วไปเกี่ยวกับการตรวจสอบความสัมพันธ์ระหว่างคะแนนที่สังเกตได้กับคะแนนจริง รวมทั้งแนวคิดการวิเคราะห์คุณภาพ โดยส่วนรวมของข้อสอบและแบบทดสอบสำหรับผู้สอบเฉพาะกลุ่ม โดยมีแนวคิดเกี่ยวกับโมเดลการวัด ข้อตกลงเบื้องต้นและการวิเคราะห์ ดังต่อไปนี้

โมเดลของทฤษฎีการทดสอบแบบดั้งเดิม

สเปียร์แมน (C. Spearman, 1904 อ้างถึงใน ศิริชัย กาญจนวาสี,2535: 99) ได้เสนอข้อคิดเชิงเหตุผลและข้อโต้แย้งทางคณิตศาสตร์ที่น่าตื่นเต้นว่า คะแนนจากแบบทดสอบมีความคลาดเคลื่อนไปจากคุณลักษณะที่แท้จริงของบุคคล ดังนั้นสหสัมพันธ์ระหว่างคะแนนจากแบบทดสอบมีค่าต่ำกว่าสหสัมพันธ์ระหว่างคะแนนที่แท้จริง ด้วยความพยายามที่จะอธิบายปรากฏการณ์ทางการทดสอบดังกล่าว สเปียร์แมน (1907, 1913อ้างถึงใน ศิริชัย กาญจนวาสี,2535: 99) ได้เสนอโมเดลคะแนนจริงแบบดั้งเดิม ไว้ว่า คะแนนที่ได้จากการวัดหรือสังเกต เกิดจากผลบวกขององค์ประกอบที่สังเกตไม่ได้ 2 ส่วน คือ คะแนนจริงและคะแนนความคลาดเคลื่อน ดังนั้น คะแนนที่สังเกตได้จากแบบทดสอบจึงมีความสัมพันธ์กับคะแนนจริงตามสมควร

$$X = T + E$$

(คะแนนจากแบบทดสอบ) (คะแนนจริง) (คะแนนนความคลาดเคลื่อน)

สมมติว่าในการทำข้อสอบ 10 ข้อ คะแนนเต็ม 10 คะแนน ก. มีความรู้จริงที่จะทำข้อสอบได้ 7 ข้อ แต่กาเครื่องหมายผิดพลาดทำให้ทำผิดไป 2 ข้อ ข. มีความรู้จริงที่จะทำข้อสอบได้ 4 ข้อ แต่โชคดีเดาถูกเพิ่มมาอีก 1ข้อ ส่วน ค.มีความรู้จริง 5 ข้อ แต่อ่านโจทย์เข้าใจผิดไป1 ข้อ ดังนั้น คะแนนที่สังเกตได้จากการทำแบบทดสอบของทั้ง 3 บุคคลเป็นดังนี้

$X_n = 7 - 2 = 5$

$X_u = 4 + 1 = 5$

$X_n = 5 + 0 = 5$

จากตัวอย่างแสดงให้เห็นว่า คะแนนที่ได้จากแบบทดสอบมีความคลาดเคลื่อนระดับหนึ่ง ผู้สอบบางคนมีความคลาดเคลื่อนของการวัดเป็นลบ บางคนมีความคลาดเคลื่อนเป็นบวก และบางคนอาจไม่มีความคลาดเคลื่อนของการวัดเกิดขึ้นเลยก็ได้

ข้อตกลงเบื้องต้น

ทฤษฎีการทดสอบแบบดั้งเดิมตั้งอยู่บนพื้นฐานของความเชื่อตามข้อตกลงเบื้องต้นเกี่ยวกับโมเดลการวัด ดังนี้

1.1 คะแนนที่ได้จากกการวัด มีความสัมพันธ์แบบเส้นตรงและเชิงบวกกับคะแนนจริง และคะแนนความคลาดเคลื่อน

โมเดล X = T+ E เป็นโมเดลเชิงเส้นตรง และเป็นโมเดลเชิงบวก การรวมกันของ T และ E จึงเป็นไปตามกฎการบวกทางคณิตศาสตร์

1.2 คะแนนจริงมีสภาวะคงที่ซึ่งเท่ากับค่าเฉลี่ยของคะแนนที่ได้จากการวัดซ้ำ ๆ หลาย ๆ ครั้ง

คะแนนของแต่ละกลุ่มมีสภาวะที่คงที่ และมีค่าเท่ากับค่าเฉลี่ยของ X ที่ได้จากการสอบซ้ำด้วย แบบทดสอบคู่ขนานจำนวน ∞ ครั้ง โดยการสอบแต่ละครั้งเป็นอิสระจากกัน

นั่นคือ $T_p = E(X_{pi})$

เมื่อ p = บุคคล

i = ข้อสอบ

1.3 คะแนนความคลาดเคลื่อนไม่มีความสัมพันธ์กับคะแนนจริง นั่นคือ

$$\rho_{E_i T_i} = 0 \text{ และ } \rho_{E_j T_j} = 0$$

ทฤษฎีการทดสอบแบบดั้งเดิมถือว่าคะแนนความคลาดเคลื่อนเป็นอิสระจากคะแนนจริง กล่าวคือ ผู้สอบที่มีคะแนนจริงสูงหรือต่ำ ย่อมไม่มีความสัมพันธ์กับคะแนนความคลาดเคลื่อนที่เกิดขึ้น และคะแนนความคลาดเคลื่อนจากแบบทดสอบฉบับหนึ่งจะไม่มีความสัมพันธ์กับคะแนนจริงจากการทำแบบทดสอบฉบับอื่น ๆ

1.4 คะแนนความคลาดเคลื่อนของบุคคลต่างกันหรือแบบทดสอบต่างชุด ไม่มีความสัมพันธ์ต่างกัน

นั่นคือ $\rho_{E_i E_j} = 0$

ข้อตกลงเบื้องต้นนี้ แสดงว่าความคลาดเคลื่อนของต่างบุคคลหรือต่างแบบทดสอบเป็นอิสระจากกันกล่าวคือ คะแนนความคลาดเคลื่อนของผู้สอบคนหนึ่งย่อมไม่มีความสัมพันธ์กับคะแนนความคลาดเคลื่อนของผู้สอบคนอื่นที่ทำแบบทดสอบเดียวกัน หรือถ้าผู้สอบคนหนึ่งทำแบบทดสอบฉบับที่ 1ย่อมมีความคลาดเคลื่อนเป็นอิสระจากการทำแบบทดสอบฉบับที่ 2

ข้อสรุปที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม

จากโมเดลการวัดและข้อตกลงเบื้องต้นของทฤษฎีการทดสอบแบบดั้งเดิม สามารถนำไปสู่ข้อสรุปต่าง ๆ ได้โดยถือว่าข้อตกลงเบื้องต้นเหล่านั้นเชื่อถือได้ ถ้าข้อตกลงเบื้องต้นนั้นสมเหตุสมผล ข้อสรุปต่อไปนี้ย่อมสะท้อนให้เห็นความจริงได้ ดังนี้

1. ค่าเฉลี่ยประชากรคะแนนความคลาดเคลื่อนของผู้สอบใดก็ตามจะมีค่าเท่ากับศูนย์

$$E(E_i) = 0$$

2. ความแปรปรวนร่วมระหว่างคะแนนความคลาดเคลื่อนกับคะแนนจริงจะเท่ากับศูนย์

$$\sigma_{ET} = 0$$

3. ความแปรปรวนของคะแนนที่สังเกตได้จะเท่ากับผลบวกของความแปรปรวนของคะแนนจริงกับคะแนนความคลาดเคลื่อน

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

4. ค่าสัมประสิทธิ์สหสัมพันธ์กำลังสองระหว่างคะแนนที่สังเกตได้กับคะแนนจริง จะเท่ากับอัตราส่วนของความแปรปรวนของคะแนนจริงต่อความแปรปรวนของคะแนนที่สังเกตได้

$$\rho_{XT}^2 = \frac{\sigma_T^2}{\sigma_X^2}$$

5. ค่าสัมประสิทธิ์สหสัมพันธ์กำลังสองระหว่างคะแนนที่สังเกตได้กับคะแนนจริงจะเท่ากับหนึ่งลบด้วยอัตราส่วนของความแปรปรวนของคะแนนความคลาดเคลื่อนต่อความแปร ปรวนของคะแนนที่สังเกตได้

$$\rho_{XT}^2 = 1 - \frac{\sigma_E^2}{\sigma_X^2}$$

6. เมื่อแบบทดสอบ 2 ฉบับคู่ขนาน ถ้า X และ X' เป็นคะแนนที่สังเกตได้จากแบบทดสอบทั้งสองและให้ Y เป็นคะแนน หรือตัวแปรเกณฑ์ภายนอกอื่น ๆ แล้ว

1) $\bar{X} = \bar{X'}$

2) $\sigma_X^2 = \sigma_{X'}^2$

3) $\rho_{xy} = \rho_{x'y}$

7. ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่สังเกตได้จากแบบทดสอบคู่ขนาน 2 ฉบับ จะเท่ากับอัตรส่วนของความแปรปรวนของคะแนนจริงต่อความแปรปรวนของคะแนนที่สังเกตได้

$$\rho_{xx'} = \frac{\sigma_T^2}{\sigma_x^2} = \frac{\sigma_{T'}^2}{\sigma_{x'}^2}$$

8. ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่สังเกตได้จากแบบทดสอบคู่ขนาน 2 ฉบับ จะเท่ากับหนึ่งลบด้วยอัตราส่วนของความแปรปรวนของคะแนนความคลาดเคลื่อนต่อความแปร ปรวนของคะแนนที่สังเกตได้

$$\rho_{xx'} = 1 - \frac{\sigma_E^2}{\sigma_x^2}$$

9. ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่สังเกตได้จากแบบทดสอบคู่ขนาน 2 ฉบับ จะเท่ากับ หนึ่งลบด้วยค่าสัมประสิทธิ์สหสัมพันธ์กำลังสองระหว่างคะแนนที่สังเกตได้กับคะแนนความคลาดเคลื่อน

$$\rho_{xx'} = 1 - \rho^2_{xE}$$

10. ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่สังเกตได้จากแบบทดสอบคู่ขนาน 2 ฉบับจะเท่ากับ ค่าสัมประสิทธิ์สหสัมพันธ์กำลังสองระว่างคะแนนที่สังเกตได้กับคะแนนจริง

$$\rho_{xx'} = \rho^2_{xT}$$

11. ความแปรปรวนของคะแนนจริง จะเท่ากับ ความแปรปรวนร่วมระหว่างคะแนนที่สังเกตได้จากแบบทดสอบคู่ขนาน

$$\sigma^2_T = \sigma_{xx'}$$

12. ความแปรปรวนของคะแนนความคลาดเคลื่อนจะเท่ากับความแปรปรวนของคะแนนที่ สังเกตได้คูณด้วยหนึ่งลบค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่สังเกตได้จากแบบทดสอบคู่ขนาน

$$\sigma^2_E = \sigma^2_x(1-\rho_{xx'})$$

รากกำลังสอง ของ σ^2_E เรียกว่า ความคลาดเคลื่อนมาตรฐานของการวัด (σ_E)

13. ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจริงจากแบบทดสอบ2 ฉบับ จะเท่ากับค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่สังเกตได้จากแบบทดสอบทั้งสองฉบับหารด้วยรากกำลังสองของผลคูณของค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่สังเกตได้จากแบบทดสอบคู่ขนานของแต่ละฉบับ

$$\rho_{T_xT_y} = \frac{\rho_{xx'}}{\sqrt{\rho_{xx'}}\sqrt{\rho_{yy'}}}$$

14. ถ้า S เป็นผลบวกของคะแนนที่สังเกตได้จากแบบทดสอบคู่ขนาน N และ X_i เป็นคะแนนที่สังเกตได้จากแบบทดสอบคู่ขนานแต่ละฉบับ ($S = \sum_{i=1}^{N} X_i$) แล้วจะพบว่า

1) $\sigma^2_{T_s} = N^2 \sigma^2_{T_x}$

$$2)\quad \sigma^2_{E_x} = N^2\,\sigma^2_{E_y}$$

จากการรวมแบบทดสอบคู่ขนานเข้าด้วยกันเป็นฉบับใหม่ที่มีความยาวมากกว่าเดิม N เท่า จะทำให้ความแปรปรวนของคะแนนจริงแบบทดสอบฉบับใหม่จะเพิ่มเป็น N^2 เท่าของฉบับเดิม ในขณะที่ความแปรปรวนของคะแนนความคลาดเคลื่อนของแบบทดสอบฉบับใหม่จะเพิ่มเป็นเพียง N เท่าของฉบับเดิม จึงเห็นได้ว่าความแปรปรวนของคะแนนจริงจะมีค่าเพิ่มขึ้นมากกว่าการเพิ่มขึ้นของความแปรปรวนของคะแนนความคลาดเคลื่อน ดังนั้น การเพิ่มจำนวนข้อสอบคู่ขนานให้แบบทดสอบยาวขึ้นจะช่วยให้การวัดมีความเชื่อมั่นหรือน่าเชื่อถือได้มากขึ้นผลที่ได้นี้เป็นที่มาของการพัฒนาสูตรของสเปียร์แมน-บราวน์

