

การเปรียบเทียบวิธีชิบเทสต์และดีเอฟไอทีในการตรวจสอบการทำหน้าที่เบี่ยงเบน ของข้อสอบ
หมวดข้อสอบและแบบทดสอบจากข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ

ปริญญานิพนธ์
ของ
สิริรัตน์ วิภาสศิลป์

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุริยางค์บัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา

2545

บทที่ 1

บทนำ

ภูมิหลัง

ในการจัดการศึกษาทุกระดับ องค์ประกอบสำคัญประการหนึ่ง คือ การวัดและประเมินผล แม้ว่าในยุคปฏิรูปการศึกษา ที่กระบวนการเรียนการสอนมุ่งเน้นผู้เรียนเป็นสำคัญ จะนิยมใช้การวัดและประเมินการปฏิบัติ (performance-based assessment) มากขึ้น แต่แบบทดสอบก็ยังได้รับความนิยมนำมาใช้เป็นเครื่องมือสำคัญอย่างหนึ่งในการวัดและประเมินผลการศึกษา โดยเชื่อว่าแบบทดสอบที่มีคุณภาพสามารถวัดคุณลักษณะของบุคคลได้อย่างถูกต้องและน่าเชื่อถือ คุณลักษณะภายในของบุคคลเป็นสิ่งที่ไม่สามารถวัดหรือสังเกตได้โดยตรง แต่มีความสำคัญและจำเป็นต้องศึกษาเพราะคุณลักษณะภายในมีอิทธิพลต่อการแสดงออกทางพฤติกรรมภายนอกของบุคคลที่สามารถสังเกตได้ ถ้าสามารถวัดคุณลักษณะภายในของบุคคลได้จะทำให้มีความรู้และสามารถทำความเข้าใจถึงสาเหตุแห่งการเกิดพฤติกรรมของบุคคล อันจะเกิดประโยชน์อย่างยิ่งต่อการพัฒนาและควบคุมพฤติกรรมของบุคคลไปในทิศทางที่สังคมต้องการ นักจิตวิทยา นักสถิติ และนักวัดผลจึงได้ร่วมกันพัฒนาทฤษฎีการทดสอบอย่างต่อเนื่องมาเป็นเวลานาน จนเป็นศาสตร์ที่ประกอบด้วยองค์ความรู้อันเป็นที่ยอมรับกันอย่างกว้างขวางในวงการศึกษ จิตวิทยา และสาขาวิชาต่าง ๆ ทฤษฎีการทดสอบเป็นองค์ความรู้ที่มีนัยทั่วไปซึ่งนักทฤษฎีการวัดได้เสนอไว้เกี่ยวกับแบบจำลองการวัด ข้อตกลงเบื้องต้น การพัฒนาเครื่องมือ การวิเคราะห์และการนำไปใช้เพื่อแก้ปัญหาการวัด และพัฒนาแบบทดสอบให้มีคุณภาพ ทฤษฎีการทดสอบจำแนกได้เป็นสองแนวทางกว้าง ๆ คือทฤษฎีการทดสอบแบบมาตรฐานเดิม (Classical Test Theory) และทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory) ทฤษฎีการทดสอบแนวใหม่พัฒนาขึ้นเพื่อแก้ไขจุดอ่อนเกี่ยวกับข้อตกลงเบื้องต้นของทฤษฎีการทดสอบแบบมาตรฐานเดิม คือ ทฤษฎีการตอบข้อสอบ (Item Response Theory: IRT) ซึ่งเป็นที่ยอมรับกันว่ามีการอบความคิดเชิงทฤษฎีที่สมเหตุสมผล ยอมรับให้ความคลาดเคลื่อนจากการวัดมีความแตกต่างกันไปขึ้นอยู่กับระดับความสามารถของผู้ตอบข้อสอบแต่ละคนและคุณลักษณะของข้อสอบแต่ละข้อ พร้อมทั้งพยายามวัดคุณลักษณะที่แท้จริงของแต่ละบุคคล (ศิริชัย กาญจนวาสี. 2535 : 94-95) ผู้เชี่ยวชาญหลายท่านได้ร่วมกันพัฒนาแบบจำลองการวัดที่อาศัยกรอบความคิดของทฤษฎีการตอบข้อสอบหลายแบบจำลอง ที่สำคัญได้แก่ แบบจำลองการตอบข้อสอบมิติเดียว (unidimensional item response theory) และแบบจำลองการตอบข้อสอบหลายมิติ (multidimensional item response theory) ที่สามารถใช้กับข้อสอบที่ให้คะแนนรายข้อแบบสองค่า (dichotomous item) และข้อสอบที่ให้คะแนนรายข้อแบบหลายค่า (polytomous item)

ทฤษฎีการตอบข้อสอบมิติเดียวซึ่งมีข้อตกลงเบื้องต้นเกี่ยวกับความเป็นมิติเดียวของแบบทดสอบนั้น นักวัดผลกลุ่มหนึ่งในยุคปัจจุบันมีความเห็นว่าเป็นข้อตกลงที่ไม่สมเหตุสมผล เนื่องจากในการตอบข้อสอบข้อใดข้อหนึ่งนั้นผู้ตอบข้อสอบต้องใช้ความสามารถหลายอย่าง (multiple traits) ในระดับที่แตกต่างกัน (Ackerman. 1992, 1996, 1997; etc.) เป็นต้นว่า แบบทดสอบวิชาเลขคณิต อาจวัดทั้งทักษะทางคณิตศาสตร์ และทักษะการอ่านในขณะเดียวกัน แวง วิลสัน และอาดัมส์ (Oshima, Raju & Flowers. 1997 : 253 ; citing Wang, Wilson & Adams. 1995) กล่าวว่า แบบทดสอบหลายมิติมีสองประเภท คือ แบบทดสอบหลายมิติระหว่างข้อสอบ (multidimensional between-item test) และแบบทดสอบหลายมิติภายในข้อสอบ (multidimensional within-item test)

การใช้แบบทดสอบเป็นเครื่องมือในการวัดนั้น สิ่งสำคัญคือ ต้องการคะแนนที่มีความหมายสามารถใช้เป็นหลักฐานในการแปลความหมายคุณลักษณะภายในที่ต้องการวัดของแต่ละบุคคลได้อย่างถูกต้องตามความเป็นจริง ในกรอบความคิดของทฤษฎีการตอบข้อสอบหลายมิตินั้น ถ้าในการตอบข้อสอบข้อใดหรือแบบทดสอบฉบับใดได้ถูกต้องผู้ตอบข้อสอบต้องใช้ความสามารถอย่างอื่นนอกเหนือจากความสามารถที่แบบทดสอบต้องการวัด หมายความว่าข้อสอบแสดงการทำหน้าที่เบี่ยงเบน (Differential Item Functioning : DIF) หรือแบบทดสอบแสดงการทำหน้าที่เบี่ยงเบน (Differential Test Functioning : DTF) ซึ่งจะส่งผลกระทบต่อความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบ ดังนั้นถ้ามีข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนจำนวนมากขึ้นความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบจะลดลงตามไปด้วย (Ackerman. 1992 : 69)

การคำนวณดัชนีแสดงการทำหน้าที่เบี่ยงเบนของข้อสอบโดยทั่วไปแบ่งเป็นสองกลุ่ม คือ กลุ่มที่ใช้เกณฑ์ภายนอกและกลุ่มที่ใช้เกณฑ์ภายในเปรียบเทียบกลุ่มอ้างอิง (reference group) และกลุ่มสนใจ (focal group) ในการศึกษาโดยทั่วไปส่วนใหญ่ใช้คะแนนจากแบบทดสอบเป็นตัวแทนความสามารถในการเปรียบเทียบ ซึ่งเป็นการคำนวณโดยอาศัยเกณฑ์ภายในที่มีกระบวนการในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ แบ่งเป็นสามประเภทใหญ่ ๆ คือ วิธีการของทฤษฎีการทดสอบแบบมาตรฐานเดิม (traditional test theory methods) วิธีการไคสแควร์ (chi – square methods) และวิธีการของทฤษฎีคุณลักษณะแฝงหรือทฤษฎีการตอบข้อสอบ (latent trait theory methods) ซึ่งเป็นวิธีการที่ได้รับการยอมรับเชิงทฤษฎีมากกว่าวิธีอื่น ๆ (Whitmore & Schumacker. 1999 : 911)

ในระยะแรก วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ใช้ตรวจสอบข้อมูลการตอบข้อสอบมิติเดียว โดยตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบครั้งละ 1 ข้อ ในปี ค.ศ. 1992 ซีลลีและสเตาต์ (Shealy & Stout, 1992) เสนอโปรแกรมคอมพิวเตอร์สำเร็จรูป SIBTEST ที่สามารถตรวจสอบการทำหน้าที่เบี่ยงเบนได้ทั้งระดับข้อสอบ (Differential Item Functioning : DIF) ระดับหมวดข้อสอบ (Differential Bundle Functioning : DBF) และระดับแบบทดสอบ (Differential Test Functioning : DTF) โดยใช้วิธีการของสถิตินอนพาราเมตริกที่อาศัยกรอบความคิดของทฤษฎีการตอบข้อสอบ (nonparametric IRT-based Approach) เน้นแนวคิดของการตอบข้อสอบโดยใช้ความสามารถหลายมิติ (Stout & Roussos. 1992 : 1) ในปี ค.ศ. 1995 ราจู แวนเดอลินเด็น และฟลีร์ (Raju, van der Linden & Fleer. 1995) เสนอโปรแกรมคอมพิวเตอร์สำเร็จรูป DFIT ที่สามารถตรวจสอบการทำหน้าที่เบี่ยงเบนได้ทั้งระดับข้อสอบ (DIF) ระดับหมวดข้อสอบ (DBF) และระดับแบบทดสอบ (DTF) ใช้ได้กับข้อมูลการตอบข้อสอบมิติเดียวและหลายมิติที่มีการให้คะแนนรายข้อแบบสองค่าและการให้คะแนนรายข้อแบบหลายค่า โดยใช้วิธีการของสถิติพาราเมตริกที่อาศัยกรอบความคิดของทฤษฎีการตอบข้อสอบ (parametric IRT-based approach) รายงานการวิจัยเกี่ยวกับการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบครั้งละ 1 ข้อหรือมากกว่า 1 ข้อมีดังนี้ ปี ค.ศ. 1996 ดักกลัส รุสโซ และ สเตาต์ (Douglas, Roussos & Stout. 1996 : 465-484) เสนอรายงานการวิจัยการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบจากข้อมูลการตอบข้อสอบมิติเดียว มีการให้คะแนนแบบสองค่า ใช้ข้อมูลเชิงประจักษ์วิเคราะห์ด้วยโปรแกรม SIBTEST พบว่าการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบครั้งละหลาย ๆ ข้อมีประสิทธิภาพสูงกว่าการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบครั้งละ 1 ข้อ โอชิม่า ราจู และ ฟลาวเวอร์ส (Oshima, Raju & Flowers. 1997 : 253-272) ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) และการทำหน้าที่เบี่ยงเบนของแบบทดสอบ (DTF) จากข้อมูลการตอบข้อสอบหลายมิติที่มีการให้คะแนนรายข้อแบบสองค่า ใช้ข้อมูลจำลองวิเคราะห์ด้วยโปรแกรม DFIT พบว่าทำให้สามารถอธิบายสาเหตุที่มีการทำหน้าที่เบี่ยงเบนได้ชัดเจนขึ้น โอชิม่าและคนอื่น ๆ

(Oshima, et al. 1998 : 353-369) ได้ตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ (DBF) ใช้ข้อมูลเชิงประจักษ์ที่เป็นข้อมูลการตอบข้อสอบมิติเดียวที่มีการให้คะแนนรายข้อแบบสองค่า วิเคราะห์ด้วยโปรแกรม DFIT ฟลาวเวอร์ส โอชิมา และ ราจู (Flowers, Oshima & Raju. 1999) เสนอผลการตรวจสอบการทำหน้าที่เบี่ยงเบนกับข้อมูลจำลองที่เป็นข้อมูลการตอบข้อสอบมิติเดียวที่มีการให้คะแนนแบบหลายค่า วิเคราะห์ด้วยโปรแกรม DFIT พบว่าให้ผลเช่นเดียวกับการวิจัยที่ใช้ข้อมูลการตอบข้อสอบมิติเดียวและหลายมิติที่มีการให้คะแนนรายข้อแบบสองค่า

นักวิจัยหลายท่าน (Oshima & Miller. 1992 ; citing Traub. 1983 ; Harrison. 1986 ; Reckase, Ackerman & Carlson. 1988) กล่าวว่าในการตอบแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนส่วนใหญ่ผู้ตอบข้อสอบต้องใช้ความสามารถหลายมิติ เป็นต้นว่า งานวิจัยของเร็คเคส (Reckase. 1985) ศึกษาโดยใช้ข้อมูลการตอบข้อสอบวิชาการใช้คณิตศาสตร์ (Mathematic Usage Test) ที่ดำเนินการโดย American College Testing (ACT) ใน ค.ศ. 1983 พบว่าข้อมูลการตอบข้อสอบวิชาดังกล่าวเหมาะกับแบบจำลองการตอบข้อสอบหลายมิติชนิดสองพารามิเตอร์ (2-parameter multidimensional logistic model: M2PL) สำหรับการวิจัยในประเทศไทย กาญจนา วัธนสุนทร (2538) ได้ศึกษาเพื่อพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ โดยใช้แบบทดสอบคัดเลือกบุคคลเพื่อศึกษาต่อในสถาบันอุดมศึกษาของทบวงมหาวิทยาลัยประจำปีการศึกษา 2535 วิชาภาษาอังกฤษ กข และวิชาคณิตศาสตร์ กข ในการตรวจสอบความเป็นมิติเดียวของแบบทดสอบพบว่า ค่าไอเกนสูงสุดของวิชาคณิตศาสตร์ คือ 3.39 และของวิชาภาษาอังกฤษ คือ 8.70 โดยที่ผลการวิเคราะห์แยกกลุ่มตามเพศชายและเพศหญิง มีผลใกล้เคียงกับผลที่ได้จากข้อมูลรวมและคิดเป็นความแปรปรวนร้อยละ 8.5 และ 10.9 ของความแปรปรวนทั้งหมดแต่ละวิชา ซึ่งมีค่าน้อยเกินกว่าที่จะสรุปถึงความ เป็นมิติเดียวชนิดมีตัวประกอบหลักตัวแรกเด่นกว่าตัวประกอบอื่น ๆ (essential unidimensionality) นอกจากนั้นยังต่ำกว่าค่าความแปรปรวน ซึ่งเร็คเคส (Reckase ; กาญจนา วัธนสุนทร. 2538 : 88 ; อ้างอิงจาก Raju. 1993 : 21-22) เสนอแนะว่าในการใช้ทฤษฎีการตอบข้อสอบวิเคราะห์ข้อสอบนั้น ค่าความแปรปรวนจากตัวประกอบหลักแรกของแบบทดสอบควรมีค่าไม่ต่ำกว่าร้อยละ 20 ข้อมูลการตอบข้อสอบวิชาคณิตศาสตร์ กข และ ภาษาอังกฤษ กข ดังกล่าวอาจไม่เหมาะสมกับการวิเคราะห์ด้วยทฤษฎีการตอบข้อสอบมิติเดียว (กาญจนา วัธนสุนทร. 2538 : 88) และโอชิมา (กาญจนา วัธนสุนทร. 2538 : 61 ; อ้างอิงจาก Oshima. 1989) ได้ศึกษาผลของข้อมูลการตอบข้อสอบหลายมิติที่มีต่อการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) โดยใช้ข้อมูลจำลอง 42 ชุด จากข้อสอบ 40 ข้อ สามารถตรวจสอบข้อสอบหลายมิติที่สร้างให้แสดงการทำหน้าที่เบี่ยงเบนและสร้างให้ไม่แสดงการทำหน้าที่เบี่ยงเบนได้อย่างถูกต้อง แต่อำนาจในการตรวจสอบลดลงถ้าจำนวนข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนในแบบทดสอบมีมากขึ้น และได้พบว่าเมื่อมีข้อสอบหลายมิติในแบบทดสอบร้อยละ 20 ทำให้การทดสอบความเป็นมิติเดียวของแบบทดสอบได้รับการปฏิเสธมากกว่าครึ่งหนึ่งของการทดสอบ โอชิมาได้เสนอให้ตรวจสอบความเป็นมิติเดียวของแบบทดสอบทุกครั้งเมื่อต้องการวิเคราะห์ด้วยทฤษฎีการตอบข้อสอบมิติเดียว แบบทดสอบที่ไม่เป็นไปตามข้อตกลงของความ เป็นมิติเดียวไม่ควรวิเคราะห์ด้วยกรอบความคิดของทฤษฎีการตอบข้อสอบมิติเดียว แต่อาจพิจารณาใช้กรอบความคิดของทฤษฎีการตอบข้อสอบหลายมิติแทน มีนักวิจัยกลุ่มหนึ่ง (Nandakumar. 1993 : 309 ; citing Humphreys. 1986; Drasgow. 1987; Roznowski. 1987; Reith & Roznowski, 1991) เชื่อว่าแบบทดสอบที่ประกอบด้วยข้อสอบที่วัดคุณลักษณะหลายอย่าง (multiple determinants) ทำให้แบบทดสอบมีความเที่ยงตรงเชิงพยากรณ์และความเที่ยงตรงเชิงโครงสร้างสูงขึ้น

การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบครั้งละข้อและนำไปพิจารณาตัดข้อสอบออกจากแบบทดสอบ อาจทำให้ความเที่ยงตรงของแบบทดสอบลดลง เนื่องจากกำจัดข้อสอบที่วัดคุณลักษณะที่เกี่ยวข้อง (trait-relevant) กับข้ออื่นออกไป (Nandakumar. 1993 : 294 citing Roznowski. 1987, Reith & Roznowski. 1992.) นอกจากนั้น แนนดาคูมาร์ (Nandakumar. 1993 : 294) เสนอแนะว่าการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบหลายข้อพร้อมกัน ทำให้สามารถศึกษาการทำหน้าที่เบี่ยงเบนของข้อสอบแบบขยายผล (DIF amplification) และการทำหน้าที่เบี่ยงเบนของข้อสอบแบบหักล้างกัน (DIF cancellation) ดังนั้นบางครั้งตรวจสอบไม่พบการทำหน้าที่เบี่ยงเบนของข้อสอบรายข้อ แต่เมื่อพิจารณาเป็นหมวดข้อสอบ (bundle of items) อาจพบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบได้ หรือในทำนองกลับกัน เมื่อตรวจสอบพบการทำหน้าที่เบี่ยงเบนของข้อสอบรายข้อต่อกลุ่มอ้างอิง และตรวจสอบพบข้อสอบข้ออื่น ๆ ทำหน้าที่เบี่ยงเบนต่อกลุ่มสนใจ เมื่อพิจารณาพร้อม ๆ กันทั้งหมวดข้อสอบ อาจไม่พบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบก็ได้ หลักการเดียวกันนี้สามารถขยายผลไปศึกษาการทำหน้าที่เบี่ยงเบนของแบบทดสอบได้ด้วย การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ ช่วยให้การพัฒนาแบบทดสอบทำได้อย่างมีประสิทธิภาพและทำให้แบบทดสอบที่สร้างขึ้นมีความเที่ยงตรงเชิงโครงสร้างสูงขึ้น

จากผลการวิจัยและการศึกษาข้างต้น ผู้วิจัยจึงสนใจที่จะศึกษาเพื่อเปรียบเทียบวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) หมวดข้อสอบ (DBF) และแบบทดสอบ (DTF) จากข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติที่มีการให้คะแนนรายข้อแบบสองค่า โดยวิธีซิมเพลท์และวิธีดีเอฟไอที เนื่องจากวิธีการทั้งสองตรวจสอบการทำหน้าที่เบี่ยงเบนได้ทั้งระดับข้อสอบ ระดับหมวดข้อสอบและระดับแบบทดสอบ ใช้ได้กับข้อมูลการตอบข้อสอบที่ใช้ความสามารถมิติเดียวและใช้ความสามารถหลายมิติ ที่อาศัยกรอบความคิดของทฤษฎีการตอบข้อสอบ (IRT-based approach) ต่างกันที่ SIBTEST ใช้วิธีการของสถิตินอนพาราเมตริก และ DFIT ใช้วิธีการของสถิติพาราเมตริก นอกจากนั้นแนนดาคูมาร์ (Nandakumar. 1993 : 309) กล่าวว่า SIBTEST มีประโยชน์อย่างยิ่งที่สามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบรายข้อและของข้อสอบหลายข้อพร้อมกันสารสนเทศที่ได้จากการตรวจสอบมีประโยชน์ต่อการตัดสินใจเลือกข้อสอบไว้หรือตัดข้อสอบออกจากแบบทดสอบ โดยพิจารณาทั้งอิทธิพลของการทำหน้าที่เบี่ยงเบนของข้อสอบรายข้อและพิจารณาผลกระทบต่อหมวดข้อสอบและแบบทดสอบทั้งฉบับด้วย อาจตัดสินใจเลือกข้อสอบไว้หรือตัดข้อสอบออกโดยพิจารณาความสมดุลของคุณลักษณะภายในที่ไม่ได้ตั้งใจวัดหรือความสามารถรอง (secondary traits) นอกจากนั้นการตัดสินใจใช้คะแนนของแบบทดสอบมีความจำเป็นต้องพิจารณาอิทธิพลสะสมของหมวดข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนด้วย ฟลาวเวอร์ส โอชิมา และ ราจู (Flowers, Oshima & Raju. 1999 : 309) กล่าวว่า DFIT มีประโยชน์หลายประการ ได้แก่ 1) จากการที่ใช้วิธีการของสถิติพาราเมตริกที่อาศัยกรอบความคิดของทฤษฎีการตอบข้อสอบ เมื่อใช้ตรวจสอบ DIF/DTF และนำผลมาใช้ในการพัฒนาแบบทดสอบทำให้ยังคงรักษากรอบความคิดทั่วไปในการพัฒนาแบบทดสอบไว้ได้ 2) ในการตรวจสอบ DIF ของข้อสอบข้อใดข้อหนึ่งไม่ได้กำหนดข้อตกลงเบื้องต้นว่าข้อสอบข้ออื่น ๆ ในแบบทดสอบไม่แสดง DIF 3) มีกระบวนการตรวจสอบ DTF สำหรับการตรวจสอบอิทธิพลทั้งหมดในการกำจัดข้อสอบข้อใดข้อหนึ่งจากแบบทดสอบ 4) ยินยอมให้มีการตรวจสอบ DIF/DTF ในแบบทดสอบที่มีข้อสอบหลายรูปแบบ (mixed test format) เป็นต้นว่า แบบทดสอบที่ประกอบด้วยข้อสอบที่ให้คะแนนรายข้อแบบสองค่าและข้อสอบที่ให้คะแนนรายข้อแบบหลายค่า 5) ใช้ได้กับข้อมูลการตอบข้อสอบที่ใช้ความสามารถมิติเดียวและใช้ความสามารถหลายมิติ ในการศึกษาครั้งนี้ใช้ข้อมูลเชิงประจักษ์ เนื่องจากในประเทศไทยยังไม่มีรายงานผลการศึกษาเพื่อตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบจากข้อมูลการตอบข้อสอบที่ใช้

ความสามารถหลายมิติด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที จากข้อมูลเชิงประจักษ์และในการศึกษาโดยใช้ข้อมูลจำลองนั้น แม้ว่าสามารถศึกษาการตรวจสอบการทำหน้าที่เบี่ยงเบนได้หลายเงื่อนไข แต่มีข้อจำกัดที่ไม่สามารถให้สารสนเทศได้ทัดเทียมกับข้อมูลจริง (กาญจนา วัฒนสุนทร. 2538 : 9)

ความมุ่งหมายของการวิจัย

เพื่อศึกษาวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนที่เหมาะสมกับข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ โดยมีวัตถุประสงค์เฉพาะเพื่อศึกษาวิธีชิบเทสต์และวิธีดีเอฟไอที ในด้านต่าง ๆ ดังนี้

1. เปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบวิธีเดียวกันในเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน
2. เปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ ด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที ในระดับต่าง ๆ ดังนี้
 - 2.1 การทำหน้าที่เบี่ยงเบนของข้อสอบ
 - 2.1.1 เปรียบเทียบความถูกต้องและการระบุผิดพลาดในการตรวจสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที ในเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน
 - 2.1.2 หาความสอดคล้องของการตรวจสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที ในเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน
 - 2.2 การทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ

เปรียบเทียบความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที ในเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน
 - 2.3 การทำหน้าที่เบี่ยงเบนของแบบทดสอบ

เปรียบเทียบความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที ในเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

ความสำคัญของการวิจัย

1. การศึกษาครั้งนี้เป็นการขยายขอบเขตของการศึกษาค้นคว้าเกี่ยวกับการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบให้กว้างขวางยิ่งขึ้น จากการศึกษาค้นคว้าโดยใช้ข้อมูลการตอบข้อสอบที่อาศัยแนวคิดของความเป็นมิติเดียวของแบบทดสอบและตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบรายข้อมาเป็นการศึกษาค้นคว้าโดยอาศัยแนวคิดของการใช้ความสามารถหลายมิติในการตอบข้อสอบ และตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ เนื่องจากในประเทศไทยยังไม่มีเผยแพร่ผลการศึกษาค้นคว้าที่อาศัยแนวคิดดังกล่าวมาก่อน
2. ผลจากการศึกษาครั้งนี้ทำให้สามารถเลือกใช้วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ที่แบ่งกลุ่มย่อยตามเพศในเงื่อนไขของกลุ่มตัวอย่างและความยาวของแบบทดสอบขนาดต่าง ๆ ได้อย่างเหมาะสมกับลักษณะของข้อมูลการตอบข้อสอบ สามารถคัดเลือกข้อสอบไว้ในแบบทดสอบหรือตัดข้อสอบออกจากแบบทดสอบได้อย่างสมเหตุสมผล ซึ่งจะส่งผลให้สามารถพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ให้มีคุณภาพสูงตามเกณฑ์มาตรฐานและสอดคล้องกับวัตถุประสงค์ในการสร้างแบบทดสอบ และมีสารสนเทศที่ช่วยให้การแปลความหมายของคะแนนจากการใช้แบบทดสอบมีความชัดเจนยิ่งขึ้น

ขอบเขตของการวิจัย

1. ข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลที่ได้จากการตอบข้อสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 1 ที่ผู้วิจัยสร้างขึ้นโดยมีวัตถุประสงค์ให้ข้อสอบส่วนหนึ่งในแบบทดสอบแสดงการทำหน้าที่เบี่ยงเบนต่อเพศชาย กลุ่มตัวอย่างที่ใช้ในการเก็บรวบรวมข้อมูลเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคการศึกษาที่ 2 ปีการศึกษา 2543 สังกัดกรมสามัญศึกษาในจังหวัดนนทบุรี

2. ตัวแปรที่ศึกษา

2.1 ตัวแปรอิสระ ประกอบด้วย

2.1.1 วิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ ได้แก่วิธีชิบเทสต์และวิธีดีเอฟไอที

2.1.2 ขนาดของกลุ่มตัวอย่าง ได้แก่ 50, 100, 200, 500 และ 1,000 คน

2.2 ตัวแปรตาม ประกอบด้วย

2.2.1 ความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที

2.2.2 การระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที

2.2.3 ความสอดคล้องของการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที

2.2.4 ความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที

2.2.5 ความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที

2.3 ตัวแปรควบคุมคือ ความยาวของแบบทดสอบ ได้แก่ แบบทดสอบที่ประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ

3. แบบจำลองที่ใช้ในการศึกษา คือ แบบจำลองของทฤษฎีการตอบข้อสอบหลายมิติชนิด 2 พารามิเตอร์ (M2PL) ใช้วิธีการให้คะแนนรายข้อแบบสองค่า

นิยามศัพท์เฉพาะ

1. ข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ (multidimensional response data) หมายถึงข้อมูลการตอบข้อสอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 ที่ให้คะแนนรายข้อแบบสองค่า คือ ตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน โดยที่ผู้ตอบข้อสอบจะตอบข้อสอบบางข้อได้ 1 คะแนนต้องใช้ความสามารถสองอย่าง ประกอบด้วยความสามารถในวิชาคณิตศาสตร์ (ค 101) ซึ่งเป็นความสามารถหลักที่ข้อสอบต้องการวัด (θ) และความสามารถอื่นนอกเหนือจากที่ข้อสอบต้องการวัด แต่เป็นความสามารถที่ต้องนำมาใช้ในการตอบข้อสอบได้ถูกต้องจัดเป็นความสามารถรอง (η) เป็นต้นว่า ความสามารถในการอ่านหรือแก้ไขปัญหาโจทย์ที่มีความซับซ้อน ไหวพริบในการทำข้อสอบ ความคุ้นเคยหรือมีประสบการณ์ในสถานการณ์ที่โจทย์กำหนด

2. กลุ่มอ้างอิง (Reference Groups : R) หมายถึง กลุ่มผู้ตอบข้อสอบที่คาดว่าจะจะเป็นกลุ่มที่ได้เปรียบจากการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ ในการวิจัยครั้งนี้หมายถึงกลุ่มนักเรียนชาย

3. กลุ่มสนใจ (Focal Groups : F) หมายถึง กลุ่มผู้ตอบข้อสอบที่คาดว่าจะจะเป็นกลุ่มที่เสียเปรียบจากการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ ในการวิจัยครั้งนี้หมายถึงกลุ่มนักเรียนหญิง

4. การทำหน้าที่เบี่ยงเบนของข้อสอบ (differential item functioning: DIF) หมายถึง ข้อสอบที่ผู้ตอบข้อสอบซึ่งเป็นสมาชิกกลุ่มอ้างอิง (เพศชาย) และเป็นสมาชิกกลุ่มสนใจ (เพศหญิง) ที่มีความสามารถในวิชาคณิตศาสตร์ (ค 101) ซึ่งเป็นความสามารถหลักเท่ากัน แต่การแจกแจงความสามารถรองตามเงื่อนไขของความสามารถหลัก ($\eta | \theta$) แตกต่างกัน มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องไม่เท่ากัน

5. การทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ (differential bundle functioning: DBF) หมายถึง ข้อสอบจำนวนหนึ่งจัดเป็นหมวดข้อสอบซึ่งผู้ตอบข้อสอบที่เป็นสมาชิกกลุ่มอ้างอิง (เพศชาย) และเป็นสมาชิกกลุ่มสนใจ (เพศหญิง) ที่มีความสามารถในวิชาคณิตศาสตร์ (ค 101) ซึ่งเป็นความสามารถหลักที่หมวดข้อสอบต้องการวัดเท่ากัน แต่การแจกแจงความสามารถรองตามเงื่อนไขของความสามารถหลัก ($\eta | \theta$) แตกต่างกัน มีความน่าจะเป็นในการตอบข้อสอบหมวดนั้นถูกต้องไม่เท่ากัน

6. การทำหน้าที่เบี่ยงเบนของแบบทดสอบ (differential test functioning: DTF) หมายถึง ข้อสอบในแบบทดสอบทั้งฉบับซึ่งผู้ตอบข้อสอบที่เป็นสมาชิกกลุ่มอ้างอิง (เพศชาย) และเป็นสมาชิกกลุ่มสนใจ (เพศหญิง) ที่มีความสามารถในวิชาคณิตศาสตร์ (ค 101) ซึ่งเป็นความสามารถหลักที่แบบทดสอบต้องการวัดเท่ากัน แต่การแจกแจงความสามารถรองตามเงื่อนไขของความสามารถหลัก ($\eta | \theta$) แตกต่างกัน มีความน่าจะเป็นในการตอบข้อสอบในแบบทดสอบฉบับนั้นถูกต้องไม่เท่ากัน

7. วิชีบเทสต์ (SIBTEST) หมายถึง วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบที่อาศัยกรอบความคิดของทฤษฎีการตอบข้อสอบ ใช้วิธีการของสถิตินอนพาราเมตริก พัฒนาโดยชีลลีและสแตต์ (Shealy & Stout, 1992) ซึ่งมีแนวคิดว่าการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบเป็นการตรวจสอบความแตกต่างระหว่างค่าที่คาดหวังของการแจกแจงความสามารถอื่นนอกเหนือจากที่ข้อสอบต้องการวัดอย่างมีเงื่อนไข ($\eta | \theta$) ของกลุ่มอ้างอิงและกลุ่มสนใจ อาจตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบครั้งละ 1 ข้อหรือหลายข้อพร้อมกันก็ได้

8. วิดีเอฟไอที (DFIT) หมายถึง วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบที่อาศัยกรอบความคิดของทฤษฎีการตอบข้อสอบ ใช้วิธีการของสถิติพาราเมตริก พัฒนาโดยราชู แวนเดอร์ลินเด็น และฟลีร์ (Raju, van der Linden & Fleer, 1995) ซึ่งมีแนวคิดว่าการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบเป็นการตรวจสอบความแตกต่างระหว่างคะแนนจริงของผู้ตอบข้อสอบเมื่อเป็นสมาชิกของกลุ่มอ้างอิงและเมื่อเป็นสมาชิกของกลุ่มสนใจ ถ้ามีค่าเท่ากันแสดงว่าคะแนนของผู้ตอบข้อสอบมีความเป็นอิสระ ไม่ได้ขึ้นอยู่กับว่าเป็นสมาชิกกลุ่มใดกลุ่มหนึ่ง อาจตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบครั้งละ 1 ข้อหรือหลายข้อพร้อมกันก็ได้

9. ความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง ร้อยละของจำนวนข้อสอบที่วิชีบเทสต์หรือวีดีเอฟไอทีระบุว่าแสดงการทำหน้าที่เบี่ยงเบนได้ถูกต้อง ซึ่งคำนวณได้จากร้อยละของจำนวนข้อสอบที่ระบุว่าแสดงการทำหน้าที่เบี่ยงเบนได้ถูกต้องต่อจำนวนข้อสอบที่แสดงการทำหน้าที่

เบี่ยงเบนทั้งหมดในแบบทดสอบ ในการศึกษาครั้งนี้ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนทั้งหมดในแบบทดสอบได้มาจากการพิจารณาคัดสินของผู้เชี่ยวชาญ

10. การระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง ร้อยละของจำนวนข้อสอบที่วิธีซิปเทสต์หรือวิธีดีเอฟไอที่ระบุว่าแสดงการทำหน้าที่เบี่ยงเบน ทั้งที่ความจริงเป็นข้อสอบที่ไม่ได้แสดงการทำหน้าที่เบี่ยงเบน ซึ่งคำนวณได้จากร้อยละของจำนวนข้อสอบที่ระบุว่าแสดงการทำหน้าที่เบี่ยงเบนไม่ถูกต้องต่อจำนวนข้อสอบที่ไม่ได้แสดงการทำหน้าที่เบี่ยงเบนในแบบทดสอบ ซึ่งในการศึกษาครั้งนี้ข้อสอบที่ไม่แสดงการทำหน้าที่เบี่ยงเบนในแบบทดสอบได้มาจากการพิจารณาคัดสินของผู้เชี่ยวชาญ

11. ความสอดคล้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิปเทสต์และวิธีดีเอฟไอที่สอดคล้องกัน คำนวณจากร้อยละของจำนวนข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนที่ตรวจสอบพบซ้ำกันทั้ง 2 วิธีต่อจำนวนข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนทั้งหมดที่ตรวจสอบพบด้วยวิธีซิปเทสต์และวิธีดีเอฟไอ

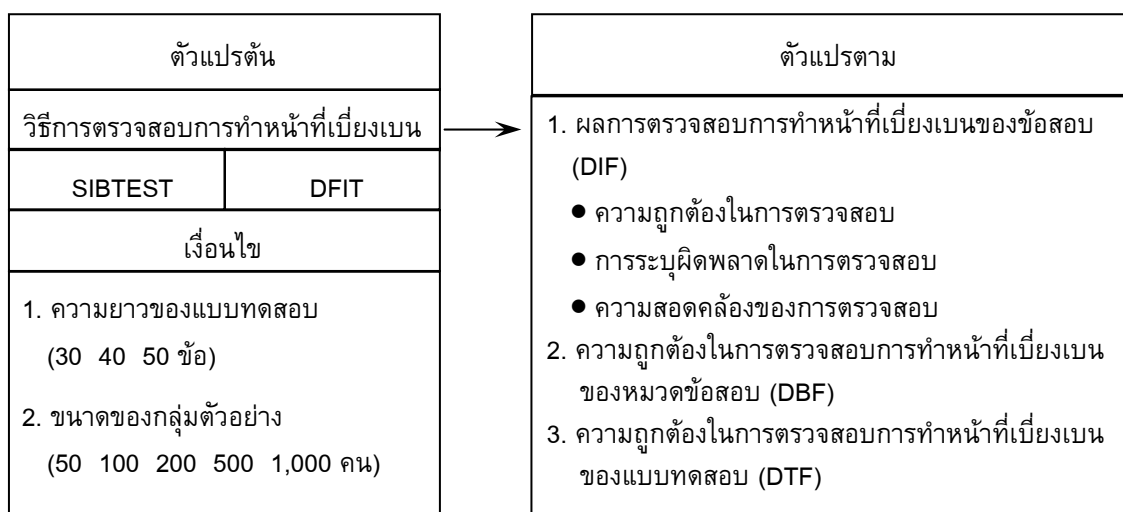
12. ความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ หมายถึง ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบด้วยวิธีซิปเทสต์หรือวิธีดีเอฟไอที่ตรงกับผลการพิจารณาของผู้เชี่ยวชาญ

13. ความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบ หมายถึง ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบด้วยวิธีซิปเทสต์หรือวิธีดีเอฟไอที่ตรงกับผลการพิจารณาของผู้เชี่ยวชาญ

14. ผู้เชี่ยวชาญ หมายถึง คณะบุคคลจำนวน 6 คน ประกอบด้วยเพศชาย 3 คน เพศหญิง 3 คน ที่สำเร็จการศึกษาตั้งแต่ระดับปริญญาโทบัณฑิตศึกษาด้านการวัดและประเมินผลการศึกษาขึ้นไป ทำหน้าที่เกี่ยวข้องกับการเรียนการสอนวิชาคณิตศาสตร์ ระดับมัธยมศึกษา เป็นเวลานานตั้งแต่ 5 ปีขึ้นไป ในการศึกษาครั้งนี้ผู้เชี่ยวชาญเป็นผู้พิจารณาคุณภาพข้อสอบในแบบทดสอบและตัดสินการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบที่ใช้เป็นเครื่องมือในการวิจัย

กรอบความคิดในการวิจัย

ประเด็นในการศึกษา : เปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ของหมวดข้อสอบ และของแบบทดสอบ จากข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ โดยอาศัยกรอบความคิด ดังนี้



ข้อตกลงเบื้องต้น

ข้อสอบวิชาคณิตศาสตร์ชั้น มัธยมศึกษาปีที่ 1 ที่ผู้วิจัยสร้างขึ้นทุกข้อมุ่งวัดความสามารถในวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) ซึ่งเป็นความสามารถหลัก (θ) โดยมีข้อสอบบางข้อผู้ตอบข้อสอบจะต้องใช้ความสามารถอย่างน้อย 2 อย่างจึงจะตอบข้อสอบแต่ละข้อได้ถูกต้อง คือ ใช้ความสามารถในวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) ซึ่งเป็นความสามารถหลัก (θ) ที่ข้อสอบต้องการวัด และความสามารถอื่นนอกเหนือจากที่ข้อสอบต้องการวัดหรือความสามารถรอง (η) แต่เป็นความสามารถที่ต้องนำมาใช้จึงจะตอบข้อสอบได้ถูกต้อง เป็นต้นว่า ความสามารถในการอ่านและแก้ไขโจทย์ปัญหาที่ซับซ้อน ไหวพริบในการทำข้อสอบ ความคุ้นเคยหรือมีประสบการณ์ในสถานการณ์ที่โจทย์กำหนด และผลการพิจารณาตัดสินการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบของผู้เชี่ยวชาญเป็นที่เชื่อถือได้

ข้อจำกัดของการวิจัย

ในการศึกษานี้เป็นการศึกษาตามกรอบความคิดของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติขั้นพื้นฐานเท่านั้น การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบมุ่งศึกษากรณีที่กำหนดให้ข้อสอบข้ออื่น ๆ ในแบบทดสอบไม่แสดงการทำหน้าที่เบี่ยงเบน (DIF free) ในขณะที่ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบข้อใดข้อหนึ่ง

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการศึกษาครั้งนี้จะนำเสนอเนื้อหาในหัวข้อต่าง ๆ เรียงลำดับดังนี้

1. แนวคิดเกี่ยวกับทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT)
2. แนวคิดเกี่ยวกับทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ (Multidimensional Item Response Theory : MIRT)
3. การทำหน้าที่เบี่ยงเบนตามแนวคิดของทฤษฎีการตอบข้อสอบ (Differential Functioning based on IRT)
4. การทำหน้าที่เบี่ยงเบนตามแนวคิดของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ (Differential Functioning based on MIRT)

1. แนวคิดเกี่ยวกับทฤษฎีการตอบข้อสอบ

การวัดและประเมินผลทางการศึกษาและทางจิตวิทยาได้อาศัยแนวคิดของทฤษฎีการทดสอบแบบมาตรฐานเดิม (Classical Test Theory : CTT) เป็นเวลานานตั้งแต่ต้นศตวรรษที่ 20 จนถึงปัจจุบัน ทฤษฎีการทดสอบแบบมาตรฐานเดิมเริ่มต้นแนวคิดที่ว่าคะแนนที่สังเกตได้เกิดจากองค์ประกอบที่สำคัญสองส่วน คือ คะแนนจริงและคะแนนความคลาดเคลื่อน ($X = T + E$) แนวคิดนี้มีการพัฒนาสืบเนื่องกันมาและได้รับความนิยมนอย่างแพร่หลาย แต่ผู้เชี่ยวชาญด้านการวัดหลายท่านพบว่า ทฤษฎีดังกล่าวมีจุดอ่อนที่สำคัญหลายประการดังต่อไปนี้

1. พารามิเตอร์ข้อสอบ (item parameter) ได้แก่ ความยาก (item difficulty) อำนาจจำแนก (item discrimination) ไม่คงที่ มีการเปลี่ยนแปลงไปขึ้นอยู่กับกลุ่มผู้ตอบข้อสอบในแต่ละครั้ง นอกจากนี้ยังไม่สามารถพิจารณาพารามิเตอร์ของข้อสอบแยกจากความสามารถของผู้ตอบข้อสอบแต่ละคนได้อย่างเป็นอิสระ คุณลักษณะทั้งสองส่วนขึ้นอยู่กับ การแจกแจงลักษณะของกลุ่มตัวอย่างที่ใช้ในการทดสอบเฉพาะคราว

