



บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การเรียนการสอนประกอบด้วยองค์ประกอบ 3 ส่วน คือ การกำหนดจุดมุ่งหมายการเรียน รู้ การจัดการเรียนการสอน และการวัดและประเมินผลการศึกษา ซึ่งการวัดและประเมินผลศึกษานับว่าเป็นกระบวนการที่สำคัญยิ่งขั้นตอนหนึ่งในการติดตามและตรวจสอบผลการเรียนการสอนว่าบรรลุตามจุดมุ่งหมายที่ตั้งไว้หรือไม่ สามารถกระทำได้โดยใช้แบบวัดเป็นสิ่งเร้าให้ผู้สอบแสดงพฤติกรรมการตอบสนองออกมา เพื่อนำพฤติกรรมเหล่านั้นมาแปลผลว่าบุคคลนั้นมีคุณลักษณะหรือความสามารถที่ต้องการวัดมากน้อยเพียงใด (อุทุมพร จามรราน, 2535; ศิริชัย กาญจนวาสี, 2538)

ในการวัดความสามารถในด้านใดด้านหนึ่งของบุคคล ผู้วัดจะต้องระบุคุณลักษณะที่ต้องการวัดให้ชัดเจน ดำเนินการสร้างข้อคำถามให้วัดได้ตรงและครอบคลุมคุณลักษณะนั้น โดยต้องสามารถแสดงให้เห็นว่าข้อคำถามนั้นให้ผลการวัดที่มีความสม่ำเสมอ คงเส้นคงวา และก่อนที่จะนำผลการวัดไปใช้ต้องพิจารณาเห็นว่า แบบแผนการตอบของผู้ตอบเป็นแบบแผนการตอบที่ตรงตามที่คาดหวังไว้หรือไม่ (Wright and Stone, 1983) ซึ่งฮาร์นิช (Harnisch, 1983) ได้กล่าวว่า การวิเคราะห์ผลการสอบด้วยทฤษฎีดั้งเดิมโดยใช้คะแนนรวมหรือคะแนนการตอบข้อสอบที่ถูกยังไม่สามารถบอกความสามารถที่แท้จริงของผู้สอบได้ เนื่องจากคะแนนรวมประกอบด้วยคะแนนความสามารถจริงกับคะแนนความคลาดเคลื่อน นอกจากนี้ ทฤษฎีดั้งเดิมยังมีจุดอ่อนอีกหลายประการ นั่นคือ ประการแรก ข้อมูลที่ได้ไม่สามารถบอกได้ว่าผู้สอบทำอะไรได้บ้างหรือมีความบกพร่องตรงไหน เนื่องจากจะนำผลการสอบไปเทียบกับกลุ่มเท่านั้น (สำเร็จ บุญเรืองรัตน์, 2529) ประการที่สอง ความลำเอียงของข้อสอบอันเนื่องมาจากภาษา เพศ ศาสนา วัฒนธรรม สิ่งแวดล้อมและถิ่นอาศัยที่แตกต่างกัน จะมีอิทธิพลต่อคะแนนรวมของผู้สอบ ตลอดจนทำให้แบบแผนการตอบข้อสอบของผู้สอบบางคนแตกต่างไปจากคนอื่น ประการที่สาม การวิเคราะห์ผลการสอบจะนำเฉพาะข้อที่ผู้สอบตอบถูกมาพิจารณา ไม่ได้นำเอาข้อที่ตอบผิดมาร่วมพิจารณาด้วย ซึ่งการตอบข้อสอบผิดก็มีผลต่อการวินิจฉัยความบกพร่องของผู้สอบ (Linn and Harnisch, 1981; Lord, 1977) ดังนั้นนักวัดผลจึงได้พัฒนาวิธีการวิเคราะห์ผลการสอบโดยใช้แบบแผนการตอบข้อสอบเป็นรายบุคคล (item response pattern) ซึ่งจะพิจารณาทั้งข้อที่ผู้สอบตอบถูกและข้อที่ตอบ

ผิด โดยมีแนวความคิดว่า คะแนนรวมเพียงอย่างเดียวไม่เพียงพอที่จะประมาณค่าความสามารถที่แท้จริงของผู้สอบได้ เพราะผู้สอบที่มีคะแนนรวมที่เท่ากันอาจมีความสามารถไม่เท่ากัน ถ้ามีแบบแผนการตอบข้อสอบไม่เหมือนกัน (Smith, 1986) กล่าวได้ว่า การวิเคราะห์ผลการสอบเพื่อตรวจสอบความสามารถ หรือวินิจฉัยข้อบกพร่องของผู้สอบให้ชัดเจนและยุติธรรมนั้น ควรพิจารณาทั้งคะแนนรวมและแบบแผนการตอบข้อสอบของผู้สอบเป็นรายบุคคล โดยพิจารณาว่าการตอบของผู้สอบเหมาะสมหรือสอดคล้องกับความสามารถที่แท้จริงหรือไม่ โดยใช้ดัชนีบ่งชี้ความเหมาะสมของบุคคล (person-fit indexes) เพื่อตรวจสอบความสอดคล้องดังกล่าว (กุสุมา สุวรรณแก้ว, 2540)