15. ถ้า S เป็นผลบวกของคะแนนที่สังเกตได้จากแบบทดสอบคู่ขนาน N ฉบับ และ X_i เป็นคะแนนที่สังเกตได้จากแบบทดสอบคู่ขนานแต่ละฉบับ แล้วจะพบว่า

$$\rho_{ss'} = \frac{N\rho_{xx'}}{1+(N-1)\rho_{xx'}} \qquad \text{(Spearman-Brown Formula)}$$

16. ถ้า $\rho_{xx'} \neq 0$, $\lim_{N \to \infty} \rho_{ss'} = 1$

นั่นคือ ถ้า $\rho_{xx'}$ ไม่เท่ากับศูนย์ และขนาดของจำนวนข้อสอบคู่ขนานเพิ่มมากขึ้นเข้าใกล้ค่าอนันต์ (∞) แล้ว $\rho_{ss'}$ จะเท่ากับหนึ่ง

ทฤษฎีการทดสอบแบบดั้งเดิมตั้งอยู่บนพื้นฐานความเชื่อของโมเดลการวัดข้อตกลงเบื้องต้นอันนำไปสู่ข้อสรุปเชิงทฤษฎีดังกล่าว อันจะเป็นพื้นฐานสำคัญสำหรับการนำไปใช้วิเคราะห์คุณภาพของแบบทดสอบและข้อสอบต่อไป

การวิเคราะห์คุณภาพของแบบทดสอบ

แบบทดสอบเป็นเครื่องมืออย่างหนึ่งสำหรับการวัดทางการศึกษาและจิตวิทยาการสร้างแบบทดสอบที่มีคุณภาพได้นั้น นักวัดผลจะต้องคำนึงถึงคุณลักษณะของแบบทดสอบที่ดี ประสิทธิภาพของแบบทดสอบสามารถประเมินได้จากการวิเคราะห์แบบทดสอบ ในที่นี้จะขอกล่าวถึงการวิเคราะห์คุณลักษณะของแบบทดสอบที่สำคัญเพียง 2 คุณลักษณะ ได้แก่ ความเชื่อมั่น และความตรง

I. ความเชื่อมั่น

1.1 ความหมายของความเชื่อมั่น ตามทฤษฎีการทดสอบแบบดั้งเดิม เราสามารถให้คำจำกัดความของคำว่า "ความเชื่อมั่น" ได้หลายลักษณะ ดังนี้

1.1.1 ความเชื่อมั่น= ρ^2_{XT}

ความเชื่อมั่นของแบบทดสอบมีค่าเท่ากับสัมประสิทธิ์สหสัมพันธ์กำลังสองระว่างคะแนนที่สังเกตได้กับคะแนนจริง ดังนั้น แบบทดสอบฉบับหนึ่งจะมีความเชื่อมั่นสูง ถ้าคะแนนที่สังเกตได้จากแบบทดสอบนั้นมีความสัมพันธ์สูงกับคะแนนจริง กล่าวคือ ถ้าเราสามารถทราบคะแนนจริงของผู้สอบแต่ละคน ค่าสัมประสิทธิ์สหสัมพันธ์กำลังสองระหว่างคะแนนที่สังเกตได้กับคะแนนจริงมีชื่อเรียกว่า "สัมประสิทธิ์ความเชื่อมั่น" ค่านี้จึงบอกถึงความสามารถของ X ในการอธิบายความผันแปรของ T ถ้า $\rho^2_{XT} = 1$ ค่าสัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบมีค่าสูงสุดเป็น 1 นั่นคือ การทราบค่า X จะสามารถทำนายหรืออธิบายความผันแปรของ T ได้ทั้งหมด (100 %) แต่ถ้า $\rho^2_{XT} = 0$ ค่าสัมประสิทธิ์เชื่อมั่นของแบบทดสอบมีค่าต่ำสุดเป็น 0 นั่นคือ ค่าของ X จะไม่สามารถใช้ทำนายหรืออธิบายความผันแปรของ T ได้เลย อย่างไรก็ตามทั้งสองกรณีเป็นกรณีในอุดมคติซึ่งมีความเป็นไปได้ในทางทฤษฎี แต่ในทางปฏิบัติมักจะพบค่าสัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบอยู่ระหว่าง 1 กับ 0

1.1.2 ความเชื่อมั่น = 1- ρ^2_{XE}

ความเชื่อมั่นของแบบทดสอบมีค่าเท่ากับหนึ่งลบด้วยค่าสัมประสิทธิ์สหสัมพันธ์กำลังสองระหว่างคะแนนที่สังเกตได้กับคะแนนความคลาดเคลื่อน ดังนั้น ถ้ามีค่าต่ำจะทำให้ความเชื่อมั่นของแบบทดสอบมีค่าสูง ถ้า $\rho_{XE} = 0$ จะทำให้ความเชื่อมั่นมีค่าสูงสุดตามอุดมคติคือ เป็น 1 แต่ถ้า $\rho^2_{XE} = 1$ จะทำให้มีค่าต่ำสุดเป็น 0

1.1.3 ความเชื่อมั่น = $\frac{\sigma^2_T}{\sigma^2_X}$

ความเชื่อมั่นของแบบทดสอบมีค่าเท่ากับอัตราส่วนของความแปรปรวนของคะแนนจริงต่อความแปรปรวนของคะแนนที่สังเกตได้ ถ้าสัดส่วนความแปรปรวนของคะแนนที่สังเกตได้เป็นความแปรปรวนอันเนื่องมาจากการผันแปรของคะแนนจริง จะทำให้ความเชื่อมั่นของแบบทดสอบมีค่าสูง(เข้าใกล้ 1) ถ้า $\sigma^2_X = \sigma^2_T$ จะทำให้ค่าความเชื่อมั่น =1 แต่ถ้า $\sigma^2_T = 0$ จะทำให้ค่าความเชื่อมั่น =0

1.1.4 ความเชื่อมั่น = $1 - \frac{\sigma^2_E}{\sigma^2_X}$

ความเชื่อมั่นของแบบทดสอบมีค่าเท่ากับหนึ่งลบด้วยอัตราส่วนของความแปรปรวนของคะแนนความคลาดเคลื่อนต่อความแปรปรวนของคะแนนที่สังเกตได้ ในกรณีที่กลุ่มผู้สอบมี

ความสามารถค่อนข้างแตกต่างกัน σ_X^2 น่าจะมีค่าสูงเมื่อเทียบกับ σ_E^2 จะทำให้ค่าความเชื่อมั่นของแบบทดสอบมีค่าสูง (เข้าใกล้ 1) แต่ในกลุ่มผู้สอบที่มีความสามารถใกล้เคียงกัน σ_X^2 น่าจะมีค่าต่ำเมื่อเทียบกับ σ_E^2 จะทำให้ค่าความเชื่อมั่นของแบบทดสอบมีค่าต่ำ (เข้าใกล้ 0) ดังนั้น ค่าความเชื่อมั่นตามทฤษฎีการทดสอบแบบดั้งเดิมจึงขึ้นอยู่กับกลุ่มผู้สอบเมื่อกลุ่มผู้สอบมีความสามารถที่หลากหลายหรือแตกต่างกันมาก ๆ ค่าความเชื่อมั่นย่อมมีแนวโน้มที่จะมีค่าสูงขึ้นกว่าความเชื่อมั่นที่ได้จากกลุ่มที่มีผู้สบความสามารถไล่เลี่ยหรือใกล้เคียงกัน

1.1.5 ความเชื่อมั่น $= \rho_{xx'}$

ความเชื่อมั่นของแบบทดสอบมีค่าเท่ากับสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่สังเกตได้จากแบบทดสอบคู่ขนาน 2 ฉบับ ถ้าคะแนนที่สังเกตได้จากแบบทดสอบคู่ขนานทั้ง 2 ฉบับมีความสัมพันธ์กันอย่างสมบูรณ์ ค่าความเชื่อมั่นของแบบทดสอบจะมีค่าเป็น 1 นั่นคือ คะแนนที่สังเกตได้จากแบบทดสอบทั้งสองมีความสอดคล้องกัน หรือตำแหน่งคะแนนตรงกันอย่างสมบูรณ์ แต่ถ้าคะแนนที่สังเกตได้จากแบบทดสอบทั้งสองฉบับไม่มีความสัมพันธ์กันเลย ค่าความเชื่อมั่นของแบบทดสอบจะเป็น 0

จากคำจำกัดความของสัมประสิทธิ์ความเชื่อมั่นทั้ง 5 นิยามนั้น คำจำกัดความที่ 1 ถึง 4 เป็นนิยามของความเชื่อมั่นในเชิงทฤษฎีไม่สามารถประมาณค่าได้โดยตรง เพราะไม่ทราบค่าของ T หรือ E ส่วนคำจำกัดความที่ 5 สามารถประเมินค่าความเชื่อมั่นได้ แต่มีความยุ่งยากในทางทฤษฎีปฏิบัติ เนื่องจากจะต้องสร้างแบบทดสอบ 2 ฉบับที่คู่ขนานกันอย่างแท้จริง ($T_i = T'_i$ และ $\sigma_E^2 = \sigma_{E'}^2$) ดังนั้น จึงได้มีนักวัดผลเสนอวิธีประมาณค่าความเชื่อมั่นของแบบทดสอบด้วยวิธีอื่น ๆ

1.2 ประเภทและวิธีการประมาณค่าความเชื่อมั่น เราอาจให้นิยามของความเชื่อมั่นเสียใหม่ว่า "ความเชื่อมั่น" เป็นความสม่ำเสมอหรือความคงเส้นคงวาของค่าที่วัดได้เมื่อทำการวัดซ้ำ ๆ ถ้าการวัดสิ่งเดียวกันหลาย ๆ ครั้งได้ค่าที่ค่อนข้างคงเส้นคงวาเพียงใดก็ถือว่าการวัดมีความเชื่อมั่นมากขึ้นเพียงนั้น แบบทดสอบที่มีความเชื่อมั่นสูงจะเป็นเครื่องมือที่ใช้วัดคุณลักษณะที่ต้องการได้ผลคงเส้นคงวา ปัญหาความเชื่อมั่นของแบบทดสอบเกี่ยวข้องกับคำถามที่ว่า "แบบทดสอบสามารถให้ผลการวัดที่คงเส้นคงวาเพียงใด ถ้าทำการวัดซ้ำ ๆ"

เราสามารถจำแนกความเชื่อมั่นออกเป็น 3 ประเภทใหญ่ ๆ โดยมีความหมายและวิธีการประมาณค่า ดังแสดงในตาราง 1

ตาราง 1 ประเภทของความเชื่อมั่น ความหมาย และวิธีการประมาณค่า

ประเภท	ความหมาย	วิธีการประมาณค่า
1. ความเชื่อมั่นแบบความคงที่	ความคงเส้นคงวาของคะแนนจากการวัดในช่วงเวลาที่ต่างกัน	คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่วัดได้จากเครื่องมือเดียวกันแต่ทำการวัดซ้ำสองครั้งในเวลาที่ต่างกัน
2. ความเชื่อมั่นแบบความทัดเทียมกัน	ความสอดคล้องกันของคะแนนจากการวัดด้วยเครื่องมือคู่ขนาน	คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่ได้จากเครื่องมือ 2 ฉบับที่คู่ขนานกัน
3. ความเชื่อมั่นแบบความสอดคล้องภายใน	ความสอดคล้องกันระหว่างคะแนนรายข้อหรือความเป็นเอกพันธ์ของเนื้อหารายข้ออันเป็นตัวแทนคุณลักษณะเด่นอันเดียวกันที่ต้องการวัด	คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ ระหว่างคะแนนรายข้อจากผลการวัดเพียงครั้งเดียว สามารถคำนวณโดยวิธีการต่าง ๆ ดังนี้ 1. แบ่งครึ่งข้อสอบ 2. ใช้สูตรการคำนวณของคูเดอร์ริชาร์ดสัน 3. ใช้สูตรการคำนวณสัมประสิทธิ์แอลฟาของครอนบาค

2. ความตรง

2.1 ความหมายของความตรง ตามทฤษฎีการทดสอบแบบดั้งเดิม เราสามารถให้คำจำกัดความของคำว่า "ความตรง" ได้ดังนี้

2.1.1 ความตรง หมายถึง ความถูกต้อง แม่นยำ ของเครื่องมือในการวัดสิ่งที่ต้องการจะวัด

ความตรงเป็นความใกล้เคียงกันระหว่างค่าที่ได้กับค่าที่แท้จริง ถ้าผลการวัดได้ค่าที่ใกล้เคียงกับค่าที่แท้จริงเพียงใด ก็ถือว่าการวัดมีความตรงมากขึ้นเพียงนั้น ดังนั้น ความตรงจึงเป็นคุณสมบัติที่สำคัญที่สุดของเครื่องมือหรือแบบทดสอบ แบบทดสอบที่มีความตรงสูงจะเป็นเครื่องมือที่

วัดคุณลักษณะที่ต้องการได้ตรงหรือใกล้เคียงกับค่าที่แท้จริงมาก คะแนนที่ได้จากแบบทดสอบที่มีความตรงสูง คะแนนนั้นสามารถบ่งบอกถึงสภาพที่แท้จริงของคนลักษณะนั้นได้เป็นอย่างดี