2. การให้คะแนนของผู้ตอบข้อสอบแต่ละคนคิดจากคะแนนรวมในการทำแบบทดสอบ เป็นการพิจารณาระดับแบบทดสอบ (test-oriented) มากกว่าการพิจารณารูปแบบการตอบข้อสอบรายข้อ (item-oriented) การแปลความหมายของคะแนนบอกไม่ได้ว่าเขามีความสามารถในด้านใด และบกพร่องในด้านใด นอกจากนี้ยังมีปัญหาในการสร้างแบบทดสอบคู่ขนาน ไม่สามารถนำคะแนนจากการทำแบบทดสอบต่างฉบับมาเปรียบเทียบกัน เพราะคะแนนจากแบบทดสอบทั้งสองมีมาตรในการวัด (scale) ต่างกัน เนื่องจากค่าสถิติในการพิจารณาคัดเลือกข้อสอบมาจัดทำแบบทดสอบแต่ละฉบับขึ้นอยู่กับกลุ่มผู้ตอบข้อสอบตามที่ได้กล่าวมาแล้ว

3. การคำนวณค่าความเชื่อมั่น (reliability) ตามแนวคิดของทฤษฎีการทดสอบแบบมาตรฐานเดิมเกี่ยวข้องกับความแปรปรวนของคะแนนรวม (total score variance) จากการทำแบบทดสอบ แนวคิดดังกล่าวเขียนเป็นสมการได้ดังนี้

$$\text{ค่าความเชื่อมั่น} = 1 - (\text{ความแปรปรวนของความคลาดเคลื่อน} / \text{ความแปรปรวนของคะแนนรวม})$$

ในการประมาณค่าคะแนนจริง (true score) ของผู้ตอบข้อสอบแต่ละคนและการกำหนดช่วงความเชื่อมั่นขึ้นอยู่กับ การแจกแจงกลุ่มตัวอย่างผู้ตอบข้อสอบแต่ละกลุ่ม

ประมาณ ค.ศ. 1936 มีผู้ริเริ่มเสนอแนวคิดเกี่ยวกับแบบจำลองการวัดคุณลักษณะของบุคคลเพื่อแก้ไขจุดอ่อนของทฤษฎีการทดสอบแบบมาตรฐานเดิม การพัฒนาเชิงคณิตศาสตร์ของทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) เริ่มต้นโดย ริชาร์ดสัน (Richardson. 1936) ทักเคอร์ (Tucker. 1936) ลอว์เลย์ (Lowley. 1944) ลอร์ด (Lord. 1952) แต่ผู้ที่ทำให้แนวคิดนี้สามารถนำไปปฏิบัติได้คือ เบิร์นบอรัม (Birnbuam. 1957 ; อุทุมพร จามรมาน. 2540 : 155)

ทฤษฎีการตอบข้อสอบมีลักษณะที่สำคัญดังนี้

- 1) คุณลักษณะของพารามิเตอร์ข้อสอบไม่ขึ้นอยู่กับกลุ่มผู้ตอบข้อสอบเฉพาะกลุ่มใดกลุ่มหนึ่ง
- 2) มีการอธิบายความสามารถของผู้ตอบข้อสอบแยกออกจากพารามิเตอร์ข้อสอบ
- 3) อธิบายความหมายการตอบข้อสอบของผู้ตอบข้อสอบในระดับข้อสอบมากกว่าระดับแบบทดสอบ
- 4) ในการคำนวณค่าสารสนเทศของแบบทดสอบไม่ใช่แบบทดสอบคู่ขนาน
- 5) อธิบายความหมายการตอบข้อสอบของผู้ตอบข้อสอบทุกคนอย่างถูกต้องที่ทุกระดับความสามารถและประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดเป็นรายบุคคล

ทฤษฎีการตอบข้อสอบเป็นทฤษฎีที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะภายในหรือความสามารถของผู้ตอบข้อสอบกับพฤติกรรมการตอบข้อสอบด้วยฟังก์ชันทางคณิตศาสตร์ที่เขียนในรูปแบบจำลองการตอบข้อสอบ และแสดงการตอบข้อสอบรายข้อด้วยโค้งคุณลักษณะข้อสอบ (Item Characteristic Curve : ICC) ที่แสดงความสัมพันธ์ระหว่างระดับความสามารถของบุคคลที่วัดด้วยข้อสอบข้อนั้นกับโอกาสที่บุคคลที่มีความสามารถระดับต่าง ๆ จะตอบข้อสอบได้ถูกต้อง

1.1 ความคิดพื้นฐานของทฤษฎีการตอบข้อสอบ

ทฤษฎีการตอบข้อสอบตั้งอยู่บนสมมุติฐาน 2 ประการคือ

- 1) พฤติกรรมการตอบข้อสอบของผู้ตอบข้อสอบซึ่งเป็นสิ่งที่สังเกตได้โดยตรงนั้น สามารถใช้ในการพยากรณ์หรืออธิบายคุณลักษณะภายใน (traits, latent traits or abilities) ของบุคคลนั้น ๆ ได้
- 2) ความสัมพันธ์ระหว่างพฤติกรรมการตอบข้อสอบของผู้ตอบข้อสอบและคุณลักษณะภายในอธิบายได้โดยฟังก์ชันการตอบข้อสอบ (item characteristic function หรือ item characteristic curve : ICC) ฟังก์ชันการตอบข้อสอบเป็นฟังก์ชันที่มีลักษณะเพิ่มขึ้นทางเดียว (monotonically increasing function) เมื่อผู้ตอบข้อสอบมีความสามารถสูงความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องก็มากขึ้นด้วย

1.2 ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อสอบ

ทฤษฎีการตอบข้อสอบมีข้อตกลงเบื้องต้นที่สำคัญดังนี้

1.2.1 ข้อตกลงเกี่ยวกับมิติของคุณลักษณะภายใน (dimensionality of the latent space)

ทฤษฎีการตอบข้อสอบถือข้อตกลงว่าความสามารถของมนุษย์มีอยู่ทั้งหมด k อย่าง ซึ่งความสามารถแต่ละอย่างนี้ต่างก็ส่งผลต่อการตอบข้อสอบที่รวมกันเป็นแบบทดสอบ ถ้าหากคะแนนของผู้ตอบข้อสอบสามารถอธิบายได้ด้วยความสามารถเพียงอย่างเดียวก็เพียงพอแล้ว เรียกข้อตกลงดังกล่าวว่าความเป็นมิติเดียว (unidimensionality) แต่ถ้าหากคะแนนของผู้ตอบข้อสอบต้องอธิบายด้วยความสามารถหลายอย่าง เรียกข้อตกลงนั้นว่า การใช้ความสามารถหลายมิติ (multidimensionality)

วิวัฒนาการของทฤษฎีการตอบข้อสอบในช่วงแรกเกี่ยวข้องกับความเป็นมิติเดียวของแบบทดสอบ นั่นคือ มีข้อตกลงว่าแบบทดสอบมุ่งวัดคุณลักษณะเด่นเพียงลักษณะเดียว ซึ่งหมายความว่าคุณลักษณะภายในหรือความสามารถของบุคคลที่เป็นตัวกำหนดพฤติกรรมการตอบข้อสอบแต่ละข้อมีลักษณะเด่นที่สำคัญเพียงลักษณะเดียว หรือกล่าวได้ว่าข้อสอบทุกข้อในแบบทดสอบมีลักษณะของความเป็นเอกพันธ์

แบบทดสอบประเภทใช้ความเร็วในการตอบ (speed test) ผู้ตอบข้อสอบต้องใช้ความสามารถอย่างน้อยสองมิติ คือ ความเร็วในการตอบ (response speed) และความสามารถที่แบบทดสอบต้องการวัด (trait being measured) ดังนั้นเพื่อให้สอดคล้องกับข้อตกลงเบื้องต้นเกี่ยวกับความเป็นมิติเดียว แบบทดสอบที่นำมาใช้จึงเป็นแบบทดสอบที่ไม่ใช้ความเร็วในการตอบ (nonspeed test) ทำให้ทุกคนมีเวลาทำข้อสอบทุกข้อ และสามารถใช้คะแนนรวมจากการทำแบบทดสอบร่วมกับลักษณะข้อสอบเป็นตัวประมาณค่าความสามารถที่แท้จริงของผู้ตอบข้อสอบโดยไม่ต้องพิจารณาเรื่องความเร็ว

1.2.2 ข้อตกลงเกี่ยวกับความเป็นอิสระในการตอบข้อสอบ (local independence)

เป็นข้อตกลงที่กำหนดว่าโอกาสที่แต่ละคนจะตอบข้อสอบแต่ละข้อได้ถูกต้องเป็นอิสระจากกัน ซึ่งมีความหมายดังนี้

1) ความเป็นอิสระระหว่างข้อสอบ หมายความว่าข้อสอบแต่ละข้อเป็นอิสระจากกัน กล่าวคือ การตอบข้อสอบข้อหนึ่งไม่มีผลกระทบต่อคำตอบข้อสอบข้ออื่น ๆ ในแบบทดสอบฉบับนั้น

2) ความเป็นอิสระระหว่างผู้ตอบข้อสอบ หมายความว่าผู้ตอบข้อสอบแต่ละคนตอบข้อสอบแต่ละข้ออย่างเป็นอิสระจากกัน นั่นคือ

$$\text{ถ้า } P_i(\theta_A) \text{ เป็นอิสระจาก } P_i(\theta_B) \rightarrow P_i(\theta_A) \cap P_i(\theta_B) = P_i(\theta_A) \cdot P_i(\theta_B)$$

อธิบายในเชิงสถิติได้ดังนี้ ถ้าให้ θ เป็นความสามารถที่มีอิทธิพลต่อการตอบข้อสอบของผู้ตอบข้อสอบ U_i เป็นผลการตอบข้อสอบข้อที่ i ของผู้ตอบข้อสอบที่สุ่มเลือกได้ ($i = 1, 2, \dots, n$) และ $P(U_i | \theta)$ เป็นความน่าจะเป็นในการตอบข้อสอบของผู้ที่มีความสามารถ θ $P(U_i = 1 | \theta)$ เป็นความน่าจะเป็นในการตอบข้อสอบถูก และ $P(U_i = 0 | \theta)$ เป็นความน่าจะเป็นในการตอบข้อสอบผิด คุณสมบัติของความเป็นอิสระในการตอบข้อสอบ (local independence) แสดงได้ดังนี้

$$\begin{aligned} \text{Prob}(U_1, U_2, \dots, U_n | \theta) &= P(U_1 | \theta) P(U_2 | \theta) \dots P(U_n | \theta) \\ &= \prod_{i=1}^n P(U_i | \theta) \end{aligned}$$

หมายความว่าสำหรับผู้ตอบข้อสอบคนใดคนหนึ่งหรือผู้ตอบข้อสอบทุกคนที่มีความสามารถตามที่กำหนด ความน่าจะเป็นของการตอบข้อสอบทั้งฉบับได้ถูกต้องเท่ากับผลคูณของความน่าจะเป็นในการตอบข้อสอบแต่ละข้อ เช่น ถ้ารูปแบบการตอบข้อสอบ 3 ข้อของผู้ตอบข้อสอบคนหนึ่งเป็น (1, 1, 0) นั่นคือ $U_1 = 1, U_2 = 1, U_3 = 0$ แล้วจะได้ว่า

$$\begin{aligned} P(U_1 = 1, U_2 = 1, U_3 = 0 | \theta) &= P(U_1 = 1 | \theta) \cdot P(U_2 = 1 | \theta) \cdot P(U_3 = 0 | \theta) \\ &= P_1 P_2 Q_3 \end{aligned}$$

$$\text{เมื่อ } P_i = P(U_i = 1 | \theta) \quad \text{และ} \quad Q_i = 1 - P_i$$

ในการทำข้อสอบแต่ละข้อผู้ตอบข้อสอบอาจใช้ความสามารถหลายอย่าง ถ้าสามารถกำจัด (partialled out) ความสามารถที่ไม่ต้องการวัดออกไป หรือทำให้มีความคงที่ (held constant) ก็จะทำให้การตอบข้อสอบแต่ละข้อของแต่ละคนมีความอิสระที่เรียกว่า ความมีอิสระอย่างมีเงื่อนไข (conditional independence) ถ้าข้อตกลงของความเป็นมิติเดียวเป็นจริงก็จะมีคุณสมบัติของความเป็นอิสระในการตอบข้อสอบด้วย (Hambleton. 1991 : 11 ; citing Lord. 1980; Lord & Novick. 1968)

ความเป็นอิสระในการตอบข้อสอบเกิดขึ้นได้แม้ว่าชุดของข้อมูลไม่ได้มีความเป็นมิติเดียว ความเป็นอิสระในการตอบข้อสอบจะเกิดขึ้นเมื่อกำหนดคุณลักษณะภายในอย่างสมบูรณ์ และความสามารถทั้งหลายเหล่านั้นต่างส่งผลต่อการตอบข้อสอบ (Hambleton. 1991 : 11)

1.2.3 ข้อตกลงเกี่ยวกับโค้งคุณลักษณะข้อสอบ (Item Characteristic Curve : ICC หรือ Item Response Function : IRF) เป็นฟังก์ชันแสดงความสัมพันธ์ระหว่างระดับความสามารถของผู้ตอบข้อสอบ (θ_p) กับโอกาสในการตอบข้อสอบได้ถูกต้อง $[P(\theta_i)]$ แสดงโดยโค้งคุณลักษณะข้อสอบของแต่ละแบบจำลองที่เลือกมาใช้ในการอธิบาย โดยที่โค้งคุณลักษณะข้อสอบมีคุณสมบัติไม่แปรเปลี่ยน (invariant) ไปตามกลุ่มผู้ตอบข้อสอบ

1.3 แบบจำลองการตอบข้อสอบ (Item Response Model)

แบบจำลองการตอบข้อสอบแบ่งตามข้อตกลงเกี่ยวกับคุณลักษณะภายในที่ใช้ในการตอบข้อสอบเป็นสองกลุ่ม คือ แบบจำลองที่ใช้ข้อตกลงเกี่ยวกับความเป็นมิติเดียว และแบบจำลองที่ใช้ข้อตกลงเกี่ยวกับการใช้ความสามารถหลายมิติ

1.3.1 แบบจำลองที่ใช้ข้อตกลงเกี่ยวกับความเป็นมิติเดียว แบ่งเป็นสองกลุ่ม

1.3.1.1 แบบจำลองที่ใช้กับการตรวจให้คะแนนรายข้อแบบสองค่า (dichotomous item response theory model)

1.3.1.2 แบบจำลองที่ใช้กับการตรวจให้คะแนนรายข้อแบบไม่ใช่สองค่า (nondichotomous item response theory model) ได้แก่ การตรวจให้คะแนนรายข้อแบบหลายค่า (polytomous item response theory model) การตรวจให้คะแนนแบบต่อเนื่อง (continuous item response theory model) เป็นต้น

1.3.2 แบบจำลองที่ใช้ข้อตกลงเกี่ยวกับการใช้ความสามารถหลายมิติ แบ่งเป็นสองกลุ่ม

1.3.2.1 แบบจำลองที่ใช้กับการตรวจให้คะแนนรายข้อแบบสองค่า (dichotomous multidimensional item response theory model) ได้แก่ แบบจำลองหลายมิตินอร์มอลอจีฟ (Normal-Ogive Multidimensional Model; McDonald. 1987) แบบจำลองหลายมิตินีร์โลจิสติกสำหรับการตรวจให้คะแนนรายข้อแบบสองค่า (A Linear Logistic Multidimensional Model for Dichotomous Item Response Data; Reckase. 1985) เป็นต้น

1.3.2.2 แบบจำลองที่ใช้กับการตรวจให้คะแนนรายข้อแบบไม่ใช่สองค่า (nondichotomous multidimensional item response model) ได้แก่ แบบจำลองหลายมิติลอกลิเนียร์สำหรับการตรวจให้คะแนนรายข้อแบบหลายค่า (Loglinear Multidimensional Item Response Models for Polytomously

Scored Items; Kelderman. 1997) แบบจำลองการตอบหลายองค์ประกอบ (Multicomponent Response Models; Embretson. 1997) เป็นต้น

ในที่นี้จะกล่าวเฉพาะแบบจำลองที่ใช้ข้อตกลงของความเป็นมิติเดียว และใช้กับการตรวจให้คะแนนรายข้อแบบสองค่า (unidimensional dichotomous IRT model) เพราะเป็นแบบจำลองที่ใช้กันมากในวงการวัดและประเมินผลการศึกษา แบบจำลองที่นิยมใช้กัน คือ แบบจำลองโลจิสติกแบบ 1 2 และ 3 พารามิเตอร์ ซึ่งมีฟังก์ชันทางคณิตศาสตร์ของแบบจำลองดังนี้

ตาราง 1 ฟังก์ชันโลจิสติกของแบบจำลองการตอบข้อสอบ

แบบจำลอง	ฟังก์ชันโลจิสติก
1 พารามิเตอร์	$P_i(\theta) = \frac{e^{D(\theta-b_i)}}{1 + e^{D(\theta-b_i)}}, i = 1, 2, \dots, n$
2 พารามิเตอร์	$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}}, i = 1, 2, \dots, n$
3 พารามิเตอร์	$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}}, i = 1, 2, \dots, n$

โดยที่ θ คือ ระดับความสามารถของผู้ตอบข้อสอบใด ๆ ที่ประมาณได้จากแบบจำลองตามทฤษฎีการตอบข้อสอบ ปรับให้เป็นคะแนนมาตรฐานที่มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 θ มีพิสัยระหว่าง $\pm \infty$ แต่ในทางปฏิบัติส่วนใหญ่ใช้ค่า θ ระหว่าง ± 3

$P_i(\theta)$ คือ ความน่าจะเป็นที่ผู้ตอบข้อสอบที่ได้มาจากการสุ่มและมีความสามารถ θ ตอบคำถามข้อที่ i ได้ถูกต้อง

b_i คือ ค่าความยากของข้อสอบข้อที่ i ซึ่งเป็นการวัดตำแหน่งของโค้งคุณลักษณะข้อสอบ (ICC) ตามแกนนอนบนสเกล θ ณ จุดที่โค้งมีความชันมากที่สุด (จุดเปลี่ยนโค้ง) หรือที่ตำแหน่งต่อไปนี้

$$b_i = \theta \text{ ที่ } P_i(\theta) = 0.5 \quad (\text{สำหรับแบบจำลอง 1 และ 2 พารามิเตอร์})$$

$$b_i = \theta \text{ ที่ } P_i(\theta) = \frac{1 + c_i}{2} \quad (\text{สำหรับแบบจำลอง 3 พารามิเตอร์})$$

ค่า b_i มีพิสัยระหว่าง $\pm \infty$ แต่ในทางปฏิบัติส่วนใหญ่ใช้ค่า b_i ระหว่าง ± 2.5

a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i ซึ่งเป็นสัดส่วน (proportion) ต่อความชันของโค้งคุณลักษณะข้อสอบ (ICC) ณ จุดเปลี่ยนโค้ง หรือที่จุด $\theta = b_i$ ค่า a_i มีพิสัยระหว่าง

$\pm \infty$ ในทางปฏิบัติส่วนใหญ่ใช้ค่า a_i ระหว่าง $+ 0.5$ ถึง $+ 2.5$ a_i มีค่าสูงแสดงว่าข้อสอบข้อนั้นมีความชัน (Slope) ที่มีค่าสูงจึงจำแนกผู้ที่มีความสามารถแตกต่างกันได้ดี
 $[\text{Slope} = .425 a_i (1 - c_i)]$

- c_i คือ ความน่าจะเป็นของการเดาได้ถูกต้อง ซึ่งเป็นความน่าจะเป็นที่ผู้ตอบข้อสอบที่มีความสามารถต่ำมาก ๆ ตอบข้อสอบข้อที่ i ได้ถูกต้อง จึงเป็นค่ากำกับต่ำสุด (lower assymtote) ของโค้งคุณลักษณะข้อสอบ (ICC) ค่า c_i มีพิสัยระหว่าง 0 ถึง 1 ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่า c_i อยู่ระหว่าง 0 ถึง 0.3
- e คือ ค่าคงที่ของลอการิทึมธรรมชาติ (natural logarithm) มีค่าประมาณ 2.71828
- D คือ ค่าองค์ประกอบของการปรับสเกลให้โลจิสติกฟังก์ชัน (logistic function) มีค่าใกล้เคียงกับนอร์มอลอจีฟฟังก์ชัน (normal ogive function) มากที่สุดเท่าที่จะเป็นไปได้ D มีค่าเท่ากับ 1.7

โค้งคุณลักษณะข้อสอบตามทฤษฎีการตอบข้อสอบแสดงดังภาพ

ก) แบบจำลองโลจิสติก

1 พารามิเตอร์

 b_i มีค่าแปรเปลี่ยนได้ a_i มีค่าคงที่ $c_i = 0$

ข) แบบจำลองโลจิสติก

2 พารามิเตอร์

 b_i ค่าแปรเปลี่ยนได้ a_i มีค่าแปรเปลี่ยนได้ $c_i = 0$

ค) แบบจำลองโลจิสติก

3 พารามิเตอร์

 b_i มีค่าแปรเปลี่ยนได้ a_i มีค่าแปรเปลี่ยนได้ c_i มีค่าแปรเปลี่ยนได้

ภาพประกอบ 2 ไค์คุณลักษณะของข้อสอบตามแบบจำลองโลจิสติก 1 2 และ 3 พารามิเตอร์

1.4 ฟังก์ชันสารสนเทศ

การวิเคราะห์ตามทฤษฎีการตอบข้อสอบ ใช้แบบแผนการตอบข้อสอบเป็นรายชื่อในการประมาณค่าความสามารถของผู้ตอบข้อสอบ การประเมินคุณภาพของแบบทดสอบจึงสามารถพิจารณาจากความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้ตอบข้อสอบ ดังนั้นจึงมีความถูกต้องดังกล่าวเรียกว่า สารสนเทศของแบบทดสอบ ซึ่งเป็นค่าฟังก์ชันสารสนเทศของแบบทดสอบ โดยมีสูตรคำนวณค่าต่าง ๆ ที่เกี่ยวข้องดังนี้

1.4.1 ความสามารถหรือคะแนนจริงของผู้ตอบข้อสอบ

$$T_p = \sum_{i=1}^n P_i(\theta_p)$$

เมื่อ T_p คือ คะแนนของผู้ตอบข้อสอบคนที่ P ซึ่งมีความสามารถระดับ θ

n คือ จำนวนข้อสอบในแบบทดสอบทั้งฉบับ

$P_i(\theta_p)$ คือ ความน่าจะเป็นที่ผู้ตอบข้อสอบที่มีความสามารถระดับ θ ตอบข้อสอบข้อที่ i ได้ถูกต้อง

1.4.2 ค่าฟังก์ชันสารสนเทศของข้อสอบ (item information function)

$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta) Q_i(\theta)}, \quad i = 1, 2, \dots,$$

เมื่อ $I_i(\theta)$ คือ สารสนเทศที่ได้จากข้อสอบข้อที่ i ที่มีความสามารถระดับ θ

$P'_i(\theta)$ คือ ความชันของฟังก์ชันการตอบข้อสอบ $[P_i(\theta)]$ ที่มีความสามารถระดับ θ
(derivative of $P_i(\theta)$ with respect to θ)

$Q_i(\theta)$ คือ $1 - P_i(\theta)$

1.4.3 ค่าฟังก์ชันสารสนเทศของแบบทดสอบ (test information function)

$$I(\theta) = \sum_{i=1}^n I_i(\theta)$$

1.4.4 ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ (standard error of estimates at ability θ)

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

จากสูตรคำนวณข้างต้นจะเห็นได้ว่าถ้าค่าฟังก์ชันสารสนเทศของแบบทดสอบมีค่าสูงที่ความสามารถของผู้ตอบข้อสอบระดับ θ ใด ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถของ
ผู้ตอบข้อสอบที่ระดับ θ นั้นจะมีค่าต่ำ ดังแสดงในภาพประกอบ 3 และ 4

ภาพประกอบ 3 โค้งสารสนเทศของข้อสอบ 6 ข้อ

ภาพประกอบ 4 โครงสร้างสารสนเทศของแบบทดสอบที่ประกอบด้วยข้อสอบ 6 ข้อ ในภาพประกอบ 3

1.5 การประยุกต์ใช้ทฤษฎีการตอบข้อสอบ

ทฤษฎีการตอบข้อสอบสามารถนำไปประยุกต์ใช้ในด้านต่าง ๆ ที่สำคัญดังนี้

1.5.1 เปรียบเทียบประสิทธิภาพของแบบทดสอบ

ตามแนวความคิดของทฤษฎีการตอบข้อสอบทำให้สามารถเปรียบเทียบประสิทธิภาพของแบบทดสอบต่างฉบับที่วัดคุณลักษณะเดียวกันที่ระดับความสามารถเดียวกันได้ด้วยการคำนวณประสิทธิภาพสัมพัทธ์ (Relative Efficiency) ของแบบทดสอบใช้สูตรดังนี้

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

เมื่อ $RE(\theta)$ คือ ประสิทธิภาพสัมพัทธ์

$I_A(\theta)$ คือ ฟังก์ชันสารสนเทศของแบบทดสอบ A ที่ระดับความสามารถ θ

$I_B(\theta)$ คือ ฟังก์ชันสารสนเทศของแบบทดสอบ B ที่ระดับความสามารถ θ

1.5.2 พัฒนาแบบทดสอบให้สอดคล้องกับจุดมุ่งหมายของการสอบ

ทำได้โดยการเขียนข้อสอบ วิเคราะห์ข้อสอบและเลือกข้อสอบที่มีสารสนเทศสูง ๆ ที่ระดับความสามารถที่ต้องการนำมารวบรวมไว้ในคลังข้อสอบ การใช้ทฤษฎีการตอบข้อสอบมาพัฒนาคลังข้อสอบทำให้ข้อสอบแต่ละข้อมีค่าสถิติต่าง ๆ ประจำข้อที่ไม่ขึ้นอยู่กับการสุ่มตัวอย่างของผู้ตอบข้อสอบ (sample-free) จึงทำให้สามารถคัดเลือกข้อสอบมาจัดทำแบบทดสอบได้ตามจุดมุ่งหมายในการทดสอบแต่ละครั้งอย่างมีประสิทธิภาพ เช่น ในการสอบคัดเลือกก็สามารถคัดเลือกข้อสอบที่มีสารสนเทศสูงที่ระดับความสามารถสูง ๆ เป็นต้น

1.5.3 พัฒนาการทดสอบที่มีการคัดเลือกข้อสอบให้เหมาะสมกับความสามารถของผู้ตอบข้อสอบ (Adaptive Testing)

ทฤษฎีการตอบข้อสอบช่วยให้สามารถคัดเลือกข้อสอบให้เหมาะสมกับความสามารถของผู้ตอบข้อสอบได้ นอกจากนั้นยังสามารถประมาณค่าความสามารถของผู้ตอบข้อสอบได้อย่างมีประสิทธิภาพ ถึงแม้จะใช้แบบทดสอบต่างชุดกัน การทดสอบที่มีการคัดเลือกข้อสอบให้เหมาะสมกับความสามารถของผู้ตอบข้อสอบเป็นความพยายามที่จะคัดเลือกข้อสอบที่มีระดับความยากเหมาะสมกับผู้ตอบข้อสอบแต่ละคน โดยที่ผู้ตอบข้อสอบไม่จำเป็นต้องทำข้อสอบเหมือนกันทุกข้อ แต่ยังคงประมาณระดับความสามารถของผู้ตอบข้อสอบได้ถูกต้องแม่นยำ หลักการของการทดสอบที่มีการคัดเลือกข้อสอบให้เหมาะสมกับความสามารถของผู้ตอบข้อสอบคือถ้าผู้ตอบข้อสอบตอบข้อสอบข้อใดได้ถูกต้อง ข้อสอบข้อถัดไปควรเป็นข้อที่ยากขึ้น แต่ถ้าตอบผิดข้อสอบข้อถัดไปควรง่ายลง ถ้านำคอมพิวเตอร์เข้ามาช่วยการทดสอบตามหลักการนี้เรียกว่า การทดสอบที่ปรับเหมาะกับความสามารถของผู้ตอบข้อสอบโดยใช้คอมพิวเตอร์ (Computerized Adaptive Testing : CAT)

1.5.4 การเทียบคะแนนของแบบทดสอบ (Test Equating)

การเทียบคะแนนตามทฤษฎีการตอบข้อสอบมีข้อดีคือ ผลการเทียบคะแนนไม่ได้ขึ้นอยู่กับระดับความสามารถของกลุ่มที่นำมาเปรียบเทียบและไม่ได้รับผลกระทบจากแบบทดสอบที่มีความยากต่างกัน จึงทำให้การพิจารณาเปรียบเทียบคะแนนบนสเกลร่วม (common scale) ทำได้อย่างมีประสิทธิภาพ

1.5.5 การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (Differential Item Functioning Detection)

นิยามของการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) ที่นักวัดผลยอมรับ คือข้อสอบแสดงการทำหน้าที่เบี่ยงเบน ถ้าบุคคลที่มีความสามารถเท่ากันแต่อยู่ต่างกลุ่มกันมีความน่าจะเป็นในการตอบข้อสอบข้อนั้น ๆ ได้ถูกต้องไม่เท่ากัน จากนิยามดังกล่าวจะเห็นได้ว่าแนวคิดของทฤษฎีการตอบข้อสอบสามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบได้อย่างมีประสิทธิภาพ

1.5.6 การประยุกต์ใช้ในด้านอื่น ๆ เช่น ใช้ศึกษาในกรณีที่มีความน่าจะเป็นในตอบคำถามได้ถูกต้องไม่ได้เป็นการเพิ่มขึ้นทางเดียว (nonmonotone items) ซึ่งใช้วัดทัศนคติ หรือความคิดเห็นเกี่ยวกับเรื่องต่าง ๆ (Andric. 1997 ; citing van der Linden & Hambleton. 1997) เป็นต้น

1.6 แนวโน้มของทฤษฎีการตอบข้อสอบ

ทฤษฎีการตอบข้อสอบช่วยให้นักวัดผลสามารถวางแผนและสร้างเครื่องมือวัด แปลความหมายผลจากการวัดและรายงานผลได้อย่างมีประสิทธิภาพ ในอนาคตทิศทางการวัดผลการศึกษาเปลี่ยนจากการใช้แบบทดสอบมาตรฐานที่มีการตรวจให้คะแนนรายข้อแบบสองค่าไปเป็นรูปแบบอื่น ๆ มากยิ่งขึ้น นักวิจัยจึงสนใจศึกษาเพื่อพัฒนาแบบจำลองให้ครอบคลุมสถานการณ์การสอบแบบต่าง ๆ เช่น ความแตกต่างของกลยุทธ์ในการตอบข้อสอบของผู้ตอบข้อสอบแต่ละคน ความเป็นไปได้ที่จะเกิดการเรียนรู้ระหว่างการทดสอบ เป็นต้น

ในอนาคตนักวิจัยจะมุ่งศึกษาเกี่ยวกับแบบจำลองการตอบข้อสอบที่ใช้ข้อตกลงเบื้องต้นเกี่ยวกับความเป็นมิติเดียวและใช้วิธีการตรวจให้คะแนนรายข้อแบบไม่ใช่สองค่า (nondichotomous unidimensional response models) และแบบจำลองการตอบข้อสอบที่ใช้ข้อตกลงเบื้องต้นเกี่ยวกับการใช้ความสามารถหลายมิติทั้งประเภทที่ใช้การตรวจให้คะแนนรายข้อแบบสองค่าและการตรวจให้คะแนนรายข้อแบบไม่ใช่สองค่า แบบจำลองการตอบข้อสอบอธิบายโดยใช้แบบจำลองทางคณิตศาสตร์ เมื่อใช้ในสถานการณ์การสอบที่ซับซ้อนมากขึ้นแบบจำลองที่ใช้ในการอธิบายการประมาณค่าพารามิเตอร์ของผู้ตอบข้อสอบและพารามิเตอร์ของข้อสอบก็มีความซับซ้อนมากขึ้นตามลำดับ

2. แนวคิดเกี่ยวกับทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ

ตามที่ได้กล่าวมาแล้วว่าการวัดและประเมินผลทางการศึกษาและทางจิตวิทยามีความซับซ้อนมากขึ้น การใช้ทฤษฎีการตอบข้อสอบโดยใช้ข้อตกลงของความเป็นมิติเดียวไม่มีประสิทธิภาพเพียงพอที่จะอธิบายคุณลักษณะของบุคคล (Goldstein, 1996) จึงมีความพยายามในการพัฒนาทฤษฎีการทดสอบที่นำไปสู่พัฒนาการของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ โดยเริ่มตั้งแต่ ฟรานซิส กาลตัน (Francis Galton, 1884) ได้แนะนำให้รู้จักคุณลักษณะภายในทางสมอง (mental trait) โดยเปรียบเทียบกับขนาดทางกายภาพ (physical size) เขาสมมุติว่าจิตใจมีมิติ (dimension) ซึ่งต้องหาวิธีวัดโครงสร้างดังกล่าวให้ได้ นักจิตวิทยาใน ช่วงต้นศตวรรษที่ 20 จึงศึกษาเรื่องดังกล่าวอย่างต่อเนื่องและมีผลต่อการวัดเชิงปริมาณเป็นอย่างมาก พัฒนาการทางการวัดเริ่มต้นอย่างจริงจังเมื่อมีการกำหนดแบบจำลองในการวิเคราะห์องค์ประกอบ (Spearman, 1904) นำไปสู่การวิเคราะห์ข้อสอบ (item analysis) สำหรับออกแบบในการสร้างแบบทดสอบวัดความสามารถทางสมอง (Gullicksen, 1950) จนถึงแบบจำลองการตอบข้อสอบ (item response model) ซึ่งพัฒนาโดยลอร์ดและโนวิก (Lord & Novick, 1968) การพัฒนาเหล่านี้ยึดถือข้อตกลงเบื้องต้นเกี่ยวกับการวัดคุณลักษณะภายในที่เป็นมิติเดียว (Unidimensionality) ตลอดมา

ในการวัดผลทางการศึกษาและทางจิตวิทยาสิ่งแรกที่นักวัดผลต้องทำคือ เลือกตัวบ่งชี้ (indicators) ที่ใช้แสดงคุณลักษณะภายในของบุคคลที่ต้องการวัด แล้วจึงสร้างแบบจำลองในการวัดซึ่งประกอบด้วยขั้นตอนในการกำหนดข้อตกลงเบื้องต้นทางคณิตศาสตร์ที่สมเหตุสมผล กำหนดหลักการเกี่ยวกับมิติที่ต้องการวัด (dimensionality) และความเป็นอิสระในการตอบข้อสอบ (conditional independence) จึงมีความจำเป็นต่อนิยามมิติของคุณลักษณะภายในของบุคคล แล้วใช้ข้อมูลการตอบข้อสอบที่สังเกตได้ประมาณค่าคุณลักษณะภายใน

ฟังก์ชันอย่างง่ายที่แสดงความน่าจะเป็นในการตอบข้อสอบแต่ละข้อได้ถูกต้องโดยใช้คุณลักษณะภายในอย่างเดียวแสดงได้ดังนี้

$$\pi_{ij} = \alpha_j - \beta_i$$

เมื่อ α_j คือ ความสามารถของบุคคล j

β_i คือ ความยากของข้อสอบข้อที่ i

กำหนดคะแนนจากการสังเกตเป็น 0 ในกรณีที่ตอบผิด เป็น 1 ในกรณีที่ตอบถูก และตั้งสมมุติฐานว่าคำตอบมีการแจกแจงแบบทวินาม (binomial distribution) นอกจากนั้นสมมุติว่าคุณลักษณะภายในไม่มีความต่อเนื่อง สมการข้างต้นสามารถขยายโดยให้ความสัมพันธ์ระหว่างคุณลักษณะที่ตอบข้อสอบและความสามารถแปรเปลี่ยนไปในการตอบข้อสอบแต่ละข้อ ดังนี้

$$\pi_{ij} = \gamma_i \alpha_j - \beta_i$$

γ_i คือ อำนาจจำแนกของข้อสอบแต่ละข้อ ซึ่งเป็นอัตราส่วนการเปลี่ยนแปลงของความน่าจะเป็นในการตอบข้อสอบถูกต้องกับความสามารถของผู้ตอบข้อสอบ

ในบริบทของแบบจำลองการตอบข้อสอบมิติเดียว (one-dimensional model) กำหนดไว้ว่าข้อสอบที่มีสหสัมพันธ์ระดับต่ำกับข้อสอบส่วนใหญ่จะถูกคัดเลือกร้างออกไป ในกรณีของแบบจำลองการตอบข้อสอบสองมิติ (two-dimensional model) แสดงความสัมพันธ์ด้วยสมการอย่างง่ายได้ดังนี้

$$\pi_{ij} = \gamma_{1i} \alpha_{1j} + \gamma_{2i} \alpha_{2j} - \beta_i$$

γ_{1i} คือ อำนาจจำแนกในการวัดมิติที่ 1 ของข้อที่ i

γ_{2i} คือ อำนาจจำแนกในการวัดมิติที่ 2 ของข้อที่ i

α_{1j} คือ ความสามารถในการมิติที่ 1 ของบุคคล j ในการตอบข้อที่ i

α_{2j} คือ ความสามารถในการมิติที่ 2 ของบุคคล j ในการตอบข้อที่ i

เรคเคส (Reckase. 1994 : 87) กล่าวว่าช่วงแรกของการศึกษาเกี่ยวกับทฤษฎีการตอบข้อสอบ (item response theory : IRT) ต้องทำการตรวจสอบมิติของข้อมูลการตอบข้อสอบ เพราะการศึกษาในช่วงดังกล่าวมีข้อตกลงของความเป็นมิติเดียวในการวัด แต่ปัจจุบันทราบดีแล้วว่าข้อสอบส่วนใหญ่ไม่ได้วัดคุณลักษณะภายในเพียงลักษณะเดียว ในการทำข้อสอบแต่ละข้อส่วนใหญ่ผู้ตอบข้อสอบต้องใช้คุณลักษณะภายในหลายอย่าง (multiple traits) จึงมีการพัฒนาทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ (multidimensional item response theory : MIRT) ขึ้น และในการศึกษามีความจำเป็นต้องให้ความสำคัญกับความเหมาะสมของจำนวนมิติของคุณลักษณะภายในที่ต้องการวัดอย่างรอบคอบ เนื่องจากการประมาณค่าความเป็นมิติของชุดข้อมูลต่ำกว่าที่เป็นจริง ทำให้เกิดผลเสียมากกว่าการประมาณค่าจำนวนมิติสูงกว่าความเป็นจริง (Reckase. 1994 : 89 citing Hirsch & Miller. 1991) ไม่ว่าในกรณีของการตรวจสอบความเที่ยงตรงเชิงโครงสร้างของเครื่องมือวัดหรือการวิเคราะห์ทักษะรายบุคคลของผู้ตอบข้อสอบเพื่อนำผลไปใช้ในการใดการหนึ่ง โดยต้องเลือกใช้วิธีการตรวจสอบจำนวนมิติให้เหมาะสม

2.1 ประวัติความเป็นมา

การวิจัยที่เป็นพื้นฐานของ MIRT มี 2 สาขา คือ การวิเคราะห์องค์ประกอบ และทฤษฎีการตอบข้อสอบโดยอาศัยข้อตกลงของความเป็นมิติเดียว

การวิเคราะห์องค์ประกอบ (FA) และ MIRT มีความคล้ายคลึงกันหลายประการ เช่น วิธีการทั้งสองพยายามนิยามมาตรวัดตามสมมุติฐาน (hypothetical scale) ซึ่งสามารถสร้าง (reproduce) ข้อมูลที่มุ่งวิเคราะห์โดยเฉพาะ พยายามนิยามมาตรวัดที่มีจุดกำเนิดไม่แน่นอน พร้อมทั้งนิยามหน่วยของการวัด MIRT ต่างจาก FA ในประเด็นที่ให้ความสำคัญกับความแปรปรวนของคุณลักษณะของตัวแปรนำเข้า ซึ่งได้แก่ข้อสอบ แต่ FA มองว่าความแตกต่างของตัวแปรนำเข้า เช่น ความแตกต่างของค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และความเชื่อมั่น เป็นความแปรปรวนที่ต้องจัดออกจากการวิเคราะห์โดยผ่านทางสถิติมาตรฐาน FA จึงเริ่มด้วยการวิเคราะห์เมตริกซ์สหสัมพันธ์และเห็นว่าจะต้องกำจัดค่าความยาก อำนาจจำแนก และการเดาออกจากการวิเคราะห์ด้วย (Caroll. 1945)

นักวิจัยที่ศึกษาค้นคว้าและมีส่วนทำให้การวิเคราะห์องค์ประกอบมีส่วนสนับสนุน MIRT มีดังนี้

1) ฮอรัส (Horst. 1965) ให้ความเห็นว่าควรวิเคราะห์เมตริกซ์ที่ประกอบด้วยข้อมูลทั้งหมดซ้ำมากกว่าการใช้เมตริกซ์สหสัมพันธ์จากตัวแปรสมมุติฐาน

2) คริสตอฟเฟอร์สัน (Christofferson. 1975) และ มูเธ็น (Muthen. 1978) ใช้แบบจำลองนอร์มอลอจีฟ (normal ogive model) ในการประมาณค่าพารามิเตอร์เทรชโฮลด์ (threshold parameters) ของข้อสอบ ซึ่งเป็นสาระสำคัญเหมือนกับพารามิเตอร์ความยากของ MIRT สิ่งที่ไม่ได้กล่าวถึงในสูตรของทั้งสองท่าน คือ เงื่อนไขความน่าจะเป็นในการตอบถูกบนเวกเตอร์ของพารามิเตอร์บุคคล

3) แมคโดแนลด์ (McDonald. 1967) เสนอวิธีการนอนลิเนียร์ (nonlinear methodology) ซึ่งใกล้เคียงกับแนวคิดของ MIRT ในปัจจุบันมาก พร้อมทั้งชี้ให้เห็นปัญหาของการวิเคราะห์องค์ประกอบที่ตัวแปรมีคะแนนแบบ 0-1 และชี้ให้เห็นว่าการให้ความสนใจกับองค์ประกอบความยากในการวิเคราะห์ข้อมูลที่มีการให้คะแนนแบบ 0-1 สามารถทำได้ด้วยวิธีง่าย ๆ ถ้าการถดถอยของข้อมูลจากการสังเกตบนองค์ประกอบได้รับการพิจารณาว่าไม่ได้เป็นเชิงเส้น (nonlinear) แมคโดแนลด์ให้ความสำคัญกับแนวคิดของความเป็นอิสระ เป็นพื้นฐานในการวิเคราะห์ข้อสอบ