การวิเคราะห์ดัชนีบ่งชี้ความเหมาะสมของบุคคล (person-fit) มีวัตถุประสงค์เพื่อบ่งชี้บุคคลที่มีแบบแผนการตอบข้อสอบที่ผิดปกติ โดยพิจารณาจากแบบแผนการตอบข้อสอบเป็นรายบุคคล (item response pattern) ซึ่งเป็นเมทริกซ์ที่แสดงการตอบข้อสอบเป็นรายข้อของผู้สอบแต่ละคน ทั้งนี้ในการวิเคราะห์จะพบทั้งผู้สอบที่มีแบบแผนการตอบที่ปกติและผิดปกติรวมอยู่ด้วยกัน ซึ่งแบบแผนการตอบที่ปกติเป็นแบบแผนการตอบที่ตอบได้สอดคล้องกับแนวคิดของกัตแมนที่เรียกว่า กัตแมนสมบูรณ์ (Perfect Gutman) (Levine and Rubin, 1979; Wright and Stone, 1979; and Van Der Fierder, 1982. cited in Meijer, 1996) ซึ่งแนวคิดของกัตแมนเชื่อว่า เมื่อผู้สอบตอบข้อสอบที่มีความยากระดับหนึ่งถูก ผู้สอบคนนั้นจะต้องตอบข้อสอบทุกข้อที่มีความยากน้อยกว่าหรือเท่ากับข้อสอบข้อนั้นได้ถูกต้องด้วย และเมื่อผู้สอบคนนั้นตอบข้อสอบที่มีความยากระดับหนึ่งผิด ผู้สอบคนนั้นจะตอบข้อสอบที่มีความยากมากกว่าข้อสอบข้อนั้นผิดด้วย ดังนั้นแบบแผนการตอบข้อสอบที่ปกติ (nonaberrant response pattern) ผู้สอบจะตอบข้อสอบได้ถูกต้องตามที่คาดหวังไว้ในโมเดล กล่าวคือ ผู้สอบสามารถตอบถูกในข้อที่มีค่าความยากต่ำกว่าหรือเท่ากับความสามารถของตน หรือตอบผิดในข้อที่มีค่าความยากสูงกว่าระดับความสามารถของตน ซึ่งผู้ที่มีแบบแผนการตอบข้อสอบที่ปกตินี้แสดงว่าเป็นบุคคลที่เหมาะสมกับโมเดล (person-fit) ส่วนแบบแผนการตอบข้อสอบที่ผิดปกติ (aberrant response pattern) ผู้สอบจะตอบข้อสอบแตกต่างไปจากที่คาดหวังไว้ในโมเดล กล่าวคือ ผู้สอบตอบผิดในข้อที่มีค่าความยากต่ำกว่าหรือเท่ากับความสามารถของตน หรือตอบถูกในข้อที่มีค่าความยากสูงกว่าระดับความสามารถของตน ซึ่งบุคคลที่มีแบบแผนการตอบที่ผิดปกติแสดงว่าเป็นบุคคลที่ไม่เหมาะสมกับโมเดล (person-misfit) โดยกัตแมนได้กล่าวว่า ความผิดปกติของแบบแผนการตอบมี 2 ลักษณะ คือ ลักษณะแรก เกิดจากการที่ผู้สอบตอบข้อสอบข้อที่มีความยากน้อยกว่าหรือเท่ากับความสามารถที่แท้จริง (true ability)

ผิด ลักษณะที่สอง เกิดจากการที่ผู้สอบตอบข้อสอบที่มีความยากสูงกว่าความสามารถที่แท้จริง ได้ถูกต้อง (Cudeck, 1980; พนิชา สังข์เพชร, 2535)

สำหรับดัชนีที่ใช้วิเคราะห์แบบแผนการตอบข้อสอบรายบุคคลนั้น มีผู้ศึกษาและพัฒนา ไว้หลายดัชนีด้วยกัน สามารถแบ่งตามเทคนิคของการวิเคราะห์ได้ 2 กลุ่ม คือ (Harnisch and Linn, 1981; สุพันธ์ ศลโกสม, 2530; กุสุมา สุวรรณแก้ว, 2540)

กลุ่มแรก วิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) โดยการหาความสัมพันธ์ของแบบแผนการตอบถูกหรือผิดเป็นรายบุคคลกับคะแนนรวม ซึ่งแบบแผนการตอบข้อสอบที่ถูกต้อง และตรงตามลักษณะของข้อสอบที่มีคุณภาพดี ผู้สอบควรจะตอบข้อสอบที่ง่ายได้และตอบข้อสอบที่ยากไม่ได้ และคนที่ได้คะแนนรวมเท่ากันควรตอบคำถามได้เหมือนกัน กล่าวคือ ต้องมีแบบแผนการตอบถูกหรือผิดเหมือนกันในกลุ่มที่มีคะแนนเท่ากัน แต่ถ้าในกลุ่มที่ได้คะแนนเท่ากันมีรูปแบบการตอบถูกและผิดไม่เหมือนกัน แสดงว่ามีความผิดปกติเกิดขึ้น อาจเนื่องมาจากข้อสอบหรือตัวผู้สอบ ดัชนีกลุ่มนี้ได้แก่ ดัชนีชี้หน้าของซาโต้ (Sato, 1970) ดัชนียู (U-index) ของเวนเดอร์ (Vender, 1977) ดัชนีความสามารถของบุคคลภายในกลุ่ม (Norm conformity index) ของทาทซูกะและทาทซูกะ (Tatsuoka and Tatsuoka, 1980) ดัชนีที่แสดงความสอดคล้องไม่สอดคล้อง (Agreement and disagreement indices) ของเคนและแบรนแนน (Kane and Brennan, 1980) และดัชนีความสอดคล้องของแบบแผนการตอบข้อสอบรายบุคคลของ พนิชา สังข์เพชร (2535)