คำจำกัดความนี้ถือเป็นนิยามของความตรงในเชิงทฤษฎี ไม่สามารถประมาณค่าได้โดยตรงเพราะไม่ทราบค่าที่แท้จริงของสิ่งที่มุ่งวัด จึงต้องหาวิธีประมาณค่าความตรงโดยวิธีอื่น ๆ

2.1.2 ความตรง หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่ได้จากเครื่องมือกับคะแนนเกณฑ์ภายนอกที่เป็นอิสระอื่น ๆ ซึ่งวสามารถวัดสิ่งที่ต้องการวัดนั้นได้

คำจำกัดความนี้ทำให้เราสามารถประมาณค่าความตรงของแบบทดสอบได้ แต่ต้องสามารถหาคะแนนเกณฑ์ภายนอกที่เป็นตัวแทนคุณลักษณะที่ต้องการวัดได้อย่างเหมาะสม เพื่อใช้เป็นเกณฑ์เทียบถึงความถูกต้องของคะแนนจากแบบทดสอบในการวัดคุณลักษณะที่ต้องากรนั้น ปัญหาเรื่องความตรงของแบบทดสอบเกี่ยวข้องกับคำถามสำคัญ 2 ประการ คือ "แบบทดสอบมุ่งวัดอะไร" และ "แบบทดสอบวัดสิ่งนั้นได้ดีเพียงอย่างไร"

2.2 ประเภทและวิธีประมาณค่าความตรง เราสามารถจำแนกความตรงที่สำคัญออกเป็น 3 ประเภท โดยมีความหมายและวิธีประมาณค่า ในขอบเขตของการสร้างเครื่องมือเพื่อวัดผลการเรียนรู้ของผู้เรียนรู้ ดังแสดงในตารางที่ 2

ตาราง 2 ประเภทของความตรง ความหมาย และวิธีการประมาณค่า

ประเภท	ความหมาย	วิธีประมาณค่า
1. ความตรงตามเนื้อเรื่อง	ความสามารถในการวัดพฤติกรรมการเรียนรู้ที่ครอบคลุมและเป็นตัวแทนของจุดมุ่งหมายและเนื้อหาการเรียนรู้	1. เปรียบเทียบพฤติกรรมที่มุ่งวัดในเครื่องมือว่ามีความสอดคล้องกับมวลประสบการณ์ทั้งหมดของการเรียนรู้หรือไม่เพียงอย่างไร 2. เปรียบเทียบสัดส่วนของข้อคำถามว่ามีความสอดคล้องกับน้ำหนักความสำคัญของแต่ละลักษณะของการเรียนรู้มากน้อยเพียงไร

ตาราง 2 (ต่อ)

ประเภท	ความหมาย	วิธีประมาณค่า
2. ความตรงตามเกณฑ์สัมพันธ์	ความสามารถในการวัดพฤติกรรมการเรียนรู้ได้สอดคล้องกับเกณฑ์ภายนอก ซึ่งวัดได้จากเครื่องมืออิสระอันอื่น	คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่ได้จากเครื่องมือกับคะแนนจากเกณฑ์ภายนอก
2.1 ความตรงร่วมสมัย	ความสามารถในการวัดผลการเรียนรู้ได้ตรงสภาพปัจจุบัน	คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่ได้จากเครื่องมือมาตรฐานอื่นซึ่งสามารถวัดสิ่งนั้นได้ในสภาพปัจจุบัน
2.2 ความตรงเชิงทำนาย	ความสามารถในการวัดผลการเรียนรู้ที่จะเกิดขึ้นในอนาคต	คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่ได้จากเครื่องมือมาตรฐานอื่นซึ่งสามารถวัดสิ่งนั้นได้ในเวลาต่อมาหรือในอนาคต
3. ความตรงตามโครงสร้าง	ความสามารถในการวัดได้ตรงตามคุณลักษณะทางจิตวิทยาของการเรียนรู้นั้น	ศึกษาความสัมพันธ์ระหว่างผลการวัดที่ได้จากเครื่องมือกับคุณลักษณะที่มุ่งวัดโดยอาศัยข้อสนับสนุนเชิงสะสมของหลักฐานจากวิธีการวิเคราะห์ต่าง ๆ เช่น 1. ทดลองเปรียบเทียบผลการเรียนรู้ระหว่างกลุ่มที่ทราบคุณลักษณะของการเรียนรู้นั้นแล้ว 2. วิเคราะห์องค์ประกอบ (FACTOR ANALYSIS) 3. ใช้เทคนิคการวัดหลายคุณลักษณะโดยใช้การวัดหลายวิธี

สำหรับรายละเอียดของวิธีการและสูตรประมาณค่าความตรงประเภทต่าง ๆ ขอให้ดูได้จากตำราเกี่ยวกับการวัดและประเมินผลทางการศึกษาทั่วๆไป

การวิเคราะห์คุณภาพของข้อสอบ

ประสิทธิภาพของข้อสอบแต่ละข้อสามารถประเมินได้จากการวิเคราะห์คำตอบของผู้สอบเป็นรายข้อ โดยวิธีที่เรียกว่า "Item Analysis" หรือ "การวิเคราะห์ข้อสอบ"

1. จุดมุ่งหมายของการวิเคราะห์ข้อสอบ การวิเคราะห์ข้อสอบเป็นเทคนิคของการตรวจสอบคุณภาพของข้อสอบเป็นรายข้อ โดยมีจุดมุ่งหมาย ดังต่อไปนี้

1.1 ข้อสอบแต่ละข้อได้ทำหน้าที่ตามที่ผู้ออกข้อสอบเป็นเทคนิคของการตรวจสอบคุณภาพของข้อสอบเป็นรายข้อ โดยมีจุดมุ่งหมาย ดังต่อไปนี้

1.1.1 ข้อสอบในแบบสอบอิงกลุ่ม มีความสามารถที่จะจำแนกผู้มีผลสัมฤทธิ์สูงและต่ำออกจากกันได้หรือไม่

1.1.2 ข้อสอบในแบบสอบอิงเกณฑ์ มีความสามารถที่จะวัดผลที่เกิดจากการเรียนการสอนตามเกณฑ์ที่ตั้งไว้เพียงไร

1.2 ข้อสอบแต่ละข้อ มีความยากที่พอเหมาะหรือไม่

1.3 ข้อสอบแต่ละข้อมีข้อบกพร่องอย่างไร

1.4 ในกรณีที่เป็นข้อสอบแบบหลากตัวเลือกจะได้ทราบว่าตัวลวงที่ใส่ไว้มีประสิทธิภาพเพียงใด

2. กระบวนการวิเคราะห์ข้อสอบสำหรับแบบทดสอบอิงกลุ่ม สิ่งที่จะต้องคำนึงในการวิเคราะห์แบบทดสอบอิงกลุ่ม มีอยู่ 3 ประการ

2.1 ระดับความยากง่ายของข้อสอบ (p)(Level of difficulty of the items) หมายถึง สัดส่วนของจำนวนคนที่ตอบข้อสอบนั้นถูก เช่น ข้อสอบหนึ่งข้อมีคนตอบ 100 คน ปรากฏว่าตอบถูกเพียง 30 คน แสดงว่าข้อสอบข้อนั้นมีระดับความยากง่ายของข้อสอบจึงมีค่าอยู่ระหว่าง 1.0 กับ 0 ถ้าข้อสอบใดมีคนตอบถูกมาก p จะมีค่าสูง (เข้าใกล้ 1) แสดงข้อนั้นง่าย ในทางตรงกันข้าม ถ้าข้อสอบข้อใดมีคนตอบถูกน้อย p จะมีค่าต่ำ (เข้าใกล้ 0) แสดงว่าข้อนั้นยาก โดยทั่วไปข้อสอบที่มีค่า p ระหว่าง 0.2-0.8 ถือว่าข้อสอบที่มีความยากง่ายพอเหมาะและข้อสอบทั้งฉบับควรมีระดับความยากง่ายเฉลี่ยประมาณ 0.5

2.2 อำนาจจำแนกข้อสอบ(D) (Dicrimination power of the items) หมายถึง ความสามารถของข้อสอบที่จะจำแนกหรือแยกให้เห็นความแตกต่างระหว่างคนที่มีความสัมฤทธิ์ต่างกัน เช่น จำแนกคนเก่งกับคนอ่อนออกจากกันได้ หรือจำแนกคนที่มีความสามารถพิเศษกับคนที่ไม่มีความสามารถออกจากกันได้ โดยถือว่าคนที่เก่งหรือมีความสามารถจะต้องทำข้อสอบนั้นได้ ส่วนผู้ที่อ่อน

หรือไม่มีความสามารถจะต้องทำข้อสอบข้อนั้นไม่ได้ การคำนวณหาค่าอำนาจจำแนกของข้อกระทง (D) อย่างง่าย ๆ สามารถกระทำได้จากผลต่างระหว่างสัดส่วนจำนวนคนตอบถูกในกลุ่มเก่งกับสัดส่วนจำนวนคนตอบถูกในกลุ่มอ่อน เช่น กลุ่มเก่ง 10 คน ตอบถูก 9 คน แต่กลุ่มอ่อน 10 คนตอบถูกเพียง 2 คน เพราะฉะนั้น r จะมีค่าเท่ากับ 0.7 ดังนั้น พิสัยของอำนาจจำแนกของข้อสอบจะมีค่าอยู่ระหว่าง 1 กับ –1 แต่อำนาจจำแนกที่ดีจะต้องมีค่าเป็นบวก ควรมากกว่า 0.2 ขึ้นไป

ค่าอำนาจจำแนกข้อสอบสามารถได้จากสัมประสิทธิ์สหสัมพันธ์ไบซีเรียล (r bis) หรือสัมประสิทธิ์สหสัมพันธ์พ้อยท์ไบซีเรียล (r pbis) ระหว่างคะแนนรวมข้อกับคะแนนรวม หรือสามารถดูได้จากตารางสำเร็จรูปของ แฟลนนาแกน (Flanagan, 1939) หรือตารางของจงตงฟาน (Chung The Fan, 1952 1954 อ้างถึงใน ศิริชัย กาญจนวาสี,2535: 107)

3. การวิเคราะห์ข้อสอบสำหรับแบบทดสอบอิงเกณฑ์ เทคนิคการวิเคราะห์ข้อสอบอิงกลุ่มไม่สามารถนำมาใช้ได้โดยตรงกับการวิเคราะห์ข้อสอบอิงเกณฑ์ เนื่องจากคุณสมบัติของข้อสอบอิงเกณฑ์ต้องการที่จะดูว่าผู้สอบมีความรู้ความสามารถที่จะปฏิบัติอะไรได้บ้าง มากกว่าที่จะจัดเรียงลำดับความสามารถของผู้สอบในกลุ่มตามที่แบบสอบอิงกลุ่มต้องการ ดังนั้นดัชนีวัดระดับความยาก (p) และอำนาจจำแนก (D) ของข้อสอบอิงเกณฑ์จึงมีหลักการคำนวณและการแปลความหมายแตกต่างไปจากข้อสอบอิงกลุ่ม

ระดับความยากง่ายของข้อสอบอิงเกณฑ์มิได้ตั้งอยู่บนฐานของการทำข้อสอบได้ของกลุ่ม สัมฤทธิ์ผลสูง-ต่ำ แต่จะขึ้นอยู่กับระดับพฤติกรรมการเรียนรู้ที่ต้องการจะวัด ถ้าพฤติกรรมนั้นง่าย ข้อสอบแบบอิงเกณฑ์จะง่าย และมีความยากปานกลาง หรือค่อนข้างยาก ถ้าพฤติกรรมการเรียนรู้ที่ต้องการจะวัดมีความยากปานกลางหรือค่อนข้างยากตามลำดับ และไม่มีความจำเป็นที่จะต้องพยายามทำให้ข้อสอบอิงเกณฑ์มีคะแนนที่ได้กระจายมาก ๆ สำหรับสูตรในการคำนวณระดับความยากง่าย (p) ก็ยังคงใช้ได้กับข้อสอบอิงเกณฑ์ โดยสามารถคำนวณระดับความยากง่ายของข้อสอบการเรียนรู้ (ควรมี % ของการทำค่าได้ต่ำ) และหลังการเรียนรู้ (ควรมี % ของการทำได้สูง) ดังนั้น แบบสอบอิงเกณฑ์จะมีความยากสูง ถ้าเปอร์เซนต์ที่ใช้ตัดสินให้ผ่านอยู่ในเกณฑ์สูงและจะมีระดับความยากน้อย ถ้าตั้งเกณฑ์ผ่านไว้ต่ำ

ถ้าการเรียนการสอนได้ดำเนินไปอย่างมีประสิทธิภาพแล้ว อำนาจจำแนกระหว่างกลุ่มสัมฤทธิ์ผลสูง-ต่ำ มิใช่องค์ประกอบสำคัญในการประเมินประสิทธิภาพของข้อสอบในแบบอิงเกณฑ์ โดยข้อสอบที่ดีอาจมีอำนาจเป็นศูนย์ (0) ก็ได้ เพราะผู้เรียนทุกคนได้เรียนรู้และสามารถตอบข้อกระทงนั้นได้ถูกต้องทั้งหมด ซึ่งอาจแปลความหมายได้ว่าทั้งการสอนและข้อสอบมีประสิทธิภาพ จึงเป็นการสะท้อนข้อมูลเกี่ยวกับการเรียนรู้ที่ผู้เรียนได้รับจากกิจกรรมการเรียนการสอนที่ผู้สอนเป็นผู้จัดให้ ดังนั้น จุดประสงค์ของข้อสอบอิงเกณฑ์จึงมุ่งที่จะบรรยายว่าผู้เรียนรู้หรือสามารถทำอะไรได้บ้างและมุ่งที่จะจำแนกผู้ที่เรียนรู้หรือสามารถทำอะไรได้บ้างและมุ่งที่จะจำแนกผู้ที่เรียนรู้กับผู้ที่ยังไม่เรียนรู้ออกจากกันมากกว่าที่จะจำแนกความแตกต่างระหว่างผู้เรียน