4) บ็อคและอิตกิน (Bock & Aitkin. 1981) ได้แสดงให้เห็นว่าการวิเคราะห์องค์ประกอบ (FA) และทฤษฎีการตอบข้อสอบ (IRT) ต่างมุ่งศึกษาในสิ่งเดียวกัน โดยผลลัพธ์สุดท้ายคือ ทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ (MIRT) พร้อมทั้งได้นิยามแบบจำลองนอร์มอลอจีฟ (normal ogive model) สำหรับสเปซคุณลักษณะภายในที่ใช้ความสามารถหลายมิติ (multidimensional trait space) ซึ่งได้รวมเอาคุณลักษณะข้อสอบที่เป็นตัวแทนแนวคิดทั้งของการวิเคราะห์องค์ประกอบและทฤษฎีการตอบข้อสอบ สิ่งที่ไม่เหมือน MIRT คือ พารามิเตอร์ข้อสอบยังคงแปลความหมายในแนวคิดของการวิเคราะห์องค์ประกอบและกำหนดให้เป็นองค์ประกอบมากกว่าการศึกษาปฏิสัมพันธ์ระหว่างบุคคลและข้อสอบ

สำหรับทฤษฎีการตอบข้อสอบ (IRT) นั้นแตกต่างจากการวิเคราะห์องค์ประกอบที่ IRT ไม่ต้องการหาองค์ประกอบน้อยที่สุดของข้อมูลในเมตริกซ์การตอบข้อสอบ แบบจำลองของทฤษฎีการตอบข้อสอบกำหนดความสัมพันธ์ระหว่างบุคคลและข้อสอบ ซึ่งลอร์ด (Lord. 1980) เสนอเป้าหมายของ IRT ว่า

ต้องการอธิบายลักษณะของข้อสอบโดยใช้พารามิเตอร์ข้อสอบ และอธิบายลักษณะของผู้ตอบข้อสอบด้วยพารามิเตอร์ผู้ตอบข้อสอบ ทำให้สามารถพยากรณ์ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้ตอบข้อสอบทุกคนในการทำข้อสอบข้อใดข้อหนึ่งได้ แม้ว่าผู้ตอบข้อสอบที่มีคุณลักษณะคล้ายคลึงกัน ไม่เคยทำข้อสอบที่คล้ายคลึงกันมาก่อน

การศึกษาเกี่ยวกับทฤษฎีการตอบข้อสอบ (IRT) ในช่วงแรกศึกษาโดยใช้ข้อตกลงเบื้องต้นที่ว่าพารามิเตอร์ที่อธิบายผู้ตอบข้อสอบแปรเปลี่ยนบนมิติเดียว (Lord & Novick. 1968; Rasch. 1960) แต่ปรากฏว่าข้อตกลงเบื้องต้นดังกล่าวมักถูกละเมิดอยู่เสมอ (Camili, Wang & Fesq. 1995) จึงมีความพยายามในการศึกษาทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ (MIRT) โดยบุคคลต่าง ๆ ดังนี้

1) ราสช์ (Rasch. 1960) ในช่วงแรกศึกษาเฉพาะแบบจำลองใช้ข้อตกลงของความเป็นมิติเดียว ในปี ค.ศ. 1962 ได้ขยายแบบจำลองที่ศึกษาในช่วงแรกโดยรวมเอาความเป็นไปได้ที่คุณลักษณะภายในของผู้ตอบข้อสอบ (trait) อยู่ในรูปของเวกเตอร์มากกว่าในรูปของสเกลาร์ ซึ่งมีแบบจำลองทั่วไปดังนี้

$$P(x | \theta_i, \sigma_i) = \frac{1}{\gamma(\theta_i, \sigma_i)} \exp [\phi(x)' \theta_i + \Psi(x)' \sigma_i + \theta_j' \chi(x) \sigma_i + \rho(x)]$$

เมื่อ ϕ , Ψ , χ และ ρ เป็นฟังก์ชันของคะแนนในการทำข้อสอบ (x) ดังนั้นค่าสถิติจากข้อมูลที่สังเกตได้จึงเหมาะกับพารามิเตอร์บุคคลและพารามิเตอร์ข้อสอบ

σ เป็นเวกเตอร์ของพารามิเตอร์ข้อสอบ

$\gamma()$ เป็น normalizing function ซึ่งรับประกันว่าค่าของฟังก์ชัน IRT มีพิสัยระหว่าง 0-1

แบบจำลองที่ราสช์พัฒนาขึ้นเพื่อใช้กับทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติมีการนำไปใช้น้อยมาก เพราะการนิยามฟังก์ชันการให้คะแนนยังไม่น่าเชื่อถือมากนักและกระบวนการในการศึกษาก็มีความซับซ้อนมาก

2) ลอร์ดและโนวิก (Lord & Novick. 1968) ในตำราทฤษฎีสถิติของคะแนนจากแบบทดสอบวัดความสามารถทางสมอง (Statistical theories of mental test scores) ได้กล่าวถึงทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ (MIRT) ไว้ในบทที่ 16 แม้ว่าจะไม่ได้นำเสนอแบบจำลองที่สมบูรณ์ แต่ได้กล่าวถึงนิยามของสเปซคุณลักษณะภายในที่สมบูรณ์ (complete latent space) และข้อตกลงเบื้องต้นของความเป็นอิสระ (local independence) โดยกล่าวถึงความหมายของความเป็นอิสระว่าในกลุ่มผู้ตอบข้อสอบกลุ่มใด ๆ ซึ่งคุณลักษณะภายในที่มีค่า $\theta_1, \theta_2, \dots, \theta_k$ การแจกแจงแบบมีเงื่อนไขของคะแนนในการตอบข้อสอบต่างมีความเป็นอิสระต่อกัน เมื่อมี k มิติ เวกเตอร์ θ หมายถึง สเปซคุณลักษณะภายในที่สมบูรณ์

3) สามิจิมา (Samejima. 1974) ได้เสนอแบบจำลองการตอบข้อสอบที่ใช้ความสามารถหลายมิติ เนื่องจากเดิมสูตรของทฤษฎีการตอบข้อสอบกำหนดข้อตกลงเบื้องต้นของการให้คะแนนรายข้อแบบสองค่า หรือการให้คะแนนรายข้อแบบหลายค่า ว่าเป็นผลของการแบ่งส่วนตัวแปรการตอบ

$$P_{zi}(\theta) = \frac{1}{2\pi} \int_{-\infty}^{a_i(\theta - b_i)} \exp\left(-\frac{u^2}{2}\right) du$$

ข้อสอบแบบต่อเนื่อง (a continuous response variable) สามีจิมาพัฒนาแบบจำลองสำหรับการตอบข้อสอบแบบต่อเนื่อง (Z_i) มีรูปแบบดังนี้

แม้ว่าแบบจำลองของสามีจิมาจะเป็นแบบจำลองการตอบข้อสอบที่ใช้ความสามารถหลายมิติที่มีความชัดเจนแบบหนึ่ง แต่ยังไม่ค่อยได้รับความนิยมมากนัก

นอกจากบุคคลที่กล่าวมาข้างต้นแล้ว ยังมีผู้ที่สนใจศึกษาเพื่อพัฒนาแบบจำลองการตอบข้อสอบที่ใช้ความสามารถหลายมิติ (MIRT) อย่างต่อเนื่องหลายท่าน ได้แก่ เร็คเคส (Reckase, 1972) มุลไคค (Mulaik, 1972) ซิมป์สัน (Simpson, 1978) ไวท์ลีย์ (Whitely, 1980) จนกระทั่งแม็คคินเลย์และเร็คเคส (McKinley & Reckase, 1982) ได้เสนอแบบจำลองที่มีการนำไปใช้มากที่สุดในปัจจุบันในรูปแบบของแบบจำลองลอจิสติก (linear logistic model) และยังมีบุคคลอื่น ๆ สนใจศึกษาค้นคว้าอย่างต่อเนื่องมาจนถึงปัจจุบัน

2.2 ความคิดพื้นฐานของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ

แบบจำลองทั่วไปของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติอธิบายปฏิกริยาร่วมระหว่างบุคคลและการให้คะแนนการตอบข้อสอบรายข้อ โดยอธิบายคุณลักษณะของผู้ตอบข้อสอบด้วยเวกเตอร์ของโครงสร้างตามสมมุติฐาน และอธิบายคุณลักษณะของข้อสอบด้วยเซตของพารามิเตอร์ข้อสอบ และรูปแบบฟังก์ชันที่เกี่ยวกับตำแหน่งในสเปซที่นิยามโดยเวกเตอร์ของพารามิเตอร์ของบุคคลต่อความน่าจะเป็นในการตอบข้อสอบรายข้อได้ถูกต้อง

ทั้งการตอบข้อสอบรายข้อและคะแนนจากแบบทดสอบทั้งฉบับเพื่อตรวจสอบความแตกต่างระหว่างแบบจำลองที่สามารถตรวจสอบได้ เมื่อกำหนดค่า p (p - value) ให้คงที่ แม้ว่าตรวจสอบความแตกต่างของข้อสอบรายข้อได้ แต่ความแตกต่างนี้ลดลงเมื่อสหสัมพันธ์ระหว่างความสามารถของผู้ตอบข้อสอบเพิ่มขึ้น นอกจากนั้นความแตกต่างที่พบก็มีขนาดเล็กมากไม่มีผลในทางปฏิบัติแต่อย่างใด และไม่พบความแตกต่างของคะแนนจากแบบทดสอบทั้งฉบับ

มิลเลอร์ (Miller. 1991) ศึกษาความคลาดเคลื่อนมาตรฐานของพารามิเตอร์จากแบบจำลองทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติแบบที่มีการชดเชยที่ประมาณค่าโดยโปรแกรมคอมพิวเตอร์สำเร็จรูป NOHARM โดยเปรียบเทียบกับโปรแกรมสำเร็จรูปอื่น (MIRTE, MULTLDIM, TESTFACT) ผลปรากฏว่า NOHARM มีประสิทธิภาพดีและเหมาะกับการนำไปประยุกต์ใช้ในการศึกษาตามกรอบความคิดของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ

เร็คเคสและแม็คคินเลย์ (Reckase & McKinley. 1991 : 361-371) ศึกษาเกี่ยวกับอำนาจจำแนกของข้อสอบที่วัดความสามารถหลายอย่าง โดยใช้ข้อมูลการตอบข้อสอบ ACT Mathematics Usage Test จำนวน 40 ข้อ ที่ดำเนินการทดสอบในปี ค.ศ. 1983 ใช้กลุ่มตัวอย่าง 1,000 คน ได้มาจากการสุ่มอย่างเป็นระบบ วิเคราะห์ข้อมูลการตอบข้อสอบด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป MAXLOG ผลการวิเคราะห์พบว่าข้อสอบข้อที่ 27 มีอำนาจจำแนกสูงสุด แสดงว่าข้อสอบข้อดังกล่าวสามารถจำแนกผู้ตอบข้อสอบที่มีความสามารถแตกต่างกันได้ดีที่สุด ถ้าพิจารณาโดยภาพรวมอาจเปรียบเทียบอำนาจจำแนกของข้อสอบ 2 ข้อได้โดยตรง ถ้าวัดในทิศทางเดียวกัน เช่น ข้อ 29 และ 36 ซึ่งมี $\alpha_{29,1} = \alpha_{36,1} = 52$ และ $\alpha_{29,2} = \alpha_{36,2} = 38$ ส่วนอำนาจจำแนกของข้อ 3 และข้อ 30 เปรียบเทียบกันไม่ได้ เพราะมีทิศทางต่างกัน แต่ถ้าพิจารณาโดยภาพรวมแล้ว ข้อ 3 มีอำนาจจำแนกมากกว่า ($\alpha_{3,1} = 13$, $\alpha_{3,2} = 77$, $MDISC_3 = 1.61$, $\alpha_{30,1} = 61$, $\alpha_{30,2} = 29$, $MDISC_{30} = 1.39$)

แอ็คเคอร์แมน (Ackerman. 1996 : 311-329) เสนอผลการวิจัยการใช้กราฟแสดงผลการวิเคราะห์ตามแนวทางทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป NOHARM ประมาณค่าข้อมูลการตอบข้อสอบของ ACT Mathematics Usage Test Form B และ Form C ซึ่งดำเนินการทดสอบในปี ค.ศ. 1989 และเสนอกราฟแสดงผลการวิเคราะห์ โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป DISSPLA (Computer Associated International, Inc., 1989) กราฟที่นำเสนอแบ่งเป็น 5 กลุ่ม คือ แสดงค่าการประมาณค่าความสามารถ (trait estimation representation) พารามิเตอร์ข้อสอบ (item representation) สารสนเทศ (information representation) การวิเคราะห์ตามเงื่อนไข (conditional analysis) และการแจกแจงความน่าจะเป็นที่ผู้ตอบข้อสอบแต่ละคนจะตอบข้อสอบในแบบทดสอบได้ถูกต้อง (ENC distribution) ผลการวิเคราะห์เมื่อนำมาแสดงด้วยกราฟทำให้เห็นภาพของข้อสอบและแบบทดสอบได้อย่างชัดเจน ถ้าผู้สร้างแบบทดสอบนำผลที่ได้มาศึกษาทำให้สามารถตรวจสอบความสอดคล้องของสาระสำคัญที่กำหนดไว้ในแผนผังการออกข้อสอบ แก้ไข และปรับปรุงคุณภาพของแบบทดสอบได้อย่างมีประสิทธิภาพ

เร็คเคส (Reckase. 1997 : 270-286) เสนอรายงานการวิจัยเกี่ยวกับแบบจำลองหลายมิติลิเนียร์โลจิสติกสำหรับการตรวจให้คะแนนรายข้อแบบทวิพันธ์ มีจุดมุ่งหมายเพื่อแสดงผลการวิเคราะห์โครงสร้างของแบบทดสอบ โดยใช้กรอบความคิดของ MIRT ซึ่งแสดงการปนกันของความยาก (confounding the difficulty) ของข้อคำถามในแบบทดสอบและเนื้อหาโดยแสดงผลของการใช้จำนวนมิติจำนวนน้อยเกินไป

ในการวิเคราะห์ข้อมูลในรูปของเมตริกซ์ ใช้ข้อมูลการตอบข้อสอบของกลุ่มตัวอย่างนักเรียนเกรด 10 ในสหรัฐอเมริกาจำนวน 1,635 คน ที่ได้รับการสุ่มให้เป็นตัวแทนของประชากร เครื่องมือที่ใช้ในการวิจัยคือแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ประกอบด้วยข้อสอบจำนวน 40 ข้อ ครอบคลุมเนื้อหาเกี่ยวกับพีชคณิตและเรขาคณิต การวิเคราะห์ข้อมูลใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป BILOG (Mislevy & Bock, 1989) เพื่อประมาณค่า c-parameter ค่าที่ประมาณได้นำไปวิเคราะห์ด้วยโปรแกรม NOHARM ซึ่งไม่มีการประมาณค่า c-parameter แล้วทำการวิเคราะห์ความสามารถ 2 มิติ (2-dimensional) และความสามารถ 6 มิติ (6-dimensional) การวิเคราะห์ความสามารถ 2 มิติสามารถเขียนกราฟแสดงผลการวิเคราะห์ได้และสามารถศึกษาอิทธิพลของการประมาณค่าจากจำนวนมิติที่น้อยเกินไป การวิเคราะห์โดยพิจารณาความสามารถถึง 6 มิติ เป็นความสามารถสูงสุดที่โปรแกรม NOHARM วิเคราะห์ได้ ผลการวิเคราะห์แสดงได้ดังภาพ

ภาพประกอบ 11 แสดงเวกเตอร์ของข้อสอบและกลุ่มข้อสอบจากการวิเคราะห์ข้อสอบ 40 ข้อ
(อ้างอิงจาก Reckase, 1997 : 281)

จากภาพจะเห็นว่าคำถามง่าย ๆ ซึ่งอยู่มุมล่างด้านซ้าย วัดเพียงมิติเดียวในที่นี้คือ θ_1 และข้อที่ยากกว่าอยู่มุมบนด้านขวาวัด θ_2 เพื่อทำให้เกิดความเข้าใจง่ายขึ้น มุมระหว่างข้อสอบทำการวิเคราะห์การจัดกลุ่ม (cluster analyze) เพื่อพิจารณาชุดของคำถามที่สามารถจำแนกอยู่ในทิศทางเดียวกันในสเปซความสามารถ คำถามที่ยากมากที่สุดอยู่ใน c_5 วัดเนื้อหาเกี่ยวกับเรขาคณิต คำถามใน c_1 วัดเนื้อหาด้านพีชคณิต ผู้ตอบข้อสอบที่ทำคะแนนได้ดีจะตอบข้อสอบใน c_5 ได้ถูกต้อง ส่วนผู้ตอบข้อสอบที่ได้คะแนนต่ำ ๆ แสดงถึงความแตกต่างของทักษะด้านพีชคณิต สำหรับการวิเคราะห์โดยใช้ความสามารถ 6 มิติ พบว่าช่วยให้ทำความเข้าใจโครงสร้างของเนื้อหาในแบบทดสอบและรู้หน้าที่ของข้อสอบแต่ละข้อได้ชัดเจนขึ้น ผลที่ได้จากการศึกษาพบว่า การวิเคราะห์โดยใช้เพียง 2 มิติ เป็นภาพสะท้อน (projection) ของการวิเคราะห์โดยใช้มิติจำนวนมากขึ้น และเมื่อวิเคราะห์ด้วยจำนวนมิติที่น้อยเกินไปทำให้สารสนเทศบางอย่างหายไป ไม่สามารถอธิบายรายละเอียดเกี่ยวกับข้อสอบและแบบทดสอบได้อย่างชัดเจน นอกจากนี้ยังพบว่ากลุ่มตัวอย่างจำนวน 1,635 คน ทำให้สามารถวิเคราะห์ถึง 6 มิติได้อย่างถูกต้อง แบบจำลองนี้ช่วยให้เข้าใจปัญหาในการวัดทางการศึกษาและทางจิตวิทยา รวมทั้งการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบและการประมาณค่าพารามิเตอร์ความสามารถของผู้ตอบข้อสอบได้

2.9.4 พัฒนารูปแบบการเทียบคะแนนของแบบทดสอบ (Test Equating) แนวคิดของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติทำให้การเทียบคะแนนของแบบทดสอบกระทำได้อย่างมีประสิทธิภาพยิ่งขึ้น (Davey & Oshima. 1994 ; Davey, Oshima & Lee. 1996)

2.9.5 อธิบายการทำหน้าที่เบี่ยงเบนของข้อสอบ (differential item functioning : DIF) ได้อย่างชัดเจน เนื่องจากทฤษฎีในการตอบข้อสอบที่ใช้ความสามารถหลายมิติมีแนวคิดเกี่ยวกับความไวของข้อสอบและตำแหน่ง (location) ของข้อสอบในสเปซคุณลักษณะแบบหลายมิติ ซึ่งทำให้ขยายแนวคิดเกี่ยวกับการตรวจสอบการทำหน้าที่เบี่ยงเบนไปยังหมวดข้อสอบ และแบบทดสอบทั้งฉบับได้อย่างมีประสิทธิภาพ (Roussos & Stout. 1996)

2.9.6 การประยุกต์ใช้ด้านอื่น ๆ เช่น ใช้เป็นกรอบความคิดในการศึกษาวิจัยเพื่อขยายองค์ความรู้เกี่ยวกับทฤษฎีการทดสอบให้มีความก้าวหน้ายิ่งขึ้น

2.10 งานวิจัยเกี่ยวกับทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ

ในที่นี้กล่าวเฉพาะงานวิจัยที่เกี่ยวข้องกับข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติตามแบบจำลองหลายมิติโลจิสติกสำหรับการตรวจให้คะแนนรายข้อแบบสองค่าดังต่อไปนี้

เร็คเคส (Reckase. 1985 : 401-412) ศึกษาเกี่ยวกับความยากของข้อสอบที่วัดความสามารถหลายอย่าง โดยใช้กลุ่มตัวอย่างเป็นนักศึกษา 1,000 คน ทำข้อสอบ ACT Assessment Mathematics Usage Test จำนวน 40 ข้อ ซึ่งทดสอบในปี ค.ศ. 1983 ประมาณค่าพารามิเตอร์ตามแบบจำลอง M2PL (multidimensional 2-parameter logistic model) โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป MAXLOG (McKinley & Rackase. 1983b) ในการศึกษาครั้งนี้ไม่มีการตรวจสอบมิติของแบบทดสอบ แต่ใช้ M2PL เพราะเป็นตัวอย่างที่อธิบายให้เข้าใจได้ง่ายและใช้ข้อมูลชุดเดียวกัน คำนวณโดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป LOGIST เพื่อเปรียบเทียบผลการวิเคราะห์ข้อสอบตามทฤษฎีการทดสอบแบบมาตรฐานเดิม และทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์ ผลการวิเคราะห์พบว่าข้อสอบที่ง่ายมีแนวโน้มที่จะวัด θ_1 (ความสามารถในการอ่านโจทย์ปัญหา) ข้อสอบที่ยากมีแนวโน้มที่จะวัด θ_2 (ทักษะในการคำนวณ) ในการเปรียบเทียบความยากของข้อสอบต้องเปรียบเทียบข้อสอบที่วัดองค์ประกอบของความสามารถเหมือนกัน นั่นคือมีทิศทางเดียวกัน ข้อสอบที่วัดในทิศทางเดียวกันนำมารวมไว้ในฉบับเดียวกันได้ และสามารถคัดเลือกข้อสอบนำไปรวมเป็นแบบทดสอบที่วัดความสามารถเพียงมิติเดียว จัดเป็นแบบทดสอบฉบับย่อยดำเนินการทดสอบเหมือนเป็นแบบทดสอบย่อยมิติเดียว เพราะทุกข้อวัดองค์ประกอบ θ_1 หรือ θ_2 เหมือนกัน ในกรณีที่ต้องการข้อสอบที่วัดทุกมิติต้องเลือกข้อสอบที่มีความชันของ IRS ไม่เป็น 0 สำหรับทิศทางใด ๆ ที่ขนานกับแกน θ

คาร์ลสัน (Carlson. 1987) ได้ศึกษาเกี่ยวกับโปรแกรมคอมพิวเตอร์สำเร็จรูป MIRTE ที่ใช้ในการประมาณค่าพารามิเตอร์ข้อสอบ และความสามารถของผู้ตอบข้อสอบที่ใช้กรอบความคิดของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ ชนิด 2 พารามิเตอร์ ศึกษาโดยใช้ข้อมูลจำลองที่มีเงื่อนไขต่าง ๆ เพื่อตรวจสอบและรับประกันว่าโปรแกรมทำงานได้อย่างมีประสิทธิภาพ

สเปรย์ และคณะ (Spray, et. al. 1990) ได้ศึกษาเปรียบเทียบแบบจำลองทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติแบบที่มีการชดเชยและแบบที่ไม่มีการชดเชย โดยใช้การจำลองข้อมูลเปรียบเทียบ

2.5.4 พารามิเตอร์ของผู้ตอบข้อสอบ (person parameter : θ_j)

พารามิเตอร์ของผู้ตอบข้อสอบในแบบจำลองการตอบข้อสอบที่ใช้ความสามารถหลายมิติ คือ สมาชิกของเวกเตอร์ θ_j มิติความสามารถ (θ -dimension) เป็นโครงสร้างเชิงสถิติซึ่งถูกสร้างขึ้นมาจัดให้เหมาะสม (fit) กับเมตริกซ์ $N \times n$

2.6 สารสนเทศที่เกี่ยวข้องกับความสามารถหลายมิติ (multidimensional information : MINF)

MINF เกี่ยวข้องกับ MDISC ในแง่ที่ว่า ถ้าข้อสอบมี MDISC ที่มีค่าสูง จะให้สารสนเทศที่มีค่ามากที่สุดที่หนึ่งในสเปซความสามารถ MINF ต่างจาก MDISC เพราะเกี่ยวข้องกับความสามารถของข้อสอบในการจำแนกที่แต่ละจุดในสเปซความสามารถมากกว่าที่จุดซึ่งมีความชันสูงสุดของ IRS

นิยามของ MINF สรุปล้างอิงจากความคิดรวบยอดของสารสนเทศข้อสอบใน IRT ซึ่งสารสนเทศที่ θ ระดับต่าง ๆ คืออัตราส่วนของ IRF ที่ θ ใด ๆ ยกกำลังสองต่อความแปรปรวนของความคลาดเคลื่อนของคะแนนรายข้อที่ θ ใด ๆ ในกรณีของ MIRT ต้องตรวจสอบความชันด้วยวิธีที่แตกต่างจาก IRT สำหรับ IRS มีความชันมากกว่า 1 ที่ความชันที่จุดใด ๆ ในสเปซขึ้นอยู่กับทิศทางที่มายังจุดนั้นในสเปซ ความชันแต่ละที่จะแตกต่างกันออกไป ความชันจากทิศทางในแนวตั้งของผิวระนาบมีค่ามากกว่าความชันตามแนวนอน ที่จุดเดียวกันทิศทางหนึ่งอาจให้ความชันเป็น 0 แต่อีกทิศทางหนึ่งอาจให้ความชันมีค่ามากพอสมควร ดังนั้นต้องตรวจสอบทิศทางในสเปซด้วย เมื่อพิจารณาสารสนเทศที่ได้จากข้อสอบซึ่งเป็นการระบุว่ามี

สารสนเทศเกี่ยวกับองค์ประกอบเฉพาะ (particular composite of abilities) ที่จุดในสเปซความสามารถมากน้อยเพียงใด ในการตรวจสอบความชันในทิศทางเฉพาะใช้กระบวนการทางคณิตศาสตร์ที่เรียกว่าการหาอนุพันธ์แบบมีทิศทาง (directional derivative)

สารสนเทศของข้อสอบที่ใช้ความสามารถหลายมิติ (multidimensional item information) กำหนดดังนี้

$$I_{i\alpha}(\theta) = \frac{[\nabla_{\alpha} P_i(\theta)]^2}{P_i(\theta)[1 - P_i(\theta)]}$$

∇_{α} คือ operator defining the directional derivative ในทิศทาง α

$$\begin{aligned} \nabla_{\alpha} P_i(\theta) &= a_{i1} P_i(\theta) Q_i(\theta) \cos \alpha_1 + a_{i2} P_i(\theta) Q_i(\theta) \cos \alpha_2 + \dots + a_{im} P_i(\theta) Q_i(\theta) \cos \alpha_m \\ &= P_i(\theta) Q_i(\theta) \sum_{k=1}^m a_{ik} \cos \alpha_k \\ \therefore I_{i\alpha}(\theta) &= \frac{[P_i(\theta) Q_i(\theta) \sum_{k=1}^m a_{ik} \cos \alpha_k]^2}{P_i(\theta) Q_i(\theta)} \end{aligned}$$

$$= P_i(\theta) Q_i(\theta) \left(\sum_{k=1}^m a_{ik} \cos \alpha_k \right)^2$$

เมื่อ $I_{i\alpha}(\theta)$ เป็นสารสนเทศที่ได้จากข้อสอบข้อที่ i ในทิศทาง α ในสเปซความสามารถ สารสนเทศที่ขึ้นกับ θ ในทิศทาง α สามารถตรวจสอบได้เช่นเดียวกับใน IRT

สารสนเทศของแบบทดสอบ (test information surface) หาได้จากผลรวมของสารสนเทศข้อสอบ เมื่อคำนวณโดยกำหนดข้อตกลงเบื้องต้นว่าข้อสอบแต่ละข้อมีทิศทางเดียวกัน

2.7 การประมาณค่าพารามิเตอร์

การประมาณค่าพารามิเตอร์ ใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป โปรแกรมที่ใช้มีดังต่อไปนี้

ชื่อโปรแกรม	วิธีการประมาณค่า
1. MAXLOG (McKinley & Reckase, 1983)	maximum likelihood
2. MIRTE (Carlson, 1987)	maximum likelihood
3. MULTIDIM (McKinley, 1987)	marginal maximum likelihood
4. TESTFACT (Wilson, Wood & Gibbons, 1987)	bayesians estimates of factor score
5. NOHARM (Fraser & McDonald, 1988)	least square

2.8 กราฟแสดงผลการวิเคราะห์

การแสดงผลการวิเคราะห์ข้อสอบรายข้อและการวิเคราะห์แบบทดสอบทั้งฉบับด้วยกราฟ ปัจจุบันสามารถนำเสนอได้เฉพาะผลการตอบข้อสอบที่ใช้ความสามารถ 2 มิติ (two-dimensional IRT) เพราะโปรแกรมคอมพิวเตอร์สำเร็จรูปที่มีอยู่สามารถแสดงกราฟได้เฉพาะกรณีที่เป็น 2 มิติ คือ graphic package DISSPLA (Computer Associates International, Inc, 1989) ในการเขียนภาพกำหนดข้อตกลงเบื้องต้นว่า มิติที่ต้องการวัด (pure dimension) ซึ่งประกอบด้วย θ_1 และ θ_2 ไม่มีสหสัมพันธ์กัน ทำให้สามารถนำเสนอโดยใช้ Cartesian Coordinate System ได้ ดังตัวอย่างในภาพประกอบต่อไปนี้

ภาพประกอบ 5 แสดงผิวระนาบการตอบข้อสอบ (IRS) ของข้อสอบ 2 ข้อ
(อ้างอิงจาก Reckase & McKinley, 1991 : 363)

ภาพประกอบ 6 แสดงเวกเตอร์ของข้อสอบรายข้อ
(อ้างอิงจาก Ackerman. 1996 : 317)

ภาพประกอบ 7 แสดงสารสนเทศหลายมิติ (MINF) ของข้อสอบข้อที่ 10
(อ้างอิงจาก Reckase & McKinley. 1991 : 371.)

ภาพประกอบ 8 แสดงสารสนเทศหลายมิติ (MINF) ของข้อสอบข้อที่ 10
(อ้างอิงจาก Reckase & McKinley. 1991 : 370.)

ภาพประกอบ 9 แสดงสารสนเทศหลายมิติ (MINF) ของแบบทดสอบ
(อ้างอิงจาก Reckase & McKinley. 1991 : 371.)

ภาพประกอบ 10 แสดงสารสนเทศหลายมิติ (MINF) ของแบบทดสอบ
(อ้างอิงจาก Reckase & McKinley. 1991 : 372.)

2.9 การประยุกต์ใช้ทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ

แนวคิดของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ (MIRT) สามารถนำไปประยุกต์ใช้ได้อย่างกว้างขวางดังนี้

2.9.1 อธิบายโครงสร้างของทักษะที่ข้อสอบต้องการวัดได้ชัดเจนยิ่งขึ้น ช่วยให้ตรวจสอบความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบได้อย่างมีประสิทธิภาพ

2.9.2 จัดทำแบบทดสอบ (Test Assembly) ได้อย่างสอดคล้องกับจุดมุ่งหมายของการทดสอบแต่ละครั้งได้เป็นอย่างดี และสร้างแบบทดสอบคู่ขนานได้อย่างมีประสิทธิภาพ (van der Linden. 1996)

2.9.3 พัฒนาการทดสอบที่มีการคัดเลือกข้อสอบให้เหมาะสมกับความสามารถของผู้ตอบข้อสอบ (Computerized Adaptive Testing) ได้อย่างมีประสิทธิภาพ (Luecht. 1996)

2.3 แบบจำลองของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ

แบบจำลองของทฤษฎีการตอบข้อสอบที่ใช้ข้อตกลงเกี่ยวกับการใช้ความสามารถหลายมิติแบ่งเป็น 2 กลุ่มใหญ่ ๆ คือแบบจำลองการตอบข้อสอบที่ใช้กับการตรวจให้คะแนนรายข้อแบบสองค่า และแบบจำลองการตอบข้อสอบที่ใช้กับการตรวจให้คะแนนรายข้อแบบอื่น ๆ ดังที่ได้กล่าวมาแล้ว แบบจำลองของ MIRT แต่ละแบบจำลองมีฟังก์ชันและเงื่อนไขที่แตกต่างกันไป ในที่นี้จะกล่าวเฉพาะแบบจำลองหลายมิติลิเนียร์โลจิสติกสำหรับการตรวจให้คะแนนรายข้อแบบสองค่า (Linear Logistic Multidimensional Model for Dichotomous Item Response Data) ของแม็คคินเลย์และเร็คเคส (McKinley & Reckase. 1982) เท่านั้น เพราะเป็นแบบจำลองที่กำหนดสิ่งที่รู้จากข้อมูลการตอบข้อสอบรายบุคคลอย่างสมเหตุสมผล สอดคล้องกับกลุ่มตัวอย่างผู้ตอบข้อสอบ สอดคล้องกับแนวคิดของทฤษฎีการตอบข้อสอบมิติเดียว และสามารถประมาณค่าพารามิเตอร์ได้จากจำนวนผู้ตอบข้อสอบและจำนวนข้อสอบที่สามารถหาได้ (Reckase. 1997 : 273)

2.3.1 รูปแบบของแบบจำลอง

ข้อมูลที่ใช้ในแบบจำลอง คือ เมตริกซ์ของคะแนนรายข้อ ทั้ง 0 และ 1 ซึ่ง 0 คือ คะแนนของคำตอบผิด และ 1 คือ คะแนนของคำตอบถูก ข้อมูลในเมตริกซ์มีดังนี้

แถว (N) แทนผู้ตอบข้อสอบ สดมภ์ (n) แทนข้อสอบ

สมาชิกแต่ละตัว หมายถึง คะแนนของบุคคล j ($j = 1, N$) ในการตอบข้อสอบข้อที่ i ($i = 1, n$)

2.3.2 ประเภทของแบบจำลอง

แบบจำลองของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติที่ใช้อธิบายข้อมูลการตอบข้อสอบที่ให้คะแนนรายข้อแบบสองค่ามีสองประเภทคือ

2.3.2.1 แบบจำลองที่มีการชดเชย (compensatory model)

แนวคิดของแบบจำลองนี้คือ θ ที่มีค่าต่ำ ๆ ของมิติใด ๆ สามารถชดเชยด้วย θ ที่มีค่าสูง ๆ ในมิติอื่น ความน่าจะเป็นในการตอบข้อสอบถูกต้องมีค่าเท่ากับ 1 อาจเกิดขึ้นได้ แม้ว่าจะมี θ ต่ำ ๆ ในบางมิติและมี θ สูง ๆ ในมิติอื่น เป็นแบบจำลองที่ขยายแบบจำลองทฤษฎีการตอบข้อสอบแบบโลจิสติก 3 พารามิเตอร์ (3PL model) ของลอร์ด (Lord, 1980) ในกรณีที่ยอธิบายคุณลักษณะของผู้ตอบข้อสอบด้วยเวกเตอร์ของพารามิเตอร์ความสามารถแทนค่าสเกลาร์ พัฒนาโดยแม็คคินเลย์และเร็คเคส (McKinley & Reckase, 1982) ความน่าจะเป็นในการตอบข้อสอบข้อที่ i ได้ถูกต้อง มีดังนี้

$$P(X_i = 1 | \theta) = c_i + (1 - c_i) \frac{\exp[1.7 a'_i \theta_j + d_i]}{1.0 + \exp[1.7 a'_i \theta_j + d_i]}$$

เมื่อ X_i คือ คะแนน (0, 1) ในการตอบคำถามข้อที่ i ($i = 1, 2, \dots, n$)

a_i คือ เวกเตอร์ของพารามิเตอร์ที่เกี่ยวข้องกับอำนาจจำแนกของข้อสอบ (อัตราการเปลี่ยนแปลงความน่าจะเป็นของการตอบถูกต้องต่อการเปลี่ยนแปลงระดับคุณลักษณะภายใน (θ) ของผู้ตอบข้อสอบ)

d_i คือ พารามิเตอร์ที่เกี่ยวข้องกับความยากของข้อสอบข้อที่ i

c_i คือ ความน่าจะเป็นของการตอบข้อสอบข้อที่ i ได้ถูกต้อง เมื่อผู้ตอบข้อสอบมีความสามารถในสิ่งที่ข้อสอบต้องการวัดในระดับต่ำมาก (เข้าสู่ $-\infty$) ปกติเรียกว่าพารามิเตอร์การเดาของข้อสอบข้อที่ i

θ_j คือ เวกเตอร์ความสามารถของผู้ตอบข้อสอบคนที่ j

2.3.2.2. แบบจำลองที่ไม่มีการชดเชย (noncompensatory model)

แนวคิดของแบบจำลองนี้คือการที่มี θ ต่ำ ๆ ในมิติใด ๆ ไม่สามารถชดเชยด้วย θ สูง ๆ ในมิติอื่น ๆ แบบจำลองนี้ได้รับการพัฒนาโดยซิมป์สัน (Simpson. 1978) และไวท์ลีย์ (Whitely. 1980) ความน่าจะเป็นในการตอบข้อสอบถูกต้องถูกกำหนดขอบเขตโดยความน่าจะเป็นที่มีขนาดเล็กกว่า (smaller probability) ของแต่ละมิติ

การศึกษาและวิจัยเกี่ยวกับแบบจำลองหลายมิติลิเนียร์โลจิสติก สำหรับข้อมูลการตอบข้อสอบที่ให้คะแนนรายข้อแบบสองค่าส่วนใหญ่ศึกษาแบบจำลองที่มีการชดเชยมากกว่า จึงมีการพัฒนาโปรแกรมคอมพิวเตอร์สำเร็จรูปในการประมาณค่าพารามิเตอร์ของแบบจำลองดังกล่าวหลายโปรแกรม เช่น NOHARM (Fraser & McDonald. 1988) TESTFACT (Wilson, Wood & Gibbon. 1987) เป็นต้น

2.4 ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ

ในการสร้างเมตริกซ์ข้อมูลการตอบข้อสอบรายข้อตามทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติมีข้อตกลงเบื้องต้นที่สอดคล้องกับแบบจำลองการตอบข้อสอบที่เกี่ยวข้องกับคุณลักษณะของผู้ตอบข้อสอบและคุณลักษณะของข้อสอบหลาย ๆ แบบจำลองดังต่อไปนี้ (Reckase. 1997 : 272)

2.4.1 ข้อตกลงของการเพิ่มขึ้นทางเดียว

เป็นข้อตกลงเบื้องต้นที่กำหนดว่าเมื่อโครงสร้างตามสมมุติฐาน (hypothetical constructs) ที่ต้องการวัดมีค่าเพิ่มขึ้น ความน่าจะเป็นในการตอบข้อสอบถูกต้องมีค่าเพิ่มขึ้นด้วย

2.4.2 ข้อตกลงเกี่ยวกับการนิยามฟังก์ชันความน่าจะเป็นในการตอบข้อสอบ

เป็นข้อตกลงเบื้องต้นที่กำหนดว่าความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้ตอบข้อสอบโดยใช้โครงสร้างตามสมมุติฐาน สอดคล้องกับฟังก์ชันที่ได้มีการนิยามขึ้น

2.4.3 ข้อตกลงเกี่ยวกับความเป็นอิสระในการตอบข้อสอบ

เป็นข้อตกลงเบื้องต้นที่กำหนดว่าความน่าจะเป็นในการตอบข้อสอบหลายข้อได้ถูกต้องพิจารณาได้จากผลคูณของความน่าจะเป็นในการตอบข้อสอบรายข้อ เมื่อความน่าจะเป็นคำนวณอย่างมีเงื่อนไขบนจุดในสเปซที่นิยามโดยโครงสร้างตามสมมุติฐาน

แฮมเบิลตัน (Hambleton. 1991 : 11-12) กล่าวว่าความเป็นอิสระในการตอบข้อสอบเกิดขึ้นได้เมื่อได้ระบุเพศคุณลักษณะภายในที่สมบูรณ์ และความสามารถทั้งหลายเหล่านั้นต่างส่งผลต่อการตอบข้อสอบ ในทำนองกลับกันความเป็นอิสระในการตอบข้อสอบจะไม่ปรากฏถ้าไม่ได้ระบุเพศคุณลักษณะภายในที่สมบูรณ์ เช่น แบบทดสอบวิชาคณิตศาสตร์ที่ต้องการความสามารถในการอ่านระดับสูง ผู้ตอบข้อสอบที่มีความสามารถในการอ่านระดับต่ำจะตอบคำถามไม่ถูกต้องโดยที่ยังไม่ได้ใช้ความสามารถด้านคณิตศาสตร์ในกรณีนี้จึงต้องระบุเพศคุณลักษณะภายในที่สมบูรณ์เป็นทั้งทักษะในการอ่านและความสามารถด้านคณิตศาสตร์ด้วย

2.5 พารามิเตอร์ในแบบจำลองการตอบข้อสอบที่ใช้ความสามารถหลายมิติ

ในที่นี้จะกล่าวถึงความยากของข้อสอบที่ใช้ความสามารถหลายมิติ (multidimensional item difficulty : MDIFF) อำนาจจำแนกของข้อสอบที่ใช้ความสามารถหลายมิติ (multidimensional item discrimination : MDISC) พารามิเตอร์การเดา (guessing parameter : c_i) และพารามิเตอร์ของผู้ตอบข้อสอบ (person parameter : θ_j)

2.5.1 ความยากของข้อสอบที่ใช้ความสามารถหลายมิติ (multidimensional item difficulty : MDIFF)

การนิยามความยากของข้อสอบที่ใช้ความสามารถหลายมิติ (MDIFF) ตั้งอยู่บนข้อตกลงเบื้องต้น 3 ประการดังนี้

1) ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องเพิ่มขึ้นเสมอ เมื่อความสามารถแต่ละมิติเพิ่มขึ้น

2) ข้อสอบแต่ละข้อมีตำแหน่งที่แน่นอนที่จุดหนึ่งในสเปซหลายมิติ

3) จุดที่สมเหตุสมผลมากที่สุดที่ใช้ในการกำหนดความยากของข้อสอบที่ใช้ความสามารถหลายมิติ (MDIFF) แต่ละข้อในสเปซหลายมิติ คือ จุดซึ่งข้อสอบมีอำนาจจำแนกสูงสุด เป็นจุดที่ข้อสอบให้สารสนเทศเกี่ยวกับผู้ตอบข้อสอบในสิ่งที่ต้องการวัดได้มากที่สุด