กลุ่มที่สอง ใช้หลักการของทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) ดัชนีกลุ่มนี้คำนวณโดยการประมาณค่าความน่าจะเป็นที่ผู้สอบแต่ละคนจะตอบข้อสอบแต่ละข้อ ได้ถูกต้อง เปรียบเทียบกับผลการตอบในข้ออื่นๆ เพื่อดูว่าจะผิดไปจากแบบแผนการตอบข้อสอบที่ปกติหรือไม่ ถ้าผิดปกติแสดงว่าผู้สอบคนนั้นไม่เหมาะสมกับโมเดลการตอบข้อสอบ (person misfit) หรือผู้สอบคนนั้นมีแบบแผนการตอบที่ผิดปกติ (aberrant response pattern) ดัชนีกลุ่มนี้ได้แก่ ดัชนีของเลวินและรูบิน (Levine and Rubin, 1979) ดัชนีที่พัฒนาโดยดราสโกว์ (Drasgow, 1978) การทดสอบไค-สแควร์ของบุคคล (Chi-Square Test of Person-fit) สถิติ U_3 (U_3 statistic) ที่พัฒนาโดยโมแคนและ เลวิส (Mokken and Lewis, 1982) ดัชนี DFC_1 และ DFC_2 ของ สาราญ มีแจ้ง (2533) และดัชนีชี้เตือนมาตรฐานหมายเลข 4 (Standardized Extended Caution Index 4, SECI4) ของทาทซูกะ (Tatsuoka, 1984) ทั้งนี้ รัดเนอร์ (Rudner, 1983) ซึ่งได้ใช้ข้อมูลมอนติคาร์โ (Monte Carlo Data) ศึกษาเปรียบเทียบดัชนีต่างๆ พบว่าดัชนีที่ใช้หลักการของทฤษฎี

การตอบสนองข้อสอบ (IRT) สามารถตรวจค้นแบบแผนการตอบข้อสอบที่ผิดปกติได้ดีกว่าดัชนีที่ใช้หลักการของทฤษฎีการทดสอบแบบดั้งเดิม (CTT) ซึ่งพิจารณาแบบแผนการตอบถูกและผิดแล้วหาค่าสหสัมพันธ์กับคะแนนรวม

สำหรับวิธีประมาณค่าดัชนีบ่งชี้ความเหมาะสมของบุคคลนั้น รอสท์และเดเวียร์ (Rost and Davier, 1994) ได้สรุปวิธีประมาณค่าดัชนีบ่งชี้ความเหมาะสมของบุคคลตามทฤษฎีการตอบสนองข้อสอบไว้ 3 วิธี คือ วิธีแรก การประมาณค่าจากคะแนนส่วนที่เหลือ (The Score Residual Approach) พัฒนาโดยไรท์และสโตน (Wright & Stone, 1979) ไรท์ (Wright, 1980) และราสช์ (Rasch, 1980) วิธีนี้สามารถตรวจค้นแบบแผนการตอบข้อสอบได้ทั้งการให้คะแนนแบบ 2 ค่าและหลายค่า ดัชนีกลุ่มนี้คือ W_1 (Wright, 1979) และ W_3 (Rudner, 1983) ซึ่งประมาณค่าดัชนีโดยประมาณค่าสัดส่วนระหว่างผลรวมของกำลังสองของผลต่างของคะแนนที่สังเกตได้กับค่าที่คาดหวังจากโมเดล ซึ่งคะแนนสองชุดนี้เป็นสัดส่วนของความแปรปรวนของคะแนนส่วนที่เหลือซึ่งไม่เป็นอิสระจากกัน มีการแจกแจงเป็นแบบ ที (t-distribution) ทดสอบความแตกต่างแบบสองทาง (two-tailed test) บุคคลซึ่งไม่เหมาะสมกับโมเดลการตอบ คือบุคคลที่มีค่าสถิติบ่งชี้ความเหมาะสมของบุคคลแตกต่างจาก 0 อย่างมีนัยสำคัญที่ระดับ 0.05 นั่นคือมีค่า $|W_1|$ มากกว่า 2.00 วิธีที่สอง เป็นการทดสอบไค-สแควร์ของบุคคล (The Chi-Square Test) พัฒนาโดย ไรท์และปัญจปากีสถาน (Wright & Panchpakesan, 1969) บอค (Bock, 1972) และวอลเลนเบอร์ก (Wollenberg, 1979) โดยเปรียบเทียบสัดส่วนของความถี่ของการตอบที่สังเกตได้กับความถี่ของการตอบที่คาดหวังของกลุ่มผู้ตอบข้อสอบ ทดสอบความแตกต่างด้วยสถิติทดสอบไค-สแควร์ (χ^2 -test) แบบทางเดียว (one-tailed test) บุคคลที่มีแบบแผนการตอบข้อสอบที่ผิดปกติ คือ บุคคลที่ค่าดัชนีบ่งชี้ความเหมาะสมแตกต่างจาก 0 อย่างมีนัยสำคัญที่ระดับ 0.05 นั่นคือ มีค่าสถิติ χ^2 มากกว่า 9.48 วิธีที่สาม การประมาณค่าด้วยฟังก์ชันไลคิลิฮูด (Likelihood-based Approach) พัฒนาโดย เลวินและรูบิน (Levine & Rubin, 1979) และดราสโกว์ (Drasgow, 1985) ดัชนีที่ประมาณค่าด้วยวิธีนี้ ได้แก่ ดัชนี L_2 เป็นการคำนวณค่าดัชนีโดยการประมาณค่าความน่าจะเป็นที่ผู้สอบแต่ละคนจะตอบข้อสอบแต่ละข้อได้ถูกต้อง เปรียบเทียบกับผลการตอบข้อสอบในข้อนั้นๆ แล้วแปลงค่าสถิติบ่งชี้ความเหมาะสมของบุคคลเป็นค่ามาตรฐาน (Z-Value) จากนั้นทดสอบความแตกต่างแบบสองทาง (two-tailed test) บุคคลที่มีแบบแผนการตอบข้อสอบที่ผิดปกติ คือบุคคลที่ค่าดัชนีบ่งชี้ความเหมาะสมมีค่าแตกต่างจาก 0 อย่างมีนัยสำคัญที่ระดับ 0.05 นั่นคือ มีค่า $|L_2|$ มากกว่า 1.96