ทฤษฎีการตอบสนองข้อสอบ

หลักการของทฤษฎีการตอบสนองข้อสอบ

ในการวิเคราะห์ข้อสอบโดยทั่วไปแล้วจะวิเคราะห์ข้อสอบโดยพิจารณาตาม ทฤษฎีดั้งเดิม(Classical Test Theory) ซึ่งเมื่อพิจารณารายข้อ (Item) เราจะดูจากค่าความยาก (p) และค่าอำนาจจำแนก (r) เมื่อพิจารณารวมทั้งฉบับ (Test) เราจะดูจากค่าความเชื่อมั่น (Reliability) และค่าความเที่ยงตรง (Validity) ซึ่งจากการพิจารณาโดยภาพรวมแล้วพบว่ายังมีจุดอ่อนอยู่หลายประการ คือ ประการแรกค่าพารามิเตอร์ของข้อสอบแปรเปลี่ยนไปตามกลุ่มของผู้สอบที่แตกต่างกัน ในด้านความสามารถ (Ability) ประการที่สอง การเปรียบเทียบความสามารถของผู้สอบจำกัดอยู่ในสถานการณ์ที่ทดสอบ และประการที่สามจะไม่สามารถบอกได้ว่าผู้เข้าสอบคนหนึ่งจะทำข้อสอบได้เพียงใด เมื่อได้เผชิญกับข้อคำถามหนึ่ง ยกเว้นเมื่อได้มีการใช้ข้อสอบนั้นแล้วกับกลุ่มตัวอย่างที่คล้ายคลึงกันกับบุคคลนั้น จากเหตุการณ์เหล่านี้ทำให้นักทดสอบทางจิตวิทยาได้สำรวจและพัฒนาทฤษฎีที่เหมาะสม ทฤษฎีที่ถูกพัฒนาขึ้นมาคือ ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) โดยใช้ข้อตกลงเบื้องต้นว่าความสามารถของบุคคล กำหนดการทำข้อสอบที่วัดความสามารถนั้นอย่างไร ซึ่งทฤษฎีการตอบข้อคำถามมีหลักการที่สำคัญอยู่ที่การใช้ผลการตอบแบบทดสอบมาอธิบายถึงความสามารถของผู้สอบในเรื่องที่ทำการทดสอบนั้น ซึ่งเมื่อใช้ทฤษฎีนี้ในการวิเคราะห์แล้วจะมีผลดี คือ

1. ค่าพารามิเตอร์ของข้อสอบจะคงที่ คือ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าสัมประสิทธิ์การเดา (c) จะเป็นค่าที่ไม่แปรเปลี่ยนไปตามกลุ่มผู้ตอบ ไม่ว่าจะนำไปสอบกับผู้ใดก็ตาม

2. เมื่อทราบลักษณะการตอบข้อสอบในแต่ละข้อคำถามของผู้เข้าสอบคนใด เราจะสามารถคำนวณหาค่าความสามารถที่แท้จริงของบุคคลนั้นได้ โดยค่านี้จะสัมพันธ์โดยตรงกับคะแนนจริง ซึ่งลักษณะเช่นนี้ถือว่าเป็นลักษณะของความเป็นอิสระของข้อสอบ

ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อคำถาม

1. แบบทดสอบที่มีมิติเดียว (Unidimention Test) หมายความว่าข้อสอบแต่ละข้อในแบบทดสอบจะต้องวัดความสามารถหรือคุณลักษณะเดียวกัน หรือมีความเป็นเอกพันธ์กัน

2. ข้อสอบแต่ละข้อจะต้องเป็นอิสระจากกัน หมายความว่าการตอบข้อสอบข้อใดข้อหนึ่งถูกจะไม่มีผลต่อการตอบข้อสอบข้ออื่น ๆ

3. โอกาสที่ผู้เข้าสอบจะตอบข้อสอบข้อนั้นถูก เป็นไปตามโค้งลักษณะข้อสอบ (Item Characteristic Curve ; ICC) ของแต่ละโมเดลที่ใช้ ไม่ขึ้นกับการแจกแจงความสามารถของกลุ่มตัวอย่าง

พารามิเตอร์ของทฤษฎีการตอบสนองข้อสอบ

ตามทฤษฎีเราแบ่งพารามิเตอร์ออกเป็น 2 ชนิด คือ

1. พารามิเตอร์ของข้อสอบ (Item Parameter) ได้แก่

1.1 ค่าความยาก (b) หมายถึง สัดส่วนของคนที่ทำข้อสอบข้อนั้นถูก หรือหมายถึงค่าที่แสดงถึงระดับความสามารถของผู้สอบ (θ) ที่จุดโค้งลักษณะข้อสอบมีความชันมากที่สุด มีค่าตั้งแต่ -∞ ถึง ∞ แต่ในทางปฏิบัติมักจะมีค่าอยู่ระหว่าง -3 ถึง +3 ค่า -3 แสดงว่าข้อสอบนั้นง่ายมาก และค่า +3 แสดงว่าข้อสอบนั้นยากมาก

1.2 ค่าอำนาจจำแนก (a) หมายถึง ความสามารถของข้อสอบที่แยกเด็กออกเป็น 2 พวก คือพวกตอบถูกกับพวกตอบผิด ในการวิจัยนี้หมายถึง ค่าที่เป็น สัดส่วนโดยตรงกับความชันของโค้งคุณลักษณะของข้อสอบ ณ จุดเปลี่ยนโค้งมีค่าตั้งแต่ -∞ ถึง +∞ แต่ในทางปฏิบัติมีค่าตั้งแต่ 0 ถึง 2 เพราะค่า a ที่เป็นลบแสดงว่าข้อสอบไม่ดี ใช้ไม่ได้ต้องตัดทิ้ง ค่า 0 แสดงว่าข้อสอบไม่มีค่าอำนาจจำแนก ค่า +2 แสดงว่าข้อสอบมีค่าอำนาจำแนกสูง ในการคัดเลือกข้อสอบ ข้อสอบที่คัดไว้จะมีค่า a ตั้งแต่ 0.5 ขึ้นไป

1.3 ค่าสัมประสิทธิ์การเดา (c) หมายถึง ความน่าจะเป็นของบุคคลหนึ่งที่มีความสามารถต่ำสุด ความสามารถจะตอบข้อสอบนั้นได้ถูกต้อง เป็นค่าที่แสดงถึงโอกาสการตอบข้อสอบถูกโดยไม่มีความรู้ในเรื่องนั้น ๆ มีค่าจาก 0 ถึง 1 จะคัดเลือกเอาข้อสอบที่มีค่า c ต่ำกว่า 0.3 ลงมาก

2. พารามิเตอร์ของผู้สอบ (Examines Parameter) ได้แก่

ระดับความสามารถของผู้สอบ (θ) หมายถึง ศักยภาพของผู้สอบที่ประมาณได้จากการกระทำข้อสอบตามทฤษฎีการตอบสนองข้อสอบส่วนใหญ่จะ มีค่าอยู่ระหว่าง -3 ถึง +3 ค่า –3 แสดงว่ามีความสามารถต่ำ และค่า +3 แสดงว่ามีความสามารถสูง

โมเดลในทฤษฎีการตอบสนองข้อสอบ

โมเดลโลจิสติกถูกพัฒนาขึ้นเพื่อให้สะดวกต่อการนำไปใช้ จึงพัฒนาขึ้นเป็น 3 รูปแบบ ดังนี้ (ล้วน สายยศ และ อังคณา สายยศ, 2539 : 202-208)

One – Parameter Logistic Model

โมเดลนี้เบอร์นบอม พัฒนาขึ้นในปี 1968 ซึ่งบังเอิญตรงกับรูปแบบของราส์ช (Rasch 1960) เป็นโมเดลที่อธิบายข้อสอบด้วยค่าพารามิเตอร์เพียงตัวเดียว คือ ค่าความยาก โดยเชื่อว่าโอกาสที่ผู้สอบจะทำข้อสอบได้ถูกหรือไม่นั้น ขึ้นอยู่กับระดับความสามารถของตนเองกับระดับความยากของข้อสอบ ดังนั้น จึงถือว่าค่าการเดาเป็นศูนย์ ($c_i = 0$) และค่าอำนาจจำแนก ($a_i = 1$) ของข้อสอบจะคงที่ทั้งฉบับ นั่นคือ กำหนดให้ทุกข้อที่ผู้สอบทำข้อสอบไม่มีการเดา และทุกข้อมีค่าอำนาจจำแนกเท่ากัน เขียนเป็นฟังก์ชันได้ดังนี้

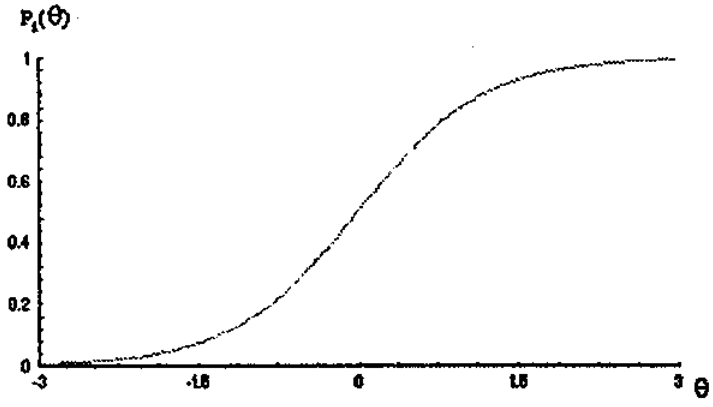
$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \quad , \quad i = 1, 2, 3, \ldots, n$$

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

θ คือ ระดับความสามารถที่แท้จริงของผู้สอบ

b_i คือ ค่าความยากของข้อสอบข้อที่ i

e คือ ค่าคงที่มีค่าเท่ากับ 2.7182818



ภาพประกอบ 5 ค่าความหมายของค่า Item Parameter (1 พารามิเตอร์) ของข้อสอบ

Two – Parameter Logistic Model

เบอร์นบอร์ม (Birnbaum) ได้พัฒนาโมเดลนี้ขึ้นมาจาก Normal Ogive Model ซึ่งรูปของโมเดลใหม่ก็มีรูปแบบไม่แตกต่างจากเดิมมากนัก จะเปลี่ยนแต่เพียงการคิดคำนวณเท่านั้น และกำหนดให้ทุกข้อไม่มีการเดาเกิดขึ้น คือ ค่า c_i มีค่าเป็น ศูนย์ทุกข้อ กล่าวคือ ผู้สอบที่มีความสามารถต่ำสุด ไม่มีโอกาสที่จะทำข้อสอบถูกในข้อสอบที่มีค่าความยากสูง ซึ่งเบอร์นบอร์ม ได้เสนอรูปแบบของสมการดังนี้

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

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

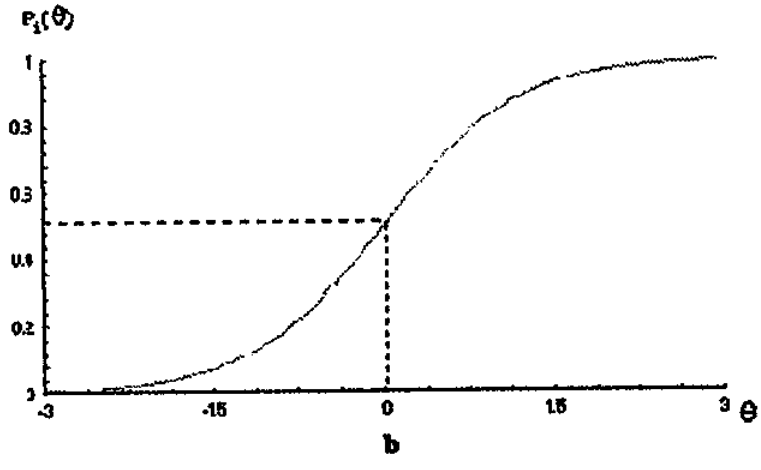
θ คือ ระดับความสามารถที่แท้จริงของผู้สอบ

a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i

b_i คือ ค่าความยากของข้อสอบข้อที่ i

e คือ ค่าคงที่มีค่าเท่ากับ 2.7182818

D คือ ค่าคงที่ ซึ่งมีค่า 1.7



ภาพประกอบ 6 แสดงความหมายของค่า Item Parameter (2 พารามิเตอร์) ของข้อสอบ

Three – parameter Logistic Model

เป็นโมเดลที่พัฒนามาจาก Two-Parameter Logistic Model เพื่อให้เหมาะกับแบบสอบที่มีอิทธิพลจากการเดาเข้ามาแฝงอยู่ด้วย และเป็นได้ลักษณะข้อสอบที่แสดงถึงลักษณะข้อสอบที่มีค่าพารามิเตอร์ของข้อสอบ 3 ตัว ซึ่งเบอร์นบอร์ม ได้เสนอรูปแบบของสมการดังนี้