จากข้อตกลงที่กล่าวมาจึงมีการกำหนดเทคนิคในการตรวจสอบความยากของข้อสอบที่ใช้ความสามารถหลายมิติ (MDIFF) การแสดงกระบวนการตรวจสอบต้องใช้แบบจำลองเฉพาะ ในที่นี้จึงเลือกใช้แบบจำลองการตอบข้อสอบโลจิสติกหลายมิติชนิด 2 พารามิเตอร์ (multidimensional 2-parameter logistic model : M2PL) ซึ่งเป็นส่วนขยายของแบบจำลองการตอบข้อสอบโลจิสติกชนิด 2 พารามิเตอร์ (two-parameter logistic model : 2 PL) เพราะ M2PL สอดคล้องกับข้อตกลงเบื้องต้นและมีกระบวนการในการประมาณค่าพารามิเตอร์แล้ว (Reckase, 1985 : 402 ; citing McKinley & Reckase, 1983b) M2PL นิยามผิวระนาบการตอบข้อสอบ (Item Response Surface : IRS) ว่าหมายถึงความน่าจะเป็นของการตอบข้อสอบได้ถูกต้องที่แต่ละจุดในสเปซความสามารถหลายมิติ (multidimensional θ space) ซึ่งเป็นความน่าจะเป็นที่เพิ่มขึ้นเสมอเมื่อสมาชิกของ θ_j เพิ่มขึ้น ในการตรวจสอบความยากของข้อสอบที่ใช้ความสามารถหลายมิติต้องตรวจสอบจุดในสเปซความสามารถที่ IRS มีความชันสูงสุด แต่ความชันตลอดผิวระนาบ (surface) แตกต่างกันไปขึ้นอยู่กับทิศทางที่สัมพันธ์กับผิวระนาบ จึงกำหนดให้ใช้ความชันที่จุดในสเปซความสามารถที่มีทิศทางจากจุดกำเนิด (origin) ของสเปซความสามารถไปยังจุดนั้น การอธิบาย MDIFF ต้องรายงานทั้งระยะทางจากจุดที่มีความชันมากที่สุดจากจุดกำเนิด และมุมหรือทิศทาง (cosine) ที่ใช้ในการอธิบายทิศทาง พารามิเตอร์ d_i ในแบบจำลอง MIRT ไม่ได้มีความหมายเหมือนพารามิเตอร์ b ใน IRT แต่มีค่าเท่ากับ $(-ab)$ พารามิเตอร์ความยากของข้อสอบที่ใช้ความสามารถหลายมิติ (MDIFF_{*i*}) ซึ่งมีความหมายเหมือน b -parameter ใน IRT กำหนดดังนี้

$$\text{MDIFF}_i = \frac{-d_i}{\text{MDISC}_i}$$

เมื่อ MDIFF_{*i*} เป็นพารามิเตอร์ความยากของข้อสอบที่ใช้ความสามารถหลายมิติ

$MDISC_i$ เป็นพารามิเตอร์อำนาจจำแนกของข้อสอบที่ใช้ความสามารถหลายมิติ

ค่าของ $MDIFF_i$ แสดงระยะทางจากจุดกำเนิดของสเปซความสามารถ ถึงจุดที่มีความชันสูงสุด ใน IRS ทิศทางของความชันสูงสุดจากจุดกำเนิด กำหนดดังนี้

$$\alpha_{ik} = \frac{a_{ik}}{MDISC_i}$$

เมื่อ α_{ik} คือ มุมซึ่งเกิดจากเส้นตรงจากจุดกำเนิดของสเปซถึงจุดที่มีความชันสูงสุดที่เกิดจาก แกนที่ k ของข้อสอบข้อที่ i

โดยสรุปความยากของข้อสอบที่ใช้ความสามารถหลายมิติเป็นการอธิบายความยากของ ข้อสอบด้วยทิศทาง (direction) และระยะทาง (distance) ในสเปซคุณลักษณะภายในที่สมบูรณ์ เมื่อข้อสอบมี อำนาจจำแนกสูงสุด นิยามดังกล่าวสามารถใช้อธิบายแบบจำลองใด ๆ ที่ความน่าจะเป็นในการตอบข้อสอบ ได้ถูกต้องเพิ่มขึ้นเมื่อความสามารถในมิติต่าง ๆ เพิ่มขึ้น

2.5.2 อำนาจจำแนกของข้อสอบที่ใช้ความสามารถหลายมิติ (multidimensional item discrimination : MDISC)

ในกรอบความคิดของทฤษฎีการตอบข้อสอบมิติเดียว ค่าความยากของข้อสอบบ่งบอกโดยจุด บนสเกลความสามารถ (θ scale) ที่ความชันของโค้งคุณลักษณะข้อสอบมีความชันสูงสุด ซึ่งเป็นจุดเดียวกับ จุดเปลี่ยนโค้ง ในกรณีของ MIRT ค่าความยากของข้อสอบที่ใช้ความสามารถหลายมิติ ($MDIFF$) เป็นจุดซึ่ง IRS มีความชันสูงสุดเมื่อพิจารณาจากทิศทางของจุดกำเนิดของสเปซ จุดนี้เป็นจุดเปลี่ยนของ IRS เช่นเดียวกัน

$MDISC$ ถูกกำหนดโดยสมาชิกของเวกเตอร์ a (a -vector) สมาชิกของเวกเตอร์ a เกี่ยวข้อง กับความชันของ IRS ในทิศทางที่สอดคล้องกับแกน θ แต่อำนาจจำแนกแตกต่างกันไปขึ้นอยู่กับทิศทางที่ถูก วัดในสเปซความสามารถ การรวมกันของมิติความสามารถจะมีอำนาจจำแนกมากกว่ามิติเดียว

อำนาจจำแนกของข้อสอบที่ใช้ความสามารถหลายมิติ ($MDISC$) กำหนดดังนี้

$$MDISC_i = \sqrt{\sum_{k=1}^P a_{ik}^2}$$

เมื่อ $MDISC_i$ คือ อำนาจจำแนกของข้อสอบที่ใช้ความสามารถหลายมิติข้อที่ i สำหรับ การรวมที่ดีที่สุดของความสามารถ (the best combination of abilities)

P คือ จำนวนมิติในสเปซความสามารถ

a_{ik} คือ สมาชิกของเวกเตอร์ a_i

2.5.3 พารามิเตอร์การเดา (guessing parameter : c_i)

พารามิเตอร์การเดา มีความหมายเหมือนกับพารามิเตอร์ c_i ในแบบจำลองโลจิสติกสาม พารามิเตอร์ ค่าของพารามิเตอร์บ่งบอกความน่าจะเป็นที่ผู้ตอบข้อสอบที่มีความสามารถต่ำมาก ๆ ในทุกมิติ ที่แบบจำลองกำหนดตอบข้อสอบได้ถูกต้อง

จากการศึกษางานวิจัยที่เกี่ยวข้องกับ MIRT พบว่านักวิจัยกลุ่มหนึ่งในต่างประเทศได้ศึกษาค้นคว้าวิจัยและพัฒนาโปรแกรมคอมพิวเตอร์เพื่อขยายแนวคิดของทฤษฎีการตอบข้อสอบที่อาศัยข้อตกลงของความเป็นมิติเดียวของแบบทดสอบมาเป็นการใช้ความสามารถหลายมิติในการตอบข้อสอบอย่างต่อเนื่องเป็นเวลาเกือบยี่สิบปี เป็นความพยายามในการค้นหาสารสนเทศจากการใช้แบบทดสอบเพื่อแปลความหมายคะแนนจากผลทดสอบได้อย่างมีความหมายมากยิ่งขึ้น

3. การทำหน้าที่เบี่ยงเบนตามแนวคิดของทฤษฎีการตอบข้อสอบ

คุณลักษณะที่สำคัญประการหนึ่งของแบบทดสอบคือต้องมีความเที่ยงตรง ซึ่งจะทำให้สามารถใช้ข้อมูลที่ได้จากการวัดแสดงหลักฐานว่าแบบทดสอบสามารถวัดได้ตรงตามวัตถุประสงค์ที่ต้องการวัดอย่างแท้จริง และทำให้เกิดความยุติธรรมในการทดสอบ ในสหรัฐอเมริกาดินแดนแห่งเสรีภาพและการเรียกร้องสิทธิของบุคคล การสร้างแบบทดสอบและการใช้แบบทดสอบอยู่ภายใต้การตรวจสอบอย่างใกล้ชิดของสังคม ผู้สร้างแบบทดสอบและผู้ใช้แบบทดสอบจะต้องแสดงให้เห็นว่าแบบทดสอบเหล่านั้นปราศจากความลำเอียงต่อผู้ตอบข้อสอบกลุ่มใดกลุ่มหนึ่ง จึงมีความพยายามในการตรวจสอบความลำเอียงของข้อสอบมาเป็นเวลานานตั้งแต่ต้นศตวรรษที่ 20 แต่วิธีการตรวจสอบอย่างเป็นระบบได้นำเสนอในปี ค.ศ. 1964 โดยคาร์ดอลล์และคอฟฟ์แมน (Cardall & Coffman) โดยประยุกต์กระบวนการวิเคราะห์ความแปรปรวน (analysis of variance) ทดสอบปฏิกริยาร่วมของข้อสอบที่ใช้กับผู้ตอบข้อสอบผิวดำและผิวขาวที่เข้ารับการทดสอบ SAT ในปี ค.ศ. 1963 (Angoff, 1993 : 3) การที่ข้อสอบอำนวยความสะดวกให้กับผู้ตอบข้อสอบที่มีความสามารถเท่าเทียมกันแต่เป็นสมาชิกของประชากรในกลุ่มย่อยที่แตกต่างกันนั้นเดิมเรียกว่า ความลำเอียงของข้อสอบ (item bias) แต่ในระยะต่อมาได้มีการโต้แย้งว่าความลำเอียงของข้อสอบเกี่ยวข้องกับการรวบรวมหลักฐานเชิงประจักษ์ที่เกี่ยวข้องกับการทำข้อสอบของผู้ตอบข้อสอบจาก 2 กลุ่มย่อย หลักฐานเชิงประจักษ์ของผลการสอบที่แตกต่างกันของผู้ตอบข้อสอบทั้ง 2 กลุ่มยังไม่เพียงพอที่จะสรุปว่าข้อสอบมีความลำเอียง เนื่องจากความลำเอียงของข้อสอบให้ความหมายอย่างน้อย 2 อย่าง คือ ความหมายในเชิงสังคมและความหมายในเชิงสถิติ ดังนั้นการตรวจสอบด้วยวิธีการทางสถิติโดยอาศัยข้อมูลเชิงประจักษ์จึงสมควรใช้คำว่าทำหน้าที่เบี่ยงเบนของข้อสอบ (differential item functioning : DIF) มากกว่า (Hambleton, 1991 : 109) ซึ่งความคิดเห็นดังกล่าวสอดคล้องกับแองกอฟฟ์ (Angoff, 1993 : 6) ซึ่งได้นิยามการทำหน้าที่เบี่ยงเบนของข้อสอบว่า หมายถึงการที่สังเกตได้ว่าข้อสอบข้อนั้นแสดงคุณสมบัติทางสถิติที่แตกต่างกัน เมื่อใช้ข้อมูลจากผู้ตอบข้อสอบที่มีความสามารถเท่ากัน แต่อยู่ในกลุ่มที่แตกต่างกัน ส่วนแฮมเบิลตัน (Hambleton, 1991 : 110) กล่าวว่า ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนเป็นข้อสอบที่เมื่อใช้กับบุคคลที่มีความสามารถเท่ากัน แต่อยู่ในกลุ่มที่แตกต่างกัน มีความน่าจะเป็นในการตอบข้อสอบข้อนั้นได้ถูกต้องไม่เท่ากัน ซึ่งนิยามดังกล่าวสามารถอธิบายโดยใช้ทฤษฎีการตอบข้อสอบ (IRT) ได้อย่างมีประสิทธิภาพ เนื่องจากพารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถของผู้ตอบข้อสอบไม่แปรเปลี่ยนไปตามกลุ่มผู้ตอบข้อสอบ

กระบวนการในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ทำได้ 2 ขั้นตอน ดังนี้ (ญาณภัทร สีหะมงคล, 2540 : 2 อ้างจาก อุทุมพร จามรمان, 2537 : 129)

1. การตรวจสอบในขั้นตอนการสร้างข้อสอบโดยให้ผู้เชี่ยวชาญพิจารณาว่าข้อสอบใดตรงตามวัตถุประสงค์และมีเนื้อหาสอดคล้องกับตารางกำหนดข้อสอบหรือไม่ รวมทั้งพิจารณาความชัดเจนของภาษาที่ใช้ และขจัดความคลุมเครือของภาษาที่อาจทำให้เกิดประโยชน์แก่ผู้ตอบข้อสอบกลุ่มใดกลุ่มหนึ่ง

2. การตรวจสอบโดยใช้วิธีการทางสถิติ ดำเนินการภายหลังการนำแบบทดสอบไปใช้แล้ว สามารถทำได้ 2 แนวทาง คือ

2.1 ใช้เกณฑ์ภายนอก (External Criterion) ใช้ข้อมูลการตอบแบบทดสอบที่สร้างขึ้นหาความสัมพันธ์กับเกณฑ์ที่เป็นมาตรฐาน เช่น เกณฑ์เฉลี่ยหรือผลการปฏิบัติงาน เป็นต้น วิธีนี้มีปัญหาในการหาเกณฑ์ภายนอก เนื่องจากถ้าเกณฑ์ไม่มีมาตรฐานแล้วจะทำให้การตรวจสอบขาดความถูกต้อง วิธีนี้จึงไม่ได้รับความนิยม

2.2 ใช้เกณฑ์ภายใน (Internal Criterion) ศึกษาจากการตรวจสอบโครงสร้างภายในของแบบทดสอบ พิจารณาจากคะแนนของผู้ตอบข้อสอบแต่ละกลุ่มว่าวัดคุณลักษณะตามโครงสร้างเดียวกันหรือไม่ถือเป็นกระบวนการในการสนับสนุนการตรวจสอบความเที่ยงตรงเชิงโครงสร้าง

ศูนย์บริการการทดสอบทางการศึกษา (Educational Testing Service) ในสหรัฐอเมริกา ใช้ค่าสถิติการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF statistics) เป็นหลักฐานในการโต้แย้งและฟ้องร้องทางศาล ในการสร้างและพัฒนาแบบทดสอบ หน่วยงานดังกล่าวจึงตรวจสอบ DIF ในทุกขั้นตอนของกระบวนการตั้งแต่การสร้างแบบทดสอบ การตรวจให้คะแนน และการรายงานผลการสอบ พร้อมทั้งนำผลจากการศึกษามาปรับตารางกำหนดข้อสอบและตรวจสอบความเที่ยงตรงเชิงโครงสร้างของผลจากการวัดด้วย (Zieky. 1993 : 337 ; Linn. 1993 : 349)

การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยใช้วิธีการทางสถิติที่อาศัยเกณฑ์ภายใน แบ่งออกเป็น 3 กลุ่มใหญ่ ๆ (Camilli and Shepard. 1994 : 122-131)

1. การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบตามแนวคิดทฤษฎีการทดสอบแบบมาตรฐานเดิม ได้แก่ วิธีการแปลงค่าความยาก (Transformation Item Difficulty) หรือ Delta-Plot วิธีการวิเคราะห์ความแปรปรวน (Analysis of Variance) วิธีพอยท์ไบเซรียล (Point Biserial)

2. การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยใช้ตารางการถ่วง (Contingency Table) ได้แก่ วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) วิธีไคสแควร์ (Chi-Square) วิธีมาตรฐาน (Standardization Method)

3. การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบตามแนวคิดของทฤษฎีการตอบข้อสอบ เป็นแนวคิดที่พิจารณาผลการตอบข้อสอบกับโอกาสการตอบข้อสอบได้อย่างถูกต้องของผู้ที่มีความสามารถระดับเดียวกัน แต่อยู่ต่างกลุ่มกัน เช่น หญิง-ชาย ถ้าเส้นโค้งคุณลักษณะข้อสอบมีความแตกต่างกันแสดงว่าข้อสอบนั้นทำหน้าที่เบี่ยงเบนไปจากปกติทำให้เกิดการได้เปรียบเสียเปรียบขึ้น

วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบมีวิวัฒนาการมาเป็นเวลานานโดยพยายามแก้ไขข้อบกพร่องของวิธีการที่มีมาก่อน ประกอบกับความก้าวหน้าของเทคโนโลยีคอมพิวเตอร์ ทำให้มีการพัฒนาวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบหลายวิธี มีผลการศึกษาค้นคว้าวิจัยโดยใช้ข้อมูลจำลองและข้อมูลเชิงประจักษ์จำนวนมาก สรุปข้อค้นพบได้ดังนี้ (ญาณภัทร สีหะมงคล. 2540 : 6)

1. วิธีที่เหมาะสมและดีที่สุดทางทฤษฎี อาจกล่าวได้ว่ารวมทั้งการปฏิบัติด้วย ได้แก่ วิธีการตรวจสอบตามแนวคิดของทฤษฎีการตอบข้อสอบ ชนิด 2 และ 3 พารามิเตอร์

2. ตัวแปรที่เกี่ยวข้องในการสอบ ได้แก่ จำนวนผู้เข้าสอบ จำนวนข้อสอบ และสัดส่วนจำนวนข้อสอบที่ทำหน้าที่เบี่ยงเบนในแบบทดสอบ ซึ่งแต่ละวิธีที่ใช้ตรวจสอบจะมีความไวในการตรวจสอบแตกต่างกัน รวมทั้งมีความไม่คงที่ในการตัดสินข้อสอบที่ทำหน้าที่เบี่ยงเบนในเงื่อนไขต่าง ๆ

3. เกณฑ์บ่งชี้ประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบหรือแบบทดสอบ ได้แก่ ความคลาดเคลื่อนในการยอมรับผิดพลาด หรือความคลาดเคลื่อนประเภทที่ 1 และการปฏิเสธผิดพลาด หรือความคลาดเคลื่อนประเภทที่ 2

3.1 วิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบตามแนวคิดทฤษฎีการตอบ

ข้อสอบ

ตามแนวคิดของทฤษฎีการตอบข้อสอบตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยการเปรียบเทียบโครงสร้างคุณลักษณะข้อสอบจากการตอบของผู้ตอบข้อสอบตั้งแต่ 2 กลุ่มขึ้นไป แบ่งเป็นกลุ่มอ้างอิง (Reference Group) หมายถึงผู้ตอบข้อสอบกลุ่มที่คาดว่าจะจะเป็นกลุ่มที่ได้ประโยชน์จากข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน และกลุ่มสนใจ (Focal Group) หมายถึงผู้ตอบข้อสอบกลุ่มที่คาดว่าจะเสียประโยชน์จากข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน ทำได้ 2 วิธี คือ ก) เปรียบเทียบพารามิเตอร์ที่อธิบายคุณลักษณะโครงสร้างข้อสอบที่ได้จากกลุ่มอ้างอิงและกลุ่มสนใจ ข) เปรียบเทียบพื้นที่ระหว่างโครงสร้างคุณลักษณะข้อสอบที่ได้จากกลุ่มอ้างอิงและกลุ่มสนใจ

3.1.1 การเปรียบเทียบค่าพารามิเตอร์ของกลุ่มที่ศึกษา

ถ้าค่าพารามิเตอร์ที่คำนวณได้ทั้ง 2 ชุดมีค่าเท่ากัน แสดงว่าฟังก์ชันเหมือนกันทุกจุด และความน่าจะเป็นในการตอบข้อสอบถูกต้องมีค่าเท่ากัน ในกรณีนี้สมมุติฐานว่างของฟังก์ชันการตอบข้อสอบคือ

$$H_0 : b_1 = b_2, a_1 = a_2, c_1 = c_2$$

ถ้าปฏิเสธสมมุติฐานว่างแสดงว่าข้อสอบทำหน้าที่เบี่ยงเบน

สถิติทดสอบสมมุติฐานว่าง คือ

$$\chi^2 = (a_{diff} \ b_{diff} \ c_{diff})' \Sigma^{-1} (a_{diff} \ b_{diff} \ c_{diff})$$

$$\text{เมื่อ } a_{diff} = a_2 - a_1, \ b_{diff} = b_2 - b_1, \ c_{diff} = c_2 - c_1$$

$$\Sigma = \text{variance - covariance matrix ของความแตกต่างระหว่าง parameter estimates}$$

แนวคิดนี้วิธีที่ได้รับความนิยม ได้แก่ วิธีที่อาศัยแนวคิดของลอร์ด (Lord's χ^2) มีขั้นตอนในการทดสอบดังนี้

1. ทดสอบข้อมูลกับแบบจำลองที่ศึกษา
2. ประมาณค่าพารามิเตอร์ข้อสอบ (a, b, c)
3. แปลงหน่วยของค่าพารามิเตอร์ให้อยู่ในหน่วยเดียวกันหรือหน่วยร่วม
4. นำค่าพารามิเตอร์ของข้อสอบมาเปรียบเทียบโดยใช้สถิติทดสอบไคสแควร์ (χ^2)

การเปรียบเทียบความแตกต่างระหว่างพารามิเตอร์ข้อสอบถูกวิจารณ์ว่ามีอัตราความคลาดเคลื่อนในการยอมรับผิดพลาด (ความคลาดเคลื่อนประเภทที่ 1) สูง (Hambleton. 1991 : 113 ; citing McLoughlin & Drasgow. 1987)

3.1.2 การเปรียบเทียบพื้นที่ระหว่างโค้งคุณลักษณะข้อสอบที่ได้จากกลุ่มที่ศึกษา

เป็นการเปรียบเทียบโค้งคุณลักษณะข้อสอบโดยตรงหลังจากทำให้ค่าพารามิเตอร์ที่ประมาณค่าได้จากกลุ่มที่ศึกษามีหน่วยเดียวกันหรือมีหน่วยรวมแล้ว ถ้าโค้งคุณลักษณะข้อสอบเท่ากันพื้นที่ระหว่างโค้งเป็นศูนย์ (0) ถ้าพื้นที่ระหว่างโค้งไม่เป็นศูนย์ก็สรุปว่าข้อสอบข้อนั้นแสดงการทำหน้าที่เบี่ยงเบนการคำนวณพื้นที่ที่มีวิธีการหลายอย่าง โดยสรุปแล้วมีขั้นตอนดังนี้

1. แปลงพิสัยของความสามารถออกเป็น k ช่วง
2. สร้างสี่เหลี่ยมผืนผ้ามีศูนย์กลางที่จุดกึ่งกลาง (midpoint) ของแต่ละช่วง
3. หาค่าความน่าจะเป็นที่จุดกึ่งกลางของแต่ละช่วง
4. หาค่าสัมบูรณ์ของความแตกต่างระหว่างความน่าจะเป็น
5. คูณความแตกต่างของความน่าจะเป็นด้วยความกว้างของช่วง แล้วนำมารวมกัน

ดังนั้น

$$A_i = \sum_{\theta=r}^s |P_{i1}(\theta) - P_{i2}(\theta)| \Delta\theta$$

เมื่อ $\Delta\theta$ เป็นความกว้างของช่วงและถูกเลือกมาโดยมีขนาดเล็กมากที่สุดเท่าที่จะเป็นไปได้ เช่น .01

ค่า r และ s แสดงพิสัยของความสามารถเหนือพื้นที่ที่นำมาคำนวณเป็นค่าที่เปลี่ยนแปลงได้ (arbitrary) โดยผู้ใช้เป็นผู้เลือก ปกติจะเลือก ± 3 SD จากค่าเฉลี่ย ซึ่งจะรับประกันพื้นที่ที่คำนวณเหนือพิสัยความสามารถของผู้ตอบข้อสอบทุกคน

แนวคิดนี้วิธีที่ได้รับความนิยมได้แก่ แนวคิดของราจู (Raju, 1990) ซึ่งเสนอให้ใช้การวัดขนาดพื้นที่บนช่วงความสามารถแบบเปิด (opened interval area) และแนวคิดของคิมและโคเฮน (Kim and Kohen, 1991) ซึ่งเสนอให้วัดขนาดพื้นที่บนช่วงความสามารถแบบปิด (closed interval area) ภายใต้การวัดขนาดพื้นที่ทั้ง 2 แนวคิดนี้ มีการวัดพื้นที่ 2 ลักษณะคือการวัดพื้นที่แบบมีเครื่องหมาย (Signed Area) และการวัดพื้นที่แบบไม่มีเครื่องหมาย (Unsigned Area)

การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบตามแนวคิดในการเปรียบเทียบพื้นที่ระหว่างโค้งคุณลักษณะข้อสอบที่ได้จากกลุ่มอ้างอิงและกลุ่มสนใจ แบ่งเป็น 2 ประเภทใหญ่ ๆ ดังนี้ (Camilli & Sheppard, 1994 : 59-60)

1. การทำหน้าที่เบี่ยงเบนอย่างสม่ำเสมอของข้อสอบ (uniform or consistent DIF) เกิดขึ้นเมื่อโค้งคุณลักษณะข้อสอบ (ICC) ของทั้ง 2 กลุ่มแตกต่างกันและไม่ตัดกัน แสดงว่าข้อสอบอำนาจประโยชน์ให้กลุ่มใดกลุ่มหนึ่งตลอดช่วงพิสัยความสามารถ เกิดขึ้นในกรณีที่ ICC มี a-parameter ไม่แตกต่างกัน

2. การทำหน้าที่เบี่ยงเบนอย่างไม่สม่ำเสมอของข้อสอบ (nonuniform or inconsistent DIF) เกิดขึ้นเมื่อโค้งคุณลักษณะข้อสอบของทั้ง 2 กลุ่มแตกต่างกันและตัดกันที่จุดใดจุดหนึ่งบนสเกลความสามารถ (θ scale) แสดงว่าข้อสอบไม่ได้อำนาจประโยชน์ให้กลุ่มใดกลุ่มหนึ่งตลอดช่วงพิสัยความสามารถ แต่อำนาจประโยชน์ให้กลุ่มหนึ่งในช่วงหนึ่ง และอำนาจประโยชน์ให้อีกกลุ่มหนึ่งในอีกช่วงหนึ่ง

3.2 งานวิจัยที่เกี่ยวข้อง

การศึกษาค้นคว้าเกี่ยวกับการทำหน้าที่เบี่ยงเบน มีทั้งการใช้ข้อมูลเชิงประจักษ์และข้อมูลจำลอง การศึกษาส่วนหนึ่งเป็นการเปรียบเทียบวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีที่แตกต่าง กัน ซึ่งจะกล่าวโดยสรุปดังนี้

งานวิจัยของต่างประเทศ

สวามินาธานและโรเจอร์ (Swaminathan & Rogers. 1990 : 361-370) ได้เปรียบเทียบผล การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีถดถอยโลจิสติกกับวิธีแมนเทิลและแฮนส์เซล โดยใช้สถานการณ์จำลอง จำลองเงื่อนไข 6 เงื่อนไข ได้แก่ กลุ่มตัวอย่าง 2 ขนาด คือ 250 คน และ 500 คน ความยาวแบบทดสอบ 3 ขนาดคือ 40 ข้อ 60 ข้อ และ 80 ข้อ ในแบบทดสอบแต่ละชุดมีข้อสอบที่ทำหน้าที่ เบี่ยงเบน 20 % เป็นข้อสอบที่ทำหน้าที่เบี่ยงเบนอย่างสม่าเสมอครึ่งหนึ่งและทำหน้าที่อย่างไม่สม่าเสมออีก ครึ่งหนึ่ง ผลการวิจัยมีดังนี้ 1) ในกรณีที่ข้อสอบทำหน้าที่เบี่ยงเบนอย่างสม่าเสมอพบว่า วิธีทั้งสองให้ผลการ วิเคราะห์ใกล้เคียงกัน แต่กรณีที่ข้อสอบทำหน้าที่เบี่ยงเบนอย่างไม่สม่าเสมอ วิธีการถดถอยโลจิสติกให้ผล ดีกว่าวิธีแมนเทิลและแฮนส์เซล 2) ระดับความคลาดเคลื่อนประเภทที่ 1 พบว่าวิธีแมนเทิลและแฮนส์เซล คลาดเคลื่อนประมาณร้อยละ 1 แต่วิธีการถดถอยโลจิสติกคลาดเคลื่อนประมาณร้อยละ 1-6 3) วิธีถดถอย โลจิสติกเสียค่าใช้จ่ายในการวิเคราะห์มากกว่าวิธีแมนเทิลและแฮนส์เซลประมาณ 3-4 เท่า

ราชูและคณะ (Raju, et. al. 1993 : 301-314) ได้ศึกษาเปรียบเทียบผลการตรวจสอบการทำ หน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีของลอร์ด (Lord's χ^2) วิธีของราชู (Z) และวิธีของแมนเทิลและแฮนส์ เซล เครื่องมือที่ใช้ในการศึกษาเป็นแบบทดสอบวัดคำศัพท์จำนวน 45 ข้อ กลุ่มตัวอย่างเป็นนักเรียนเกรด 10 และ 12 จำนวน 839 คน เป็นเพศหญิง 440 คน เพศชาย 399 คน ผิวขาว 436 คน ผิวดำ 245 คน นอกนั้น เป็นเชื้อชาติอื่น ผลการศึกษาพบว่าการตรวจสอบการทำหน้าที่เบี่ยงเบนระหว่างนักเรียนผิวขาวและผิวดำมี 1 ข้อ สำหรับการทดสอบด้วยวิธีของลอร์ดและวิธีของราชู ส่วนวิธีของแมนเทิลและแฮนส์เซลพบ ข้อสอบทำหน้าที่เบี่ยงเบนจำนวน 2 ข้อ ข้อสอบทั้ง 3 ข้อเป็นข้อสอบที่มีความยากต่างกันมากสำหรับนักเรียน ผิวขาวและผิวดำ ส่วนผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างชายและหญิง โดยวิธีของ ลอร์ดและวิธีของราชูพบข้อสอบทำหน้าที่ซ้ำกัน 4 ข้อ ส่วนวิธีของแมนเทิลและแฮนส์เซลพบ 5 ข้อ ซ้ำกัน 4 ข้อ ข้อสอบทุกข้อมีความยากสูงกว่าข้อสอบที่ไม่แสดงการทำหน้าที่เบี่ยงเบน

นารายานานและสวามินาธาน (Narayanan & Swaminathan. 1994 : 315-328) ศึกษาผล ของการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ด้วยวิธีแมนเทิลและแฮนส์เซลกับวิธีซิปเทสท์โดยใช้ ข้อมูลจำลอง แบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ ตัวแปรที่ศึกษาคือ 1) ขนาดกลุ่มตัวอย่าง โดยกลุ่ม อ้างอิงมี 3 ขนาด ได้แก่ 300, 500 และ 1,000 คน กลุ่มสนใจมี 3 ขนาด ได้แก่ 100, 200 และ 300 คน 2) การกระจายความสามารถมี 2 แบบ 3) อัตราส่วนของข้อสอบที่ทำหน้าที่เบี่ยงเบนในแบบทดสอบมี 2 ขนาด 4) ขนาดของพื้นที่ระหว่างโค้งคุณลักษณะข้อสอบของผู้ตอบข้อสอบมี 4 ขนาด 5) ความยากและอำนาจ จำแนกของข้อสอบมี 6 ระดับ ผลการศึกษาพบว่าขนาดกลุ่มตัวอย่าง อัตราส่วนของข้อสอบที่ทำหน้าที่ เบี่ยงเบน ขนาดของพื้นที่ระหว่างโค้งคุณลักษณะข้อสอบ ค่าความยากและอำนาจจำแนกของข้อสอบเป็นตัว แปรที่มีผลกระทบต่ออัตราการตรวจสอบของทั้งสองวิธีอย่างมีนัยสำคัญ วิธีแมนเทิลและแฮนส์เซลและวิธี ซิปเทสท์มีประสิทธิภาพเท่ากันในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบอย่างสม่าเสมอ เมื่อการ กระจายความสามารถระหว่างกลุ่มเท่ากัน แต่เมื่อการกระจายความสามารถระหว่างกลุ่มไม่เท่ากันวิธีซิปเทสท์ จะมีประสิทธิภาพมากกว่า

ราชู แวนเดอร์ลินเด็น และฟลีร์ (Raju, van der Linden & Fleer. 1995 : 353-368) เสนอวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบและแบบทดสอบ (internal measures of differential functioning of items and tests : DFIT) โดยอาศัยกรอบความคิดของทฤษฎีการตอบข้อสอบ ศึกษาโดยใช้ข้อมูลจำลองที่สร้างขึ้นมา ใช้แบบจำลองโลจิสติกชนิด 2 พารามิเตอร์ ด้วยโปรแกรม RANGEN (Fleer, Kiley & Raju. 1991) สร้างแบบทดสอบที่มีความยาว 40 ข้อ ประกอบด้วยข้อสอบที่แสดง DIF 0 % 5 % 10 % และ 20 % ขนาดของกลุ่มตัวอย่าง กลุ่มละ 500 คน สำหรับเงื่อนไขกลุ่มตัวอย่างขนาดเล็ก และกลุ่มละ 1,000 คน สำหรับกลุ่มตัวอย่างขนาดใหญ่ ตรวจสอบการทำหน้าที่เบี่ยงเบนในเงื่อนไขต่าง ๆ ด้วยวิธีการของลอร์ด (Lord's χ^2 test) การวัดพื้นที่แบบคิดเครื่องหมาย การวัดพื้นที่โดยไม่คิดเครื่องหมาย และวิธี DFIT ผลที่ได้พบว่าวิธีการ DFIT สามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบ (DTF) และการทำหน้าที่เบี่ยงเบนของข้อสอบทั้งแบบที่มีการชดเชยและแบบที่ไม่มีการชดเชย (CDIF และ NCDIF) ได้อย่างถูกต้อง

รูสโซและสแตต์ (Roussos & Stout. 1996 : 215-230) ศึกษาผลของการใช้กลุ่มตัวอย่างขนาดเล็ก ความแตกต่างของค่าพารามิเตอร์ข้อสอบ ความแตกต่างของค่าเฉลี่ยความสามารถของกลุ่มอ้างอิงและกลุ่มสนใจที่มีต่ออัตราความคลาดเคลื่อนชนิดที่ 1 การศึกษาที่ 1 ใช้กลุ่มอ้างอิงและกลุ่มสนใจกลุ่มละ 100, 200, 500 และ 1,000 คน เลือกข้อสอบ 1 ข้อ จาก ASVAB (Armed Services Vocational Aptitude Battery) โดยเลือกข้อที่มีค่าพารามิเตอร์ (a, b, c) ใกล้เคียงกับค่าเฉลี่ยของแบบทดสอบมากที่สุดและเป็นข้อสอบที่ไม่ลำเอียง ความสามารถ (θ) ของกลุ่มตัวอย่างทั้ง 2 กลุ่ม ถูกจำลองให้มีการกระจายเป็นปกติ และมีค่าความแปรปรวนเป็น 1 ความแตกต่างระหว่างค่าเฉลี่ยของการแจกแจงของกลุ่มถูกจำลองให้มีค่าเป็น 0.0 0.5 และ 1.0 แล้วคำนวณค่าความสามารถเฉลี่ยของแต่ละกลุ่ม ให้จุดกลางของค่าเฉลี่ยมีค่าใกล้เคียงระดับความยากเฉลี่ยของแบบทดสอบ จำลองการตอบ 400 ครั้ง และวิเคราะห์ 400 ครั้งในแต่ละสถานการณ์ ผลการใช้ SIBTEST และ MH ตรวจสอบข้อสอบที่เลือกขึ้นมาพบว่า นอกจากกรณีกลุ่มตัวอย่าง 100 คน และความแตกต่างของความสามารถเท่ากับ 0.0 แล้ว อัตราความคลาดเคลื่อนชนิดที่ 1 ของทั้ง 2 วิธีแตกต่างกันอย่างไม่มีนัยสำคัญ แต่พบว่า SIBTEST มีอัตราความคลาดเคลื่อนชนิดที่ 1 สูงกว่า MH เล็กน้อย สำหรับการวิเคราะห์เมื่อใช้กลุ่มตัวอย่างไม่เกิน 1,000 คน มีแนวโน้มว่าอัตราความคลาดเคลื่อนจะสูงขึ้นเมื่อกลุ่มตัวอย่างมากขึ้นทั้ง 2 วิธี การศึกษาที่ 2 ใช้แบบทดสอบฉบับเดิม แต่เลือกค่า a b และ c ของข้อสอบที่นำมาวิเคราะห์ต่างกันดังนี้ a = 0.4 1.0 และ 2.5 b = -1.5 -0.5 0.0 0.5 และ 1.5 c = 0.20 จำลองการกระจายความสามารถ (θ) ให้เป็นแบบปกติ ความแตกต่างของความสามารถมีค่า 0.0 และ 1.0 นอกนั้นจำลองเช่นเดียวกับการศึกษาที่ 1 แต่ใช้กลุ่มตัวอย่างขนาดใหญ่ คือ 500 1,000 และ 3,000 คน แต่ละสถานการณ์จำลองการตอบ 100 ครั้ง ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยใช้ SIBTEST และ MH พบว่าเมื่อ a มีค่าสูงขึ้นและ b มีค่าต่ำ อัตราการปฏิเสธสมมุติฐานจะสูงกว่าระดับที่กำหนดไว้คือ 0.05 โดยภาพรวม SIBTEST มีอัตราการปฏิเสธสมมุติฐานต่ำกว่า MH สรุปได้ว่าเมื่อใช้กลุ่มตัวอย่างขนาดเล็กไม่เกิน 1,000 คน MH และ SIBTEST ให้ผลในด้านความคลาดเคลื่อนชนิดที่ 1 ใกล้เคียงกันและอยู่ในระดับ 0.05 และพบว่า SIBTEST มีอัตราความคลาดเคลื่อนชนิดที่ 1 ต่ำกว่า MH ในกรณีที่กลุ่มตัวอย่างมีขนาดใหญ่มากกว่า 1,000 คน

ดักกลาส รูสโซ และ สแตต์ (Douglas, Roussos & Stout. 1996 : 465-484) ได้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยพิจารณาข้อสอบพร้อม ๆ กันครั้งละหลายข้อ โดยเสนอวิธีการตรวจสอบหมวดข้อสอบที่คาดว่าจะแสดงการทำหน้าที่เบี่ยงเบนเมื่อข้อสอบเหล่านั้นถูกนำมาพิจารณาพร้อม

กัน 2 วิธี คือ ใช้เฉพาะความคิดเห็นของผู้เชี่ยวชาญ และใช้วิธีการทางสถิติและตรวจสอบซ้ำด้วยความเห็นของผู้เชี่ยวชาญ พร้อมทั้งแสดงตัวอย่างการตรวจสอบการทำหน้าที่เบี่ยงเบน 3 ตัวอย่าง ดังนี้ ตัวอย่างที่ 1 เลือกหมวดข้อสอบโดยใช้ความเห็นของผู้เชี่ยวชาญอย่างเดียว ให้ผู้เชี่ยวชาญชาย 3 คน และหญิง 1 คน พิจารณาแบบทดสอบย่อยการใช้เหตุผลเชิงตรรกศาสตร์ (Logical Reasoning Subtest) ที่ได้จากการดำเนินการสอบในเดือนธันวาคม ค.ศ. 1991 (December 1991 administration of the Law School Admission Test : LSAT) จำนวน 49 ข้อ ในการตรวจสอบให้เพศชายเป็นกลุ่มอ้างอิงและเพศหญิงเป็นกลุ่มสนใจ ผู้เชี่ยวชาญจัดหมวดข้อสอบออกเป็น 8 หมวด และพิจารณาว่าหมวดใดน่าจะให้ประโยชน์แก่เพศชายหรือเพศหญิง หลังจากนั้นวิเคราะห์ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป SIBTEST ใช้กลุ่มตัวอย่างเพศชาย 3,000 คน เพศหญิง 3,000 คน พบว่าความคิดเห็นของผู้เชี่ยวชาญและผลการวิเคราะห์ด้วย SIBTEST สอดคล้องกัน จึงทำการวิเคราะห์ปริมาณ DIF ของแต่ละข้อ พบว่าแต่ละข้อแสดง DIF ด้วยปริมาณที่น้อยมาก จึงไม่มีการคัดเลือกข้อสอบออกจากแบบทดสอบ ตัวอย่างที่ 2 ใช้วิธีการทางสถิติในการตรวจสอบหมวดข้อสอบ คือ วิธี HCA (agglomerative hierarchical cluster analysis ; Jain & Dubes. 1988) และ DIMTEST (Nandakumar & Stout. 1993, Stout. 1987) แล้วจึงตรวจสอบด้วย SIBTEST ในการตรวจสอบใช้ข้อสอบ National Assessment of Educational Progress (NAEP) จำนวน 36 ข้อ ใช้กลุ่มตัวอย่างกลุ่มละ 500 คน ผลการตรวจสอบพบว่าการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบประสบความสำเร็จเป็นอย่างดี ตัวอย่างที่ 3 ใช้แบบทดสอบความเข้าใจในการอ่าน ที่สอบในเดือนธันวาคม ค.ศ. 1991 ซึ่งเป็นส่วนหนึ่งของการทดสอบ LSAT มีข้อสอบ 28 ข้อ แบ่งตามบทความที่กำหนดให้อ่านเป็น 4 บทความ แต่ละบทความมีข้อสอบ 5-8 ข้อ แล้วทำการตรวจสอบหมวดข้อสอบซ้ำโดยใช้ HCA และ DIMTEST หลังจากนั้นจึงวิเคราะห์ด้วย SIBTEST ใช้กลุ่มตัวอย่างกลุ่มละ 1,000 คน พบว่าทั้ง 4 บทความแสดงการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ ข้อสอบ 2 หมวดให้ประโยชน์แก่เพศชาย และ 2 หมวดให้ประโยชน์แก่เพศหญิง

โอชิม่าและคณะ (Oshima, et al. 1998 : 353-369) ใช้กรอบความคิดของ DFIT ตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ เพื่อตรวจสอบสาเหตุของการแสดงการทำหน้าที่เบี่ยงเบน ใช้ข้อมูลเชิงประจักษ์จากข้อมูลการตอบข้อสอบความเข้าใจในการอ่านของนักเรียนเกรด 4 ในการทดสอบ Metropolitan Achievement Tests ฟอรม s เป็นข้อสอบแบบเลือกตอบ 4 ตัวเลือก จำนวน 55 ข้อ ให้คะแนนรายข้อแบบสองค่า กลุ่มตัวอย่างที่ศึกษามี 2 กลุ่ม คือ เด็กชาย (R) กับเด็กหญิง (F) และกลุ่มที่มีสถานภาพทางสังคมระดับสูง (R) และกลุ่มที่มีสถานภาพทางสังคมระดับต่ำ (F) จัดหมวดข้อสอบเป็น 3 ลักษณะดังนี้ 1) จัดตามประเภทของความคิด (cognitive classification : - knowledge, understanding and thinking) 2) จัดตามวัตถุประสงค์การสอน 3) จัดตามเนื้อหา (passage) วิเคราะห์ข้อมูลการตอบของผู้ตอบข้อสอบกลุ่มอ้างอิงและกลุ่มสนใจกลุ่มละ 1,000 คน คำนวณค่าพารามิเตอร์ของแต่ละกลุ่มแยกจากกันด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป BILOG 3 ใช้แบบจำลอง 3PL แล้วหาสเกลร่วมของพารามิเตอร์ข้อสอบของแต่ละกลุ่ม โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป IPLINK หลังจากนั้นตรวจสอบ DIF ของข้อสอบด้วยโปรแกรม DFIT ถ้าพบข้อสอบที่มี DIF ขนาดใหญ่ก็ตัดออกไป นำค่าสัมประสิทธิ์ (linking coefficient) ไปใช้ในการวิเคราะห์ DIF, DTF และ DBF ต่อไป ผลปรากฏว่าการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ ทำให้สามารถตรวจสอบสาเหตุของการแสดงการทำหน้าที่เบี่ยงเบนได้เป็นอย่างดี การศึกษาในครั้งนี้พบว่าเนื้อหาในบทความที่ไม่สะท้อนชีวิตประจำวันของผู้ตอบข้อสอบเป็นสาเหตุประการหนึ่ง หากผู้สร้างแบบทดสอบใช้ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบประกอบกันคือใช้ Bundle CDIF ประกอบกับ Bundle