จากการศึกษางานวิจัยที่ผ่านมา พบว่า ในการวิเคราะห์บุคคลที่เหมาะสม (person-fit) ด้วยดัชนีบ่งชี้ความเหมาะสมของบุคคล (person-fit indices) นั้น มีข้อค้นพบที่น่าสนใจหลายประเด็นด้วยกัน อาทิเช่น รัดเนอร์ (Rudner, 1983) ได้ศึกษาเปรียบเทียบประสิทธิผลของการวิเคราะห์แบบแผนการตอบข้อสอบด้วยดัชนี L_2 ซึ่งประมาณค่าด้วยฟังก์ชันโลดิลิสต์ และดัชนี W_1 ซึ่งประมาณค่าด้วยคะแนนส่วนที่เหลือ โดยจำลองคะแนนของผู้สอบให้สูงกว่าความเป็นจริง (spuriously high) และต่ำกว่าความเป็นจริง (spuriously low) พบว่า ดัชนี L_2 สามารถตรวจค้นผู้สอบซึ่งไม่เหมาะสมกับโมเดลการตอบได้ดีกว่าดัชนี W_1 นอกจากนี้ดราสโกว์และคณะ (Drasgow et al., 1987) ได้จำลองข้อมูลคะแนนผู้สอบให้สูงกว่าความเป็นจริง (spuriously high) และต่ำกว่าความเป็นจริง (spuriously low) ในระดับที่แตกต่างกัน คือ 15 % และ 30 % แล้วใช้ดัชนีตามทฤษฎีการตอบสนองข้อสอบตรวจค้นแบบแผนการตอบที่ผิดปกติ พบว่า เมื่อจำลองให้ผู้สอบมีคะแนนสูงกว่าความเป็นจริง ดัชนี L_2 และดัชนี W_1 สามารถตรวจค้นผู้สอบซึ่งไม่เหมาะสมกับโมเดลการตอบข้อสอบได้ร้อยละ 75 และ 46 ตามลำดับ และเมื่อจำลองให้ผู้สอบมีคะแนนต่ำกว่าความเป็นจริง ดัชนี L_2 และดัชนี W_1 สามารถตรวจค้นผู้สอบซึ่งไม่เหมาะสมกับโมเดลการตอบข้อสอบได้ร้อยละ 73 และ 21 ตามลำดับ

เบอร์นบอม (Birbaum, 1986) ได้ศึกษาเปรียบเทียบดัชนีบ่งชี้ความเหมาะสมของบุคคล ซึ่งใช้หลักการของทฤษฎีการตอบสนองข้อสอบ ทั้งหมด 9 ดัชนี ได้แก่ ดัชนีชี้เตือนของทาทสโวกะที่ยังไม่ได้ปรับความเป็นมาตรฐาน (Extended Caution Index : ECI) คือ ดัชนี ECI_1 , ECI_2 , ECI_4 ดัชนีชี้เตือนมาตรฐานของทาทสโวกะ (Standardized Extended Caution Index , SECI) คือ ดัชนี ECI_2 , ECI_2 , ECI_4 ดัชนีของเลวิฟและรูบิน (L_0) ดัชนีของเลวินและดราสโกว์ (L_2) และดัชนีที่เสนอโดยไรท์ (U_2) โดยได้แบ่งกลุ่มตัวอย่างเป็น 3 กลุ่ม คือ กลุ่มนักเรียนที่ให้ความร่วมมือในการทำข้อสอบ (cooperative examinees) กลุ่มที่ไม่ให้ความร่วมมือ (uncooperative examinees) และกลุ่มที่สุ่มต่างๆไป ใช้การพิจารณาการเขียนชื่อบนกระดาษคำตอบของนักเรียนเป็นเกณฑ์ตัดสินว่านักเรียนให้ความร่วมมือหรือไม่ ถ้าเขียนชื่อแสดงว่าให้ความร่วมมือ ถ้าไม่เขียนแสดงว่าไม่ให้ความร่วมมือ ผลการเปรียบเทียบความสัมพันธ์พบว่าดัชนี L_2 มีความสัมพันธ์กับดัชนีอื่นๆ สูงมากที่สุดและเป็นดัชนีที่มีความสัมพันธ์กับคะแนนรวมน้อยที่สุด

ในประเด็นเกี่ยวกับคุณภาพของแบบสอบหลังการวิเคราะห์บุคคลที่เหมาะสมกับโมเดลการตอบนั้น ชมิตทและคณะ (Schmitt et al. 1993) ได้ศึกษาความตรงตามเกณฑ์สัมพันธ์ (criterion-related validity) ของแบบสอบวัดความสามารถทางเครื่องกล ซึ่งเป็นแบบสอบชุด