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

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

θ คือ ระดับความสามารถที่แท้จริงของผู้สอบ จะมีค่าอยู่ระหว่าง –3 ถึง +3 และค่า –3 แสดงว่ามีค่าความสามารถต่ำ +3 แสดงว่ามีค่าความสามารถสูง

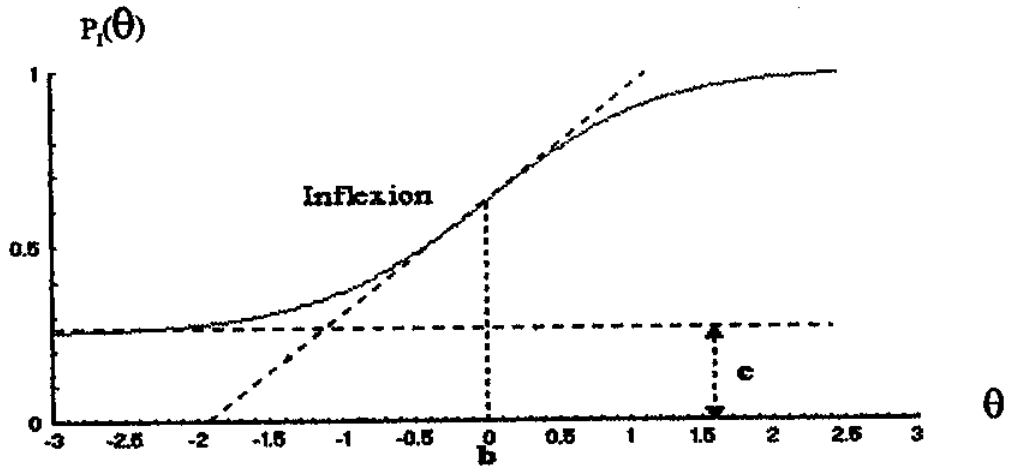
a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i ซึ่งมีค่าสัดส่วนโดยตรงกับค่าความชันของโค้ง ณ จุดเปลี่ยนโค้งและมีค่าอยู่ระหว่าง $-\alpha$ ถึง $+\alpha$ ค่าอำนาจจำแนกที่เป็นลบ แสดงว่าข้อสอบจำแนกคนเก่งอ่อนไม่ได้ ถ้าค่าอำนาจจำแนกเป็นบวก แสดงว่าข้อสอบสามารถจำแนกคนเก่งอ่อนได้ โดยทั่วไปแล้วจะเอาค่าที่ .30 ถึง 2.00

b_i คือ ค่าความยากของข้อสอบข้อที่ i มีค่าเท่ากับ ระดับความสามารถของผู้สอบ ณ จุดเปลี่ยนโค้ง และมีค่าอยู่ระหว่าง $-\alpha$ ถึง $+\alpha$ แต่ในทางปฏิบัติจะเลือกข้อสอบที่ มีค่าอยู่ระหว่าง -2 ถึง +2 ค่า –2 แสดงว่าข้อสอบง่ายมาก และค่า +2 แสดงว่าข้อสอบยากมาก

c_i คือ ค่าการเดาของข้อสอบข้อที่ i เป็นความน่าจะเป็นหรือโอกาสของคนที่มีความสามารถต่ำจะตอบข้อสอบถูก มีค่าอยู่ระหว่าง 0 ถึง 1 โดยทั่วไปแล้ว ข้อสอบที่ดีจะต้องมีค่าการเดาต่ำกว่า 0.30

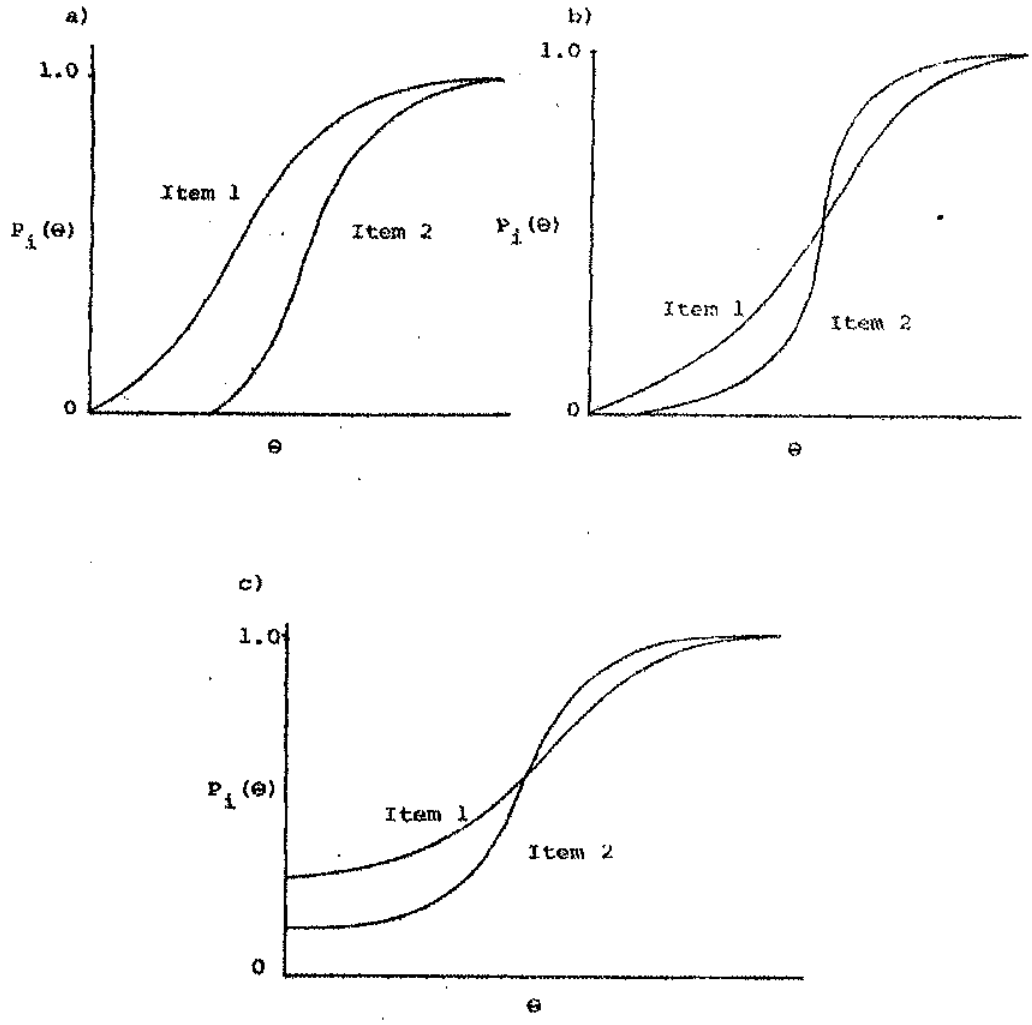
e คือ ค่าคงที่มีค่าเท่ากับ 2.7182818

D คือ ค่าคงที่ โดยปกติมีค่า 1.702 หรือ 1.7



ภาพประกอบ 7 แสดงความหมายของค่า Item Parameter (3 พารามิเตอร์) ของข้อสอบ

จากโมเดลโลจิสติกทั้ง 3 โมเดล สามารถเขียนแสดงได้ดังรูปต่อไปนี้ (Hambleton and Cook, 1977 : 79 ,อ้างถึงใน ประดิษฐ์ เรืองตระกูล, 2529 : 21-22)



ภาพ 8 แสดง ICC ของโมเดลโลจิสติก 1, 2, 3 พารามิเตอร์

จากภาพประกอบ 8 แสดงถึงโอกาสที่ผู้ทำข้อสอบจะตอบข้อสอบได้ถูกต้อง ขึ้นอยู่กับค่าพารามิเตอร์ ได้แก่

รูป a แสดงถึงโอกาสที่จะทำข้อสอบได้ถูกต้องขึ้นอยู่กับความสามารถของข้อสอบเพียงอย่างเดียว โดยถือว่าข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากันหมด คือเท่ากับหนึ่ง และค่าการเดาเท่ากับศูนย์ ในรูปแสดงให้เห็นว่าข้อสอบข้อที่ 2 ยากกว่าข้อที่ 1

รูป b แสดงถึงโอกาสที่จะทำข้อสอบได้ถูกต้อง ขึ้นอยู่กับความยากและค่าอำนาจจำแนก ในรูปแสดงให้เห็นว่าข้อสอบข้อที่ 2 ยากกว่าข้อที่ 1 และข้อสอบข้อที่ 2 จำแนกคนได้ดีกว่าข้อที่ 1

รูป c แสดงถึงโอกาสที่จะทำข้อสอบได้ถูกต้อง ขึ้นอยู่กับค่าอำนาจจำแนก ค่าความยากและค่าการเดา ในรูปแสดงให้เห็นว่าข้อสอบข้อที่ 1 ยากกว่าข้อที่ 2 และ ข้อสอบข้อที่ 2 จำแนกคนได้ดีกว่าและมีค่าการเดาน้อยกว่าข้อที่ 1

ฟังก์ชันสารสนเทศของแบบทดสอบ

ในการประมาณค่าความสามารถของผู้สอบด้วยวิธีแมกซิมั่มไลถีฮูด (Maximum Likelihood) นั้นความแน่นอนของการประมาณค่าความสามารถแสดงได้ในเทอมของค่าฟังก์ชันสารสนเทศ (Information Function) โดยที่ในทฤษฎีการทดสอบดั้งเดิม นั้นเราศึกษาเรื่องความเที่ยง (Reliability) ของคะแนนและความคลาดเคลื่อนมาตรฐานของการวัด(Standard Error of Measurement) ซึ่งค่าที่ได้จะแปรเปลี่ยนไปตามกลุ่มผู้สอบดันเป็นจุดอ่อนประการศึกษา แต่ในทฤษฎี การตอบสนองข้อสอบ จะศึกษาถึง ค่าฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function) แทนการหาค่าความเที่ยง (Hambleton 1979 : 64)

สำหรับโมเดลโลจีสติก 3 พารามิเตอร์จะสามารถหาค่าฟังก์ชันสารสนเทศของข้อสอบ (Item Information Functions) ในแต่ละระดับความสามารถ θ ได้จากสูตร (Lord 1980 : 72-74)

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

และสามารถกำหนด โค้งฟังก์ชันสารสนเทศของข้อสอบ(Item Information Curve) ได้จากสมการ

$$I\{\theta,u_i\} = \frac{(P_i')^2}{P_i\theta_i}$$

.....(2)

I $\{\theta,u_i\}$ คือ ค่าฟังก์ชันสารสนเทศของข้อสอบ (Item Information Functions)

P'_i คือ ความชันของ ICC ที่ระดับความสามารถ θ

P_i คือ ความน่าจะเป็นที่ผู้สอบที่มีความสามารถ θ จะต้องตอบข้อสอบข้อที่ i ถูก

Q_i คือ $1 - P_i$

และจะสามารถหาค่าฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function) ได้โดยหาผลรวมของค่าฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function)

$$I\{\theta\} = \sum_{i=1}^{n} I\{\theta, u_i\} \quad \text{.....(3)}$$

เมื่อ I $\{\theta\}$ คือ ค่าฟังก์ชันสารสนเทศของแบบทดสอบ(Test Information Function)

จากสมการ (2) จะเห็นว่าข้อสอบแต่ละข้อมี โค้งฟังก์ชันสารสนเทศของข้อสอบ (Item Information Curve) ซึ่งจะขึ้นอยู่กับความชันของ ICC และความแปรปรวนของการตอบข้อสอบถูกในแต่ละข้อในแต่ละระดับความสามารถ และยิ่งความชันของ ICC มีค่ามาก ๆ ประกอบกับความแปรปรวน มีค่าน้อย ๆ โค้งฟังก์ชันสารสนเทศของข้อสอบ ที่ระดับความสามารถนั้นจะยิ่งสูงขึ้นสำหรับโค้งฟังก์ชันสารสนเทศของข้อสอบ(Item Information Curve) ที่มีค่าสูงสุด ณ ระดับความสามารถใดก็จะจำแนกระดับความสามารถผู้สอบได้ดี ณ ระดับความสามารถนั้น (Hambleton, 1979 : 66)

ดังนั้นประโยชน์ที่ได้จากประเด็นข้างต้นก็คือ ถ้ามีกลุ่มของข้อสอบอยู่ชุดหนึ่งที่สามารถทราบ โค้งฟังก์ชันสารสนเทศ(Information Curve)ของแต่ละข้อ เราก็จะสามารถสร้างแบบสอบฉบับหนึ่งให้มี ค่าฟังก์ชันสารสนเทศของแบบทดสอบ(Test Information Curve) ณ ระดับหนึ่งของความสามารถตามที่เราต้องการได้ และนั่นหมายถึงว่าเราสามารถสร้างฉบับแบบทดสอบให้เป็นไปตามจุดมุ่งหมายของการสอบได้ เช่น ถ้าต้องการได้แบบสอบคัดเลือก ก็ต้องเลือกใช้ข้อสอบที่มีความสูงสุดของโค้งที่ระดับความสามารถสูง ๆ ซึ่งก็คือ เลือกข้อสอบที่จะให้ได้ โค้งฟังก์ชันสารสนเทศของแบบทดสอบ(Test Information Curve)สูงที่ระดับความสามารถสูง ๆ เป็นต้น

การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบ

ปัจจุบันได้มีการนำเอา ทฤษฎีการตอบสนองข้อสอบ โดยเฉพาะ โมเดลของราซ์ท (Rasch Model) มาใช้อย่างกว้างขวางในการวัดผลและการทดสอบโดยทั่วไป ซึ่งไรต์(Wright ,1980 : 194 – 196 อ้างถึงใน, ประดิษฐ์ เรืองตระกูล. 2529 : 26) ได้รวบรวมประโยชน์ถึงการนำโมเดลของราซ์ท(Rasch Model)มาใช้ไว้มากมาย แต่อย่างไรก็ตามในงานที่ Rasch Model สามารถทำได้นั้น โมเดลโลจิสติก 3 พารามิเตอร์ย่อมทำได้และอาจทำได้ดีกว่าในบางกรณี ดังนั้นการนำโมเดลโลจิสติก 3 พารามิเตอร์ไปใช้ประโยชน์จึงพอสรุปได้ดังนี้

1. ใช้ในการสร้างคลังข้อสอบ (Item Bank) เนื่องจากข้อสอบที่วิเคราะห์แล้วค่าพารามิเตอร์มีลักษณะคงที่ ดังนั้น ข้อสอบเหล่านี้จึงนำมาใช้สร้างข้อสอบชุดใหม่ตามเกณฑ์ที่ต้องการวัด

2. ใช้วิเคราะห์ข้อสอบ (Item Analysis) การวิเคราะห์ข้อสอบด้วยโมเดล โลจิสติก 3 พารามิเตอร์ จะสามารถแก้ปัญหาที่โมเดลคลาสสิกคอลแก้ไม่ได้ (Hambleton, 1979 : 14 – 15) อันได้แก่

2.1 ค่าสถิติและค่าพารามิเตอร์ต่างขึ้นอยู่กับสภาพกลุ่มตัวอย่างที่ทำการสอบ

2.2 การเปรียบเทียบความสามารถหรือคุณลักษณะ (Trait) ใด ๆ ในแต่ละบุคคลจะเปรียบเทียบกันได้ก็ต่อเมื่อต้องสอบด้วยแบบสอบฉบับเดียวกัน

2.3 ความคลาดเคลื่อนมาตรฐานในการวัดของแต่ละคนเท่ากัน ซึ่ง ลอร์ดและโนวิค(Lord and Novick :1968) ได้แสดงให้เห็นว่าความคลาดเคลื่อนมาตรฐานในการวัดสำหรับคะแนนปานกลางจะมีขนาดเล็กกว่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนสูงหรือคะแนนต่ำ

3. ใช้ในการกำหนดเกณฑ์ของระดับความสามารถของแบบทดสอบอิงเกณฑ์ ผลของการวิเคราะห์ข้อสอบถูกดัดแปลงให้เป็นคะแนนความสามารถของแต่ละกลุ่มบุคคลจึงสามารถเปรียบเทียบกับคะแนนความสามารถซึ่งเป็นเกณฑ์คงที่ได้ ทำให้เราทราบได้ว่า ระดับความสามารถอิงเกณฑ์แต่ละชุด (Minimum Mastery Level) ของข้อสอบแต่ละข้อควรเป็นอย่างไร

4. ใช้วินิจฉัยความสามารถของผู้สอบ (Diagnostic) ในกรณีที่ ICC ของข้อสอบไม่เหมาะสมกับโค้งของโมเดล แสดงว่าบางสิ่งบางอย่างผิดปกติในตัวผู้สอบที่เราควรสนใจแก้ไข

5. ใช้ในการค้นหาความเป็นอคติของข้อสอบ (Item Bias) เมื่อเกิดเหตุ ดังกล่าวในข้อ 3 แสดงว่า มีความสามารถอื่นแฝงเข้ามาในความสามารถที่ต้องการวัด เราก็สามารถที่จะทำการตรวจสอบความลำเอียงของข้อสอบได้

6. ใช้ในการวัดระดับความสามารถของแต่ละบุคคล (Tailoring Test) เราอาจสุ่มข้อสอบที่วิเคราะห์แล้ว และมีความยากแยกตามลำดับมาสอบระดับความสามารถของแต่ละบุคคลได้

7. ใช้ในการจัดชั้นเรียน (Grade – Placement Tailoring) ค่าความสามารถของผู้สอบที่ได้จากการวิเคราะห์ข้อสอบอาจใช้ในการจัดชั้นเรียนที่เหมาะสมกับระดับความสามารถของผู้เรียนเป็นกลุ่ม ๆ ได้

8. ใช้ในการสร้างแบบทดสอบที่ดีที่สุด (Best Test Design) ผลจากการวิเคราะห์ข้อสอบโดย Latent Trait Models สามารถนำไปใช้ในการสร้างข้อสอบที่มีคุณลักษณะต่าง ๆ ตามที่ต้องการได้

9. ใช้ในการเปรียบเทียบคะแนนต่างชุด (Equating Score) ในแบบสอบที่วิเคราะห์แล้ว 2 ชุดที่ต่างกัน แต่วัดในสิ่งเดียวกัน จะสามารถนำคะแนนของผู้สอบในแบบสอบฉบับหนึ่งเปรียบเทียบกับคะแนนของผู้สอบในแบบสอบอีกฉบับหนึ่งได้ ทั้งนี้เพราะคะแนนแต่ละชุดของทฤษฎี IRT นั้นถูกแปลงเป็นคะแนนมาตรฐานที่สามารถเปรียบเทียบกันได้

10. ใช้หาค่าการเดา (Guessing) ในทฤษฎีคลาสสิคอลได้มีการใช้สูตรแก้การเดา แต่ในทางปฏิบัติจริงแล้วถ้าผู้สอบมีเวลาเพียงพอ มีตัวเลือกที่มีประสิทธิภาพ และผู้สอบได้แสดงความสามารถอย่างเต็มที่ การเดาแบบสุ่มน่าจะเกิดขึ้นได้น้อยมาก ถ้าเป็นเช่นนี้สูตรแก้การเดาจึงเป็นเสมือนการลงโทษผู้สอบผู้นั้น สำหรับในทฤษฎี IRT จะไม่ใช้สูตรแก้การเดา แต่จะใช้การพิจารณาค่าการเดา ซึ่งหาได้จากการนำโมเดลทางคณิตศาสตร์เข้ามาช่วยสนับสนุนทำให้หลักการพิจารณาค่าการเดามีความสมเหตุสมผลยิ่งขึ้น

สรุปแนวคิดทฤษฎีการตอบสนองข้อสอบ

จากแนวคิดทฤษฎีตอบสนองข้อสอบมีผู้สรุปไว้ดังนี้ (ชัยพจน์ รักงาม, 2538 : 12-13) แนวคิดทฤษฎีการตอบสนองข้อสอบ (Item Response Theory ; IRT) มีความเชื่อว่าความสามารถของผู้สอบและโอกาสในการตอบข้อสอบถูก สามารถเขียนเป็น สมการได้ และมีความสัมพันธ์กันเป็นโค้งโลจิสติก (Logistic Curve) นั่นคือ มีลักษณะเป็นฟังก์ชันที่เพิ่มขึ้นทางเดียว

จากแนวคิดทฤษฎี IRT สามารถเขียนเป็นโมเดลทางคณิตศาสตร์ได้ทำให้เห็นความสัมพันธ์ระหว่างความสามารถของผู้สอบและโอกาสที่จะตอบข้อสอบถูกชัดเจน เรียกว่า ทฤษฎีแกร่ง (Strength Theory หรือ Hard Theory)

สาระสำคัญของทฤษฎีนี้ ก็คือ ค่าคุณลักษณะของข้อสอบ (a,b และ c) เป็นค่าพารามิเตอร์ประจำข้อสอบ นั่นคือค่าคุณลักษณะข้อสอบจะเป็นค่าคงที่ ไม่ว่าจะนำไปสอบกับผู้สอบกลุ่มใด ต่างกับทฤษฎีดั้งเดิม (Classical Test Theory) ซึ่งความยากของข้อสอบขึ้นอยู่กับกลุ่มนักเรียนที่เรานำข้อสอบนั้นไปสอบ ถ้ากลุ่มนักเรียนนั้นเก่ง ข้อสอบนั้นก็จะถูกวิเคราะห์ว่าข้อสอบนั้นง่าย เพราะคนส่วนใหญ่ในกลุ่มทำได้ ในทำนองเดียวกันถ้านำข้อสอบข้อเดียวกันมาสอบกลุ่มนักเรียนที่อ่อน ข้อสอบข้อนั้นจะถูกวิเคราะห์ออกมาว่ายาก เพราะคนส่วนใหญ่ในกลุ่มทำไม่ได้ แต่สำหรับทฤษฎี IRT นั้น ถ้านำไปสอบกลุ่มนักเรียนอ่อนก็จะได้ความสัมพันธ์ของสมการในตำแหน่งความสามารถต่ำ แต่ถ้าไปสอบกับนักเรียนเก่งก็จะได้ความสัมพันธ์ของสมการในตำแหน่งความสามารถสูง ซึ่งเป็นโค้งความสัมพันธ์อันเดียวกัน

ค่าพารามิเตอร์ของข้อสอบในทฤษฎี IRT แต่ละตัวมีความหมาย ดังนี้

θ : คือ ระดับความสามารถของผู้สอบในสเกลคะแนนมาตรฐาน อยู่ระหว่าง -3 ถึง +3 โดยประมาณ ผู้ที่มีความสามารถปานกลางจะมีค่า θ อยู่ระหว่าง -1.0 ถึง +1.0 ถ้าค่า θ มากกว่า 1.0 แสดงว่า ค่าความสามารถค่อนข้างสูงถ้า (น้อยกว่า -1.0 แสดงว่า ค่าความสามารถค่อนข้างต่ำ)

a : คือ ค่าความชัน (Slope) หรือค่าอำนาจจำแนก (Item Discriminating)ของสมการที่จุดบนโค้งที่ตรงกับค่าโอกาสที่จะทำข้อสอบถูกเท่ากับ 0.5 (ค่าแกนตั้งที่ 0.5) เป็นค่าพารามิเตอร์ที่แสดงถึงความสามารถของข้อสอบในการจำแนกบุคคลที่มีความสามารถต่างไปออกได้มากน้อยเพียงใด ถ้าค่าความชันสูงคนที่มีความสามารถต่างกันเพียงเล็กน้อยก็ส่งผลให้โอกาสที่ทำข้อสอบถูกต่างกันมาก โดยปกติค่าควรจะมากกว่า 0.3 ขึ้นไป

$a > 0.5$	การจำแนกดี
$0.3 < a < 0.5$	การจำแนกปานกลาง
$a < 0.3$	การจำแนกไม่ดี

b : คือ ค่าความยาก (Item Difficulty) ก็คือตำแหน่งบนสเกลของค่าความสามารถ (แกนนอน) ที่ตรงกับ ค่าโอกาสที่จะทำข้อสอบถูกเท่ากับ 0.5 นั้นคือค่าความยากก็คือ ค่าความสามารถของผู้สอบที่มีโอกาสทำข้อสอบข้อนั้นถูกร้อยละ 50 ค่าความยากเป็นค่าที่บอกถึงความเหมาะสมในการที่จะเอาข้อสอบข้อสอบนี้ไปใช้ ถ้าค่าความยากสูงกว่า 1.0 ก็ควรนำไปใช้กับผู้มีความสามารถสูง (กลุ่มเก่ง) เช่น นำไปใช้ในการสอบแข่งขัน ถ้าค่าความยากต่ำกว่า -1.0 ก็ควรนำไปใช้กับกับผู้ที่มีความสามารถต่ำ (กลุ่มอ่อน) เช่น นำไปใช้กับนักเรียนอ่อนเพื่อวินิจฉัยข้อบกพร่องในการเรียนวิชานั้น ส่วนข้อที่มีค่า b อยู่ระหว่าง -1.0 ถึง 1.0 เหมาะสำหรับไปใช้กับนักเรียนที่มีความสามารถปานกลาง เช่น ใช้ในการเรียนการสอนในชั้นปกติ หรือเรียกว่า วัดผลสัมฤทธิ์ นั่นเอง

c : คือ ค่าจุดตัดบนแกนตั้ง (ค่าโอกาสที่จะทำข้อสอบข้อนั้นถูก) เรียกค่าโอกาสการเดา (Item Guessing) เพราะเป็นโอกาสที่จะทำข้อสอบนั้นถูกโดยไม่มีความสามารถ ,มีค่าที่เป็นไปได้อยู่ระหว่าง 0 ถึง 1 ถ้าข้อสอบข้อนั้นไม่สามารถที่จะตอบถูกด้วยการเดาแล้ว $C_i = 0$ และปกติควรจะมีค่าน้อยกว่า 0.3