NCDIF จะทำให้สามารถสร้างแบบทดสอบที่มีความสมดุล ทำให้ผู้ตอบข้อสอบทุกกลุ่มได้รับความยุติธรรมในการทดสอบเท่ากัน

ไรอันและชิว (Ryan, Katherine E. & Chiu, Shuwan. 2001 : 73-90.) ได้ศึกษาเพื่อตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบเมื่อมีการเปลี่ยนแปลงตำแหน่งของข้อสอบ ใช้แบบทดสอบวิชาคณิตศาสตร์ระดับอุดมศึกษาชั้นปีที่ 1 จำนวน 40 ข้อ จัดทำเป็น 2 ฟอรัม ฟอรัมที่ 1 เรียงลำดับข้อสอบตามเนื้อหาจากง่ายไปยาก ฟอรัมที่ 2 เรียงลำดับข้อสอบแบบสุ่ม กลุ่มตัวอย่างที่ตอบข้อสอบฟอรัม 1 เป็นเพศชาย 546 คน เพศหญิง 520 คน กลุ่มตัวอย่างที่ตอบข้อสอบฟอรัม 2 เป็นเพศชาย 554 คน เพศหญิง 511 คน วิเคราะห์โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป SIBTEST พบว่าการเปลี่ยนแปลงตำแหน่งข้อสอบในแบบทดสอบไม่มีผลต่อการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ

งานวิจัยภายในประเทศ

งานวิจัยเกี่ยวกับการทำหน้าที่เบี่ยงเบนของข้อสอบภายในประเทศส่วนใหญ่เป็นการศึกษาค้นคว้าเพื่อทำวิทยานิพนธ์ในระดับบัณฑิตศึกษา เริ่มทำการศึกษาประมาณ พ.ศ. 2527 (ชัชชัย เผ่าพงศ์. 2527) ศึกษาโดยการเปรียบเทียบวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบวิธีการต่าง ๆ ในช่วงแรก (พ.ศ. 2527-2537) ใช้วิธีการที่อาศัยแนวคิดของทฤษฎีการทดสอบแบบมาตรฐานเดิม วิธีการทดสอบไคสแควร์ และวิธีการที่อาศัยแนวคิดทฤษฎีการตอบข้อสอบ (ทัศนีย์ พิรมนตรี. 2530, พัชรีย์ ปิยะกันต์. 2531, สุรศักดิ์ อมรรัตนศักดิ์. 2531, สุพัฒน์ สุขุมลันต์. 2534, นิรมล ชัยชลิต. 2537) ประมาณ พ.ศ. 2538 เริ่มศึกษาค้นคว้าโดยใช้วิธีการของแมนเทลและแฮนส์เซล และวิธีซิบเทสต์ สำหรับการศึกษาค้นคว้าเพื่อทำวิทยานิพนธ์ระดับดุษฎีบัณฑิตกิตติมศักดิ์ต่อไปนี้

กาญจนา วัธนสุนทร (2538) ได้พัฒนาเกณฑ์ในการตัดสินข้อสอบลำเอียงทางเพศโดยใช้ข้อมูลเชิงประจักษ์จากการสอบคัดเลือกเข้าศึกษาในระดับอุดมศึกษา ปีการศึกษา 2535 ใช้วิธีการตรวจสอบ 3 วิธี คือ วิธีตามแนวคิดทฤษฎีการตอบข้อสอบ วิธีแมนเทลและแฮนส์เซล และวิธีซิบเทสต์ ดัชนีที่พัฒนาเพื่อเป็นเกณฑ์ในการตัดสินข้อสอบลำเอียง คือ SA , UA , α_{MH} , β_{SIB} ตามลำดับ ซึ่งในการวิจัยครั้งนี้ได้จากการวิเคราะห์ค่าเฉลี่ยของดัชนีแต่ละตัว ตัวแปรที่ใช้ในการศึกษา ได้แก่ ความยาวของแบบทดสอบขนาด 20 30 40 ข้อสำหรับวิชาคณิตศาสตร์ และ 50 60 70 และ 80 ข้อ สำหรับวิชาภาษาอังกฤษ ขนาดกลุ่มตัวอย่างผู้ตอบข้อสอบ 100 200 400 600 800 และ 1,000 คน ผลการวิจัยพบว่าขนาดผู้ตอบข้อสอบมีอิทธิพลต่อค่าเฉลี่ยของดัชนีทุกตัว ความยาวของแบบทดสอบมีอิทธิพลต่อค่าเฉลี่ยของดัชนี SA และ UA แต่ไม่มีอิทธิพลต่อค่าเฉลี่ยของดัชนี α_{MH} และ β_{SIB} เกณฑ์ที่พัฒนาขึ้นเพื่อใช้ตัดสินความลำเอียงระหว่างผู้ตอบข้อสอบเพศหญิงและเพศชายมีดังนี้ 1) $SA > .80$ และ $UA > .50$ เมื่อความยาวแบบทดสอบน้อยกว่า 50 ข้อ 2) $SA > .40$ และ $UA > 1.20$ เมื่อความยาวข้อสอบมากกว่า 50 ข้อขึ้นไป 3) $\alpha_{MH} < .60$ และ $\alpha_{MH} > 1.40$ สำหรับทุกขนาดของผู้ตอบข้อสอบและความยาวแบบทดสอบ 4) $\beta_{SIB} > .06$ สำหรับทุกขนาดผู้ตอบข้อสอบและความยาวแบบทดสอบ ทั้งนี้ในการใช้ดัชนี SA หรือ UA ควรใช้ผู้ตอบข้อสอบขนาด 800 คนขึ้นไป ส่วนดัชนี α_{MH} และ β_{SIB} ควรใช้ขนาดผู้ตอบข้อสอบอย่างน้อย 600 คน

เรวดี อินทะสระ (2539) ศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่วิเคราะห์ความลำเอียงต่อเพศ ด้วยวิธีทฤษฎีการตอบข้อสอบ (IRT) วิธีแมนเทลและแฮนส์เซล (MH) และวิธีซิบเทสต์ (SIBTEST) พร้อมทั้งศึกษาการตัดสินผลสอบที่ใช้คะแนนมาตรฐานที่ปกติและคะแนนน้ำหนักความสามารถและสาเหตุความลำเอียงของข้อสอบ โดยศึกษาความลำเอียงของข้อสอบคัดเลือกเข้าศึกษาชั้นปี

ที่ 1 ประเภทรับตรง ปีการศึกษา 2538 ของมหาวิทยาลัยสงขลานครินทร์ ผลการศึกษาพบว่าวิธีตรวจสอบความลำเอียงทั้ง 3 วิธี ตัดสินจำนวนข้อสอบที่ลำเอียงแตกต่างกันในวิชาภาษาไทย ก. ฉบับที่ 2 และวิชาสังคมศึกษา ก. ฉบับที่ 1 ที่ระดับนัยสำคัญทางสถิติ .05 นอกนั้นแตกต่างกันที่ระดับ .01 โดยวิธีทฤษฎีการตอบข้อสอบตัดสินจำนวนข้อสอบที่ลำเอียงได้มากที่สุด ความสัมพันธ์ของลำดับที่การสอบไม่ว่าจะคิดคะแนนมาตรฐานที่ปกติหรือคิดคะแนนน้ำหนักความสามารถและใช้ข้อสอบทั้งหมดหรือใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง ต่างมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

เสรี ชัดแฉ่ม (2539) เปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันแบบไม่สม่ำเสมอของข้อสอบระหว่างวิธีแมนเทิลและแฮนส์เซล แบบแบ่งกลุ่มความสามารถผู้ตอบข้อสอบและความยากของข้อสอบและแบบปกติ ใช้แบบทดสอบความสามารถในการอ่านภาษาไทยที่ผู้วิจัยสร้างขึ้น ศึกษาในกลุ่มตัวอย่างนักเรียนชั้นมัธยมศึกษาปีที่ 1 จำนวน 1,200 คน จำแนกตามเพศ ผลการศึกษาพบว่าวิธีแมนเทิลและแฮนส์เซลแบบแบ่งกลุ่มความสามารถและความยากของข้อสอบ ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกันแบบไม่สม่ำเสมอได้มากกว่าวิธีแมนเทิลและแฮนส์เซลแบบปกติอย่างมีนัยสำคัญระดับ .05 เมื่อใช้การตรวจสอบตามวิธี IRT Area เป็นเกณฑ์

ญาณภัทร สีหะมงคล (2540) เปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบทำหน้าที่ต่างกันระหว่างวิธี Lord's χ^2 วิธี Raju's Area Measures และวิธี Closed Interval Area เมื่อความยาวของแบบทดสอบและขนาดของกลุ่มตัวอย่างต่างกัน ข้อมูลที่ใช้เป็นข้อมูลผลการประเมินคุณภาพและความก้าวหน้าทางการศึกษาปีการศึกษา 2536 วิชาคณิตศาสตร์ชั้นประถมศึกษาปีที่ 4 พบว่าจำนวนข้อสอบที่ทำหน้าที่ต่างกันจากการตรวจสอบด้วยวิธีการทั้งสามมีจำนวนแตกต่างกัน แต่เมื่อคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างวิธีทั้งสามพบว่ามีค่าสูงมากและมีนัยสำคัญทางสถิติเกือบทุกเงื่อนไขของการศึกษา สำหรับความสอดคล้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบส่วนใหญ่จะมีค่าปานกลางถึงต่ำในเกือบทุกเงื่อนไขของการศึกษา

วลีมาศ แซ่อึ้ง (2543) เปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอนุกรมระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทิลและแฮนส์เซล และวิธีการถดถอยโลจิสติก โดยการศึกษาจำลองด้วยโมเดลโลจิสติกชนิด 3 พารามิเตอร์ในเงื่อนไขต่าง ๆ 324 เงื่อนไข ผลการศึกษาพบว่าอำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าเท่าเทียมกันเกือบทุกเงื่อนไข และทั้งสองวิธีดังกล่าวมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์และวิธีแมนเทิลและแฮนส์เซลเกือบทุกเงื่อนไข อัตราความคลาดเคลื่อนประเภทที่ 1 ที่ตรวจสอบด้วยวิธีการทั้ง 4 ข้างต้นมีค่าอยู่ภายในเกณฑ์ของอัตราความคลาดเคลื่อนที่ระดับ 10 % เกือบทุกเงื่อนไข

สำหรับการศึกษาค้นคว้าเกี่ยวกับการทำหน้าที่เบี่ยงเบนของข้อสอบเพื่อทำวิทยานิพนธ์ระดับบัณฑิตศึกษาในประเทศไทยสรุปโดยสังเขปได้ดังนี้

ตาราง 2 สรุปการศึกษาเกี่ยวกับการทำหน้าที่เบี่ยงเบนของข้อสอบในประเทศไทย

ลำดับที่	ผู้วิจัย	ปีการศึกษา	วิธีการตรวจสอบที่ใช้
1	ซัชชัย เผ่าพงศ์	2527	ทฤษฎีการตอบข้อสอบชนิด 3 พารามิเตอร์
2	ทัศนีย์ พิรมนตรี	2530	1. กำหนดจุดเดสตั 2. ทดสอบความแตกต่างระหว่างกลุ่มด้วย χ^2 ในแบบจำลองลอกลิเนียร์ 3. ทฤษฎีการตอบข้อสอบชนิด 3 พารามิเตอร์
3	พัชรีย์ ปิยภักดิ์	2531	1. แปลงค่าความยาก 2. χ^2 3. ทฤษฎีการตอบข้อสอบชนิด 1 พารามิเตอร์
4	สุศักดิ์ อมรรัตนศักดิ์	2531	1. การวิเคราะห์ความแปรปรวน 2. แปลงค่าความยากของข้อสอบ 3. ทฤษฎีการตอบข้อสอบชนิด 1 และ 3 พารามิเตอร์
5	สุพัฒน์ สุขกุลสันต์	2534	1. กำหนดจุดเดสตั 2. $\chi^2 - 3$ 3. ทฤษฎีการตอบข้อสอบชนิด 3 พารามิเตอร์
6	นิรมล ชัยชวลิต	2537	1. แปลงค่าความยาก 2. χ^2 3. การวิเคราะห์ความแปรปรวน
7	กาญจนา วัฒนสุนทร	2538	1. การตอบข้อสอบชนิดคิดพื้นที่ (SA) 2. การตอบข้อสอบชนิดไม่คิดพื้นที่ (UA) 3. MH 4. SIBTEST
8	จิตติมา วรณศรี	2539	MH, SIBTEST
9	เสรี ชัดแจ้ง	2539	1. MH แบบปกติ 2. MH แบบแบ่งกลุ่มความสามารถและความยากของข้อสอบ
10	ญาณภัทร สีหะมงคล	2540	1. Lord' s χ^2 2. Raju' s Area Measure 3. Closed Interval Area
11	เรวดี อินทะสระ	2540	1. IRT 2. MH 3. SIBTEST

ตาราง 2 (ต่อ)

ลำดับที่	ผู้วิจัย	ปีการศึกษา	วิธีการตรวจสอบที่ใช้
12	พรรณี จิตมาศ	2540	1. แปลงค่าความยาก 2. MH 3. SIBTEST
13	นพมาศ พิพัฒน์สุข	2541	1. MH 2. การถดถอยโลจิสติก
14	อารี วัชรโสติกุล	2543	1. SIBTEST 2. การถดถอยโลจิสติก
15	ทองอยู่ สาระ	2543	1. MH 2. การถดถอยโลจิสติก
16	วลีมาศ แซ่อึ้ง	2543	1. MH 2. SIBTEST และ Modified SIBTEST 3. การถดถอยโลจิสติก

งานวิจัยเกี่ยวกับการทำหน้าที่เบี่ยงเบนของข้อสอบในประเทศไทยในตารางข้างต้น ดำเนินการโดยอาศัยกรอบความคิดของความเป็นมิติเดียวของแบบทดสอบ ตัวแปรที่ศึกษาได้แก่ขนาดของแบบทดสอบ ขนาดของกลุ่มตัวอย่าง อัตราส่วนของกลุ่มอ้างอิงและกลุ่มสนใจ อัตราส่วนของข้อสอบที่ทำหน้าที่เบี่ยงเบนในแบบทดสอบ วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ตัวแปรแบ่งกลุ่มประชากรส่วนมากแบ่งโดยใช้เพศ นอกจากนี้มีการแบ่งตามภูมิสำเนา ประสบการณ์ในการสอบ สังกัดของสถานศึกษา เป็นต้น ผลจากการตรวจสอบไม่พบความสอดคล้องถึงร้อยละ 100 ของวิธีการตรวจสอบแต่ละวิธีกับวิธีการอื่น ๆ การศึกษาค้นคว้าเกี่ยวกับการทำหน้าที่เบี่ยงเบนของข้อสอบยังคงเป็นประเด็นที่น่าสนใจศึกษาค้นคว้าในแง่มุมอื่น ๆ อีกต่อไป

4. การทำหน้าที่เบี่ยงเบนตามแนวคิดของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ

เป้าหมายสำคัญของการสร้างแบบทดสอบมาตรฐานส่วนใหญ่คือการจำแนกระดับความสามารถของผู้ตอบข้อสอบแล้วนำมาเรียงลำดับตามระดับความสามารถของผู้ตอบข้อสอบ การเรียงลำดับจะมีความถูกต้องถ้าข้อสอบทุกข้อในแบบทดสอบวัดความสามารถตามวัตถุประสงค์ของแบบทดสอบอย่างแท้จริง ปัญหาเกิดขึ้นเมื่อแบบทดสอบประกอบด้วยข้อสอบที่วัดความสามารถแตกต่างกันออกไป หรือวัดความสามารถหลายอย่างที่แตกต่างกัน นอกจากนั้นเป็นการเรียงลำดับความสามารถตามแนวคิดความเป็นมิติเดียว จึงไม่สามารถเรียงลำดับความสามารถของผู้ตอบข้อสอบตั้งแต่ 2 มิติขึ้นไปในขณะเดียวกันได้อย่างถูกต้อง (Ackerman. 1992 : 67) ในการสร้างแบบทดสอบจึงต้องศึกษาอย่างรอบคอบ เนื่องจากมีสมมุติฐานว่าแบบทดสอบมาตรฐานที่วัดผลสัมฤทธิ์ทางการเรียนส่วนใหญ่เป็นแบบทดสอบที่ใช้ความสามารถหลายมิติ

(Ackerman. 1994 : 105 ; citing Traub. 1983, Harisson. 1986) ซึ่งสอดคล้องกับความเห็นของเร็คเคส (Reckase. 1997 : 278) ที่กล่าวว่าแบบทดสอบวัดผลสัมฤทธิ์ทางวิชาการส่วนใหญ่ต้องการทักษะที่ใช้ความสามารถหลายมิติในการตอบข้อสอบให้ถูกต้อง และความสามารถดังกล่าวแปรเปลี่ยนไปข้อต่อข้อ หากสมมุติฐานดังกล่าวเป็นจริงและมีการรายงานผลสอบโดยใช้คะแนนเดียว (single score) แสดงว่าอาจมีการแสดงการทำหน้าที่เบี่ยงเบนข้อสอบ (Shealy & Stout. 1992, 1993 ; Ackerman. 1992) เพราะการเรียงลำดับความสามารถของผู้ตอบข้อสอบ 2 กลุ่มอาจไม่ตรงตามความเป็นจริง จึงมีการศึกษาค้นคว้าเพื่ออธิบายและตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยอาศัยแนวคิดของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ ในที่นี้จะนำเสนอในประเด็นต่อไปนี้

- 1) กรอบความคิดเชิงทฤษฎีของการทำหน้าที่เบี่ยงเบนของข้อสอบตามแนวคิดของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ
- 2) ขั้นตอนในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบตามแนวคิดของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ
- 3) วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบตามแนวคิดของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ
- 4) งานวิจัยเกี่ยวกับการตรวจสอบการทำหน้าที่เบี่ยงเบนตามแนวคิดของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ

4.1 กรอบความคิดเชิงทฤษฎีของการทำหน้าที่เบี่ยงเบนของข้อสอบตามแนวคิดของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ

ตามแนวคิดของทฤษฎีการตอบข้อสอบที่ใช้ข้อตกลงเบื้องต้นของความเป็นมิติเดียวของแบบทดสอบ ได้นิยามการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) ว่าหมายถึงข้อสอบที่บุคคลซึ่งมีความสามารถเท่ากันแต่เป็นสมาชิกของกลุ่มที่แตกต่างกันมีความน่าจะเป็นในการตอบข้อสอบข้อนั้นได้ถูกต้องไม่เท่ากัน (Hambleton. 1991 : 110)

การทำหน้าที่เบี่ยงเบนของข้อสอบตามแนวคิดของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติอธิบายได้ดังนี้

มิติ (dimension) หมายถึงคุณลักษณะใด ๆ ที่มีอิทธิพลต่อความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง (Shealy & Stout. 1996 : 356) โครงสร้างหลักที่แบบทดสอบต้องการวัดเรียกว่า มิติหลัก (primary dimension) ใช้สัญลักษณ์ θ ส่วนมิติที่แบบทดสอบไม่ได้ตั้งใจวัดเรียกว่า มิติรอง (secondary dimension) แบ่งเป็นมิติรองที่มีส่วนสนับสนุนมิติหลัก เรียกว่า auxillary dimension และมิติรองที่ไม่เกี่ยวข้องกับมิติหลัก เรียกว่า nuisance dimension ใช้สัญลักษณ์ η เป็นที่ยอมรับกันว่าสาเหตุที่ทำให้ข้อสอบแสดงการทำหน้าที่เบี่ยงเบนนั้นเป็นเพราะข้อสอบวัดหลายมิติโดยวัดมิติรองอย่างน้อย 1 มิติ

ธรรมชาติของการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติ มีการศึกษากันมานานแล้ว แต่เพิ่งมีการพัฒนาแบบจำลองเชิงคณิตศาสตร์อย่างเป็นทางการ โดยซีลลีและสตาต์ (Shealy & Stout) ประมาณ ค.ศ. 1993 อธิบายว่าการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติมีองค์ประกอบ 2 ประการคือ 1) ข้อสอบวัดมิติรอง 2) มิติรองมีทั้งส่วนที่ช่วยให้ผู้ตอบข้อสอบทั้งกลุ่มอ้างอิงและกลุ่มสนใจตอบข้อสอบได้ถูกต้องมากขึ้นและทำให้ตอบข้อสอบถูกต้องน้อยลง การที่มีมิติรองอำนาจประโยชน์ต่อกลุ่มอ้างอิงหรือกลุ่มสนใจนั้น ในกรณีที่ข้อสอบประกอบด้วยเนื้อหาที่กลุ่มใดกลุ่มหนึ่งมีความรู้

หรือคุ้นเคยมากกว่า เช่น ข้อสอบวัดความคิดเชิงตรรกศาสตร์ที่มีบริบทเกี่ยวกับการเล่นฟุตบอล อาจทำให้เพศชายซึ่งเป็นกลุ่มอ้างอิงได้รับประโยชน์มากกว่าเพศหญิงซึ่งเป็นกลุ่มสนใจ เนื่องจากเพศชายมีความคุ้นเคยและมีความรู้เกี่ยวกับฟุตบอลมากกว่าเพศหญิง ในกรณีที่อาจทำให้เกิดข้อจำกัดหรือไม่เป็นผลดีต่อกลุ่มใดกลุ่มหนึ่ง เช่น สถานการณ์ในข้อสอบวิชาฟิสิกส์กำหนดว่า ถ้าท่านอยู่ในแถวหน้าของผู้เดินขบวนเรียกร้องสิทธิมนุษยชน เมื่อตำรวจตัดสินใจสลายฝูงชนโดยใช้แรงน้ำ ถ้าหัวบีบฉีดน้ำมีแรงดัน 1,000 ลิตร/นาที่ แรงน้ำที่กระทำต่อท่านเป็นเท่าไร เงื่อนไขในสถานการณ์ดังกล่าวอาจทำให้ผู้ตอบข้อสอบที่มีอารมณ์รุนแรง รู้สึกหงุดหงิด และไม่มีสมาธิทำข้อสอบเท่าที่ควรจะเป็น

4.1.1 แบบจำลองการทำหน้าที่เบี่ยงเบนของข้อสอบของซีลลีและสเตาต์ (Shealy and Stout's Multidimensional Model for DIF)

เป็นแบบจำลองที่ได้มีการพัฒนาต่อเนื่องกันมา ซึ่งซีลลีและสเตาต์ (Shealy & Stout. 1993) ได้พัฒนาต่อจากคอค (Kok. 1988) กล่าวว่าการทำหน้าที่เบี่ยงเบนของข้อสอบเกิดจากองค์ประกอบ 2 ประการ คือ 1) ข้อสอบมีความไวทั้งกับคุณลักษณะที่ต้องการวัดโดยตรง หรือมิติหลัก (primary dimension : θ) และคุณลักษณะที่ไม่ต้องการวัดโดยตรง หรือมิติรอง (secondary dimension : η) โดยทั่วไปจัดเป็นข้อสอบที่วัดความสามารถหลายมิติ (multidimensional) 2) กลุ่มอ้างอิงและกลุ่มสนใจมีการแจกแจงตามเงื่อนไขของ η แตกต่างกันเมื่อกำหนดให้ θ คงที่ ($\eta | \theta$) ถ้ามีองค์ประกอบทั้ง 2 ประการข้อสอบจึงจะแสดงการทำหน้าที่เบี่ยงเบน ดังแสดงในตาราง

ตาราง 3 แสดงอิทธิพลขององค์ประกอบทั้ง 2 ประการต่อการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) ตามแนวคิดของซีลลีและสเตาต์

Conditional Probability Density function	Probability Correct	
	$P(\theta, \eta)$	$P(\theta)$
$f_R(\eta \theta) \neq f_F(\eta \theta)$	DIF	No DIF
$f_R(\eta \theta) = f_F(\eta \theta)$	No DIF	No DIF

อ้างอิงจาก Roussos & Stout. 1996 : 358

จากตารางแสดงให้เห็นว่าข้อสอบจะแสดงการทำหน้าที่เบี่ยงเบน (DIF) ในกรณีที่ฟังก์ชันความน่าจะเป็นตามเงื่อนไขเมื่อกำหนดให้ θ มีค่าคงที่ของกลุ่มอ้างอิงและกลุ่มสนใจไม่เท่ากัน และในการตอบคำถามได้ถูกต้องต้องใช้ความสามารถทั้ง θ และ η

ฟังก์ชัน $p(\theta, \eta)$ ไม่ได้ขึ้นอยู่กับการเป็นสมาชิกของกลุ่ม การเป็นสมาชิกกลุ่มใดกลุ่มหนึ่งจึงไม่ได้เป็นตัวพยากรณ์ผลการสอบอีกต่อไป ดังนั้นข้อสอบจะแสดง DIF ในกรณีที่กลุ่มสนใจและกลุ่มอ้างอิงถูกปรับเทียบ (equated) บนมิติที่ต้องการวัด (θ) แต่มิติที่ไม่ต้องการวัด (η) มีการแจกแจงแตกต่างกัน และฟังก์ชันการตอบข้อสอบ (IRF) มีความไวต่อมิติที่ไม่ต้องการวัด (η) ถ้าข้อสอบมีความไวต่อ η แต่การแจก

แจกแจงความสามารถอย่างมีเงื่อนไขของ η ของกลุ่มอ้างอิงและกลุ่มสนใจไม่แตกต่างกัน ข้อสอบจะไม่แสดง DIF ถ้าผู้ตอบข้อสอบทั้งสองกลุ่มมีการแจกแจงความสามารถอย่างมีเงื่อนไขของ η แตกต่างกัน แต่ข้อสอบไม่มีความไวต่อ η แม้ว่าข้อสอบจะวัด η ข้อสอบข้อนั้นก็ไม่ได้แสดง DIF เช่นกัน ตัวอย่างสถานการณ์โจทย์ข้อสอบการวัดเหตุผลเชิงตรรกศาสตร์ กล่าวว่า “ถ้าแมลงเป็นผึ้งมันจะต่อยได้ 1 ครั้งเท่านั้น มันต่อยได้ 1 ครั้ง ดังนั้นมันคือผึ้ง” แล้วให้ผู้ตอบข้อสอบหาข้อความที่มีแบบแผนการให้เหตุผลเหมือนโจทย์ ข้อสอบข้อนี้ η คือ กวีวิทยาหรือความรู้เกี่ยวกับผึ้ง แต่ข้อความในโจทย์ไม่ช่วยให้ผู้ที่มีความรู้ในเรื่องดังกล่าวได้ประโยชน์แต่อย่างใด โจทย์ข้อนี้อาจสรุปได้ว่า $p(\theta, \eta)$ มีค่าใกล้เคียง $p(\theta)$ เป็นไปได้ว่าทั้ง 2 กลุ่มมีการแจกแจงของ η เหมือนกัน แต่มี $\eta|\theta$ แตกต่างกัน หรือมีการแจกแจง η แตกต่างกัน แต่มี $\eta|\theta$ เหมือนกัน

4.1.2 การใช้ทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติทำความเข้าใจกับการแสดงการทำหน้าที่เบี่ยงเบนของข้อสอบ

ตามที่กล่าวมาข้างต้นว่าข้อสอบที่มีศักยภาพจะแสดง DIF นั้นมีสมมุติฐานว่ามีอิทธิพลจากความสามารถ 2 มิติ คือ θ และ η ซึ่งสามารถแปรเปลี่ยนจากข้อหนึ่งไปยังอีกข้อหนึ่ง และ joint distribution ของเวกเตอร์ของความสามารถ (θ, η) สมมุติให้เป็น bivariate normal distribution ที่มีค่าเฉลี่ยเท่ากับ μ_θ และ μ_η ส่วนเบี่ยงเบนมาตรฐานเท่ากับ σ_θ และ σ_η สหสัมพันธ์เท่ากับ ρ พารามิเตอร์ทั้งหมดของ bivariate distribution ของกลุ่มอ้างอิงและกลุ่มสนใจแตกต่างกัน

จากข้อตกลงเบื้องต้นดังกล่าว ค่าคาดหวังของความแตกต่างของค่าเฉลี่ยของ η (expected difference in the means of η) สำหรับผู้ตอบข้อสอบที่กำหนดให้ค่า θ คงที่ทั้ง 2 กลุ่ม สามารถเขียนได้ดังนี้ (Roussos & Stout. 1996 : 361 citing Johnson & Wichern. 1992)

$$E_R(\eta|\theta) - E_F(\eta|\theta) = (\mu_{\eta_R} - \mu_{\eta_F}) + \theta(\rho_R \frac{\sigma_{\eta_R}}{\sigma_{\theta_R}} - \rho_F \frac{\sigma_{\eta_F}}{\sigma_{\theta_F}}) + (\mu_{\theta_F} \rho_F \frac{\sigma_{\eta_F}}{\sigma_{\theta_F}} - \mu_{\theta_R} \rho_R \frac{\sigma_{\eta_R}}{\sigma_{\theta_R}}) \quad (1)$$

ตามทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ การแสดง DIF แสดงผ่าน marginalized IRF [$P_R(\theta)$ และ $P_F(\theta)$]

$$P_R(\theta) = \int P(\theta, \eta) f_R(\eta|\theta) d\eta \quad (2)$$

$$P_F(\theta) = \int P(\theta, \eta) f_F(\eta|\theta) d\eta \quad (3)$$

ความน่าจะเป็นในการตอบคำถามได้ถูกต้องเป็นฟังก์ชันของ θ ซึ่งได้มาโดยค่าเฉลี่ยของฟังก์ชันการตอบข้อสอบ $P(\theta, \eta)$ สำหรับแต่ละค่า θ ที่คงที่ ระดับของ DIF ที่แต่ละ θ เกิดจากความแตกต่างของการแจกแจงอย่างมีเงื่อนไขของ $\eta|\theta$ ของกลุ่มอ้างอิงและกลุ่มสนใจ จะเห็นได้ว่าสำหรับค่าคงที่ θ ใด ๆ $P_R(\theta)$ อาจไม่เท่ากับ $P_F(\theta)$ เมื่อการแจกแจงอย่างมีเงื่อนไข $f_R(\eta|\theta) \neq f_F(\eta|\theta)$ ในการตรวจสอบ bivariate normal model สำหรับ joint distribution ของ η และ θ จะมีประโยชน์ในการพิจารณากรณีเฉพาะกรณีที่ทั้งกลุ่มอ้างอิงและกลุ่มสนใจจะมีส่วนเบี่ยงเบนมาตรฐานเท่ากัน ($\sigma_{\theta_R} = \sigma_{\theta_F}$ และ

สัมประสิทธิ์สหสัมพันธ์เท่ากัน $\rho_R = \rho_F$ และ $\sigma_\eta = \sigma_\theta$ (กรณีที่เป็น totally nonrestrictive assumption) แล้วทั้ง 2 กลุ่มแตกต่างกันเฉพาะค่าเฉลี่ยเท่านั้น (μ_η และ μ_θ) กรณีเฉพาะดังกล่าว สมการ (1) เขียนใหม่ได้ดังนี้

$$E_R(\eta|\theta) - E_F(\eta|\theta) = (\mu_{\eta_R} - \mu_{\eta_F}) - \rho(\mu_{\theta_R} - \mu_{\theta_F})$$

ความแตกต่างของค่าเฉลี่ยอย่างมีเงื่อนไข (conditional means) ในสมการที่ (4) ประกอบด้วยการสนทนาคู่ทั้งหมดที่เป็นผลจากความแตกต่างระหว่าง conditional densities f_R และ f_F ตัวอย่างส่วนใหญ่เน้นที่ความแตกต่างของค่าเฉลี่ยของ $\eta|\theta$ ของกลุ่มอ้างอิงและกลุ่มสนใจเพื่อแสดงให้เห็นว่ามีมิติที่แบบทดสอบไม่ได้ต้องการวัด (η) ทำให้เกิด DIF ได้อย่างไร ในกรณีเฉพาะที่แสดงด้วยสมการ (4) ความแตกต่างดังกล่าวเป็นความแตกต่างของค่าเฉลี่ยไม่มีเงื่อนไข (unconditional mean) ของ η ลบด้วยค่าเฉลี่ยของ θ ซึ่งเป็นกรณีที่ถูกกำหนดน้ำหนักโดยสัมประสิทธิ์สหสัมพันธ์ของคุณลักษณะภายใน 2 มิติ และในกรณีของความสามารถทางสมองเป็นกรณีที่มีสัมประสิทธิ์สหสัมพันธ์เป็นบวก ($\rho > 0$) ซึ่งเป็นบทบาทสำคัญในการวิเคราะห์ DIF

4.1.3 สาเหตุที่ข้อสอบแสดงการทำหน้าที่เบี่ยงเบน

รูสโซและสตาต์ (Roussos & Stout. 1996) กล่าวถึงสาเหตุที่อาจทำให้ข้อสอบแสดงการทำหน้าที่เบี่ยงเบนมีดังนี้

กรณีที่ 1 ข้อสอบแสดง DIF เมื่อมีความไวต่อมิติที่ข้อสอบไม่ต้องการวัด (η) และผู้ตอบข้อสอบมีความแตกต่างของการแจกแจงอย่างมีเงื่อนไขของ η [$f_R(\eta|\theta) \neq f_F(\eta|\theta)$]

กรณีที่ 2 ข้อสอบแสดง DIF แม้ว่า $\mu_{\eta_R} \approx \mu_{\eta_F}$ มิติที่ข้อสอบไม่ต้องการวัด (η) ที่ต่างกันเพียงเล็กน้อยอาจทำให้ข้อสอบแสดง DIF ได้ ถ้ามีสหสัมพันธ์กับมิติที่ข้อสอบต้องการวัด (θ) และทั้ง 2 กลุ่มมีค่าเฉลี่ยของ θ แตกต่างกัน

กรณีที่ 3 กรณีที่ข้อสอบไม่แสดง DIF เป็นกรณีที่ข้อสอบไม่มีความไวต่อ η ฟังก์ชันการตอบข้อสอบ (IRF) ขึ้นอยู่กับ θ เท่านั้น

กรณีที่ 4 กรณีที่ข้อสอบไม่แสดง DIF เป็นกรณีที่ข้อสอบมีความไวต่อ η แต่ทั้ง 2 กลุ่มมีการแจกแจงของ η ไม่แตกต่างกันเมื่อได้รับการปรับเทียบบน θ

กรณีที่ 5 กรณีที่ข้อสอบแสดงการทำหน้าที่เบี่ยงเบนแบบไม่สม่ำเสมอ ได้แก่กรณีที่ $\rho_R = \rho_F$ แต่ $\sigma_{\theta_F} < \sigma_{\theta_R}$ รวมกับ $\sigma_{\eta_F} > \sigma_{\eta_R}$ เป็นต้น

4.2 ขั้นตอนในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติ

4.2.1 วิธีที่เสนอโดยเอ็คเคอร์แมน (Ackerman. 1994)

เอ็คเคอร์แมน (Ackerman. 1994) เสนอขั้นตอนในการตรวจสอบ DIF ดังนี้

1) ศึกษาทำความเข้าใจตารางกำหนดข้อสอบที่ผู้สร้างข้อสอบกำหนดไว้ ตารางดังกล่าวสร้างขึ้นโดยคณะผู้เชี่ยวชาญ ระบุเนื้อหาและคุณลักษณะที่ต้องการวัด

2) ตรวจสอบข้อมูลการตอบข้อสอบ โดยมีสมมุติฐานเบื้องต้นว่าข้อสอบทุกข้อที่สร้างตามตารางกำหนดข้อสอบมีความเที่ยงตรงเชิงเนื้อหา จึงต้องตรวจสอบความเที่ยงตรงเชิงสถิติ (statistically valid) ด้วย เพื่อตรวจสอบว่าปฏิกริยาร่วมระหว่างข้อสอบและผู้ตอบข้อสอบเป็นสิ่งที่ต้องการวัดจริง ในกรณีที่ใช้คะแนนรวมต้องตรวจสอบมิติของข้อมูลการตอบข้อสอบ อาจใช้วิธีการของสเตาต์ (Stout, 1990) หรือใช้วิธีการวิเคราะห์หองค์ประกอบ (factor analysis)

3) ประเมินค่าพารามิเตอร์ข้อสอบ ของแต่ละกลุ่ม อาจใช้โปรแกรมสำเร็จรูป NOHARM (Fraser & McDonald, 1988) ประเมินค่าพารามิเตอร์ข้อสอบที่เหมาะสมกับข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติชนิด 2 พารามิเตอร์ (M2PL) หลังจากประเมินค่าพารามิเตอร์ข้อสอบแล้วก็ประมาณค่าความสามารถหลายมิติโดยวิธีการ Newton–Raphson iteration ประมาณค่าความสามารถของแต่ละกลุ่มแยกจากกัน หลังจากนั้นเขียนเวกเตอร์ของข้อสอบ โดยใช้วิธีการของเร็คเคส (Reckase's formulation) ทำให้มองเห็นภาพการจับกลุ่มข้อสอบ ถ้าจับกลุ่มอยู่ในบริเวณแคบ ๆ ก็อาจจะประมาณได้ว่าข้อสอบมีลักษณะเป็นเอกพันธ์ มีความแตกต่างกันเฉพาะค่าความยาก ควรตรวจสอบ DIF รายข้อ แต่ถ้าข้อสอบมีการกระจายในลักษณะกลุ่ม (cluster) ควรตรวจสอบ DIF แบบเป็นกลุ่มข้อสอบจะได้ผลดีกว่า กลุ่มข้อสอบที่มีขนาดใหญ่ที่สุด พิจารณาได้ว่าเป็นกลุ่มข้อสอบที่มีความเที่ยงตรง (validity sector) เวกเตอร์ข้อสอบที่อยู่นอกกลุ่มนี้ควรตั้งข้อสังเกตว่าอาจจะแสดง DIF

4) ตรวจสอบข้อสอบในกลุ่มที่สงสัยว่าจะแสดง DIF ด้วยวิธีการต่าง ๆ ที่เหมาะสมต่อไป เช่น SIBTEST เป็นต้น

4.2.2 วิธีที่เสนอโดยรูสโซและสเตาต์ (Roussos & Stout, 1996)

รูสโซและสเตาต์ (Roussos & Stout, 1996) เสนอวิธีการตรวจสอบ DIF ที่เรียกว่า The Dimensionality Based Approach มีขั้นตอนดังนี้

1) ตรวจสอบโครงสร้างมิติ (dimensionality structure) ของแบบทดสอบก่อนการวิเคราะห์ DIF

2) ใช้กลุ่มข้อสอบที่วัด θ เป็นเกณฑ์การจับคู่ (matching criterion) หรือมีจะนั้นก็ใช้ข้อสอบที่วัด η ที่ไม่แสดงการทำหน้าที่เบี่ยงเบน (DIF-free items) ในกรณีที่ต้องใช้คะแนน 2 ค่าเป็นเกณฑ์การจับคู่ ควรตรวจสอบ DIF โดยใช้โปรแกรม SIBTEST หรือวิธีถดถอยโลจิสติก (Mazor, et. al. 1995) เพราะยินยอมให้ใช้การจับคู่จากคะแนนหลายค่า (multiple scoring matching)

3) จำแนกมิติรองที่มีส่วนสนับสนุน θ ที่เรียกว่า auxilliary dimension และมิติรองที่ไม่มีส่วนสนับสนุน θ ที่เรียกว่า nuisance dimension

ข้อดีของวิธีการที่เสนอโดยรูสโซและสเตาต์

1. ตรวจสอบข้อสอบมากกว่า 1 ข้อในขณะเดียวกันทำให้อำนาจในการตรวจสอบ DIF สูงขึ้นเมื่อข้อสอบวัดมิติรอง

2. ช่วยให้ทำความเข้าใจสาเหตุของการเกิด DIF ได้ดียิ่งขึ้น จึงสามารถวางแผนการสร้างแบบทดสอบได้อย่างรัดกุม และทำความเข้าใจกับความแตกต่างของความรู้ของผู้ตอบข้อสอบกลุ่มต่าง ๆ ได้

4.3 วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติ

ในที่นี้จะนำเสนอ 2 วิธี คือ SIBTEST และ DFIT เนื่องจากทั้ง 2 วิธีอาศัยกรอบความคิดของทฤษฎีการตอบข้อสอบ (IRT) ที่สามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบรายข้อ หมวดข้อสอบ และแบบทดสอบทั้งฉบับ นอกจากนั้นยังสามารถตรวจสอบการตอบข้อสอบที่มีข้อตกลงของความเป็นมิติเดียวและการใช้ความสามารถหลายมิติที่มีการตรวจให้คะแนนรายข้อแบบสองค่าและแบบหลายค่า แตกต่างกันที่ SIBTEST ใช้วิธีการของสถิตินอนพาราเมตริก แต่ DFIT ใช้วิธีการของสถิติพาราเมตริก

4.3.1 การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธี SIBTEST

SIBTEST (Simutaneous Items Bias TEST) เป็นวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่พัฒนาโดยซีลลีและสแตต์ (Shealy & Stout. 1992) ตรวจได้ทั้งการทำหน้าที่เบี่ยงเบนอย่างสม่ำเสมอและการทำหน้าที่เบี่ยงเบนอย่างไม่สม่ำเสมอ กรณีของการตรวจสอบการทำหน้าที่เบี่ยงเบนอย่างสม่ำเสมอใช้ β_{uni} เป็นดัชนีแสดงการทำหน้าที่เบี่ยงเบนของข้อสอบ ในบริบทของ IRT เป็น vertical distance ระหว่าง the marginal ICC_S ของข้อสอบที่กำลังศึกษา (สอดคล้องกับ θ ที่เป็น valid subtest ability) ดัชนีดังกล่าวบอกให้รู้ว่าข้อสอบข้อนั้นมีการขจัดความแตกต่างที่เกิดจาก θ ของกลุ่มที่ศึกษา (item impact) แล้ว β_{uni} จึงถือว่าเป็นดัชนีมาตรฐาน (standardized index) (Ackerman. 1992 : 77 ; citing Dorans & Kulick. 1986) ซึ่งคำนวณได้ดังนี้

$$\beta_{uni} = \int_{-\infty}^{+\infty} [T_R(\theta) - T_F(\theta)] f_F(\theta) d\theta \quad (1)$$