(Batteries Test) โดยวิเคราะห์แบบแผนการตอบข้อสอบรายบุคคลด้วยดัชนี L_{2m} ซึ่งเป็นดัชนี L_2 ที่พัฒนาขึ้นให้สามารถใช้วิเคราะห์แบบสอบชุด เมื่อวิเคราะห์หาคุณภาพของแบบสอบ พบว่าแบบสอบซึ่งวิเคราะห์โดยใช้แบบแผนการตอบข้อสอบรายบุคคล มีความตรงตามเกณฑ์สัมพัทธ์สูงกว่าแบบสอบซึ่งวิเคราะห์โดยไม่พิจารณาแบบแผนการตอบ การศึกษาครั้งนี้ของซมิทและคณะได้สรุปสอดคล้องกับดราสโกว์ (Drasgow, 1987) กล่าวคือเมื่อวิเคราะห์แบบแผนการตอบข้อสอบรายบุคคล เพื่อวิเคราะห์บุคคลที่เหมาะสมกับโมเดลการตอบข้อสอบแล้วแยกบุคคลที่ไม่เหมาะสมกับโมเดลการตอบข้อสอบออกจากการวิเคราะห์ จะทำให้ค่าความตรงของแบบสอบที่วิเคราะห์เฉพาะบุคคลที่มีแบบแผนการตอบปกติเพิ่มขึ้น ดังนั้น การวิเคราะห์ความเหมาะสมของบุคคลกับโมเดล (analysis of person-fit) จึงมีบทบาทสำคัญในการสร้างและศึกษาคุณภาพแบบสอบ ซึ่งดัชนีบ่งชี้ความเหมาะสมของบุคคล (person-fit index) ที่มีคุณภาพจะต้องสามารถบ่งชี้บุคคลซึ่งมีแบบแผนการตอบผิดปกติได้อย่างถูกต้องแม่นยำ เป็นผลให้แบบสอบที่วิเคราะห์หลังจากนำบุคคลที่ไม่เหมาะสมกับโมเดลการตอบข้อสอบออกจากการวิเคราะห์มีคุณภาพเพิ่มขึ้น (กุสุมา สุวรรณแก้ว, 2540)

จากการศึกษาของเนอริง (Nering, 1995) ซึ่งได้จำลองข้อมูลการตอบข้อสอบของผู้สอบให้เป็นแบบทวนเดียว (uniform) และมีการแจกแจงเป็นโค้งปกติ เพื่อศึกษาลักษณะการแจกแจงของค่าสถิติบ่งชี้ความเหมาะสมของบุคคลซึ่งประมาณค่าด้วยฟังก์ชันโลดิสติก (L_2 index) พบว่าลักษณะการแจกแจงของสถิติ L_2 มีการแจกแจงเป็นโค้งปกติ และการประมาณค่าความสามารถของผู้สอบส่งผลต่อดัชนีมากกว่าระดับความสามารถที่ทราบค่าแล้ว วิธีการประมาณค่าพารามิเตอร์โดยวิธีของเบย์ให้ผลดีกว่าวิธีแมกซิมัมไลดิสติก แต่เนื่องจากในความเป็นจริงไม่สามารถทราบค่าการแจกแจงเริ่มแรกของกลุ่มตัวอย่างได้ ดังนั้นเนอริงจึงไม่ได้เสนอแนะให้ใช้วิธีการประมาณค่าด้วยวิธีนี้ และการเปลี่ยนค่าการเดาจาก 0.20 เป็น 0.00 ไม่มีผลต่อลักษณะการแจกแจงของค่าสถิติ L_2 นอกจากนี้ เนอริง (Nering, 1996) ยังได้ศึกษาเปรียบเทียบการกระจายของดัชนีบ่งชี้ความเหมาะสมของบุคคล คือดัชนี L_2 กับ ดัชนี ECI_4 ในบริบทการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (Computerized Adaptive Testing) โดยจำลองข้อมูลการตอบข้อสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (CAT) เมื่อจำนวนข้อสอบในกลุ่มข้อสอบ (item pool) การกระจายของค่าอำนาจจำแนกของข้อสอบ (a) และเกณฑ์การยุติการสอบซึ่งใช้ความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบเป็นเกณฑ์มีค่าต่างกัน พบว่า ในบริบทการทดสอบแบบ CAT ทั้งดัชนี L_2 และ ดัชนี ECI_4 มีการกระจายเข้าใกล้โค้งปกติเมื่อจำนวนข้อสอบในกลุ่ม

ข้อสอบ (item pool) มีจำนวนน้อย ค่าอำนาจจำแนกของข้อสอบ (a) มีค่าน้อย และเกณฑ์การยุติ การสอบซึ่งใช้ความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบเป็นเกณฑ์มีค่ามาก