$c < 0.2$	ตัวเลือกดีมาก
$0.2 < c < 0.3$	ตัวเลือกดี
$c > 0.3$	ตัวเลือกเดาง่าย

งานวิจัยที่เกี่ยวข้องกับทฤษฎีการตอบสนอง

สมพร บุญอิ่ม (2529 :90-92) ได้ศึกษาความคงที่ของพารามิเตอร์ความยากในการวิเคราะห์ข้อสอบด้วยราส์ชโมเดลกับประชากรที่มีความสามารถต่างกัน 3 กลุ่ม คือกลุ่มที่มีความสามารถสูง กลุ่มปานกลางและกลุ่มต่ำ ศึกษาโดยใช้เทคนิคมอนติคาร์โลซิมูเลชันสร้างสถานการณ์จำลองจากเครื่องคอมพิวเตอร์กับกลุ่มตัวอย่างกลุ่มละ 500 คน ข้อสอบ 30 ข้อ ทดลองซ้ำในแต่ละสถานการณ์ 100 ครั้งการศึกษาพบว่า

1. ค่าพารามิเตอร์ความยากของข้อสอบมีความคงที่ในข้อที่มีความยากไม่สูง หรือไม่ต่ำมากกลุ่มที่มีความสามารถต่ำและกลุ่มที่มีความสามารถสูงพารามิเตอร์ความยากมีความคงที่ 19 ข้อ กลุ่มที่มีความสามารถปานกลางมีความคงที่ 22ข้อ

2. การกระจายของค่าความคลาดเคลื่อนมาตรฐานของค่าความยากของข้อสอบเมื่อผู้สอบมีความสามารถต่ำและผู้ที่มีความสามารถสูงมีแนวโน้มไม่เป็นสมมาตร แต่เมื่อผู้สอบมีความสามารถปานกลางจะมีการกระจายเป็นสมมาตร

3. ค่าความคลาดเคลื่อนมาตรฐานของค่าความยากของข้อสอบทั้ง 3 กลุ่ม แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

บุญชิต รอดแก้ว (2533 : 71- 73) ได้เปรียบเทียบค่าพารามิเตอร์ของข้อสอบได้แก่ ค่าอำนาจจำแนก ค่าความยาก และค่าการเดาของข้อสอบ ของกลุ่มที่มีความสามารถทางคณิตศาสตร์แตกต่างกันคือกลุ่มสูง กลาง ต่ำ และกลุ่มผู้สอบที่มีจำนวนไม่เท่ากัน 3 กลุ่ม คือ 1,052 คน กลุ่ม 2,104 คน และกลุ่ม 3,155 คน ซึ่งวิเคราะห์ด้วยโมเดลโลจิสติคสามพารามิเตอร์ โดยศึกษากับกลุ่มตัวอย่างจำนวน 3,155 คน ผลการศึกษาพบว่า ค่าอำนาจจำแนกและค่าการเดาของข้อสอบของกลุ่มผู้สอบที่มีจำนวนไม่เท่ากัน แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ แต่ค่าความยากของข้อสอบกลุ่มผู้สอบที่มีจำนวนไม่
เท่ากัน แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อวยพร วิบูลย์กาญจน์ (2526 : 60) ได้เปรียบเทียบผลการวิเคราะห์แบบทดสอบอุปมาอุปไมยด้วยวิธีคลาสสิคคอลโมเดลและราส์ชโมเดล โดยกลุ่มตัวอย่างเป็นนักเรียนที่สอบคัดเลือกเข้าเรียนต่อในชั้นมัธยมศึกษาปีที่ 1 จำนวน 1,884 คนเครื่องมือที่ใช้เป็นแบบทดสอบอุปมาอุปไมยและคะแนนจากแบบทดสอบมาตรฐานวิชาวิทยาศาสตร์ของสำนักทดสอบทางการศึกษา และ จิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร พบว่าข้อสอบที่ได้จาก
คลาสสิคคอลโมเดลมีมากกว่าราส์ชโมเดลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ค่าความเชื่อมั่นข้อสอบที่ได้รับการคัดเลือกจากคลาสสิคคอลโมเดลสูงกว่าราส์ชโมเดลอย่างมีนัยสำคัญทางสถิติที่

ระดับ .01 และแบบทดสอบที่ประกอบด้วยข้อสอบที่ได้รับการคัดเลือกจากคลาสสิคคอลโมเดลกับราส์ชโมเดล มีความสัมพันธ์กันสูงคือเท่ากับ .9468 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001

กิ่งกาญจน์ เมฆา (2540 :86)ได้ศึกษาความสัมพันธ์ระหว่างค่าพารามิเตอร์ ได้แก่ความยากของข้อสอบ และค่าความสามารถของผู้สอบที่วิเคราะห์จากกลุ่มตัวอย่างในแต่ละขนาดคือ 250 คน 500 คน และ 1,000 คนกับกลุ่มตัวอย่างที่มีขนาดเป็นมาตรฐานคือ 2,000 คน และเปรียบเทียบความสัมพันธ์ของค่าพารามิเตอร์ที่คำนวณจากแต่ละขนาด กับขนาดมาตรฐานของกลุ่มตัวอย่างระหว่างกลุ่มตัวอย่างที่มีขนาดต่างกัน ซึ่งวิเคราะห์ด้วยวิธีราส์ชโมเดล โดยศึกษากับกลุ่มตัวอย่างเป็นนักเรียนมัธยมศึกษาปีที่ 3จำนวน 2,000 คน ผลการศึกษาสรุปได้ดังนี้

1.ค่าความยากของข้อสอบที่วิเคราะห์จากกลุ่มตัวอย่างในแต่ละขนาดกับกลุ่มตัวอย่างที่มีขนาดเป็นมาตรฐานมีความสัมพันธ์กันทางบวกสูง อย่างมีนัยสำคัญทางสถิติที่ระดับ.01ทุกค่า

2.ค่าความสามารถของผู้สอบที่วิเคราะห์จากกลุ่มตัวอย่างในแต่ละขนาดกับกลุ่มตัวอย่างที่มีขนาดเป็นมาตรฐานมีความสัมพันธ์กันทางบวกสูงมากอย่าง มีนัยสำคัญทางสถิติที่ระดับ.01ทุกค่า

3.ค่าสหสัมพันธ์ระหว่างค่าความยากของข้อสอบที่วิเคราะห์จากกลุ่มตัวอย่างที่มีขนาด 250 คนกับกลุ่มตัวอย่างที่มีขนาดเป็นมาตรฐาน กับค่าสหสัมพันธ์ระหว่างค่าความยากของข้อสอบที่วิเคราะห์จากกลุ่มตัวอย่างที่มีขนาด 1,000 คน กับกลุ่มตัวอย่างที่มีขนาดเป็นมาตรฐานแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.01 ค่าสหสัมพันธ์ระหว่างค่าความยากของข้อสอบที่วิเคราะห์จากกลุ่มตัวอย่างที่มีขนาด 250 คนกับกลุ่มตัวอย่างที่มีขนาดเป็นมาตรฐาน กับค่าสหสัมพันธ์ระหว่างค่าความยากของข้อสอบที่วิเคราะห์จากกลุ่มตัวอย่างที่มีขนาด 500 คน กับกลุ่มตัวอย่างที่มีขนาดเป็นมาตรฐานแตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ.05 ส่วนค่าค่าสหสัมพันธ์ระหว่างค่าความยากของข้อสอบที่วิเคราะห์จากกลุ่มตัวอย่างที่มีขนาด 500 คน กับกลุ่มตัวอย่างที่มีขนาดเป็นมาตรฐาน กับค่าสหสัมพันธ์ระหว่างค่าความยากของข้อสอบที่วิเคราะห์จากกลุ่มตัวอย่างที่มีขนาด 1,000 คน กับกลุ่มตัวอย่างที่มีขนาดเป็นมาตรฐานแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

4. ค่าสหสัมพันธ์ระหว่างค่าความสามารถของผู้สอบที่คำนวณจากกลุ่มตัวอย่างในแต่ละขนาดกับขนาดมาตรฐานของกลุ่มตัวอย่างระหว่างกลุ่มตัวอย่างที่มีขนาดต่างกันแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

จงกลณี จันทรังษ์ (2540:บทคัดย่อ) ได้ศึกษาและเปรียบเทียบค่าพารามิเตอร์ของข้อสอบ ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบ และการจำแนกความสามารถของผู้สอบ เมื่อโมเดลและความยาวแบบทดสอบต่างกัน จากการวิเคราะห์ด้วยโมเดลโลจิสติก1, 2 และ 3 พารามิเตอร์โดยใช้แบบทดสอบวัดผลปลายปี กลุ่มทักษะภาษาไทย จำนวน 40 ข้อ เป็นเครื่องมือในการเก็บรวบรวม

ข้อมูล กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2538 ของโรงเรียนสังกัดสำนักงานการประถมศึกษาจังหวัดอุทัยธานี จำนวน 1,909 คน ผลการวิจัยพบว่า

1. ค่าความยากสูงสุดที่ได้จากการวิเคราะห์ด้วยโมเดลโลจิสติก 2 พารามิเตอร์มีค่าสูงกว่าค่าสูงสุดที่ได้จากการวิเคราะห์ด้วยโมเดลโลจิสติก 1 และ3 พารามิเตอร์ ค่าความยากต่ำสุดที่ได้จากการวิเคราะห์ด้วยโมเดลโลจิสติก 2 พารามิเตอร์ มีค่าต่ำกว่าค่าต่ำสุดที่ได้จากการวิเคราะห์ด้วยโมเดลโลจิสติก 1 และ3 พารามิเตอร์ ค่าอำนาจจำแนกที่ได้จากการวิเคราะห์ด้วยโมเดลโลจิสติก 3 พารามิเตอร์ สูงกว่าค่าอำนาจจำแนกที่ได้จากการวิเคราะห์ด้วยโมเดลโลจิสติก 2 พารามิเตอร์ ค่าการเดาที่ได้จากการวิเคราะห์ด้วยโมเดลโลจิสติก 3 พารามิเตอร์ จากแบบทดสอบที่มีความยาว 20, 25, 30, 35 และ 40 ข้อมีค่าต่ำสุดระหว่าง 0.088 ถึง 0.110 และมีค่าสูงสุดระหว่าง 0.423 ถึง 0.484 อันดับที่ค่าความยากที่ได้จากการวิเคราะห์ด้วยโมเดลโลจิสติก 1,2 และ 3 พารามิเตอร์ มีความสอดคล้องกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.1 อันดับที่ค่าความยากที่ได้จากการวิเคราะห์ด้วยโมเดลโลจิสติก 2 และ 3 พารามิเตอร์ มีความสอดคล้องกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในแบบทดสอบที่มีความยาว 20, 35 และ 40 ข้อ และไม่มีความสอดคล้องกันในแบบทดสอบที่มีความยาว 25 และ 30 ข้อ พิสัยของค่าความยากมีค่าระหว่าง 0.072 ถึง 8.049 พิสัยของค่าอำนาจจำแนกมีค่าระหว่าง 0.002 ถึง 3.495

2. ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบที่ได้จากการวิเคราะห์ด้วยโมเดลโลจิสติก 1 และ 2 พารามิเตอร์มีค่าสูงสุดตรงระดับความสามารถปานกลาง ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบที่ได้จากการวิเคราะห์ด้วยโมเดลโลจิสติก 3 พารามิเตอร์มีค่าสูงสุดตรงระดับความสามารถสูงและค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบจะมีค่าสูงขึ้นทุกระดับความสามารถเมื่อเพิ่มขนาดความยาวแบบทดสอบ

3. เมื่อแบบทดสอบยาว 20, 25, 30, 35 และ 40 ข้อ ผลการจำแนกความสามารถของผู้สอบที่ได้จากการวิเคราะห์ด้วยโมเดลโลจิสติก 1 พารามิเตอร์ ไม่แตกต่างกัน ผลการจำแนกความสามารถของผู้สอบที่ได้จากการวิเคราะห์ด้วยโมเดลโลจิสติก 2 และ 3 พารามิเตอร์ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 ในแต่ละโมเดลตามลำดับ เมื่อแบบทดสอบยาว 25, 30, 35 และ 40 ข้อ ผลการจำแนกความสามารถของผู้ตอบที่ได้จากการวิเคราะห์ด้วยโมเดลโลจิสติก 2 และ 3 พารามิเตอร์ไม่แตกต่างกันในแต่ละโมเดล ส่วนการจำแนกความสามารถของผู้สอบที่ได้จากการวิเคราะห์ด้วยโมเดลต่างกัน ผลการจำแนกความสามารถไม่แตกต่างกันในแต่ละขนาดความยาวแบบทดสอบ

วัฒนา ขัดสี (2533: บทคัดย่อ) ได้ศึกษาความสัมพันธ์ระหว่างค่าความยากอำนาจจำแนกของข้อสอบและคะแนนความสามารถในการสอบที่ได้จากการวิเคราะห์โดยทฤษฎีดั้งเดิมกับทฤษฎีการตอบสนองข้อสอบอบเลือกตอบวิชาคณิตศาสตร์ กลุ่มตัวอย่างเป็นนักเรียนชั้นประถม