เมื่อ $T_R(\theta)$ และ $T_F(\theta)$ เป็น marginal ICC_S ของข้อสอบที่สงสัยสำหรับกลุ่มอ้างอิงและกลุ่มสนใจ กำหนดโดยสมการต่อไปนี้

$$P(X_i = 1 | \theta = \theta) = \int P_i[\theta, \eta] f(\eta | \theta) d\eta \quad (2)$$

$f_F(\theta)$ คือ θ - marginal density ของกลุ่มสนใจ

$P(X_i = 1 | \theta = \theta)$ หาได้จากค่าเฉลี่ยของ $P_i[\theta, \eta]$ เหนือ η (averaging $P_i[\theta, \eta]$ over η)

แนวคิดของซีลลีและสแตต์ (Shealy & Stout) นั้นการตรวจสอบ DIF เป็นการตรวจสอบความแตกต่างระหว่างค่าที่คาดหวังของการแจกแจง $\eta|\theta$ ของกลุ่มอ้างอิงและกลุ่มสนใจ ถ้าความแตกต่างนี้ไม่เท่ากับ 0 ที่ทุก ๆ θ แสดงว่าข้อสอบนั้นอาจแสดง DIF ความแตกต่างของค่าที่คาดหวังของการแจกแจงอย่างมีเงื่อนไข ณ θ ที่กำหนด แสดงได้ดังนี้

$$\begin{aligned} \xi[\eta_R|\theta] - \xi[\eta_F|\theta] &= (\mu_{\eta_R} - \mu_{\eta_F}) + (\rho_R \frac{\sigma_{\eta_R}}{\sigma_{\theta_R}}) - \\ &\quad (\rho_F \frac{\sigma_{\eta_F}}{\sigma_{\theta_F}}) (\theta - \mu_{\theta_F}) \end{aligned} \quad (3)$$

จากสมการที่ (3) แสดงให้เห็นว่า DIF อาจเกิดขึ้นได้เมื่อการแจกแจงความสามารถของกลุ่มที่ศึกษาไม่เหมือนกัน

จะเห็นว่า SIBTEST พัฒนามาจากแนวคิดของการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบบนพื้นฐานของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ โดยมีข้อตกลงว่ามีมิติการวัด 2 มิติ คือ มิติที่ต้องการวัดหรือมิติหลัก (primary dimension : θ) และมิติที่ไม่ได้ตั้งใจวัดหรือมิติรอง (secondary dimension : η) ในการตอบข้อสอบทุกข้อผู้ตอบข้อสอบต้องใช้ θ ขณะที่การตอบข้อสอบบางข้อผู้ตอบข้อสอบต้องใช้ทั้ง θ และ η

ในการตรวจสอบ DIF ด้วยวิธี SIBTEST นั้นจะแบ่งแบบทดสอบฉบับเดิมออกเป็น 2 ฉบับย่อย คือแบบทดสอบย่อยที่มีความเที่ยงตรง (valid subtest) และแบบทดสอบย่อยที่ต้องการศึกษา (studied subtest) หลังจากนั้นจึงเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจดังนี้

$$X = \sum_{i=1}^n U_i \quad \text{เมื่อ } X \text{ คือ คะแนนรวมจากแบบทดสอบที่มีความเที่ยงตรง}$$

U_i คือ ผลการตอบข้อสอบข้อที่ i (ตอบถูกได้ 1 ตอบผิดได้ 0)

$$Y = \sum_{i=n+1}^N U_i \quad \text{เมื่อ } Y \text{ คือ คะแนนรวมจากแบบทดสอบที่ต้องการศึกษา}$$

$\bar{Y}_{RK}$ และ $\bar{Y}_{FK}$ คือคะแนนเฉลี่ยที่ได้จากแบบทดสอบย่อยที่ต้องการศึกษาสำหรับผู้เข้าสอบทุกคนในกลุ่มอ้างอิงและกลุ่มสนใจที่มีคะแนนรวมจากการทำแบบทดสอบที่มีความเที่ยงตรง (X) เท่ากับ k การทำหน้าที่เบี่ยงเบนของข้อสอบวัดได้จาก $\bar{Y}_{RK} - \bar{Y}_{FK}$ เนื่องจากเป็นความแตกต่างของกลุ่มผู้ตอบข้อสอบในกลุ่มอ้างอิงและกลุ่มสนใจที่มีระดับความสามารถหลักที่ต้องการวัด (θ) เท่ากัน ในกรณีที่ข้อสอบไม่แสดงการทำหน้าที่เบี่ยงเบน $\bar{Y}_{RK} - \bar{Y}_{FK} = 0$

สมมุติฐานในการทดสอบ

$$H_0 : \beta_{uni} = 0$$

$$H_1 : \beta_{uni} > 0$$

ขั้นตอนในการวิเคราะห์

1. ประเมินค่า β_{uni} ซึ่งเป็นดัชนีบ่งชี้การทำหน้าที่ต่างกันของข้อสอบจากสูตร

$$\beta_{uni} = \sum_{k=0}^u P_K (\bar{Y}_{RK} - \bar{Y}_{FK})$$

เมื่อ P_K คือ สัดส่วนของผู้สอบกลุ่มเปรียบเทียบที่ตอบแบบทดสอบที่มีความเที่ยงตรงได้ถูกต้อง k ข้อ

2. ทดสอบนัยสำคัญทางสถิติ

$$B = \frac{\beta_{uni}}{\sigma(B_{uni})}$$

เมื่อ $\sigma(\beta_{uni})$ คือ ความคลาดเคลื่อนมาตรฐานในการประมาณค่า β_{uni}

$$\sigma(\beta_{uni}) = \left\{ \sum_{k=0}^n P_k^2 \left[\frac{1}{J_{R_K}} \sigma^2(Y/K, R) + \frac{1}{J_{F_K}} \sigma^2(Y/K, F) \right] \right\}^{1/2}$$

$\sigma^2(Y/K, R)$ และ $\sigma^2(Y/K, F)$ เป็นความแปรปรวนของคะแนนของผู้ตอบข้อสอบในกลุ่ม

อ้างอิงและกลุ่มสนใจที่ระดับคะแนน k

J_{R_K} และ J_{F_K} คือจำนวนผู้ตอบข้อสอบในกลุ่มอ้างอิงและกลุ่มสนใจที่มีคะแนนจาก

แบบทดสอบที่มีความเที่ยงตรง k คะแนน

การทดสอบนัยสำคัญของ β_{uni} จะปฏิเสธ H_0 ที่ว่า $\beta_{uni} = 0$ เมื่อ β_{uni} ที่คำนวณได้มีค่ามากกว่าค่า Z ที่กำหนดจาก $P[N(0, 1)]$ (Nandakumar. 1993)

4.3.2 การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธี DFIT

DFIT (Differential Functioning of Items and Tests) พัฒนาโดยราชู แวนเดอร์ลินเด็น และฟลีร์ (Raju, van der Linden & Fleer) เสนอเป็นครั้งแรกใน ค.ศ. 1992 และมีการพัฒนาต่อไปเรื่อย ๆ โดยใช้บริบทของทฤษฎีการตอบข้อสอบ มีสาระโดยสังเขปดังนี้

4.3.2.1 การทำหน้าที่เบี่ยงเบนของแบบทดสอบ (DTF)

จากแบบจำลองการตอบข้อสอบที่ใช้ความสามารถหลายมิติชนิด 2 พารามิเตอร์ (M2PL) ความน่าจะเป็นในการตอบคำถามข้อที่ i ได้ถูกต้อง กำหนดดังนี้

$$P_i(\theta) = \frac{1}{1 + e^{-1.7(a_i' \theta + b_i)}} \quad (1)$$

เมื่อ a_i คือ $m \times 1$ vector ของพารามิเตอร์อำนาจจำแนก

b_i คือ สเกลาร์พารามิเตอร์ (scalar parameter) เกี่ยวข้องกับความยากของข้อสอบ

θ คือ $m \times 1$ vector ของความสามารถของผู้ตอบข้อสอบ

m คือ จำนวนมิติของความสามารถ

ให้แบบทดสอบประกอบด้วยข้อสอบ k ข้อ และผู้ตอบข้อสอบจากกลุ่มอ้างอิงและกลุ่มสนใจมีพารามิเตอร์ข้อสอบกลุ่มละชุด สมมุติว่าอยู่บนสเกลร่วม (common scale)

ให้ $P_{iR}(\theta)$ แทนความน่าจะเป็นในการตอบข้อสอบข้อที่ i ได้ถูกต้องของผู้ตอบข้อสอบคนหนึ่งในกลุ่มอ้างอิง

$P_{iF}(\theta)$ แทนค่าความน่าจะเป็นในการตอบข้อสอบข้อที่ i ได้ถูกต้องของผู้ตอบข้อสอบคนเดียวกันเมื่อเป็นสมาชิกกลุ่มสนใจ

ในการคำนวณ $P_{iR}(\theta)$ และ $P_{iF}(\theta)$ ใช้พารามิเตอร์ (a_i, b_i) จากกลุ่มอ้างอิงและกลุ่มสนใจตามลำดับ ใช้เวกเตอร์ของ θ เดียวกันในการคำนวณทั้ง 2 ครั้ง ถ้าข้อสอบทำหน้าที่เบี่ยงเบนแล้ว P_{iR} และ $P_{iF}(\theta)$ ของผู้ตอบข้อสอบแต่ละคนจะแตกต่างกัน

ในบริบทของ IRT คะแนนจริงของผู้ตอบข้อสอบแต่ละคนกำหนดดังนี้

$$T = \sum_{i=1}^k P_i(\theta) \quad (2)$$

ในกระบวนการ DFIT ผู้ตอบข้อสอบแต่ละคนจะมีคะแนนจริง 2 ค่า คือ เมื่อเป็นสมาชิกของกลุ่มอ้างอิง (T_R) และเมื่อเป็นสมาชิกกลุ่มสนใจ (T_F) ถ้า T_R และ T_F ของผู้ตอบข้อสอบคนเดียวกันมีค่าเท่ากัน แสดงว่าคะแนนของผู้ตอบข้อสอบมีความเป็นอิสระ ไม่ได้ขึ้นอยู่กับความเป็นสมาชิกกลุ่มใดกลุ่มหนึ่ง ความแตกต่างของคะแนนจริงยิ่งมากเท่าไร แสดงว่ามีการทำหน้าที่เบี่ยงเบนของแบบทดสอบ DTF ที่วัดในระดับของผู้ตอบข้อสอบ (examinee level) นิยามว่ามีค่าเท่ากับ $(T_F - T_R)^2$ ให้ $D = T_F - T_R$ การวัด DTF จากผู้ตอบข้อสอบทุกคนกำหนดดังนี้

$$DTF = E_F (T_F - T_R)^2 = E_F D^2 = \sigma_D^2 + (\mu_{TF} - \mu_{TR})^2 = \sigma_D^2 + \mu_D^2 \quad (3)$$

ค่าที่คาดหวัง (Expectation : E) อาจได้จากกลุ่มอ้างอิงหรือกลุ่มสนใจ ในที่นี้สมมุติว่าได้จากกลุ่มสนใจ

4.3.2.2 การทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF)

ในที่นี้จะกล่าวถึงการทำหน้าที่เบี่ยงเบนของข้อสอบ 2 ประเภท คือ การทำหน้าที่เบี่ยงเบนของข้อสอบที่มีการชดเชย (Compensatory DIF : CDIF) ซึ่งมีคุณสมบัติของการบวก (additive) CDIF เชิงบวกของข้อหนึ่งอาจชดเชย CDIF เชิงลบของข้ออื่น ๆ ได้ และมีผลต่อ DTF ส่วนการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่มีการชดเชย (Noncompensatory DIF : NCDIF) กำหนดข้อตกลงว่าข้อสอบข้ออื่นในแบบทดสอบไม่แสดงการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF-free)

การทำหน้าที่เบี่ยงเบนของข้อสอบที่มีการชดเชย (CDIF)

จากสมการ (3) อาจเขียนใหม่ได้ดังนี้

$$DTF = E_F (D^2) = E_F \left[\sum_{i=1}^k (d_i D) \right] = \sum_{i=1}^k E_F (d_i D) = \sum_{i=1}^k [Cov(d_i, D) + \mu_{d_i} \mu_D] \quad (4)$$

$$\text{เมื่อ } d_i = P_{iF}(\theta) - P_{iR}(\theta), \quad \sum_{i=1}^k d_i = D = T_F - T_R$$

$Cov(d_i, D)$ คือ ความแปรปรวนร่วมระหว่างความแตกต่างของคะแนนจะเป็นในการตอบข้อสอบข้อที่ i ได้ถูกต้อง (d_i) และความแตกต่างของคะแนนจริง 2 ค่า

ค่าคาดหวังคำนวณจากกลุ่มสนใจ ดังนั้น

$$CDIF_i = E_F (d_i D) = Cov(d_i, D) + \mu_{d_i} \mu_D$$

รวมสมการ (4) + (5) ได้ดังนี้

$$DTF = \sum_{i=1}^k CDIF_i$$

สำหรับข้อสอบข้อที่ i

$$Cov(d_i, D) = \sigma_{d_i}^2 + \sum Cov(d_i, d_j), i \neq j$$

CDIF ของข้อสอบข้อที่ i รวมความสัมพันธ์ของ DIF ระหว่างข้อที่ i และข้ออื่น ๆ ในแบบทดสอบ สมการ (3) เขียนใหม่ได้ดังนี้

$$DTF = E_F \left[\sum_{i=1}^k (P_{iF} - P_{iR})^2 \right] = E_F \left[(P_{1F} - P_{1R}) + (P_{2F} - P_{2R}) + \dots + (P_{kF} - P_{kR}) \right]^2$$

การทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่มีการชดเชย (NCDIF)

การแสดง DIF ของข้อสอบ ในกรณีที่กำหนดให้ข้อสอบข้ออื่น ๆ ในแบบทดสอบไม่แสดง DIF ดังนั้น $d_j = 0$ สำหรับทุก $j \neq i$ ดังนั้นสมการ (5) เขียนได้ใหม่ดังนี้

$$NCDIF_i = \sigma_{d_i}^2 + \mu_{d_i}^2 \quad (8)$$

การทดสอบนัยสำคัญ

ในการทดสอบนัยสำคัญของ DTF, CDIF และ NCDIF ใช้ค่าประมาณของ θ และค่าประมาณของพารามิเตอร์ข้อสอบ ดังนี้

$$D = \sum_{i=1}^k d_i \quad (9)$$

$$d_i = P_{iF} - P_{iR} \quad (10)$$

$$DTF = \text{Mean of } D^2 = \sigma_D^2 + \mu_D^2 \quad (11)$$

$$CDIF_i = Cov(d_i, D) + \mu_{d_i} \mu_D \quad (12)$$

$$NCDIF_i = \sigma_{d_i}^2 + \mu_{d_i}^2$$

การทดสอบ χ^2 สำหรับ DTF

สมมุติว่า D มีการแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_D และส่วนเบี่ยงเบนมาตรฐานเท่ากับ σ_D คะแนน Z ของผู้ตอบข้อสอบ s หาได้ดังนี้

$$Z_s = \frac{D_s - \mu_D}{\sigma_D} \quad (14)$$

ดังนั้น Z_s^2 มีการแจกแจงแบบไคสแควร์ที่มีองศาแห่งความเป็นอิสระ (df) เท่ากับหนึ่ง ผลรวมของ Z_s^2 ของผู้ตอบข้อสอบ N_F คนในกลุ่มสนใจมีการแจกแจงแบบไคสแควร์ที่มี $df = N_F$ กำหนดดังนี้

$$\chi^2_{N_F} = \sum_{s=1}^{N_F} Z_s^2 = \frac{\sum_{s=1}^{N_F} (D_s - \mu_D)^2}{\sigma_D^2} \quad (15)$$

ในที่นี้สมมุติฐานว่างคือ

$$E(DTF) = \mu_D^2 = 0 \quad (16)$$

แทนค่า $\mu_D = 0$ ในสมการ (15)

$$\chi^2_{N_F} = \frac{\sum_{s=1}^{N_F} D_s^2}{\sigma_D^2} \quad (17)$$

$$\text{จากสมการ (3)} \quad \chi^2_{N_F} = \frac{N_F (DTF)}{\sigma_D^2}, \quad df = N_F - 1 \quad (18)$$

เมื่อใช้การประมาณค่าความแปรปรวนของ D จากกลุ่มตัวอย่าง

$$\chi^2_{N_F-1} = \frac{N_F (DTF)}{\sigma_D^2}, \quad df = N_F - 1 \quad (19)$$

เมื่อ DTF มีนัยสำคัญ ต้องเริ่มต้นค้นหาข้อสอบที่ทำให้ χ^2 มีนัยสำคัญ เมื่อตรวจพบและกำจัดออกไปจากแบบทดสอบแล้ว คำนวณ χ^2 โดยใช้ข้อสอบที่เหลือ แสดงว่าค่าของ $Cov(d_i, D)$ ขึ้นอยู่กับองค์ประกอบอื่น ๆ ได้แก่ จำนวนข้อสอบที่ยังเหลืออยู่ในแบบทดสอบ และควรกำจัดข้อสอบครั้งละ 1 ข้อ และคำนวณซ้ำจนกระทั่ง DTF ไม่มีนัยสำคัญ เมื่อเพิ่มข้อสอบที่มี CDIF เข้าไปในแบบทดสอบ เมื่อ DTF มีนัยสำคัญควรตัดข้อที่มี CDIF ขนาดใหญ่และมีค่าเป็นบวกครั้งละ 1 ข้อ จนกระทั่ง DTF ของข้อสอบที่เหลือไม่มีนัยสำคัญ ข้อสอบที่ถูกตัดออกไปแสดง CDIF

การทดสอบไคสแควร์สำหรับ NCDIF

กำหนดให้ d_i มีการแจกแจงแบบปกติ มี finite variance การทดสอบไคสแควร์สำหรับ NCDIF กำหนดดังนี้

$$\chi^2_{N_F-1} = \frac{N_F (NCDIF)}{\sigma_{d_i}^2}, \quad df = N_F - 1 \quad (20)$$

4.3.2.3 การทำหน้าที่ย่อยของหมวดข้อสอบ (DBF)

การวิเคราะห์ DBF เริ่มโดยคำนวณค่าพารามิเตอร์ข้อสอบทุกข้อในแบบทดสอบด้วยแบบจำลองของทฤษฎีการตอบข้อสอบที่เหมาะสมสำหรับ 2 กลุ่มที่ศึกษา แล้วหาสเกลรวมของพารามิเตอร์เหล่านั้นหลังจากนั้นจึงจัดหมวดข้อสอบ (bundle) ตามที่นิยามอย่างเหมาะสม หมวดข้อสอบมี 2 ประเภทคือ หมวดข้อสอบที่มีการชดเชย (bundle CDIF) เป็นผลรวมของ CDIF และผลรวมของ bundle CDIF ก็คือ DTF และหมวดข้อสอบที่ไม่มีการชดเชย (bundle NCDIF) ค่าของ bundle NCDIF ไม่ใช่ผลรวมของ NCDIF แต่จะกลุ่มทำการวิเคราะห์แยกจากกัน สรุปได้ดังตาราง 4

ตาราง 4 สรุปดัชนีการทำหน้าที่เบี่ยงเบนของข้อสอบ (CDIF, NCDIF) ของหมวดข้อสอบ (Bundle CDIF, Bundle NCDIF) และของแบบทดสอบ (DTF)

CDIF	NCDIF
- ไม่ได้กำหนดให้ข้อสอบข้ออื่น ๆ ในแบบทดสอบเป็น DIF free - มีคุณสมบัติของการบวก (Additive) - มีการชดเชย (Compensatory)	- กำหนดให้ข้อสอบข้ออื่น ๆ ในแบบทดสอบเป็น DIF free - ไม่มีคุณสมบัติของการบวก (Nonadditive) - ไม่มีการชดเชย (Noncompensatory)
Bundle CDIF	Bundle NCDIF
$\sum_{i=1}^n \text{CDIF}$ - เมื่อ n คือจำนวนข้อสอบในหมวดข้อสอบ - ระบุผลกระทบ (impact) ในการกำจัดหมวดข้อสอบกลุ่มใดกลุ่มหนึ่ง	$\left[\frac{1}{N_F} \sum_{s=1}^{N_F} \left(\sum_{i=1}^n P_{iF}(\theta_s) - \sum_{i=1}^n P_{iR}(\theta_s) \right)^2 \right] / n$ - เมื่อ n คือจำนวนข้อสอบในหมวดข้อสอบ และ N_F คือจำนวนผู้ตอบข้อสอบในกลุ่มสนใจ - ใช้ในการตรวจสอบสาเหตุที่ทำให้เกิดการทำหน้าที่เบี่ยงเบน
DTF	
$\sum_{i=1}^{N_i} \text{CDIF}$ เมื่อ N_i คือจำนวนข้อสอบในแบบทดสอบ	$\frac{1}{N_F} \sum_{s=1}^{N_F} \left(\sum_{i=1}^{N_i} P_{iF}(\theta_s) - \sum_{i=1}^{N_i} P_{iR}(\theta_s) \right)^2$ เมื่อ N_i คือจำนวนข้อสอบในแบบทดสอบทั้งฉบับ N_F คือจำนวนผู้ตอบข้อสอบในกลุ่มสนใจ

การนำดัชนีที่วิเคราะห์ตามวิธีการ DFIT ไปใช้ขึ้นอยู่กับวัตถุประสงค์ เช่น กรณีที่ต้องใช้คะแนนรวมจากแบบทดสอบพิจารณาประสิทธิภาพของโปรแกรมการสอน การจัดตำแหน่ง (placement) หรือ

การคัดเลือก (selection) ควรเลือกใช้ดัชนี DTF ส่วนดัชนี CDIF ใช้ในกรณีที่ต้องการรวมข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนในแบบทดสอบด้วยเหตุผลใดเหตุผลหนึ่ง เพราะบางครั้งการกำจัด (exclude) ข้อสอบทุกข้อที่แสดง DIF ออกจากแบบทดสอบเป็นสิ่งที่กระทำได้ยาก ในกรณีที่ต้องการตรวจสอบข้อสอบที่แสดง DIF ในแบบทดสอบเฉพาะกับกลุ่มเฉพาะ หรือตรวจสอบสาเหตุที่ข้อสอบประเภทใดประเภทหนึ่ง แสดงการทำหน้าที่เบี่ยงเบนมากกว่าประเภทอื่น ควรใช้ดัชนี NCDIF (Raju, van der Linden and Fleer. 1995 : 356)

การเปรียบเทียบพารามิเตอร์ข้อสอบให้อยู่บนสเกลร่วม

ในกรณีของข้อมูลการตอบข้อสอบมิติเดียว ก่อนตรวจสอบ DIF ค่าพารามิเตอร์ข้อสอบที่ประมาณจากกลุ่มสนใจถูกแปลงให้อยู่บนสเกลเดียวกับค่าพารามิเตอร์ที่ประมาณจากกลุ่มอ้างอิง ในกรณีของข้อมูลการตอบข้อสอบหลายมิติก็ต้องปรับให้อยู่บนสเกลร่วม (common scale) เช่นเดียวกัน โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป เช่น IPLINK (Lee & Oshima. 1996)

4.4 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่เบี่ยงเบนที่อาศัยกรอบความคิดของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติมีการศึกษากันในต่างประเทศ ดังนี้

แม็คคอกเลย์และเม็นโดซา (McCauley & Mendoza. 1985 : 389-400) ใช้การจำลองข้อมูลศึกษาการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยใช้แบบจำลองทฤษฎีการตอบข้อสอบชนิด 2 พารามิเตอร์ (two-parameter Item Response Model) เพื่อตอบคำถาม 2 ประการคือ 1) ข้อสอบที่ผู้ตอบข้อสอบต้องใช้ความสามารถในการตอบข้อสอบ แสดงการทำหน้าที่เบี่ยงเบนของข้อสอบมากกว่าข้อสอบที่ผู้ตอบข้อสอบไม่ต้องใช้ความสามารถหรือไม่ 2) ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนเป็นข้อสอบที่วัดโครงสร้างที่แตกต่างกันของกลุ่มที่แตกต่างกันมากกว่าข้อสอบที่วัดโครงสร้างที่คล้ายคลึงกันหรือไม่ ผลที่ได้พบว่าข้อสอบที่ผู้ตอบข้อสอบต้องใช้ความสามารถ ซึ่งกลุ่มอ้างอิงและกลุ่มสนใจมีระดับความสามารถเฉลี่ยแตกต่างกันจะแสดงการทำหน้าที่เบี่ยงเบนมากกว่าข้อสอบที่ผู้ตอบข้อสอบไม่ต้องใช้ความสามารถ และข้อสอบที่วัดองค์ประกอบโครงสร้างของความสามารถที่แตกต่างกันระหว่างกลุ่มอ้างอิงกับกลุ่มสนใจไม่มีความคงเส้นคงวาในการแสดงการทำหน้าที่เบี่ยงเบนของข้อสอบมากกว่าข้อสอบที่มีโครงสร้างองค์ประกอบคล้ายคลึงกัน นอกจากนี้ยังพบว่ามีความสอดคล้องกันระหว่างดัชนีที่ใช้ในการศึกษา คือ ดัชนีพื้นที่ (area indices) และการวัดความแตกต่างของพารามิเตอร์ (parameter difference measures) สรุปได้ว่าผลการวิจัยสนับสนุนว่าเมื่อกลุ่มใดกลุ่มหนึ่งที่ศึกษามีค่าเฉลี่ยของความสามารถที่ต้องใช้ในการตอบข้อสอบมากกว่าอีกกลุ่มหนึ่งเป็นสาเหตุทำให้ข้อสอบแสดงการทำหน้าที่เบี่ยงเบน

คามิลลี (Camilli. 1992 : 129-147) ใช้การจำลองข้อมูลศึกษาการทำหน้าที่เบี่ยงเบนของข้อสอบโดยอาศัยกรอบความคิดของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติชนิด 2 พารามิเตอร์ (two-dimensional item response model) ที่กล่าวว่า การแสดง DIF ของข้อสอบเกิดจากความสามารถที่มีอิทธิพลต่อความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง โดยที่กลุ่มใดกลุ่มหนึ่งมีความสามารถดังกล่าวมากกว่าอีกกลุ่มหนึ่ง เช่น ไหวพริบในการทำข้อสอบ (test-wiseness) การแจกแจงพารามิเตอร์ของความสามารถถูกนำไปรวมกับพารามิเตอร์ข้อสอบจากวิธีการตรวจสอบ DIF ตามแนวคิดความเป็นมิติเดียวและทำให้เกิดความแตกต่างของการประมาณค่าพารามิเตอร์ข้อสอบ นอกจากนี้ DIF อาจถูกนำไปรวมกับความแตกต่าง

ต่างของความสามารถหลักที่ต้องการวัดของทั้งสองกลุ่ม ซึ่งเป็นข้อจำกัดของวิธีการตรวจสอบ DIF ตามแนวคิดของ IRT ในบางสถานการณ์ กรอบความคิดของวิธี MIRT พิจารณาว่า DIF เป็นฟังก์ชันของภูมิหลังทางการศึกษาของผู้ตอบข้อสอบ ดังนั้นการทำความเข้าใจ DIF ได้ดีควรศึกษาด้วยวิธีการของแบบจำลองเชิงโครงสร้าง (structural modeling) โดยใช้เกณฑ์ภายนอกที่อธิบายถึงภูมิหลังและประสบการณ์ในโรงเรียนของผู้ตอบข้อสอบ

โอชิม่าและมิลเลอร์ (Oshima & Miller. 1992 : 237-248) ได้ศึกษาเกี่ยวกับความเป็นหลายมิติและการทำหน้าที่เบี่ยงเบนของข้อสอบตามกรอบความคิดของทฤษฎีการตอบข้อสอบ โดยมีวัตถุประสงค์ในการศึกษา 2 ประการ คือ 1) เพื่อยืนยันว่าข้อสอบที่วัดความสามารถหลายมิติไม่จำเป็นต้องแสดงการทำหน้าที่เบี่ยงเบน นอกจากกลุ่มที่ศึกษามีค่าเฉลี่ยของความสามารถแตกต่างกัน 2) เพื่อยืนยันว่าความแตกต่างของความสามารถหลักไม่ใช่สาเหตุที่ทำให้ข้อสอบแสดงการทำหน้าที่เบี่ยงเบนถ้าไม่มีสหสัมพันธ์ระหว่างความสามารถหลักและความสามารถรอง วิธีการศึกษาใช้การจำลองข้อมูล โดยมีเงื่อนไขที่ศึกษา 12 ($2 \times 2 \times 3$) เงื่อนไข คือ 2 ระดับของ $\theta_{2A} - \theta_{2B}$ (ไม่มี DIF และมี DIF) 2 ระดับของ $\theta_{1A} - \theta_{1B}$ (0.0, 0.5) 3 ระดับของข้อสอบที่วัดความสามารถหลายมิติในแบบทดสอบ (5% 10% และ 20%) จำลองข้อมูลการตอบข้อสอบกลุ่มละ 1,000 คน ทำข้อสอบคนละ 40 ข้อ คำนวณดัชนีการทำหน้าที่เบี่ยงเบนของข้อสอบ 4 แบบ คือ พื้นที่แบบคิดเครื่องหมาย (signed area : SA) พื้นที่แบบไม่คิดเครื่องหมาย (unsigned area : UA) signed sum of squares (SSOS) และ unsigned sum of squared (USOS) ผลการศึกษาพบว่ายืนยันตามวัตถุประสงค์ของการศึกษาทั้ง 2 ประการ

แอ็คเคอร์แมน (Ackerman. 1992 : 67-91) ได้ศึกษาเกี่ยวกับการทำหน้าที่เบี่ยงเบนของข้อสอบและความเที่ยงตรงของข้อสอบ โดยใช้กรอบความคิดของ MIRT ศึกษาโดยการจำลองข้อมูล 2 ชุด ชุดละ 1,000 คน ข้อมูลทั้ง 2 ชุดมีการแจกแจงความสามารถจำนวน 2 มิติแตกต่างกัน (different underlying two-dimensional ability distribution) ประมาณค่าพารามิเตอร์จากการเปรียบเทียบข้อสอบของ ACT (Form 26 A of the ACT Assessment Programs Math Usage Test) ตามที่เร็คเคส (Reckase) เสนอไว้ในปี ค.ศ. 1985 ภายหลังการเปรียบเทียบข้อสอบโดยใช้แบบจำลอง M2PL เลือกกลุ่มข้อสอบที่มีความเที่ยงตรง (valid sector) และกลุ่มข้อสอบที่สงสัยว่าจะแสดง DIF (suspected of being bias) เพื่อนำไปจำลองข้อมูลเพื่อศึกษาในกรณีต่าง ๆ ดังนี้

กรณีที่ 1 : equal θ , η Distributions : No DIF, No impact

กรณีที่ 2 : unequal θ means : uniform DIF

กรณีที่ 3 : unequal η means : uniform DIF

กรณีที่ 4 : unequal η variance : nonuniform DIF

กรณีที่ 5 : unequal $p_{\theta, \eta}$: nonuniform DIF

ใช้วิธีการตรวจสอบ DIF 2 วิธี คือ Mantel-Haenszel (MH) และ SIBTEST ผลจากการศึกษาพบว่าวิธี MH และ SIBTEST สามารถตรวจสอบ DIF ในกรณีต่าง ๆ ได้ แต่วิธี MH ต้องแปลความหมายด้วยความระมัดระวัง ผู้วิจัยให้ข้อเสนอแนะว่าผู้สร้างแบบทดสอบและผู้ใช้แบบทดสอบควรกำหนดทักษะที่ต้องการวัดให้ชัดเจน และกำจัดข้อสอบที่วัดทักษะที่ไม่ต้องการออกไป พร้อมทั้งกล่าวว่าข้อสอบตั้งแต่ 2 ข้อขึ้นไปทำให้เกิดความเป็นหลายมิติ (multidimensionality) และควรประมาณค่าพารามิเตอร์โดยใช้แบบจำลองหลายมิติ จึงจะทำให้การแปลความหมายผลที่ได้มีความถูกต้อง

โอชิมะและคณะ (Oshima, et. al. 1997 : 253-272) ได้แสดงวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบและแบบทดสอบที่ใช้ความสามารถหลายมิติโดยวิธี DFIT ที่พัฒนาโดยราชูและคณะ (Raju, et. al. 1995) ซึ่งเป็นกระบวนการในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบและของแบบทดสอบทั้งฉบับ (DIF/DTF) ที่อาศัยกรอบความคิดของทฤษฎีการตอบข้อสอบโดยใช้วิธีการของสถิติพาราเมตริก ใช้ได้กับแบบทดสอบที่สร้างขึ้นโดยอาศัยข้อตกลงของความเป็นมิติเดียวและอาศัยข้อตกลงของความเป็นหลายมิติ ที่มีการให้คะแนนรายข้อแบบสองค่าและแบบมากกว่าสองค่า ในการศึกษาครั้งนี้เป็นการศึกษาโดยใช้ข้อมูลจำลองที่มีแบบจำลองเป็น M2PL (multidimensional two-parameter logistic model) ใช้แบบทดสอบที่มีความยาว 40 ข้อ กลุ่มอ้างอิงและกลุ่มสนใจมีสมาชิกกลุ่มละ 1,000 คน ตัวแปรที่ศึกษาคือประเภทของ DIF (uniform DIF กับ nonuniform DIF และ unidirectional กับ balanced-bidirectional DIF) และความแตกต่างของการแจกแจงความสามารถที่ต้องการวัด (θ distribution difference) ระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ องค์ประกอบที่คงที่ได้แก่ จำนวนข้อสอบที่แสดง DIF (มี 4 ข้อ) และขนาดของ DIF ข้อสอบวัด θ_1 และ θ_2 ในระดับที่แตกต่างกันทั้งฉบับ (multidimensional within-item test) ประมาณค่าพารามิเตอร์ข้อสอบด้วยโปรแกรม NOHARM แล้วคำนวณค่าสัมประสิทธิ์การเปรียบเทียบ โดยใช้โปรแกรม IPLINK แล้วจึงใช้โปรแกรม DFIT คำนวณดัชนี CDIF NCDIF และ DTF การศึกษาครั้งนี้ดำเนินการเป็น 2 ส่วน คือ วิเคราะห์ DFIT ตามเงื่อนไขที่เป็นจริง (true condition) วิเคราะห์ด้วยค่าพารามิเตอร์ข้อสอบที่เป็นจริง (true item parameter) แล้วทำซ้ำด้วยเงื่อนไขที่เป็นค่าประมาณ (estimated condition) วิเคราะห์ด้วยค่าพารามิเตอร์ข้อสอบที่เป็นค่าประมาณ (estimated item parameter) ผลที่ได้ปรากฏว่าการวิเคราะห์ทั้ง 2 ส่วนให้ผลใกล้เคียงกันในทุกเงื่อนไข แสดงว่าวิธี DFIT ตรวจสอบ DIF/DTF ได้ดีในกรณีต่อไปนี้ 1) มี DIF แต่การแจกแจง θ ของทั้ง 2 กลุ่มเหมือนกัน 2) ไม่มี DIF แต่ทั้ง 2 กลุ่มมีการแจกแจง θ แตกต่างกัน

สมมุติฐานในการวิจัย

จากการค้นคว้าพบว่ายังไม่มีการศึกษาเพื่อเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนด้วยวิธีซิปเทสท์และวิธีดีเอฟไอที แต่ตามที่ได้กล่าวมาแล้วว่าวิธีการทั้งสองอาศัยกรอบความคิดของทฤษฎีการตอบข้อสอบ แตกต่างกันว่าวิธีซิปเทสท์ใช้วิธีการของสถิติพาราเมตริกและวิธีดีเอฟไอทีใช้วิธีการของสถิติพาราเมตริก ในการศึกษาครั้งนี้ใช้กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน ซึ่งรูสโซและสแตตต์ (Roussos & Stout. 1996.) จัดเป็นกลุ่มตัวอย่างขนาดเล็ก ผู้วิจัยจึงตั้งสมมุติฐานในการวิจัยดังนี้

1. กลุ่มตัวอย่างขนาดต่าง ๆ ส่งผลต่อความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิปเทสท์แตกต่างกันทุกความยาวแบบทดสอบ
2. กลุ่มตัวอย่างขนาดต่าง ๆ ส่งผลต่อความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีดีเอฟไอทีแตกต่างกันทุกความยาวแบบทดสอบ
3. เมื่อพิจารณาความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบพร้อมกัน วิธีซิปเทสท์ให้ผลดีกว่าวิธีดีเอฟไอทีในทุกเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน
4. ความสอดคล้องของผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิปเทสท์และวิธีดีเอฟไอที มีค่าต่ำกว่าร้อยละ 50 ทุกเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน
5. วิธีซิปเทสท์มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบมากกว่าวิธีดีเอฟไอที ทุกเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

6. วิธีชิบเทสที่มีความถูกต้องในการตรวจสอบการทำหน้าที่เป็ยงเบนของแบบทดสอบมากกว่าวิธีดีเอฟไอที ทุกเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

บทที่ 3

วิธีดำเนินการวิจัย

การศึกษานี้มีจุดมุ่งหมายเพื่อศึกษาวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนที่เหมาะสมกับข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ โดยเปรียบเทียบวิธี SIBTEST และวิธี DFIT ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบในเงื่อนไขความยาวแบบทดสอบ และกลุ่มตัวอย่างขนาดแตกต่างกัน

แหล่งข้อมูล

ข้อมูลที่ใช้ในการศึกษาคือ ข้อมูลการตอบข้อสอบวิชาคณิตศาสตร์ (ค 101) ของนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคการศึกษาที่ 2 ปีการศึกษา 2543 ในโรงเรียนสังกัดกรมสามัญศึกษา จังหวัดนนทบุรี

ประชากร

ประชากรที่ใช้ในการศึกษาเป็นประชากรเทียม(pseudo population)ที่กำหนดจากนักเรียนชายและนักเรียนหญิง ชั้นมัธยมศึกษาปีที่ 1 ภาคการศึกษาที่ 2 ปีการศึกษา 2543 ของโรงเรียนสังกัดกรมสามัญศึกษา ในจังหวัดนนทบุรี

กลุ่มตัวอย่าง

ตัวอย่างที่ใช้ในการศึกษาเป็นกลุ่มตัวอย่างที่สุ่มจากประชากรเทียมซึ่งกำหนดจากนักเรียนชายและนักเรียนหญิง ชั้นมัธยมศึกษาปีที่ 1 ภาคการศึกษาที่ 2 ปีการศึกษา 2543 ของโรงเรียนสังกัดกรมสามัญศึกษา จังหวัดนนทบุรี ที่ได้มาโดยการสุ่มแบบสองขั้นตอน (two-stage random sampling) มีขนาดโรงเรียนเป็นชั้น (strata) และมีห้องเรียนเป็นหน่วยการสุ่ม (sampling unit) โดยดำเนินการตามลำดับ ดังนี้

ขั้นที่ 1 ประมาณขนาดกลุ่มตัวอย่างที่ใช้ในการศึกษาด้วยความเชื่อมั่น 95% ($\alpha = .05$) เมื่อเทียบจากตารางขนาดกลุ่มตัวอย่างที่ควรเลือกจากประชากร พบว่าต้องใช้กลุ่มตัวอย่างอย่างน้อย 383 คน (Yamane, 1974 : 403) แต่ในการศึกษานี้ต้องการนักเรียนเพศชายอย่างน้อย 1,000 คน และนักเรียนเพศหญิงอย่างน้อย 1,000 คน ใช้เป็นประชากรเทียมในการสุ่มกลุ่มตัวอย่างขนาดต่าง ๆ

ขั้นที่ 2 แบ่งโรงเรียนสังกัดกรมสามัญศึกษาจังหวัดนนทบุรี โดยพิจารณาจากจำนวนนักเรียนออกเป็น 3 ขนาดคือ โรงเรียนขนาดใหญ่ ขนาดกลาง และขนาดเล็ก ซึ่งจากการสำรวจพบว่ามีโรงเรียนขนาดใหญ่ 10 โรงเรียน โรงเรียนขนาดกลาง 6 โรงเรียน และโรงเรียนขนาดเล็ก 4 โรงเรียน

ขั้นที่ 3 สุ่มโรงเรียนตามสัดส่วนของโรงเรียนแต่ละขนาดโดยใช้วิธีการสุ่มอย่างง่าย ได้โรงเรียนขนาดใหญ่ 3 โรงเรียน โรงเรียนขนาดกลาง 2 โรงเรียน และโรงเรียนขนาดเล็ก 1 โรงเรียน หลังจากนั้นจึงสุ่มห้องเรียนจากแต่ละโรงเรียน โดยวิธีสุ่มอย่างง่ายตามสัดส่วนของจำนวนนักเรียน เพื่อให้ได้นักเรียนเพศชายและเพศหญิงตามจำนวนที่ต้องการ รายชื่อโรงเรียนและจำนวนนักเรียนแสดงในตาราง 5

ตาราง 5 รายชื่อโรงเรียนและจำนวนนักเรียนจากแต่ละโรงเรียน

โรงเรียน	จำนวน ห้องเรียน	จำนวนนักเรียน		
		ชาย	หญิง	รวม
ขนาดใหญ่				
1. โรงเรียนวัดเขมาภิรตาราม	13	369	220	589
2. โรงเรียนปากเกร็ด	11	251	277	528
3. โรงเรียนสวนกุหลาบวิทยาลัย	9	218	237	455
ขนาดกลาง				
4. โรงเรียนบางบัวทอง	8	175	216	391
5. โรงเรียนโพธิ์นิมิตรวิทยา	6	106	84	190
ขนาดเล็ก				
6. โรงเรียนไทรน้อย	3	54	52	106
รวม	50	1,173	1,086	2,259

ใช้นักเรียนทั้งหมดของแต่ละเพศคือ เพศชายจำนวน 1,173 คน เพศหญิงจำนวน 1,086 คน เป็นประชากรเทียม ทำการสุ่มกลุ่มตัวอย่างขนาดต่าง ๆ จำแนกตามเพศ โดยใส่คั่นตามเงื่อนไขความยาวแบบทดสอบที่ประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน แต่ละเงื่อนไขทำการสุ่ม 50 ครั้ง ดังแสดงในภาพประกอบ 12