กุสุมา สุวรรณแก้ว (2540) ได้ศึกษาเปรียบเทียบผลการตรวจสอบความเหมาะสมของ บุคคลระหว่างดัชนีซึ่งประมาณค่าด้วยวิธีที่แตกต่างกัน ได้แก่ ดัชนี L_2 ประมาณค่าด้วย ฟังก์ชันโลดัลลิสต์ และ ดัชนี W_1 ประมาณค่าดัชนีจากคะแนนส่วนที่เหลือ เมื่อระดับความ สามารถของผู้สอบ และความยาวของแบบสอบแตกต่างกัน ผลการตรวจสอบความเหมาะสม ของบุคคลวัดจากคุณภาพของแบบสอบหลังการตรวจสอบความเหมาะสมของบุคคลกับโมเด ลการตอบข้อสอบในด้าน ค่าสารสนเทศ ความเที่ยง และความตรงตามทฤษฎีของแบบสอบ ผลการวิจัยพบว่า ผลการตรวจสอบความเหมาะสมของบุคคลด้วยดัชนี L_2 ด้านค่าสารสนเทศ ของแบบสอบสูงกว่าดัชนี W_1 ในทุกระดับความสามารถของผู้สอบและทุกระดับความยาวของ แบบสอบ และผลการตรวจสอบความเหมาะสมของบุคคลด้วยดัชนี L_2 และ ดัชนี W_1 ด้าน ความเที่ยงและความตรงตามทฤษฎีไม่แตกต่างกัน ในทุกระดับความสามารถของผู้สอบ และทุก ระดับความยาวของแบบสอบ

จากแนวคิดและรายงานการวิจัยที่ผ่านมาได้มีการศึกษาและสร้างดัชนีสำหรับใช้ตรวจค้น บุคคลที่มีแบบแผนการตอบที่ผิดปกติในเงื่อนไขต่างๆ หลายดัชนี ซึ่งดัชนี L_2 ก็เป็นอีกดัชนีหนึ่ง ที่มีรายงานการวิจัยที่ผ่านมาสนับสนุนว่ามีประสิทธิภาพในการตรวจค้นบุคคลที่มีแบบแผนการ ตอบที่ผิดปกติ และปัจจุบันได้มีการพัฒนาโปรแกรมสำเร็จรูปที่สามารถวิเคราะห์ได้โดยสะดวก และมีประสิทธิภาพ ผู้วิจัยจึงสนใจศึกษาเปรียบเทียบสัดส่วนของบุคคลที่มีแบบแผนการตอบที่ผิด ปกติ โดยมีคำถามวิจัยคือ เมื่อให้ผู้สอบที่มีระดับความสามารถแตกต่างกันทำข้อสอบที่มีค่า อำนาจจำแนกของข้อสอบ (a), ค่าความยากของข้อสอบ (b) และค่าการเดา (c) ที่แตกต่างกัน แล้วใช้ดัชนีบ่งชี้ความเหมาะสมของบุคคลที่ประมาณค่าด้วยฟังก์ชันโลดัลลิสต์ คือดัชนี L_2 ตรวจค้น แบบแผนการตอบข้อสอบของผู้สอบ จะทำให้พบสัดส่วนของผู้ที่มีแบบแผนการตอบที่ผิดปกติใน แต่ละเงื่อนไขมากน้อยเพียงใด อันจะเป็นประโยชน์ในการเลือกข้อสอบให้เหมาะสมกับระดับ ความสามารถของผู้สอบเพื่อให้การวัดผลการศึกษามีความคลาดเคลื่อนต่ำ โดยพิจารณาจากสัดส่วนของแบบแผนการตอบผิดปกติในแต่ละเงื่อนไข

วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบสัดส่วนของผู้ที่มีแบบแผนการตอบที่ผิดปกติ เมื่อระดับความสามารถของผู้สอบ (θ) และคุณภาพของข้อสอบ ได้แก่ ค่าอำนาจจำแนกของข้อสอบ (a) ค่าความยาก (b) และค่าการเดา (c) แตกต่างกัน โดยการวิเคราะห์ด้วยดัชนีบ่งชี้ความเหมาะสมของบุคคลที่ประมาณค่าด้วยฟังก์ชันโลดิสติก (L) โดยมีวัตถุประสงค์เฉพาะดังนี้

1. เพื่อศึกษาสัดส่วนของผู้ที่มีแบบแผนการตอบที่ผิดปกติ เมื่อระดับความสามารถของผู้สอบ (θ) และคุณภาพของข้อสอบ ได้แก่ ค่าอำนาจจำแนกของข้อสอบ (a) ค่าความยาก (b) และค่าการเดา (c) มีระดับสูง ปานกลางและต่ำ
2. เพื่อเปรียบเทียบสัดส่วนของผู้ที่มีแบบแผนการตอบที่ผิดปกติ เมื่อระดับความสามารถของผู้สอบ (θ) และคุณภาพของข้อสอบ ได้แก่ ค่าอำนาจจำแนกของข้อสอบ (a) ค่าความยาก (b) และค่าการเดา (c) มีระดับสูง ปานกลางและต่ำ

สมมติฐานในการวิจัย

จากการศึกษาของแฮมเบิลตันและสวามินาธาน (Hambleton and Swaminathan, 1985) ได้แนะนำวิธีเลือกข้อสอบให้เหมาะสมกับผู้สอบเพื่อให้การวัดผลมีความคลาดเคลื่อนต่ำไว้ว่า นอกจากจะพิจารณาในแง่ความตรงตามเนื้อหา (content validity) แล้วจะต้องเลือกข้อที่มีค่าอำนาจจำแนกสูง ส่วนค่าความยากขึ้นอยู่กับความสามารถของกลุ่มผู้สอบที่จะนำข้อสอบไปใช้โดยคาดการณ์ล่วงหน้าถึงความสามารถของกลุ่มผู้สอบที่จะไปสอบ และเนื่องจากค่าการเดาในโมเดลของทฤษฎีการตอบสนองข้อสอบหมายถึงโอกาสที่ผู้สอบที่มีความสามารถต่ำจะตอบถูก ดังนั้นข้อสอบที่เหมาะสมกับกลุ่มผู้สอบที่มีความสามารถสูงน่าจะมีค่าการเดาต่ำ ข้อสอบที่เหมาะสมกับกลุ่มผู้สอบที่มีความสามารถปานกลางน่าจะมีค่าการเดาปานกลาง และข้อสอบที่เหมาะสมกับกลุ่มผู้สอบที่มีความสามารถต่ำน่าจะมีค่าการเดาสูง (Worm; 1978 อ้างถึงในคณิต ไร่มุฑ, 2533) ซึ่งหากให้ผู้สอบทำข้อสอบที่เหมาะสมกับความสามารถของตนแล้วผู้สอบน่าจะตอบข้อสอบได้ถูกต้องตามแบบแผนการตอบข้อสอบที่ปกติ