ศึกษาปีที่ 6 สังกัดสำนักงานการประถมศึกษาจังหวัดสกลนคร ปีการศึกษา 2533 จำนวน 1008 คน ซึ่งเลือกมาโดยการสุ่มแบบแบ่งชั้น พบว่าค่าความยากของข้อสอบที่ได้จากการวิเคราะห์ข้อสอบโดยทฤษฎีดั้งเดิมกับทฤษฎีการตอบข้อคำถามมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ค่าอำนาจจำแนกของข้อสอบที่ได้จากการวิเคราะห์ข้อสอบโดยทฤษฎีดั้งเดิมกับทฤษฎีการตอบข้อคำถามมีความสัมพันธ์กันอย่างไม่มีนัยสำคัญทางสถิติ คะแนนสามารถในการสอบได้จากการวิเคราะห์ข้อสอบโดยทฤษฎีดั้งเดิมกับทฤษฎีการตอบข้อคำถามมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ฟรานซิส (Francis, 1976 อ้างถึงใน อวยพร วิบูลย์กาญจน์, 2526 : 25) ได้ศึกษาความไม่แปรปรวนของค่าความยากของข้อสอบโดยใช้สถานการณ์จำลอง ซึ่งทดลองกับกลุ่มตัวอย่างที่มีระดับความสามารถของผู้สอบ 3 ระดับที่แต่ละกลุ่มมีจำนวนไม่เท่ากัน วิเคราะห์ในแต่ละสถานการณ์ด้วยวิธีโลจิสติคโมเดล 30 ครั้ง ผลการศึกษาพบว่า การประมาณค่าความยากมีความคงที่สำหรับทุกระดับความสามารถ แต่จะให้ผลด้านความเหมาะสมกับโมเดลดีที่สุดกับกลุ่มตัวอย่างที่มีการแจกแจงความสามารถในทางบวก (Positively Skewed Distribution of Ability) และพบว่าขนาดของกลุ่มตัวอย่างที่เหมาะสมกับโลจิสติคโมเดล คือ กลุ่มที่มีจำนวน 1,200 คน

ทินส์เล่ย์และเดวิส (Tinsley and Dawis, 1975 : 325 – 339) ได้ศึกษาความคงที่ของความยากของข้อสอบด้วยราส์ชโมเดล โดยใช้แบบทดสอบ 4 ฉบับ คือ Word Analogy, Number Analogy, Picture Analogy และ Symbol Analogy มีจำนวนข้อ 60, 60, 50 และ 40 ข้อตามลำดับ นำไปทดสอบกับนักศึกษาในมหาวิทยาลัยและนักเรียนระดับมัธยมศึกษาวิเคราะห์ค่าความยากด้วยราส์ชโมเดล 10 ครั้ง แล้วนำค่าความยากของแต่ละครั้งมาหาความสัมพันธ์ โดยวิธี Product – moment correlation ผลปรากฏว่า แบบทดสอบทั้ง 4 ฉบับ ที่มีจำนวนข้อมากกว่า 30 ข้อขึ้นไป มีค่าความยากสัมพันธ์กันสูงมากและถ้าตัดข้อที่ไม่เหมาะสม (misfit) กับโมเดลออก จะทำให้ค่าความยากมีความสัมพันธ์กันสูงมากขึ้น

โกลด์แมนและราชู (Goldman and Raju, 1986 : 11 – 19) ได้ศึกษาผลของขนาดกลุ่มตัวอย่างที่แตกต่างกัน 3 กลุ่ม คือ 250 คน, 500 คน และ 1,000 คน เมื่อเทียบกับจำนวนของกลุ่มตัวอย่างที่เป็นมาตรฐาน 3,000 คน ในการกะประมาณค่าพารามิเตอร์ของผู้สอบ และพารามิเตอร์ของข้อสอบ ซึ่งวิเคราะห์โดยราส์ชโมเดลและโมเดล 2 พารามิเตอร์ ผลการศึกษาพบว่ากลุ่มตัวอย่าง 250 คน มีความเหมาะสมสำหรับราส์ชโมเดล แต่ถ้าสำหรับโมเดล 2 พารามิเตอร์จะต้องใช้กลุ่มตัวอย่าง 1,000 คนขึ้นไป จึงจะสามารถกะประมาณค่าพารามิเตอร์ได้ถูกต้อง

งานวิจัยที่เกี่ยวข้องกับความถนัดด้านเหตุผล

เกี่ยวกับความสัมพันธ์ของความถนัดด้านเหตุผล กับความสามารถด้านอื่น ๆ นั้น บราวน์ และจอนห์สัน (Brown and Johnson. 1952 : 3 – 4) พบว่าเหตุผลเชิงนามธรรมและเหตุผลเชิงภาษา เป็นองค์ประกอบที่สำคัญที่ส่งผลต่อการเรียนวิชาวิทยาศาสตร์ และวิศวกรรม

ฮูห์ลแมน และแอนเดอร์สัน (Johnson. 1955 : 411 ; citing Kuhlman and Anderson. n.d. : unpaged) ได้ศึกษาความสัมพันธ์ระหว่างการคิดหาเหตุผลกับสติปัญญา พบว่ามีความสัมพันธ์กันเท่ากับ 0.22 และการคิดหาเหตุผลกับองค์ประกอบด้านภาษา เท่ากับ 0.52 ส่วน จาคอบ (Jacob. n.d. : unpaged อ้างถึงใน Johnson. 1955 : 411) ได้ศึกษาเด็กเกรด 10 ในเรื่องเดียวกัน พบว่าการคิดหาเหตุผลมีความสัมพันธ์กับสติปัญญา เท่ากับ 0.67 และ องค์ประกอบด้านภาษาเท่ากบ 0.58

เธอร์สโตน (Thurstone. n.d. : unpaged อ้างถึงใน Johnson. 1955 : 410) ศึกษาค่าสหสัมพันธ์ของการคิดหาเหตุผล กับความสามารถด้านต่าง ๆ โดยศึกษากับเด็กอายุ 10 – 18 ปี จำนวน 1,000 คน โดยใช้แบบทดสอบการคิดหาเหตุผลสัมพันธ์กับความสามารถด้านจำนวนเท่ากับ 0.54 ด้านความคล่องแคล่วในการใช้คำเท่ากับ 0.48 ด้านภาษาเท่ากับ 0.54 ด้านมิติสัมพันธ์ เท่ากับ 0.38 ด้านความจำเท่ากับ 0.39 และด้านความสามารถทั่ว ๆ ไป เท่ากับ 0.84

คราวเดอร์ (Crowder. 1957 : 512 – 517) ได้ศึกษาความตรงของแบบทดสอบ โฮลซิงเจอร์ – คราวเดอร์ ยูนิ แฟคเตอร์ (Holzinger Crowder Uni Factor Test) ข้อสอบชุดนี้ประกอบด้วยข้อสอบวัดความสามารถทางสมองย่อยๆ 4 ชนิด คือ ความสามารถด้านภาษา ด้านมิติสัมพันธ์ ด้านตัวเลข และด้านเหตุผล ปรากฏว่าข้อสอบชุดนี้มีความตรงต่อวิชาคณิตศาสตร์ดังนี้ คือ 0.51, 0.35, 0.53 และ 0.60 ตามลำดับ)

ฮารูตูเนียน (Harootunion. 1959 : 203 – 204) ได้ศึกษาความสัมพันธ์ระหว่างการคิดหาเหตุผล การเรียนรู้ และสติปัญญาของเด็กเกรด 8 จำนวน 88 คน โดยใช้แบบทดสอบการคิดหาเหตุผล 3 ชุด คือ The Davis – Eells Games, Number Series และ Maw Critical Thinking พบว่าความสัมพันธ์ระหว่างการเรียนกับสติปัญญาอยู่ระหว่าง 0.14 – 0.65 สติปัญญากับการคิดหาเหตุผล 0.43 – 0.63 และการเรียนรู้กับการคิดหาเหตุผล 0.06 – 0.58 ในปีเดียวกัน สตินสัน (Stinson. 1959 : 103 – 104) ได้ใช้แบบทดสอบ DAT เป็นตัวพยากรณ์เกรดเฉลี่ยของนักเรียนจาก เมเปิลวู้ด มิสซูรี (Maplewood Missuri) พบว่าแบบทดสอบด้านเหตุผลกับเกรดเฉลี่ยมีความสัมพันธ์เท่ากับ 0.34

กูดแมน (Goodman. 1961 : 436) ได้ทำการศึกษาค้นคว้าเกี่ยวกับความสามารถทางสมองด้านต่าง ๆ ที่มีผลต่อการเรียนวิชาวิทยาศาสตร์ สาขาเคมี โดยศึกษากับนักเรียนระดับวิทยาลัย 113 คน ปรากฏว่า ค่าสหสัมพันธ์ระหว่างวิชาเคมี กับความสามารถด้านเหตุผลเท่ากับ 0.43 มิติสัมพันธ์เท่ากับ 0.25 ภาษาเท่ากับ 0.28 และความจำเท่ากับ 0.25 ซึ่งจากการศึกษาของกูดแมน จะเห็นว่า วิชาเคมีมีความสัมพันธ์กับความสามารถทางสมองด้านเหตุผลมากที่สุด

อดัมส์ และทอร์เกอร์สัน (Adams and Torgerson. 1964 : 124 – 134) ได้วิเคราะห์องค ประกอบของการคิดหาเหตุผล และพบว่าประกอบด้วยความสามารถของสมองด้านการคิดคำนวณ ด้านเหตุผลทั่วไป ด้านการคิดหาเหตุผลแบบอนุมาน และด้านความสัมพันธ์ของการคิด

มาร์ติน (Martin. 1964 : 2547 – 2548) ซึ่งศึกษาสมรรถภาพทางสมองด้านการคิดหาเหตุ ผล ความเข้าใจในการอ่าน และความคล่องแคล่วในการคำนวณ ที่มีต่อการแก้ปัญหาโจทย์เลขคณิต วิชาเลขคณิตกับความสามารถทางสมองด้านเหตุผล เท่ากับ 0.61 ความเข้าใจในการอ่านเท่ากับ 0.64 ความคล่องแคล่วในการคำนวณเท่ากับ 0.60

คอร์แมน (Dorman. 1990 : 1492 A) ได้ศึกษาผลกระทบของการเข้าเรียน แนวทางในการ แก้ปัญหา ความสามารถในการจำแนก ความสามารถด้านเหตุผลและความถนัดทางการเรียน ของ นักเรียนชั้นมัธยมศึกษา โดยใช้แผนการสอนที่เขียนขึ้นมาอย่างดีให้นักเรียนเข้าเรียน มีกลุ่มที่เข้าเรียนกับ กลุ่มที่ไม่ได้เข้าเรียน ผลการทดลองปรากฏว่า ผู้เข้าเรียนตามแผนการสอนนี้มีการเปลี่ยนแปลงคะแนน อย่างมีนัยสำคัญทางสถิติ ในการทำแบบทดสอบการจำแนกแยกแยะในการแก้ปัญหา และแบบทดสอบ เหตุผลทางตรรกศาสตร์ ค่า ANOVA ของนักเรียนที่ได้รับการสอนจะมีการเปลี่ยนแปลงคะแนนอย่างมีนัย สำคัญทางสถิติ โดยค่าเฉลี่ยเท่ากับ 119 ส่วนคะแนนของกลุ่มเปรียบเทียบเท่ากับ85 ANCOVA ควบคุม ความแตกต่างของเชื้อชาติกับเพศ ความสามารถในการอ่าน และค่าเฉลี่ย แสดงให้เห็นว่าผู้ที่ได้รับการสอน ตามแผนการสอน คะแนนของแบบทดสอบ SAT ด้านภาษาจะมีผลในทางบวก ส่วนคะแนนของแบบ ทดสอบ SAT ด้านคณิตศาสตร์มีผลเพียงเล็กน้อย สรุปได้ว่าผู้เข้าร่วมในโครงการจะมีผลต่อการแก้ปัญหา ความสามารถในการจำแนกแยกแยะ ความสามารถด้านเหตุผล และความถนัดทางการเรียน ในทางบวกสูง

เมย์ตา (Mayta. 1991 : 2604 A) ได้ศึกษาความสัมพันธ์ระหว่างการคำนวณคณิตศาสตร์ กับความสามารถด้านเหตุผลที่ประจักษ์ชัด ความถนัดทางคณิตศาสตร์ ระดับการศึกษาของญาติ การพูดภาษาอังกฤษ และภาษาสเปน ของคนในครอบครัว ขนาดของครอบครัว วิชาทางการศึกษา ของวัยรุ่นชาย พบว่า ความสามารถด้านเหตุผลเชิงประจักษ์ วิชาทางการศึกษา และขนาดของ ครอบครัวมีผลต่อความแปรปรวนของความสามารถในการคำนวณทางคณิตศาสตร์ ความสามารถ ด้านเหตุผลเชิงประจักษ์ และวิชาทางการศึกษาสามารถพยากรณ์การคำนวณคณิตศาสตร์ ระดับการ ศึกษาของญาติเป็นตัวพยากรณ์ที่ไม่ดีนักต่อการคำนวณคณิตศาสตร์ การพูดภาษาอังกฤษของคนใน ครอบครัว และญาติจะส่งผลต่อการเล่าเรียนมากกว่ากลุ่มที่พูดภาษาสเปน ส่วนตัวแปรอื่น ๆ ไม่มีผล ต่อการพยากรณ์การคำนวณคณิตศาสตร์สำหรับครอบครัวที่มีความสามารถทางการคำนวณ คณิตศาสตร์ต่ำ ปรากฏว่าส่งผลต่อการคำนวณคณิตศาสตร์อยู่ในขั้นก่อนเชิงประจักษ์ ครอบครัวทั้ง หมดไม่เห็นความสำคัญของภาษา หรือความสามารถทางคณิตศาสตร์