ขนาดกลุ่ม ตัวอย่าง ความยาว แบบทดสอบ	50		100		200		500		1,000	
	ชาย	หญิง	ชาย	หญิง	ชาย	หญิง	ชาย	หญิง	ชาย	หญิง
30	← สุ่ม 50 ครั้ง →									→
40	← สุ่ม 50 ครั้ง →									→
50	← สุ่ม 50 ครั้ง →									→

ภาพประกอบ 12 แสดงการสุ่มกลุ่มตัวอย่างในเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 (ค 101) ที่ผู้วิจัยสร้างขึ้น มีลักษณะเป็นข้อสอบแบบเลือกตอบชนิด 5 ตัวเลือก จำนวน 50 ข้อ มีข้อสอบจำนวน 16 ข้อที่ผู้เชี่ยวชาญพิจารณาตัดสินว่าแสดงการทำหน้าที่เบี่ยงเบนต่อเพศชาย ทำการตรวจให้คะแนนรายข้อแบบสองค่า (ตอบถูกให้ 1 คะแนน ตอบผิดให้ 0 คะแนน)

สำหรับการศึกษาในเงื่อนไขความยาวของแบบทดสอบ 40 และ 30 ข้อ ใช้ข้อมูลการตอบข้อสอบจากแบบทดสอบที่ประกอบด้วยข้อสอบ 50 ข้อของนักเรียนกลุ่มตัวอย่างที่กำหนดเป็นประชากรเทียมมาคัดเลือกข้อสอบตามสัดส่วนที่กำหนดในตารางกำหนดข้อสอบ(รายละเอียดในภาคผนวก ง)และพิจารณาคะแนนเฉพาะข้อสอบที่คัดเลือกไว้ จำนวน 40 และ 30 ข้อตามลำดับ

การสร้างเครื่องมือในการวิจัย

1. ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบและแบบทดสอบ เพื่อศึกษาลักษณะของข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน ซึ่งมีข้อสรุปที่สำคัญดังนี้

ลักษณะของข้อสอบโดยทั่วไปที่แสดงการทำหน้าที่เบี่ยงเบน

1. เนื้อหาหรือภาษาในข้อสอบยั่วให้ผู้ตอบข้อสอบกลุ่มสนใจโกรธ เกิดการโต้แย้งหรือเกิดอารมณ์ไม่พอใจ
2. เนื้อหาหรือภาษาในข้อสอบมีความหมายไปในทางลบ ดูถูกเหยียดหยามหรือก้าวร้าวต่อผู้ตอบข้อสอบกลุ่มสนใจ
3. เนื้อหาหรือภาษาในข้อสอบแสดงว่าผู้ตอบข้อสอบกลุ่มสนใจมีปมด้อยเกี่ยวกับอำนาจหรือความเป็นผู้นำ
4. เนื้อหาหรือภาษาในข้อสอบหลาย ๆ ข้อให้ความสนใจ เน้นความสำคัญและยกย่องผู้ตอบข้อสอบกลุ่มอ้างอิง
5. เนื้อหาในข้อสอบมีสารสนเทศที่เป็นประโยชน์กับกลุ่มอ้างอิงมากกว่ากลุ่มสนใจ

ลักษณะของข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนต่อเพศ

1. รูปแบบหรือโครงสร้างของข้อสอบเป็นปัญหากับผู้ตอบข้อสอบเพศใดเพศหนึ่งมากกว่าผู้ตอบข้อสอบอีกเพศหนึ่ง
2. เนื้อหาในข้อสอบใช้สรรพนามเฉพาะเพศใดเพศหนึ่ง
3. เนื้อหาในข้อสอบกำหนดสถานการณ์ที่ผู้ตอบข้อสอบเพศใดเพศหนึ่งได้รับการฝึกฝนเฉพาะทาง มีความสนใจ และมีโอกาสพบเห็นในชีวิตประจำวันมากกว่า

ลักษณะของข้อสอบวิชาคณิตศาสตร์ที่แสดงการทำหน้าที่เบี่ยงเบนต่อเพศ

1. ลักษณะของตัวเลือก เช่น เพศหญิงชอบเลือกตัวเลือกที่ระบุว่ามีคำตอบถูกมากกว่าเพศชาย (Gipps & Murphy. 1994 : 182)
2. ความซับซ้อนของโจทย์ เพศชายชอบโจทย์ที่ซับซ้อนและท้าทายความสามารถมากกว่าเพศหญิง
3. ลักษณะโจทย์ที่เป็นโจทย์ปัญหา (problem solving) เพศชายจะตอบได้ดีกว่าเพศหญิง

4. ความยากของข้อสอบ เพศชายมีความพยายามในการหาคำตอบข้อสอบที่ยากมากกว่าเพศหญิง

จากข้อสรุปข้างต้น ผู้วิจัยจึงกำหนดเงื่อนไขในการสร้างแบบทดสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค101) ที่แสดงการทำหน้าที่เบี่ยงเบนต่อเพศชาย มีลักษณะดังนี้

1. เป็นข้อสอบที่มีค่าความยากสูงมาก ($1.5 \leq P \leq 2.5$)
2. เป็นโจทย์ปัญหาที่มีความซับซ้อน
3. เนื้อหาในโจทย์เป็นสิ่งที่เพศชายสนใจ ได้รับการฝึกฝน มีความคุ้นเคย มีความรู้ มีโอกาสได้

พบเห็นหรือเกี่ยวข้องกับชีวิตประจำวันมากกว่าเพศหญิง

ในการสร้างแบบทดสอบมุ่งวัดความสามารถด้านคณิตศาสตร์ตามวัตถุประสงค์ที่กำหนดในหลักสูตรชั้นมัธยมศึกษาปีที่ 1 (ค 101) เป็นความสามารถหลัก (θ) ส่วนคุณลักษณะอื่น ๆ เป็นต้นว่าความสามารถในการอ่านและแก้ปัญหาโจทย์ที่มีความซับซ้อน ไหวพริบในการทำข้อสอบ ความคุ้นเคยหรือประสบการณ์ในสถานการณ์ของโจทย์ที่ผู้ตอบข้อสอบนำมาใช้ในการตอบข้อสอบได้ถูกต้อง จัดเป็นความสามารถรอง (η)

2. ศึกษาคู่มือครูและแบบเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 (ค 101) เพื่อพิจารณาโครงสร้างหลักสูตร กำหนดเนื้อหาและจำนวนข้อสอบแต่ละเนื้อหาตั้งรายละเอียดต่อไปนี้

บทที่ 1 การเขียนตัวเลขแทนจำนวน (4 คาบ)

- 1.1 จำนวนและตัวเลข (1 คาบ)
- 1.2 ระบบตัวเลขฐานสิบ (1 คาบ) (1)
- 1.3 เลขยกกำลัง (2 คาบ) (2)

บทที่ 2 สมบัติของจำนวนนับ (7 คาบ)

- 2.1 ตัวประกอบ (1 คาบ)
- 2.2 จำนวนเฉพาะ (1 คาบ) (1)
- 2.3 การแยกตัวประกอบ (1 คาบ) (1)
- 2.4 ตัวหารร่วมมาก (2 คาบ) (2)
- 2.5 ตัวคูณร่วมน้อย (2 คาบ) (2)

บทที่ 3 ทศนิยม (10 คาบ)

- 3.1 ค่าประจำหลักของทศนิยม (1 คาบ) (1)
- 3.2 การเปรียบเทียบทศนิยม (1 คาบ) (1)
- 3.3 การบวกและการลบทศนิยม (3 คาบ) (2)
- 3.4 การคูณและการหารทศนิยม (4 คาบ) (4)
- 3.5 การแทนเศษส่วนด้วยทศนิยม (1 คาบ) (1)

บทที่ 4 การวัดและการประมาณ (5 คาบ)

- 4.1 ค่าประมาณจากการวัด (1 คาบ) (1)
- 4.2 การวัดเศษ (2 คาบ) (2)
- 4.3 การประมาณค่า (2 คาบ) (2)

บทที่ 5 เส้นตรงและมุม (8 คาบ)

- 5.1 จุด เส้นตรง ส่วนของเส้นตรง รังสี และมุม (1 คาบ)

5.2 ส่วนของเส้นตรงและความยาวส่วนของเส้นตรง (1 คาบ) (1)

5.3 มุมและขนาดของมุม (3 คาบ) (3)

5.4 มุมฉากและมุมตรง (3 คาบ) (3)

บทที่ 6 เศษส่วน (12 คาบ)

6.1 เศษส่วน (1 คาบ) (1)

6.2 การเปรียบเทียบเศษส่วน (2 คาบ) (2)

6.3 การบวกและการลบเศษส่วน (2 คาบ) (2)

6.4 การคูณและการหารเศษส่วน (3 คาบ) (3)

6.5 โจทย์ปัญหาเศษส่วน (4 คาบ) (4)

บทที่ 7 ความยาว พื้นที่ และปริมาตร (8 คาบ)

7.1 ความยาว (1 คาบ) (1)

7.2 มาตรการส่วน (2 คาบ) (2)

7.3 พื้นที่ (3 คาบ) (3)

7.4 ปริมาตรของทรงสี่เหลี่ยมมุมฉาก (2 คาบ) (2)

หมายเหตุ ตัวเลขในวงเล็บคือจำนวนข้อสอบในแบบทดสอบที่มีความยาว 50 ข้อ

การวิเคราะห์ระดับพฤติกรรมการเรียนรู้ด้านสติปัญญาวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1

(ค 101) ยึดตามแนวคิดของวิลสัน (Wilson. 1971 : 643-685) ซึ่งแบ่งเป็น 4 ระดับ คือ ความรู้หรือทักษะการคำนวณ (computation) ความเข้าใจ (comprehension) การนำไปใช้ (application) และการวิเคราะห์ (Analysis) รายละเอียดแสดงในตารางกำหนดข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101)

ตาราง 6 ตารางกำหนดข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101)

เนื้อหา	วัตถุประสงค์	พฤติกรรม/จำนวนข้อ				รวม
		ทักษะการ คำนวณ	เข้าใจ	นำไปใช้	วิเคราะห์	
บทที่ 1						(3) 2 2
เรื่องที่ 1.2	1. น.ร.เขียนตัวเลขที่กำหนดให้อยู่ในรูปการกระจายได้	(1) 1				1
เรื่องที่ 1.3	1. น.ร.เขียนตัวเลขให้อยู่ในรูปการคูณของจำนวนนับที่น้อยที่สุดกับเลขยกกำลังที่มีฐานเป็นสิบ และเลขชี้กำลังเป็นจำนวนนับได้	(1)				1
	2. น.ร.บวกเลขและลบจำนวนที่อยู่ในรูปเลขยกกำลังอย่างง่ายได้			(1) 1		1

ตาราง 6 (ต่อ)

เนื้อหา	วัตถุประสงค์	พฤติกรรม/จำนวนข้อ			รวม
		ทักษะการ คำนวณ	เข้าใจ	นำไปใช้ วิเคราะห์	
บทที่ 2					(6) 5 4
เรื่องที่ 2.2	1. น.ร.บอกได้ว่าจำนวนที่กำหนดให้จำนวนใดบ้างเป็นจำนวนเฉพาะ และจำนวนใดบ้างไม่เป็นจำนวนเฉพาะ	(1) 1 1			
เรื่องที่ 2.3	1. น.ร.เขียนการแยกตัวประกอบในรูปการคูณของเลขยกกำลังได้	(1) 1 1			(6) 5 4
เรื่องที่ 2.4	1. เมื่อกำหนดจำนวนนับตั้งแต่ 2 จำนวนขึ้นไป น.ร. หา ห.ร.ม. ของจำนวนเหล่านั้นได้		(2) 1 1		
เรื่องที่ 2.5	1. เมื่อกำหนดจำนวนนับตั้งแต่ 2 จำนวนขึ้นไป น.ร. หา ค.ร.น. ของจำนวนเหล่านั้นได้		(2) 2 1		
บทที่ 3					(9) 7 5
เรื่องที่ 3.1	1. น.ร.บอกค่าประจำหลักของทศนิยมตำแหน่งต่าง ๆ ได้	(1) 1 1			
เรื่องที่ 3.2	1. น.ร.เปรียบเทียบจำนวนที่เขียนในรูปทศนิยมได้	(1) 1 1			
เรื่องที่ 3.3	1. น.ร.บวกและลบทศนิยมได้อย่างแม่นยำ รวดเร็ว		(1)		
	2. น.ร.แก้โจทย์ปัญหาที่มีการบวกและการลบทศนิยมได้			(1) 1 1	
เรื่องที่ 3.4	1. น.ร.คูณและหารทศนิยมได้อย่างแม่นยำ และรวดเร็ว		(2) 1 0		
	2. น.ร.แก้โจทย์ปัญหาที่มีการคูณและการหารทศนิยมได้			(2) 2 1	
เรื่องที่ 3.5	1. น.ร.เขียนจำนวนที่อยู่ในรูปเศษส่วนให้อยู่ในรูปทศนิยมซ้ำได้	(1) 1 1			
บทที่ 4					(5) 4 3
เรื่องที่ 4.1	1. น.ร.บอกปริมาณที่ได้จากการวัดเป็นทศนิยมที่ใกล้เคียงตามจำนวนตำแหน่งที่กำหนดให้ได้	(1) 1 1			

ตาราง 6 (ต่อ)

เนื้อหา	วัตถุประสงค์	พฤติกรรม/จำนวนข้อ				รวม
		ทักษะการ คำนวณ	เข้าใจ	นำไปใช้	วิเคราะห์	
เรื่องที่ 4.2	2. น.ร. บัดเศษใกล้เคียงจำนวนตามความต้องการได้			(2) 1 1		
เรื่องที่ 4.3	1. น.ร. ใช้การประมาณค่าหาคำตอบเกี่ยวกับ โจทย์การซื้อขาย การลงทุน การทำงาน ประมาณ การเตรียมการอย่างใดอย่าง หนึ่งได้				(2) 2 1	
บทที่ 5						(7) 6 4
เรื่องที่ 5.2	1. น.ร. หาจุดกึ่งกลางส่วนของเส้นตรงโดยใช้ วงเวียนและสันตรงได้			(1) 1 1		
เรื่องที่ 5.3	1. น.ร. สร้างมุมให้มีขนาดเท่ากับมุมที่ กำหนดให้ได้			(1) 1 1		
	2. น.ร. แบ่งมุมที่กำหนดให้ออกเป็น 2 ส่วน หรือ 4 ส่วนที่มีขนาดเท่ากันได้			(1) 0 0		
	3. น.ร. สร้างรูปสามเหลี่ยมบางรูปโดยใช้ ความรู้เรื่องการสร้างส่วนของเส้นตรงและ การสร้างมุมให้มีขนาดเท่ากับมุมที่ กำหนด			(1) 1 1		
เรื่องที่ 5.4	ให้			(1) 1 0		
	1. น.ร. สร้างมุมฉาก มุมที่มีขนาด 45° 60° และนำความรู้ไปสร้างมุมขนาดอื่น ๆ ได้			(1) 1 1		
	2. น.ร. สร้างเส้นตั้งฉากจากจุดภายนอก			(1) 1 0		
	3. น.ร. สร้างเส้นตรงตั้งฉากที่จุด ๆ หนึ่งบน เส้นตรงที่กำหนดให้ได้			(1) 1 0		
บทที่ 6						(12) 10 7
เรื่องที่ 6.1	1. น.ร. บอกเศษส่วนที่แทนด้วยจำนวนบน เส้นจำนวนที่แบ่งส่วนไวชัดเจนได้			(1) 1 1		
เรื่องที่ 6.2	1. น.ร. บอกได้ว่าเศษส่วนที่กำหนดให้มาก กว่า น้อยกว่า หรือเท่ากันได้			(2) 1 1		

ตาราง 6 (ต่อ)

เนื้อหา	วัตถุประสงค์	พฤติกรรม/จำนวนข้อ				รวม
		ทักษะการ คำนวณ	เข้าใจ	นำไปใช้	วิเคราะห์	
เรื่องที่ 6.3	1. น.ร.หาผลบวกและผลลบของเศษส่วนได้ อย่างแม่นยำและรวดเร็ว			(1) 0 0		(12) 10 7
	2. น.ร.ใช้การบวกและการลบเศษส่วนแก้ ปัญหาเศษส่วนได้			(1) 1 1		
เรื่องที่ 6.4	1. น.ร.หาผลคูณและผลหารของเศษส่วนได้ อย่างแม่นยำและรวดเร็ว			(1) 1 0		
	2. น.ร.ใช้การคูณและการหารแก้ปัญห โจทย์ได้				(2) 2 1	
เรื่องที่ 6.5	1. น.ร.แก้โจทย์ปัญหาเกี่ยวกับเศษส่วนได้				(4) 4 3	
บทที่ 7						(8) 6 5
เรื่องที่ 7.1	1. น.ร.หาความยาวของสิ่งต่าง ๆ โดยการ วัดและประมาณความยาวที่ได้เป็นจำนวน เต็มหน่วยหรือเป็นทศนิยมไม่เกิน 2 ตำแหน่ง	(1) 1 1				
เรื่องที่ 7.2	1. น.ร.เขียนแผนผังของส่วนต่าง ๆ ได้			(1) 0 0		
	2. น.ร.เปลี่ยนความยาวที่ได้จากแผนผังเป็น ความยาวจริงได้ถูกต้อง		(1) 1 1			
เรื่องที่ 7.3	1. น.ร.หาพื้นที่ของรูปสี่เหลี่ยมมุมฉาก สี่ เหลี่ยมคางหมู สามเหลี่ยม และวงกลมได้			(1) 1 1		
	2. น.ร.นำความรู้เกี่ยวกับการประมาณพื้นที่ ไปใช้ในชีวิตประจำวันได้			(2) 2 1		
เรื่องที่ 7.4	1. น.ร.หาปริมาตรของทรงสี่เหลี่ยมได้อย่าง ถูกต้องรวดเร็ว และนำไปแก้โจทย์ ได้				(2) 1 1	
รวม		(7) 6 6	(6) 5 5	(24) 17 11	(13) 12 8	(50) 40 30

หมายเหตุ ตัวเลขในเครื่องหมาย ()
ตัวเลขที่ขีดเส้นใต้
ตัวเลขไม่ขีดเส้นใต้

หมายถึงจำนวนข้อสอบในแบบทดสอบที่มีความยาว 50 ข้อ
หมายถึงจำนวนข้อสอบในแบบทดสอบที่มีความยาว 40 ข้อ
หมายถึงจำนวนข้อสอบในแบบทดสอบที่มีความยาว 30 ข้อ

3. เขียนข้อสอบตามตารางกำหนดข้อสอบที่วิเคราะห์ไว้ในขั้นตอนที่ 2 โดยเขียนข้อสอบมากกว่าจำนวนที่ต้องการใช้จริง และคัดเลือกตามสัดส่วนเนื้อหาจำนวน 75 ข้อ จัดพิมพ์ให้ผู้เชี่ยวชาญ (รายชื่อในภาคผนวก ก) ตรวจสอบความเที่ยงตรงเชิงเนื้อหาและการแสดงการทำหน้าที่เบี่ยงเบนของข้อสอบหมวดข้อสอบ และแบบทดสอบ ซึ่งการแบ่งหมวดข้อสอบแบ่งตามเนื้อหาเป็น 7 หมวด คือ การเขียนเลขแทนจำนวน สมบัติของจำนวนนับ ทศนิยม การวัดและการประมาณ เส้นตรงและมุม เศษส่วน ความยาว พื้นที่ และปริมาตร (รายละเอียดในภาคผนวก ง)

4. ปรับปรุงแก้ไขข้อสอบตามข้อเสนอแนะของผู้เชี่ยวชาญและวิเคราะห์ดัชนีความสอดคล้องของความเห็นเกี่ยวกับการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ โดยใช้ดัชนีความสอดคล้องของโรวินेलลีและแฮมเบิลตัน (บุญเชิด ภิญญอนันตพงษ์. 2527 : 67-70 ; อ้างอิงจาก Rovinelli and Hambleton. 1972) ซึ่งมีสูตรดังนี้

$$IOC = \frac{\sum R}{N}$$

เมื่อ	IOC	แทน	ดัชนีความสอดคล้องของความเห็นของผู้เชี่ยวชาญ
	$\sum R$	แทน	ผลการตัดสินของผู้เชี่ยวชาญ
	N	แทน	จำนวนผู้เชี่ยวชาญ

โดยผู้เชี่ยวชาญมีเกณฑ์ในการพิจารณาการทำหน้าที่เบี่ยงเบน ดังนี้

+1 แทน แน่ใจว่าข้อสอบ หมวดข้อสอบ หรือแบบทดสอบ แสดงการทำหน้าที่เบี่ยงเบน

0 แทน ไม่แน่ใจว่าข้อสอบ หมวดข้อสอบ หรือแบบทดสอบ แสดงการทำหน้าที่เบี่ยงเบน

-1 แทน แน่ใจว่าข้อสอบ หมวดข้อสอบ หรือแบบทดสอบ ไม่แสดงการทำหน้าที่เบี่ยงเบน

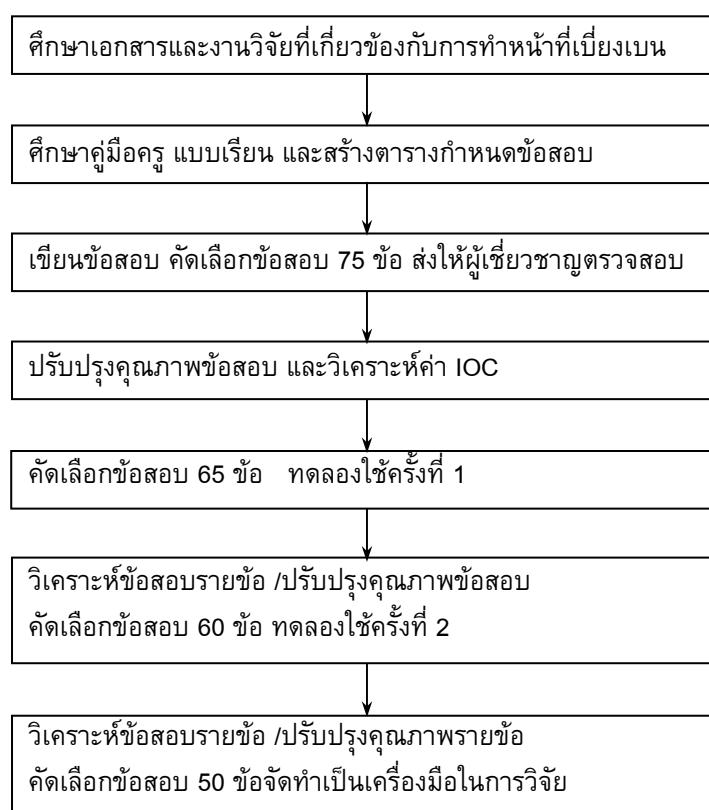
การแปลความหมายดัชนีที่คำนวณได้ถือว่า ถ้า IOC มีค่าตั้งแต่ 0.5 ขึ้นไป แสดงว่าข้อสอบ หมวดข้อสอบ หรือแบบทดสอบ แสดงการทำหน้าที่เบี่ยงเบน (รายละเอียดในภาคผนวก ง)

5. คัดเลือกข้อสอบจำนวน 65 ข้อ ให้สอดคล้องกับตารางกำหนดข้อสอบ และประกอบด้วยข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนตามเกณฑ์การพิจารณาในข้อ 4 จำนวน 20 ข้อ จัดพิมพ์และนำไปทดลองใช้กับนักเรียนชั้นมัธยมศึกษาปีที่ 1 โรงเรียนเตรียมอุดมศึกษาน้อมเกล้า นนทบุรี จำนวน 1 ห้องเรียน มีนักเรียนเข้ารับการทดสอบจำนวน 34 คน วิเคราะห์คุณภาพข้อสอบรายข้อด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป CTIA (สุพรรณ สุกมลสันต์. 2538) คัดเลือกข้อสอบที่ไม่ได้แสดงการทำหน้าที่เบี่ยงเบนที่มีค่าความยากระหว่าง 0.2 ถึง 0.8 ($0.2 \leq P \leq 0.8$) อำนาจจำแนกตั้งแต่ 0.2 ขึ้นไป ($r \geq 0.2$) และคัดเลือกข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนที่มีค่าความยากระหว่าง 0.15 ถึง 0.25 ($0.15 \leq P \leq 0.25$) อำนาจจำแนกตั้งแต่ 0.2 ขึ้นไป ($r \geq 0.2$) นำข้อสอบที่คัดเลือกมาพิจารณาค่าสถิติจากผลการวิเคราะห์รายข้ออีกครั้ง และปรับปรุงความชัดเจนของโจทย์ ปรับสถานการณ์ และตัวเลือกให้เหมาะสมยิ่งขึ้น

6. คัดเลือกข้อสอบจากขั้นตอนในข้อ 5 ให้สอดคล้องกับตารางกำหนดข้อสอบ จำนวน 60 ข้อ ประกอบด้วยข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนจำนวน 17 ข้อ (รายละเอียดในภาคผนวก ง) จัดพิมพ์แบบทดสอบ นำไปทดลองใช้ครั้งที่ 2 กับนักเรียนชั้นมัธยมศึกษาปีที่ 1 โรงเรียนศรีบุญยานนท์ จำนวน 1 ห้องเรียน และโรงเรียนนวมินทราชูทิศ หอวัง นนทบุรี จำนวน 2 ห้องเรียน มีกระดาษคำตอบของนักเรียนที่คัดเลือกนำมาวิเคราะห์คุณภาพรายข้อจำนวน 110 คน วิเคราะห์คุณภาพข้อสอบรายข้อด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป CTIA (สุพรรณ สุกมลสันต์. 2538) ตรวจสอบค่าสถิติของข้อสอบรายข้อ และปรับปรุงข้อสอบให้เหมาะสมยิ่งขึ้น

7. คัดเลือกข้อสอบโดยใช้เกณฑ์ที่กำหนดในข้อ 5 ให้สอดคล้องตามสัดส่วนที่ระบุในตารางกำหนดข้อสอบ จำนวน 50 ข้อ ประกอบด้วยข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนจำนวน 16 ข้อ (รายละเอียดในภาคผนวก ง) จัดพิมพ์แบบทดสอบนำไปใช้เป็นเครื่องมือในการเก็บรวบรวมข้อมูลในการศึกษาต่อไป

กระบวนการในการสร้างแบบทดสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) แสดงดังภาพประกอบ 13



ภาพประกอบ 13 สรุปรขั้นตอนการสร้างแบบทดสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101)

การเก็บรวบรวมข้อมูล

ผู้วิจัยดำเนินการดังนี้

1. ขอหนังสือจากบัณฑิตวิทยาลัยมหาวิทยาลัยศรีนครินทรวิโรฒ ถึงอธิบดีกรมสามัญศึกษา เพื่อขอความร่วมมือในการเก็บรวบรวมข้อมูลจากโรงเรียนกลุ่มตัวอย่าง

2. นำหนังสือขอความร่วมมือในการเก็บรวบรวมข้อมูลจากอธิบดีกรมสามัญศึกษาไปติดต่อประสานงานกับโรงเรียนกลุ่มตัวอย่าง โดยติดต่อกับผู้อำนวยการโรงเรียน ผู้ช่วยผู้อำนวยการฝ่ายวิชาการ หรือผู้ที่ได้รับมอบหมาย เพื่อบันทึกหมายเรื่องเวลาในการทดสอบและการดำเนินการสอบ ซึ่งแบ่งเป็น 2 กรณี

2.1 การทดลองใช้แบบทดสอบ ผู้วิจัยเป็นผู้เก็บรวบรวมข้อมูลด้วยตนเอง

2.2 การเก็บรวบรวมข้อมูลจริง ขอความร่วมมือให้ทางโรงเรียนแต่งตั้งคณาจารย์ในโรงเรียนเป็นกรรมการคุมสอบ

3. ผู้วิจัยนำแบบทดสอบไปทดลองใช้ครั้งที่ 1 และครั้งที่ 2 ด้วยตนเอง ส่วนการเก็บรวบรวมข้อมูลจริง ผู้วิจัยนำแบบทดสอบ กระดาษคำตอบและกระดาษทดตามจำนวนที่แต่ละโรงเรียนระบุไปให้ก่อนวันสอบหรือในวันสอบแล้วแต่กรณี และนัดหมายเรื่องการรวบรวมแบบทดสอบและกระดาษคำตอบคืน ซึ่งโรงเรียนกลุ่มตัวอย่างให้ความร่วมมือเป็นอย่างดี

การเตรียมข้อมูลในการศึกษาและการวิเคราะห์ข้อมูล

1. การเตรียมข้อมูลในการศึกษา

1.1 คัดเลือกกระดาษคำตอบของนักเรียนที่ตอบข้อสอบทุกข้ออย่างสมบูรณ์แบบ บันทึกลงการตอบข้อสอบวิชาคณิตศาสตร์ (ค 101) จากแบบทดสอบซึ่งประกอบด้วยข้อสอบจำนวน 50 ข้อของนักเรียนกลุ่มตัวอย่างที่กำหนดให้เป็นประชากรเทียมจำนวน 2,259 คน เป็นเพศชาย 1,173 คน เพศหญิง 1,086 คน แล้วตรวจให้คะแนนโดยแปลงคำตอบข้อที่ตอบถูกให้เป็น 1 คะแนน และข้อที่ตอบผิดให้เป็น 0 คะแนน โดยใช้โปรแกรมคอมพิวเตอร์ SPSS-PC for windows บันทึกแฟ้มข้อมูลการตอบข้อสอบ 50 ข้อแล้วพิจารณาคะแนนของนักเรียนเฉพาะข้อสอบในแบบทดสอบที่ประกอบด้วยข้อสอบจำนวน 40 ข้อและ 30 ข้อ บันทึกแฟ้มข้อมูลการตอบข้อสอบแบบทดสอบที่มีความยาว 40 ข้อและ 30 ข้อ ตามลำดับ

1.2 ตรวจสอบความเหมาะสมของข้อมูลการตอบข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) กับแบบจำลองการตอบข้อสอบที่ใช้ความสามารถหลายมิติชนิดสองพารามิเตอร์ (M2PL) โดยดำเนินการดังนี้

1.2.1 ตรวจสอบว่าข้อมูลการตอบข้อสอบของกลุ่มตัวอย่างที่กำหนดเป็นประชากรเทียมไม่ได้เป็นไปตามข้อตกลงเบื้องต้นของความเป็นมิติเดียวของแบบทดสอบ โดยใช้การวิเคราะห์องค์ประกอบ (factor analysis) ด้วยวิธีวิเคราะห์ตัวประกอบสำคัญ (principal component analysis) และหมุนแกนด้วยวิธี varimax ตามที่ลอร์ด (Lord, 1980) เสนอแนะไว้ โดยใช้ข้อมูลการตอบข้อสอบจำนวน 50 ข้อและ 30 ข้อของนักเรียนกลุ่มตัวอย่างที่กำหนดให้เป็นประชากรเทียม 2,259 คน จำแนกเป็นเพศชายจำนวน 1,173 คน และเพศหญิงจำนวน 1,086 คน ผลวิเคราะห์ดังแสดงในตาราง 7 -9

ตาราง 7 ค่าไอเกนและร้อยละของความแปรปรวนของตัวประกอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 (ค 101) เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ

ตัวประกอบที่	ชาย		หญิง		รวม	
	ค่าไอเกน	ค่าร้อยละ	ค่าไอเกน	ค่าร้อยละ	ค่าไอเกน	ค่าร้อยละ
1	4.675	9.351	5.514	11.028	5.075	10.150
2	1.504	3.009	1.578	3.157	1.480	2.959
3	1.430	2.861	1.487	2.974	1.370	2.741
4	1.379	2.757	1.300	2.601	1.241	2.482
5	1.342	2.684	1.267	2.534	1.226	2.451

ผลการวิเคราะห์ตามตาราง 7 พบว่าค่าไอเกนสูงสุดของวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ คือ 5.075 โดยที่ผลวิเคราะห์แยกตามเพศชายและเพศหญิงมีค่าใกล้เคียงกับผลที่ได้จากข้อมูลรวมและคิดเป็นค่าความแปรปรวนร้อยละ 10.150 ของความแปรปรวนทั้งหมด ซึ่งมีค่าน้อยเกินกว่าที่จะสรุปถึงความเป็นมิติเดียวชนิดที่มีตัวประกอบหลักตัวแรกเด่นกว่าตัวประกอบอื่น ๆ (essential unidimensionality) ตามที่เร็คเคส (Reckase. ; กาญจนนา วัธนสุนทร. 2538 : 88 ; อ้างอิงจาก Raju. 1993 : 21-22.) เสนอแนะว่าการใช้ทฤษฎีการตอบข้อสอบวิเคราะห์ข้อสอบนั้น ค่าความแปรปรวนจากตัวประกอบหลักตัวแรกของแบบทดสอบควรมีค่าไม่ต่ำกว่าร้อยละ 20 ด้วยเหตุนี้จึงสรุปว่าข้อมูลการตอบข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) ชุดนี้ไม่เหมาะสมกับแบบจำลองการตอบข้อสอบที่มีข้อตกลงเบื้องต้นของความเป็นมิติเดียว และเมื่อพิจารณาค่าไอเกนสูงสุดของวิชาคณิตศาสตร์ เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ และ 30 ข้อ ก็ได้ผลในทำนองเดียวกันดังแสดงในตาราง 8 -9

ตาราง 8 ค่าไอเกนและร้อยละของความแปรปรวนของตัวประกอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ

ตัวประกอบที่	ชาย		หญิง		รวม	
	ค่าไอเกน	ค่าร้อยละ	ค่าไอเกน	ค่าร้อยละ	ค่าไอเกน	ค่าร้อยละ
1	3.152	7.879	3.526	8.814	3.392	8.479
2	1.457	3.642	1.539	3.847	1.1572	3.931
3	1.387	3.469	1.491	3.726	1.377	3.442
4	1.386	3.464	1.351	3.378	1.247	3.116
5	1.283	3.206	1.316	3.289	1.215	3.038

ผลการวิเคราะห์ตามตาราง 8 พบว่าค่าไอเกนสูงสุดของวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ คือ 3.392 โดยที่ผลวิเคราะห์แยกตามเพศชายและเพศหญิง มีค่าใกล้เคียงกับผลที่ได้จากข้อมูลรวมและคิดเป็นค่าความแปรปรวนร้อยละ 8.479 ของความแปรปรวนทั้งหมด

ตาราง 9 ค่าไอเกนและร้อยละของความแปรปรวนของตัวประกอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ

ตัวประกอบที่	ชาย		หญิง		รวม	
	ค่าไอเกน	ค่าร้อยละ	ค่าไอเกน	ค่าร้อยละ	ค่าไอเกน	ค่าร้อยละ
1	2.764	9.214	3.130	10.432	2.612	8.707
2	1.344	4.479	1.426	4.755	1.367	4.556
3	1.286	4.285	1.310	4.367	1.320	4.400
4	1.223	4.078	1.192	3.972	1.316	4.388
5	1.210	4.035	1.145	3.815	1.163	3.876

ผลการวิเคราะห์ตามตาราง 9 พบว่าค่าไอเกนสูงสุดของวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ คือ 2.612 โดยที่ผลวิเคราะห์แยกตามเพศชายและเพศหญิง มีค่าใกล้เคียงกับผลที่ได้จากข้อมูลรวมและคิดเป็นค่าความแปรปรวนร้อยละ 8.707 ของความแปรปรวนทั้งหมด

1.2.2 ประมาณค่าพารามิเตอร์ข้อสอบจากข้อมูลการตอบข้อสอบของกลุ่มตัวอย่างที่กำหนดให้เป็นประชากรเทียม เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ 40 ข้อ และ 30 ข้อ ตามแบบจำลองการตอบข้อสอบที่ใช้ความสามารถหลายมิติชนิดสองพารามิเตอร์ (M2PL) โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป NOHARM (Normal Ogive Harmonic Analysis Robust Method : Fraser & McDonald. 1987) ผลปรากฏว่าสามารถวิเคราะห์หาค่าพารามิเตอร์ของข้อสอบได้ทุกข้อ (ผลการวิเคราะห์ในภาคผนวก ฉ)

1.3 การจำลองคะแนนของกลุ่มตัวอย่าง

1.3.1 การเตรียมข้อมูลจากการตอบข้อสอบในแบบทดสอบที่ประกอบด้วยข้อสอบ 50 ข้อ โดยการกำหนดให้กลุ่มตัวอย่างทั้งหมดแต่ละเพศเป็นประชากรเทียม แล้วสุ่มกลุ่มตัวอย่างจากประชากรเทียมแต่ละเพศเป็นกลุ่มตัวอย่างที่มีขนาดต่างกันคือ 50 100 200 500 และ 1,000 คน โดยวิธีสุ่มแบบใส่คืนขนาดละ 50 ครั้ง ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป SPSS-PC for windows แล้วบันทึกเป็นแฟ้มข้อมูลการตอบแบบทดสอบที่มีความยาว 50 ข้อ จำนวน $2 \times 5 \times 50 = 500$ แฟ้ม

1.3.2 การเตรียมข้อมูลจากการตอบข้อสอบในแบบทดสอบที่มีความยาว 40 ข้อ และ 30 ข้อ โดยใช้ข้อมูลการตอบข้อสอบจำนวน 50 ข้อ ของนักเรียนเพศชายและเพศหญิงที่กำหนดให้เป็นประชากรเทียม มาคัดเลือกราคาตอบเฉพาะข้อสอบที่ต้องการตามสัดส่วนของเนื้อหาในแผนผังกำหนดข้อสอบให้เหลือเป็นข้อมูลการตอบข้อสอบจำนวน 40 ข้อ และ 30 ข้อ (รายละเอียดในภาคผนวก ง) หลังจากนั้นสุ่มตัวอย่างแต่ละเพศจากประชากรเทียมให้มีขนาด 50 100 200 500 และ 1,000 คน โดยวิธีสุ่มแบบใส่คืน ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป SPSS-PC for windows ขนาดละ 50 ครั้ง บันทึกเป็นแฟ้มข้อมูลการตอบแบบทดสอบที่มีความยาว 40 ข้อ จำนวน $2 \times 5 \times 50 = 500$ แฟ้ม และบันทึกเป็นแฟ้มข้อมูลการตอบแบบทดสอบที่มีความยาว 30 ข้อ จำนวน $2 \times 5 \times 50 = 500$ แฟ้ม

1.4 ประเมินค่าพารามิเตอร์ข้อสอบตามแบบจำลองการตอบข้อสอบที่ใช้ความสามารถหลายมิติ (M2PL) จากข้อมูลการตอบข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) ที่มีความยาว 30 40 และ 50 ข้อ ของกลุ่มตัวอย่าง เพศหญิงและเพศชายที่มีขนาดต่าง ๆ คือ 50 100 200 500 และ 1,000 คน โดยทำซ้ำเงื่อนไขละ 50 ครั้ง ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป NOHARM

การวิเคราะห์ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป NOHARM ต้องดำเนินการดังนี้

1.4.1 วิเคราะห์ด้วยโปรแกรมย่อย PRODMOM เพื่อเตรียม Product Moment Matrix เป็นไฟล์ข้อมูลป้อนเข้าสำหรับโปรแกรม NOHARM จึงต้องทำการวิเคราะห์และเตรียมไฟล์ข้อมูลจำนวน $2 \times 3 \times 5 \times 50 = 1,500$ แฟ้ม

1.4.2 วิเคราะห์เพื่อหาค่าพารามิเตอร์ข้อสอบของกลุ่มตัวอย่างจำแนกตามเพศในเงื่อนไขต่าง ๆ ด้วยโปรแกรมคอมพิวเตอร์ NOHARM บันทึกแฟ้มข้อมูลจากการวิเคราะห์ด้วยโปรแกรม NOHARM จำนวน $2 \times 3 \times 5 \times 50 = 1,500$ แฟ้ม ผลการวิเคราะห์ด้วยโปรแกรม NOHARM ในแต่ละแฟ้ม ประกอบด้วยค่าต่อไปนี้ λ_{i1} คือ promax factor loading ของความสามารถมิติที่ 1 ของข้อสอบข้อที่ i λ_{i2} คือ promax factor loading ของความสามารถมิติที่ 2 ของข้อสอบข้อที่ i และ τ_i คือ threshold ของข้อสอบข้อที่ i

1.4.3 แปลงค่า λ_{i1} , λ_{i2} และ τ_i ในแต่ละแฟ้มให้เป็นค่าพารามิเตอร์ข้อสอบดังนี้

a_{i1} หมายถึง อำนาจจำแนกของความสามารถมิติที่ 1 ของข้อสอบข้อที่ i

a_{i2} หมายถึง อำนาจจำแนกของความสามารถมิติที่ 2 ของข้อสอบข้อที่ i

b_i หมายถึง ความยากของข้อสอบข้อที่ i โดยใช้สูตรดังนี้

$$a_{i1} = \lambda_{i1} / \psi_i, \quad a_{i2} = \lambda_{i2} / \psi_i, \quad b_i = -\tau_i / \psi_i$$