ดังนั้นผู้วิจัยจึงตั้งสมมติฐานของสัดส่วนแบบแผนการตอบข้อสอบของผู้สอบที่มีความสามารถแตกต่างกัน เมื่อแบ่งระดับของค่าพารามิเตอร์ของข้อสอบเป็นระดับสูง ปานกลาง และต่ำ ดังต่อไปนี้

1. การใช้ข้อสอบที่มีค่าอำนาจจำแนกสูง ($a > 0.85$) ค่าความยากอยู่ในระดับสูง ($b > 0.67$) และค่าการเดาอยู่ในระดับต่ำ ($c < 0.15$) สำหรับกลุ่มผู้สอบที่มีความสามารถสูงน่าจะให้มีแบบแผนการตอบที่ผิดปกติน้อยกว่ากลุ่มผู้สอบที่มีความสามารถปานกลางและต่ำ

2. การใช้ข้อสอบที่มีค่าอำนาจจำแนกสูง ($a > 0.85$) ค่าความยากอยู่ในระดับปานกลาง ($-0.67 \leq b \leq 0.67$) และค่าการเดาอยู่ในระดับปานกลาง ($0.15 \leq c \leq 0.25$) สำหรับกลุ่มผู้สอบที่มีความสามารถปานกลางน่าจะให้มีแบบแผนการตอบที่ผิดปกติน้อยกว่ากลุ่มผู้สอบที่มีความสามารถสูงและต่ำ

3. การใช้ข้อสอบที่มีค่าอำนาจจำแนกสูง ($a > 0.85$) ค่าความยากอยู่ในระดับต่ำ ($b < -0.67$) และค่าการเดาอยู่ในระดับสูง ($c > 0.15$) สำหรับกลุ่มผู้สอบที่มีความสามารถต่ำน่าจะให้มีแบบแผนการตอบที่ผิดปกติน้อยกว่ากลุ่มผู้สอบที่มีความสามารถสูงและปานกลาง

ขอบเขตของการวิจัย

1. วิธีการประมาณดัชนีบ่งชี้ความเหมาะสมของบุคคลที่ใช้ในการวิจัยครั้งนี้ ใช้การประมาณค่าจากฟังก์ชันไลค์ลิฮูด (Likelihood-Based Approach) ซึ่งเป็นโมเดลโลจิสติกแบบ 3 พารามิเตอร์ ภายใต้ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

2. การวิจัยครั้งนี้ศึกษาค่าความสามารถและคุณลักษณะของข้อสอบ ที่มีค่าเป็นไปตามโมเดลโลจิสติกแบบ 3 พารามิเตอร์ของทฤษฎีการตอบสนองข้อสอบ ซึ่งค่าความสามารถและคุณลักษณะข้อสอบที่ยอมรับได้ว่าเป็นข้อสอบที่มีคุณภาพ คือ ค่าความสามารถ (θ) อยู่ในช่วง -3 ถึง 3 , ค่าอำนาจจำแนกของข้อสอบ (a) อยู่ในช่วง 0.5 ถึง 2.5 , ค่าความยากของข้อสอบ (b) อยู่ในช่วง -3 ถึง +3 และค่าการเดา (c) อยู่ในช่วง 0 ถึง 0.3

คำจำกัดความที่ใช้ในการวิจัย

แบบแผนการตอบข้อสอบ (item response pattern) หมายถึง เมทริกซ์ของแบบสอบของผู้สอบแต่ละคน ที่นำมาจัดเรียงลำดับตามแนวแถวจากข้อที่ง่ายที่สุดซึ่งเป็นข้อที่มีจำนวนผู้สอบภายในกลุ่มตอบได้ถูกต้องมากที่สุด ไปจนถึงข้อที่ยากที่สุดซึ่งเป็นข้อที่มีจำนวนผู้สอบภายในกลุ่มตอบได้ถูกต้องน้อยที่สุด

แบบแผนการตอบที่ปกติ (nonaberrant response pattern) หมายถึง แบบแผนการตอบข้อสอบของผู้สอบ ซึ่งตอบข้อสอบที่มีค่าความยากต่ำกว่าหรือเท่ากับความสามารถของตนได้ถูกต้อง และตอบข้อสอบผิดในข้อที่มีค่าความยากสูงกว่าระดับความสามารถของตน

แบบแผนการตอบผิดปกติ (aberrant response pattern) หมายถึง แบบแผนการตอบข้อสอบของผู้สอบ ซึ่งตอบผิดในข้อที่มีค่าความยากต่ำกว่าหรือเท่ากับความสามารถของตน หรือเกิดจากการที่ผู้สอบตอบข้อสอบในข้อที่มีค่าความยากสูงกว่าระดับความสามารถของตนได้ถูกต้อง

โมเดลการตอบข้อสอบ (item response model) หมายถึง โมเดลโลจิสติกแบบ 3 พารามิเตอร์ซึ่งเป็นแบบจำลองทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างผลการตอบข้อสอบกับความสามารถของผู้สอบ

ค่าพารามิเตอร์ของข้อสอบ หมายถึง ค่าสถิติที่บ่งบอกคุณภาพของข้อสอบตามทฤษฎีการตอบสนของข้อสอบ มี 3 ค่า ได้แก่ ค่าอำนาจจำแนก (a) ,ค่าความยาก (b) และค่าการเดา (c)

ค่าอำนาจจำแนก (a) หมายถึง ค่าความชันของโค้งคุณลักษณะข้อสอบ ณ จุดเปลี่ยนโค้ง มีค่าตั้งแต่ $-\alpha$ ถึง $+\alpha$ แต่ในทางปฏิบัติจะใช้ค่าอำนาจจำแนกที่มีค่าตั้งแต่ 0.5 ถึง 2.5 ซึ่งในที่นี้แบ่งเป็น 3 ระดับ คือ $a > 0.85$ เป็นค่าอำนาจจำแนกระดับสูง $0.55 \leq a \leq 0.85$ เป็นค่าอำนาจจำแนกระดับปานกลาง และ $a < 0.55$ เป็นค่าอำนาจจำแนกระดับต่ำ

ค่าความยากของข้อสอบ (b) หมายถึง ระดับความยากของข้อสอบซึ่งมีค่าเท่ากับระดับความสามารถของผู้สอบ (θ) ที่มีโอกาสทำข้อสอบนั้นถูกต้องมีค่าตั้งแต่ $-\alpha$ ถึง $+\alpha$ แต่ในทางปฏิบัตินิยมใช้ข้อสอบที่ค่าความยากระหว่าง -3 ถึง +3 ซึ่งในที่นี้จะแบ่งเป็น 3 ระดับ คือ $b > .67$ เป็นค่าความยากระดับสูง , $-0.67 \leq b \leq 0.67$ เป็นค่าความยากระดับปานกลาง และ $b < -0.67$ เป็นค่าความยากระดับต่ำ

ค่าการเดา (c) หมายถึง ค่าโอกาสที่ผู้สอบที่มีความสามารถต่ำจะสามารถทำข้อสอบข้อนั้นได้ถูกต้องอย่างน้อยเพียงใด โดยปกตินิยมใช้ข้อสอบที่มีค่าการเดาระหว่าง 0 ถึง 0.3 ซึ่งในที่นี้จะแบ่งเป็น 3 ระดับ คือ $c > 0.25$ เป็นค่าการเดาระดับสูง , $0.15 \leq c \leq 0.25$ เป็นค่าการเดาระดับปานกลาง และ $c < .15$ เป็นค่าการเดาระดับต่ำ

ระดับความสามารถของผู้สอบ (θ) หมายถึง ศักยภาพของผู้สอบที่ประมาณได้จากการทำแบบสอบตามทฤษฎีการตอบสนองข้อสอบ โมเดลโลจิสติกแบบ 3 พารามิเตอร์ ซึ่งในที่นี้แบ่งเป็น 3 ระดับ คือ

ผู้สอบที่มีความสามารถสูง	หมายถึง	กลุ่มผู้สอบที่มีระดับความสามารถเฉลี่ย เท่ากับ 1.50
ผู้สอบที่มีความสามารถปานกลาง	หมายถึง	กลุ่มผู้สอบที่มีระดับความสามารถเฉลี่ย เท่ากับ 0.0
ผู้สอบที่มีความสามารถต่ำ	หมายถึง	กลุ่มผู้สอบที่มีระดับความสามารถเฉลี่ย เท่ากับ -1.50

ดัชนีบ่งชี้ความเหมาะสมของบุคคล (person-fit index) หมายถึง ค่าสถิติที่บ่งบอกว่าบุคคลนั้นมีแบบแผนการตอบสอดคล้องหรือแตกต่างจากค่าที่คาดหวังในโมเดลการตอบข้อสอบตามทฤษฎีการตอบสนองข้อสอบสำหรับการวิจัยนี้ ได้แก่ ค่าดัชนีแอลซัด (L_2)

ดัชนีแอลซัด (L_2 index) หมายถึง ดัชนีบ่งชี้ความเหมาะสมของบุคคล ซึ่งประมาณค่าด้วยฟังก์ชันไลค์ลิฮูด (Likelihood-based Approach) (Levine & Rubin, 1979)

บุคคลที่เหมาะสมกับโมเดล (person fit) หมายถึง บุคคลที่มีแบบแผนการตอบข้อสอบสอดคล้องกับรูปแบบของการตอบที่ควรจะเป็นในการตอบตามที่คาดหวังไว้ในโมเดลการตอบข้อสอบ ตามทฤษฎีการตอบสนองข้อสอบ คือมี $|L_2|$ น้อยกว่า 1.96

ประโยชน์ที่คาดว่าจะได้รับ

1. ทำให้ทราบว่าคุณลักษณะข้อสอบแบบใดบ้างที่ทำให้ผู้สอบที่มีระดับความสามารถแต่ละระดับมีแบบแผนการตอบที่ผิดปกติ
2. เป็นแนวทางในการคัดเลือกข้อสอบให้เหมาะสมกับระดับความสามารถของผู้สอบ เนื่องจากปัจจุบันการทดสอบแบบปรับเหมาะ (Adaptive Testing) เริ่มเป็นที่แพร่หลายในวงการวัดและประเมินผลการศึกษา