เมื่อความสามารถทั้งสองมิติมีสหสัมพันธ์กัน $\text{Cor}(F_1, F_2) = \phi_{12}$ และ

$$\psi_i = \sqrt{1 - \lambda_i^2 - \lambda_k^2 - 2\lambda_i\lambda_k\phi_{12}}$$

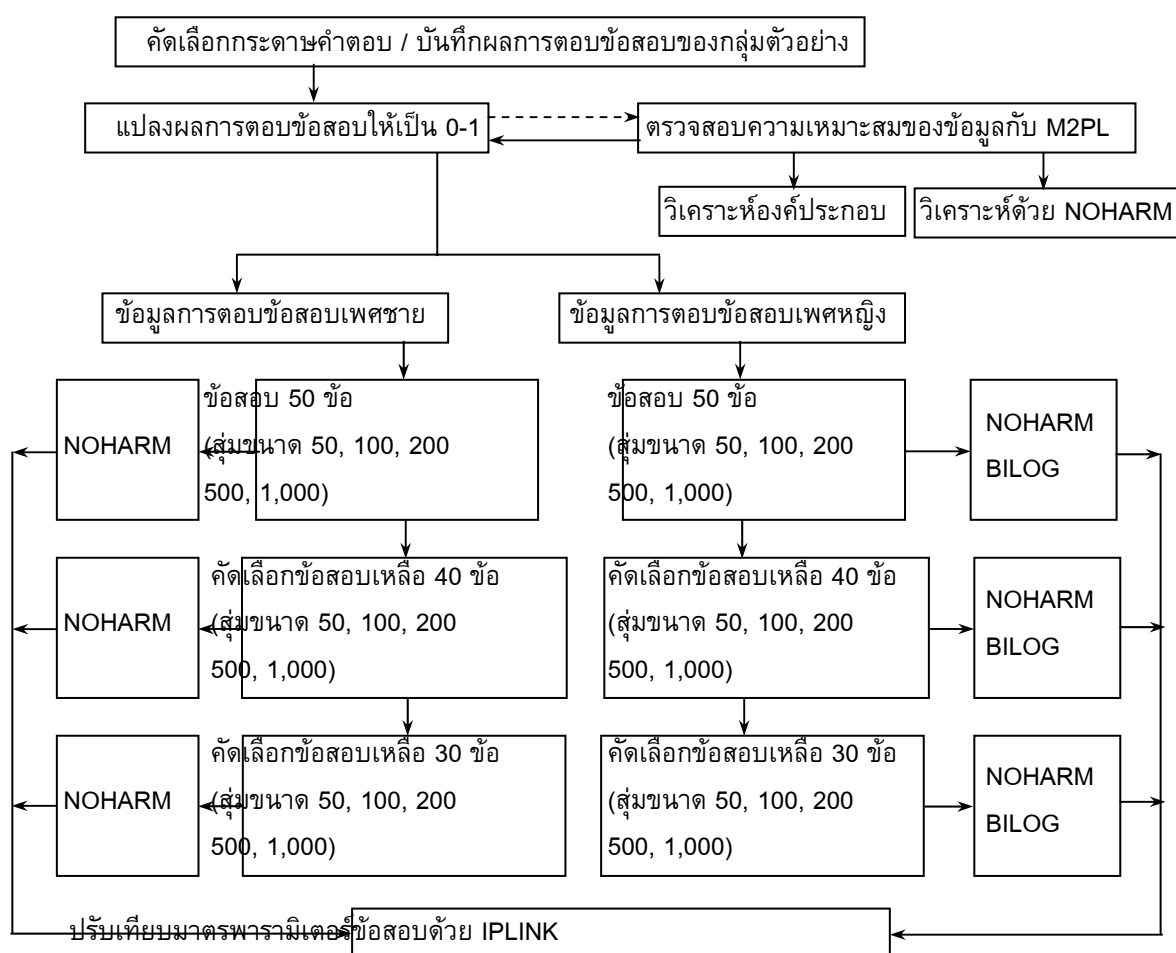
ดังนั้นต้องเตรียมแฟ้มข้อมูลพารามิเตอร์ข้อสอบจำแนกตามเพศจำนวน $2 \times 3 \times 5 \times 50 = 1,500$ แฟ้ม

1.4.4 ประเมินค่าสัมประสิทธิ์การเทียบมาตรฐาน (linking coefficient) เพื่อปรับเทียบมาตรฐาน (equate) ค่าพารามิเตอร์ข้อสอบของเพศชายและเพศหญิง โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป IPLINK (IPLINK : Multidimensional and Unidimensional Parameter Linking in Item Response Theory ; Lee & Oshima. 1996) บันทึกแฟ้มค่าสัมประสิทธิ์การเทียบมาตรฐาน (linking coefficient) จำนวน $3 \times 5 \times 50 = 750$ แฟ้ม

1.5 ประเมินค่าพารามิเตอร์ความสามารถหลัก(θ)ของกลุ่มตัวอย่างขนาดต่าง ๆ คือ 50 100 200 500 และ 1,000 คน เมื่อแบบทดสอบมีความยาวขนาดต่างกัน คือ 30 40 และ 50 ข้อ โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป BILOG Version 3.04 คำนวณเฉพาะค่าพารามิเตอร์ความสามารถหลัก(θ) ของเพศหญิง ซึ่งเป็นกลุ่มสนใจ (focal group) จึงต้องวิเคราะห์ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป BILOG

จำนวน $3 \times 5 \times 50 = 750$ รอบ แล้วนำผลวิเคราะห์ที่ได้มาจัดเตรียมเป็นแฟ้มข้อมูลพารามิเตอร์ความสามารถหลัก (θ) ของเพศหญิง จำนวน $3 \times 5 \times 50 = 750$ แฟ้ม

การจัดเตรียมข้อมูลในการศึกษาสรุปได้ดังภาพประกอบ 14



ภาพประกอบ 14 แสดงกระบวนการในการเตรียมข้อมูลในการศึกษา

2. การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลแบ่งเป็น 2 ส่วน คือ การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป SIBTEST และ DFIT และเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบด้วยวิธีการทั้งสอง

2.1 การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป SIBTEST และ DFIT

2.1.1 ตรวจสอบด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป SIBTEST ซึ่งใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อมูลที่ได้จากตอบข้อสอบที่ตรวจให้คะแนนรายข้อแบบสองค่า โดยเตรียมแฟ้มข้อมูล

การตอบข้อสอบของกลุ่มตัวอย่างที่มีการตรวจให้คะแนนรายข้อในการวิเคราะห์ให้มีรูปแบบตามที่โปรแกรมต้องการ แล้วดำเนินการตามขั้นตอนดังนี้

2.1.1.1 ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (Differential Item Functioning : DIF)

ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) ตามเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน แต่ละเงื่อนไขทำซ้ำ 50 ครั้ง โดยตรวจสอบการทำหน้าที่เบี่ยงเบนที่ระดับนัยสำคัญ .05

ก) เตรียมข้อมูลป้อนเข้า (input parameter file) ด้วยโปรแกรมย่อย SIBIN โดยใช้ข้อมูลการตอบข้อสอบรายข้อของกลุ่มตัวอย่างเพศชายเป็นกลุ่มอ้างอิง และกลุ่มตัวอย่างเพศหญิงเป็นกลุ่มสนใจ ต้องเตรียมแฟ้มข้อมูลป้อนเข้า $3 \times 5 \times 50 = 750$ แฟ้ม

ข) ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป SIBTEST โดยใช้ข้อมูลป้อนเข้าที่เตรียมด้วยโปรแกรม SIBIN และบันทึกเป็นแฟ้มข้อมูลผลลัพธ์จำนวน $3 \times 5 \times 50 = 750$ แฟ้ม

2.1.1.2 ตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ (Differential Bundle Functioning : DBF)

จัดหมวดข้อสอบตามเนื้อหาบทเรียนที่กำหนดในหลักสูตรเป็น 7 หมวด คือ การเขียนตัวเลขแทนจำนวน สมบัติของจำนวนนับ ทศนิยม การวัดและการประมาณเส้นตรงและมุม เศษส่วน ความยาว พื้นที่และปริมาตร โดยมีจำนวนข้อสอบในแต่ละหมวดในแบบทดสอบที่มีความยาว 50 40 และ 30 ข้อตามลำดับดังนี้

การเขียนตัวเลขแทนจำนวน	3	2	2	ข้อ
สมบัติของจำนวนนับ	6	5	4	ข้อ
ทศนิยม	9	7	5	ข้อ
การวัดและการประมาณ	5	4	3	ข้อ
เส้นตรงและมุม	7	6	4	ข้อ
เศษส่วน	12	10	7	ข้อ
ความยาว พื้นที่ และปริมาตร	8	6	5	ข้อ
รวม	50	40	30	ข้อ

ทำการวิเคราะห์การทำหน้าที่เบี่ยงเบนของหมวดข้อสอบในเงื่อนไข ความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดต่างแตกต่างกัน ที่ระดับนัยสำคัญ .05 โดยสุ่มกลุ่มตัวอย่างมาขนาดละ 1 กลุ่ม ต้องทำการตรวจสอบทั้งหมด $7 \times 3 \times 5 = 105$ ครั้ง

2.1.1.3 ตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบ (Differential Test Functioning : DTF)

ทำการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบในเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดต่างแตกต่างกัน โดยทำซ้ำเงื่อนไขละ 50 ครั้ง มีขั้นตอนดังนี้

- ก. คัดเลือกข้อสอบที่แสดง DIF ที่ระดับนัยสำคัญ .05 ออกจากแบบทดสอบ
- ข. ตรวจสอบการแสดงผล DIF ของข้อสอบที่เหลือในแบบทดสอบที่ระดับนัยสำคัญ .10 จำนวน $3 \times 5 \times 50 = 750$ รอบ และคัดเลือกข้อสอบที่แสดง DIF ที่ระดับ .10 ออกจากแบบทดสอบ

ค. ตรวจสอบการแสดงผลการทำหน้าที่เบี่ยงเบนของแบบทดสอบ (DTF) โดยพิจารณาข้อสอบที่แสดง DIF ที่ระดับนัยสำคัญ .05 และ .10 พร้อมกัน บันทึกแฟ้มผลการตรวจสอบ DTF ในเงื่อนไขต่าง ๆ จำนวน $3 \times 5 \times 50 = 750$ แฟ้ม

2.1.2 ตรวจสอบด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป DFIT ซึ่งใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อมูลที่ได้จากการตอบข้อสอบที่ตรวจให้คะแนนรายข้อทั้งแบบสองค่าและให้คะแนนแบบหลายค่า ในการวิเคราะห์ต้องเตรียมข้อมูลเบื้องต้น คือ พารามิเตอร์ข้อสอบของกลุ่มสนใจและกลุ่มอ้างอิง และพารามิเตอร์ความสามารถของกลุ่มสนใจ ดังรายละเอียดที่กล่าวในขั้นตอนการเตรียมข้อมูลในการศึกษา ใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป DFIT ตรวจสอบการทำหน้าที่เบี่ยงเบน 3 ระดับ คือ การทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) หมวดข้อสอบ (DBF) และแบบทดสอบ (DTF)

2.1.2.1 การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (Differential Item Functioning : DIF)

ตรวจสอบนัยสำคัญที่ระดับนัยสำคัญ .05 โดยการทำซ้ำเงื่อนไขละ 50 ครั้ง มีขั้นตอนดังนี้

1. เตรียมแฟ้มข้อมูลป้อนเข้า โดยเขียนคำสั่งตามรูปแบบที่โปรแกรมกำหนดพร้อมกับระบุค่าสัมประสิทธิ์การเทียบมาตรฐาน (linking coefficient) ของแต่ละกลุ่มในเงื่อนไขต่าง ๆ ต้องเตรียมแฟ้มข้อมูลป้อนเข้า $3 \times 5 \times 50 = 750$ แฟ้ม

2. ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป DFITP5 และบันทึกแฟ้มข้อมูลผลลัพธ์จำนวน $3 \times 5 \times 50 = 750$ แฟ้ม

2.1.2.2 การตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ (Differential Bundle Functioning : DBF)

จัดหมวดข้อสอบเพื่อวิเคราะห์เช่นเดียวกับการวิเคราะห์ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป SIBTEST ในขั้นตอนที่ 2.1.1.2 แล้วตรวจสอบด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป DFIT ที่ระดับนัยสำคัญ .05 ใช้กลุ่มตัวอย่างเดียวกับที่วิเคราะห์การทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป SIBTEST แต่ต้องเตรียมแฟ้มข้อมูลเพื่อการวิเคราะห์ โดยเตรียมแฟ้มข้อมูลพารามิเตอร์ข้อสอบของเพศชายและเพศหญิงซึ่งวิเคราะห์ด้วยโปรแกรม NOHARM จำแนกตามหมวดข้อสอบ และบันทึกแฟ้มข้อมูลจำนวน $2 \times 7 \times 3 \times 5 = 210$ แฟ้ม คำนวณสัมประสิทธิ์การเทียบมาตรฐานด้วยโปรแกรม IPLINK และบันทึกแฟ้มข้อมูลจำนวน $7 \times 3 \times 5 = 105$ แฟ้ม เตรียมแฟ้มข้อมูลป้อนเข้าตามที่โปรแกรมต้องการจำนวน $7 \times 3 \times 5 = 105$ แฟ้ม โดยระบุค่าสัมประสิทธิ์การเทียบมาตรฐานด้วยแล้ววิเคราะห์ด้วยโปรแกรม DFITP5 พร้อมกับบันทึกแฟ้มข้อมูลผลลัพธ์จำนวน 105 แฟ้ม

2.1.2.3 การตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบ (Differential Test Functioning : DTF)

โปรแกรมคอมพิวเตอร์สำเร็จรูป DFIT วิเคราะห์การทำหน้าที่เบี่ยงเบนของแบบทดสอบพร้อมกับการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ระดับนัยสำคัญ .05 จึงไม่ต้องวิเคราะห์ซ้ำแต่พิจารณาได้จากแฟ้มผลวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบในข้อ 2.1.2.1

2.2 เปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป SIBTEST และ DFIT

2.2.1 เปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีเดียวกัน

2.2.1.1 เปรียบเทียบความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธี SIBTEST เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน โดยประยุกต์เทคนิคการวิเคราะห์ความแปรปรวนแบบตัวแปรพหุ (multivariate analysis of variance : MANOVA) ทดสอบในแต่ละระดับความยาวของแบบทดสอบ วิเคราะห์โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป SPSS-PC for windows

2.2.1.2 เปรียบเทียบความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธี DFIT เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน โดยประยุกต์เทคนิคการวิเคราะห์ความแปรปรวนแบบตัวแปรพหุ (multivariate analysis of variance : MANOVA) ทดสอบในแต่ละระดับความยาวของแบบทดสอบ วิเคราะห์โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป SPSS-PC for windows

2.2.2 เปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบด้วยวิธีต่างกัน

2.2.2.1 เปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

2.2.2.1.1 เปรียบเทียบความถูกต้องและการระบุผิดพลาดในการตรวจสอบด้วยวิธี SIBTEST และวิธี DFIT โดยประยุกต์เทคนิคการวิเคราะห์ความแปรปรวนแบบตัวแปรพหุ กรณีมีกลุ่มที่ศึกษา 2 กลุ่ม (Hotelling's T^2) วิเคราะห์โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป SPSS-PC for windows

2.2.2.1.2 คำนวณความสอดคล้องในการตรวจสอบด้วยวิธี SIBTEST และวิธี DFIT

$$\text{ใช้สูตร ความสอดคล้อง} = \frac{R}{T} \times 100 \quad (\text{กาญจนา วัธนสุนทร. 2538 : 104})$$

เมื่อ R แทน จำนวนข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนที่ตรวจสอบพบซ้ำกันทั้งสองวิธี

T แทน จำนวนข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนที่ตรวจสอบพบทั้งหมดจากการตรวจสอบทั้งสองวิธี

2.2.2.2 เปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ

ทดสอบความแตกต่างของความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบด้วยวิธี SIBTEST และวิธี DFIT โดยสถิติ Z-test

$$\text{สูตร } z = \frac{p_1 - p_2}{\sqrt{pq \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (\text{Guilford and Fruchter. 1978 : 159})$$

เมื่อ p_1 แทน สัดส่วนความถูกต้องในการตรวจสอบ DBF ด้วยวิธี SIBTEST

p_2 แทน สัดส่วนความถูกต้องในการตรวจสอบ DBF ด้วยวิธี DFIT

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} \quad ; \quad q = 1 - P$$

2.2.2.3 เปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบ
ทดสอบความแตกต่างของความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของ
แบบทดสอบด้วยวิธี SIBTEST และ DFIT โดยใช้สถิติ Z-test

$$\text{สูตร } z = \frac{p_1 - p_2}{\sqrt{pq \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (\text{Guilford and Fruchter. 1978 : 159})$$

เมื่อ p_1 แทน สัดส่วนความถูกต้องในการตรวจสอบ DTF ด้วยวิธี SIBTEST

p_2 แทน สัดส่วนความถูกต้องในการตรวจสอบ DTF ด้วยวิธี DFIT

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} ; q = 1 - P$$

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อศึกษาวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนที่เหมาะสมกับข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ โดยมีวัตถุประสงค์เฉพาะเพื่อเปรียบเทียบวิธีซิปเทสต์และวิธีดีเอฟไอทีเมื่อความยาวของแบบทดสอบและขนาดของกลุ่มตัวอย่างแตกต่างกัน เครื่องมือที่ใช้เป็นแบบทดสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) ที่ผู้วิจัยสร้างขึ้นจำนวน 50 ข้อ มีข้อสอบจำนวน 16 ข้อ แสดงการทำหน้าที่เบี่ยงเบนต่อเพศชาย แล้วคัดเลือกข้อสอบตามสัดส่วนของเนื้อหาในตารางกำหนดข้อสอบ จัดทำเป็นแบบทดสอบที่มีความยาว 40 ข้อ และ 30 ข้อ ซึ่งแต่ละฉบับมีข้อสอบแสดงการทำหน้าที่เบี่ยงเบนต่อเพศชายจำนวน 16 ข้อ และ 14 ข้อ ตามลำดับ นำแบบทดสอบฉบับที่ประกอบด้วยข้อสอบ 50 ข้อไปทดสอบกับนักเรียนกลุ่มตัวอย่างเพศชายจำนวน 1,173 คน เพศหญิงจำนวน 1,086 คน ใช้นักเรียนทั้งหมดของแต่ละเพศเป็นประชากรเทียบ สุ่มกลุ่มตัวอย่างแยกเพศชายและเพศหญิงให้มีขนาด 50 100 200 500 และ 1,000 คน สุ่มกลุ่มตัวอย่างแบบใส่คืนขนาดละ 50 ครั้ง การศึกษาในเงื่อนไขที่แบบทดสอบประกอบด้วยข้อสอบ 40 ข้อและ 30 ข้อใช้ข้อมูลการตอบข้อสอบจำนวน 50 ข้อของกลุ่มตัวอย่างที่กำหนดให้เป็นประชากรเทียบมาคัดเลือกเฉพาะคำตอบข้อสอบที่ต้องการจำนวน 40 ข้อและ 30 ข้อ แล้วสุ่มกลุ่มตัวอย่างขนาดต่าง ๆ โดยการสุ่มแบบใส่คืนขนาดละ 50 ครั้ง หลังจากนั้นวิเคราะห์ข้อมูลการตอบข้อสอบของกลุ่มตัวอย่างที่เตรียมไว้ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป SIBTEST และ DFIT ผลการวิเคราะห์ข้อมูลครั้งนี้ผู้วิจัยนำเสนอเป็น 3 ตอน คือ

ตอนที่ 1 ค่าสถิติพื้นฐานของข้อมูลการตอบข้อสอบของนักเรียนเพศชายและเพศหญิง เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ

ตอนที่ 2 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบด้วยวิธีซิปเทสต์และวิธีดีเอฟไอทีในเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

ตอนที่ 3 เปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีเดียวกัน และเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบด้วยวิธีต่างกัน

เพื่อให้การนำเสนอผลการวิเคราะห์ข้อมูลมีความเข้าใจตรงกัน จึงกำหนดสัญลักษณ์ดังนี้

$\bar{X}$	แทน ค่าเฉลี่ย
S.D.	แทน ค่าส่วนเบี่ยงเบนมาตรฐาน
R	แทน กลุ่มอ้างอิง (เพศชาย)
F	แทน กลุ่มสนใจ (เพศหญิง)
DIF	แทน การทำหน้าที่เบี่ยงเบนของข้อสอบ
DBF	แทน การทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ
DTF	แทน การทำหน้าที่เบี่ยงเบนของแบบทดสอบ
θ	แทน ความสามารถหลักในการตอบข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1
P	แทน ความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ
E	แทน การระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

ตอนที่ 1 ค่าสถิติพื้นฐานของข้อมูลการตอบข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษา ปีที่ 1 (ค 101)

หลังจากตรวจให้คะแนนการตอบข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) จากแบบทดสอบที่ประกอบด้วยข้อสอบ 50 ข้อ ของนักเรียนกลุ่มตัวอย่างเพศชายจำนวน 1,173 คน และนักเรียนเพศหญิงจำนวน 1,086 คน ที่กำหนดให้เป็นประชากรเรียบร้อยแล้ว คัดเลือกข้อสอบตามสัดส่วนในแผนผังกำหนดข้อสอบให้เหลือข้อสอบในแบบทดสอบจำนวน 40 ข้อและ 30 ข้อตามลำดับ แล้วคำนวณค่าสถิติพื้นฐานของการตอบข้อสอบคะแนนสอบของนักเรียนที่เป็นประชากรเทียมเมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อและประมาณค่าพารามิเตอร์ความสามารถหลัก (θ) ในการตอบข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101)

1.1 ค่าสถิติพื้นฐานของคะแนนจากการตอบข้อสอบของนักเรียนกลุ่มตัวอย่างที่กำหนดให้เป็นประชากรเทียม

ค่าสถิติพื้นฐานของคะแนนการตอบข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 ของนักเรียนกลุ่มตัวอย่างที่กำหนดให้เป็นประชากรเทียม เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ แสดงในตารางที่ 10-12

ตาราง 10 ค่าสถิติพื้นฐานของคะแนนจากการตอบข้อสอบของนักเรียนเพศชายและเพศหญิง เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ

เพศ	จำนวน (คน)	คะแนนสูงสุด	คะแนนต่ำสุด	พิสัย	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	ความโด่ง	ความเบ้
ชาย	1,173	30	1	29	8.691	3.619	2.065	1.077
หญิง	1,086	28	1	27	8.700	3.796	2.379	1.210
รวม	2,259	30	1	29	8.472	3.537	2.149	1.114

ผลวิเคราะห์ตามตาราง 10 แสดงให้เห็นว่าค่าสถิติพื้นฐานของคะแนนจากการตอบข้อสอบ เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ ของนักเรียนเพศชายและเพศหญิงมีค่าใกล้เคียงกัน

ตาราง 11 ค่าสถิติพื้นฐานของคะแนนจากการตอบข้อสอบของนักเรียนเพศชายและเพศหญิง เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ

เพศ	จำนวน (คน)	คะแนนสูงสุด	คะแนนต่ำสุด	พิสัย	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	ความโด่ง	ความเบ้
ชาย	1,173	37	1	36	11.425	4.621	2.370	1.260
หญิง	1,086	36	2	34	11.904	5.069	2.116	1.261
รวม	2,259	37	1	36	11.656	4.846	2.285	1.273

ผลวิเคราะห์ตามตาราง 11 แสดงให้เห็นว่าค่าสถิติพื้นฐานของคะแนนจากการตอบข้อสอบ เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ ของนักเรียนเพศชายและเพศหญิงมีค่าใกล้เคียงกัน

ตาราง 12 ค่าสถิติพื้นฐานของคะแนนจากการตอบข้อสอบของนักเรียนเพศชายและเพศหญิง เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ

เพศ	จำนวน (คน)	คะแนนสูงสุด	คะแนนต่ำสุด	พิสัย	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	ความโด่ง	ความเบ้
ชาย	1,173	46	3	43	14.111	5.872	2.720	1.433
หญิง	1,086	45	3	42	14.890	6.460	2.229	1.388
รวม	2,259	46	3	43	14.482	6.171	2.508	1.423

ผลวิเคราะห์ตามตาราง 12 แสดงให้เห็นว่าค่าสถิติพื้นฐานของคะแนนจากการตอบข้อสอบ เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ ของนักเรียนเพศชายและเพศหญิงมีค่าใกล้เคียงกัน

1.2 ค่าสถิติพื้นฐานของพารามิเตอร์ความสามารถหลัก (θ) ในการตอบข้อสอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 (ค 101) ของนักเรียนกลุ่มตัวอย่างที่กำหนดให้เป็นประชากรเทียม

หลังจากประมาณค่าพารามิเตอร์ความสามารถหลัก(θ)ในการตอบข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) ของนักเรียนกลุ่มตัวอย่างที่กำหนดให้เป็นประชากรเทียมแล้วคำนวณค่าสถิติพื้นฐานของพารามิเตอร์ความสามารถหลัก (θ) ในการตอบข้อสอบวิชาคณิตศาสตร์เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50ข้อ ได้ผลดังแสดงในตาราง 13-15

ตาราง 13 ค่าสถิติพื้นฐานของพารามิเตอร์ความสามารถหลัก (θ) ในการตอบข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) ของนักเรียนกลุ่มตัวอย่าง เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ

เพศ	จำนวน (คน)	θ สูงสุด	θ ต่ำสุด	พิสัย	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	ความโด่ง	ความเบ้
ชาย	1,173	4.179	-2.440	6.619	0.000	1.000	0.722	0.829
หญิง	1,086	3.980	-2.186	6.166	0.000	1.000	0.834	0.867
รวม	2,259	4.179	-2.440	6.619	0.000	1.000	0.806	0.845

ผลวิเคราะห์ตามตาราง 13 แสดงให้เห็นว่าค่าสถิติพื้นฐานของพารามิเตอร์ความสามารถหลัก (θ) ของนักเรียนเพศชายและเพศหญิง เมื่อตอบแบบทดสอบที่ประกอบด้วยข้อสอบ 30 ข้อมีค่าใกล้เคียงกัน

ตาราง 14 ค่าสถิติพื้นฐานของพารามิเตอร์ความสามารถหลัก (θ) ในการตอบข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) ของนักเรียนกลุ่มตัวอย่าง เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ

เพศ	จำนวน (คน)	θ สูงสุด	θ ต่ำสุด	พิสัย	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	ความโด่ง	ความเบ้
ชาย	1,173	4.090	-2.230	6.320	0.000	1.000	1.409	0.994
หญิง	1,086	3.945	-1.960	5.026	0.000	1.000	1.081	1.026
รวม	2,259	4.090	-2.230	6.320	0.000	1.000	1.123	1.018

ผลวิเคราะห์ตามตาราง 14 แสดงให้เห็นว่าค่าสถิติพื้นฐานของพารามิเตอร์ความสามารถหลัก (θ) ของนักเรียนเพศชายและเพศหญิง เมื่อตอบแบบทดสอบที่ประกอบด้วยข้อสอบ 40 ข้อมีค่าใกล้เคียงกัน

ตาราง 15 ค่าสถิติพื้นฐานของพารามิเตอร์ความสามารถหลัก (θ) ในการตอบข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) ของนักเรียนกลุ่มตัวอย่าง เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ

เพศ	จำนวน (คน)	θ สูงสุด	θ ต่ำสุด	พิสัย	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	ความโด่ง	ความเบ้
ชาย	1,173	4.132	-2.290	6.422	0.000	1.000	1.447	1.147
หญิง	1,086	4.013	-2.099	6.112	0.000	1.000	1.33	1.149
รวม	2,259	4.132	-2.290	6.422	0.000	1.000	1.374	1.540

ผลวิเคราะห์ตามตาราง 15 แสดงให้เห็นว่าค่าสถิติพื้นฐานของพารามิเตอร์ความสามารถหลัก (θ) ของนักเรียนเพศชายและเพศหญิง เมื่อตอบแบบทดสอบที่ประกอบด้วยข้อสอบ 50 ข้อมีค่าใกล้เคียงกัน

ตอนที่ 2 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ ด้วยวิธีชิบเทสต์และวิธีดีเอฟไอทีในเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

หลังจากเตรียมข้อมูลการตอบข้อสอบในเงื่อนไขความยาวแบบทดสอบ 30 40 และ 50 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คนแล้วตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป SIBTEST และ DFIT ได้ผลดังต่อไปนี้

2.1 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ จากกลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน

เมื่อตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอทีในเงื่อนไขต่าง ๆ เรียบร้อยแล้ว ผู้วิจัยคำนวณความถูกต้องในการตรวจสอบและการระบุผิดพลาดในการตรวจสอบของแต่ละวิธีในเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน ได้ผลดังรายละเอียดในภาคผนวก ฅ และผลสรุปดังแสดงในตาราง 16

ตาราง 16 สรุปผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอทีใน
เงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

ขนาดกลุ่ม ตัวอย่าง ความ ยาวแบบ ทดสอบ (ข้อ)	50		100		200		500		1,000	
	P		E		P		E		P	
	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.
30 SIBTEST	3.86	5.04	3.75	5.21	3.71	5.63	3.75	4.03	4.43	6.27
DFIT	38.09	25.75	50.14	25.51	33.78	27.68	51.28	23.95	29.73	31.61
40 SIBTEST	4.25	5.71	4.17	3.74	3.84	4.56	4.50	4.44	4.31	4.75
DFIT	44.01	18.81	61.59	17.69	38.28	15.12	60.25	13.72	41.38	18.12
50 SIBTEST	5.38	6.44	3.23	63.92	4.33	7.65	5.12	4.28	6.63	5.98
DFIT	43.70	16.58	2.86	12.98	48.34	18.43	65.36	17.62	4.21	18.51

ผลการวิเคราะห์ตามตาราง 16 แสดงให้เห็นว่าการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบใน
เงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกันได้ผลดังนี้

กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ เมื่อใช้วิธีชิบเทสต์กับกลุ่มตัวอย่าง 50 คน พบว่า
ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 3.86 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 5.04 การ
ระบุผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 3.75 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 5.21
เมื่อใช้กลุ่มตัวอย่าง 100 คน พบว่าความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 3.71 ส่วนเบี่ยงเบน
มาตรฐานคิดเป็นร้อยละ 5.63 การระบุผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 3.75 ส่วน
เบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 4.03 เมื่อใช้กลุ่มตัวอย่าง 200 คน พบว่าความถูกต้องในการตรวจสอบมี
ค่าเฉลี่ยคิดเป็นร้อยละ 4.43 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 6.27 การระบุผิดพลาดในการตรวจสอบ
มีค่าเฉลี่ยคิดเป็นร้อยละ 4.85 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 4.93 เมื่อใช้กลุ่มตัวอย่าง 500 คน
พบว่าความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 7.57 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ
7.56 การระบุผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 10.00 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อย
ละ 5.50 เมื่อใช้กลุ่มตัวอย่าง 1,000 คน พบว่าความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 11.57
ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 5.93 การระบุผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 13.13
ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 5.96 เมื่อใช้วิธีดีเอฟไอทีกับกลุ่มตัวอย่าง 50 คน พบว่าความถูกต้อง
ในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 38.09 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 25.75 การระบุ
ผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 50.14 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 25.51 เมื่อ
ใช้กลุ่มตัวอย่าง 100 คน พบว่าความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 33.78 ส่วนเบี่ยงเบน
มาตรฐานคิดเป็นร้อยละ 27.68 การระบุผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 51.28 ส่วน
เบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 23.95 เมื่อใช้กลุ่มตัวอย่าง 200 คน พบว่าความถูกต้องในการตรวจสอบมี
ค่าเฉลี่ยคิดเป็นร้อยละ 29.73 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 31.61 การระบุผิดพลาดในการ
ตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 41.03 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 25.76 เมื่อใช้กลุ่มตัวอย่าง

[illegible]

กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ เมื่อใช้วิธีซิปเทสท์กับกลุ่มตัวอย่าง 50 คน พบว่า ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 5.38 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 6.44 การ ระบุผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 3.23 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 6.44 เมื่อใช้กลุ่มตัวอย่าง 100 คน พบว่าความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 4.33 ส่วนเบี่ยงเบน มาตรฐานคิดเป็นร้อยละ 7.65 การระบุผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 5.12 ส่วน เบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 4.28 เมื่อใช้กลุ่มตัวอย่าง 200 คน พบว่าความถูกต้องในการตรวจสอบมี ค่าเฉลี่ยคิดเป็นร้อยละ 6.63 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 5.98 การระบุผิดพลาดในการตรวจสอบ

มีค่าเฉลี่ยคิดเป็นร้อยละ 5.76 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 4.32 เมื่อใช้กลุ่มตัวอย่าง 500 คน พบว่าความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 4.75 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 5.73 การระบุผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 8.53 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 4.13 เมื่อใช้กลุ่มตัวอย่าง 1,000 คน พบว่าความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 8.75 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 3.79 การระบุผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 10.23 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 2.99 เมื่อใช้วิธีดีเอฟไอทีกับกลุ่มตัวอย่าง 50 คน พบว่าความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 43.70 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 16.58 การระบุผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 63.92 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 12.98 เมื่อใช้กลุ่มตัวอย่าง 100 คน พบว่าความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 48.34 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 18.43 การระบุผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 65.36 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 17.62 เมื่อใช้กลุ่มตัวอย่าง 200 คน พบว่าความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 47.21 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 18.51 การระบุผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 68.58 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 15.26 เมื่อใช้กลุ่มตัวอย่าง 500 คน พบว่าความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 42.15 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 18.46 การระบุผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 66.27 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 12.14 เมื่อใช้กลุ่มตัวอย่าง 1,000 คน พบว่าความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 41.84 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 13.52 การระบุผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 61.18 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 13.29

2.2 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ จากกลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ จัดหมวดข้อสอบตามเนื้อหาที่กำหนดในหลักสูตรเป็น 7 หมวดแล้วตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบแต่ละหมวดด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที ในเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน ได้ผลตั้งรายละเอียดในภาคผนวก ญ และผลสรุปดังแสดงในตาราง 17

ตาราง 17 สรุปผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอทีในเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

ความยาว แบบทดสอบ (ข้อ)		จำนวนหมวด	ขนาดกลุ่มตัวอย่าง									
			50		100		200		500		1,000	
			(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
30	SIBTEST	7	-	5	-	5	-	4	-	5	-	6
	DFIT	7	-	5	-	5	-	3	-	4	-	0
40	SIBTEST	7	-	5	-	5	-	5	-	6	-	6
	DFIT	7	-	4	-	3	-	3	-	2	1	2
50	SIBTEST	7	-	6	-	5	-	5	-	5	-	6
	DFIT	7	-	3	-	4	-	3	1	3	-	5

หมายเหตุ (1) หมายถึง จำนวนหมวดข้อสอบที่คำนวณค่าไม่ได้

(2) หมายถึง จำนวนหมวดข้อสอบที่ตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบได้ถูกต้อง

ผลการวิเคราะห์ตามตาราง 17 แสดงให้เห็นว่าการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบในเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกันได้ผลดังนี้

กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ เมื่อใช้วิธีชิบเทสต์พบว่า กลุ่มตัวอย่างขนาด 50 คน ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 5 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 100 คน ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 5 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 200 คน ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 4 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 500 คน ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 5 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 1,000 คน ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 6 หมวด จาก 7 หมวด เมื่อใช้วิธีดีเอฟไอทีพบว่า กลุ่มตัวอย่างขนาด 50 คน ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 5 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 100 คน ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 5 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 200 คน ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 3 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 500 คน ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 4 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 1,000 คน ตรวจสอบได้ไม่ตรงกับความเห็นของผู้เชี่ยวชาญเลย

กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ เมื่อใช้วิธีชิบเทสต์พบว่า กลุ่มตัวอย่างขนาด 50 คน ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 5 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 100 คน ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 5 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 200 คน ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 4 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 500 คน ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 6 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 1,000 คน ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 6 หมวด จาก 7 หมวด เมื่อใช้วิธีดีเอฟไอทีพบว่า กลุ่มตัวอย่างขนาด 50 คน ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 4 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 100 คน ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 3 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 200 คน ตรวจสอบได้ตรง

กับความเห็นของผู้เชี่ยวชาญ 3 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 500 คน ตรวจสอบได้ตรงกับ
ความเห็นของผู้เชี่ยวชาญ 2 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 1,000 คน คำนวณค่าไม่ได้ 1 หมวด
ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 2 หมวด จาก 6 หมวด

กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ เมื่อใช้วิธีซิปเทสต์พบว่า กลุ่มตัวอย่างขนาด 50
คน ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 6 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 100 คน
ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 5 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 200 คน
ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 5 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 500 คน ตรวจสอบ
ได้ตรงกับความเห็นของผู้เชี่ยวชาญ 5 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 1,000 คน ตรวจสอบได้ตรง
กับความเห็นของผู้เชี่ยวชาญ 6 หมวด จาก 7 หมวด เมื่อใช้วิธีดีเอฟไอทีพบว่า กลุ่มตัวอย่างขนาด 50 คน
ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 3 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 100 คน ตรวจสอบ
ได้ตรงกับความเห็นของผู้เชี่ยวชาญ 4 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 200 คน ตรวจสอบได้ตรง
กับความเห็นของผู้เชี่ยวชาญ 3 หมวด จาก 7 หมวด กลุ่มตัวอย่างขนาด 500 คน คำนวณค่าไม่ได้ 1 หมวด
ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 3 หมวด จาก 6 หมวด กลุ่มตัวอย่างขนาด 1,000 คน
ตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ 5 หมวด จาก 7 หมวด

2.3 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบด้วยวิธีซิปเทสต์และวิธีดีเอฟไอที
เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ จากกลุ่มตัวอย่างขนาด 50 100 200 500 และ
1,000 คน

การตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบเป็นการตรวจสอบการทำหน้าที่
เบี่ยงเบนของข้อสอบทุกข้อในแบบทดสอบทั้งฉบับพร้อมกัน ในเงื่อนไขความยาวแบบทดสอบและกลุ่ม
ตัวอย่างขนาดแตกต่างกันแต่ละเงื่อนไขทำการตรวจสอบ 50 ครั้ง ได้ผลดังรายละเอียดในภาคผนวก ก และ
ผลสรุปดังแสดงในตาราง 18

ตาราง 18 สรุปผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบด้วยวิธีซิปเทสต์และวิธีดีเอฟไอทีใน
เงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

ความยาว แบบทดสอบ (ข้อ)	วิธี	จำนวนครั้ง	ขนาดกลุ่มตัวอย่าง									
			50		100		200		500		1,000	
			(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
30	SIBTEST	50	-	37	-	33	-	39	-	46	-	49
	DFIT	50	5	35	2	39	1	42	1	45	-	47
40	SIBTEST	50	-	43	-	40	-	47	-	21	-	1
	DFIT	50	-	35	2	35	-	47	1	44	-	46
50	SIBTEST	50	-	43	-	44	-	49	-	50	-	50
	DFIT	50	-	44	1	36	3	39	1	46	-	46

หมายเหตุ (1) หมายถึง จำนวนครั้งที่คำนวณค่าไม่ได้

(2) หมายถึง จำนวนครั้งที่ตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบได้ถูกต้อง

ผลการวิเคราะห์ตามตาราง 18 แสดงให้เห็นว่าการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบในเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกันได้ผลดังนี้

กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ เมื่อใช้กลุ่มตัวอย่าง 50 คน วิธีชิบเทสต์ตรวจพบ DTF จำนวน 37 ครั้ง วิธีดีเอฟไอที่คำนวณค่าไม่ได้ 5 ครั้ง ตรวจพบ DTF จำนวน 35 ครั้ง กลุ่มตัวอย่าง 100 คน วิธีชิบเทสต์ตรวจพบ DTF จำนวน 33 ครั้ง วิธีดีเอฟไอที่คำนวณค่าไม่ได้ 2 ครั้ง ตรวจพบ DTF จำนวน 39 ครั้ง กลุ่มตัวอย่าง 200 คน วิธีชิบเทสต์ตรวจพบ DTF จำนวน 39 ครั้ง วิธีดีเอฟไอที่คำนวณค่าไม่ได้ 1 ครั้ง ตรวจพบ DTF จำนวน 42 ครั้ง กลุ่มตัวอย่าง 500 คน วิธีชิบเทสต์ตรวจพบ DTF จำนวน 46 ครั้ง วิธีดีเอฟไอที่คำนวณค่าไม่ได้ 1 ครั้ง ตรวจพบ DTF จำนวน 45 ครั้ง กลุ่มตัวอย่าง 1,000 คน วิธีชิบเทสต์ตรวจพบ DTF จำนวน 49 ครั้ง วิธีดีเอฟไอที่ตรวจพบ DTF จำนวน 47 ครั้ง

กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ เมื่อใช้กลุ่มตัวอย่าง 50 คน วิธีชิบเทสต์ตรวจพบ DTF จำนวน 43 ครั้ง วิธีดีเอฟไอที่ตรวจพบ DTF จำนวน 35 ครั้ง กลุ่มตัวอย่าง 100 คน วิธีชิบเทสต์ตรวจพบ DTF จำนวน 40 ครั้ง วิธีดีเอฟไอที่คำนวณค่าไม่ได้ 2 ครั้ง ตรวจพบ DTF จำนวน 35 ครั้ง กลุ่มตัวอย่าง 200 คน วิธีชิบเทสต์ตรวจพบ DTF จำนวน 47 ครั้ง วิธีดีเอฟไอที่ตรวจพบ DTF จำนวน 47 ครั้ง กลุ่มตัวอย่าง 500 คน วิธีชิบเทสต์ตรวจพบ DTF จำนวน 21 ครั้ง วิธีดีเอฟไอที่คำนวณค่าไม่ได้ 1 ครั้ง ตรวจพบ DTF จำนวน 44 ครั้ง กลุ่มตัวอย่าง 1,000 คน วิธีชิบเทสต์ตรวจพบ DTF จำนวน 1 ครั้ง วิธีดีเอฟไอที่ตรวจพบ DTF จำนวน 46 ครั้ง

กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ เมื่อใช้กลุ่มตัวอย่าง 50 คน วิธีชิบเทสต์ตรวจพบ DTF จำนวน 43 ครั้ง วิธีดีเอฟไอที่ตรวจพบ DTF จำนวน 44 ครั้ง กลุ่มตัวอย่าง 100 คน วิธีชิบเทสต์ตรวจพบ DTF จำนวน 44 ครั้ง วิธีดีเอฟไอที่คำนวณค่าไม่ได้ 1 ครั้ง ตรวจพบ DTF จำนวน 36 ครั้ง กลุ่มตัวอย่าง 200 คน วิธีชิบเทสต์ตรวจพบ DTF จำนวน 49 ครั้ง วิธีดีเอฟไอที่คำนวณค่าไม่ได้ 3 ครั้ง ตรวจพบ DTF จำนวน 39 ครั้ง กลุ่มตัวอย่าง 500 คน วิธีชิบเทสต์ตรวจพบ DTF จำนวน 50 ครั้ง วิธีดีเอฟไอที่คำนวณค่าไม่ได้ 1 ครั้ง ตรวจพบ DTF จำนวน 46 ครั้ง กลุ่มตัวอย่าง 1,000 คน วิธีชิบเทสต์ตรวจพบ DTF จำนวน 50 ครั้ง วิธีดีเอฟไอที่ตรวจพบ DTF จำนวน 46 ครั้ง

ตอนที่ 3 เปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ

และแบบทดสอบ ด้วยวิธีชิบเทสต์และวิธีดีเอฟไอ

หลังจากตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบและแบบทดสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที่ได้นำผลการตรวจสอบของแต่ละวิธีมาเปรียบเทียบกันดังต่อไปนี้

3.1 การเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

3.1.1 การเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีเดียวกัน

การเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีเดียวกัน

ผู้วิจัยประยุกต์เทคนิคการวิเคราะห์ความแปรปรวนแบบตัวแปรพหุ (multivariate analysis of variance : MANOVA) เพื่อศึกษาผลของกลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน ที่มีต่อตัวแปรตาม 2 ตัวพร้อมกัน คือ ความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (P) และการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (E) เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ โดยกำหนดเกณฑ์ในการพิจารณาดังนี้ เงื่อนไขที่ต้องการมากที่สุดคือ กรณีที่มีความถูกต้องในการ

ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบสูง แต่มีการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบต่ำ ดังแสดงในภาพประกอบ 15

		ความถูกต้อง	
		สูง	ต่ำ
การระบุผิดพลาด	สูง	X	X
	ต่ำ	✓✓	X

ภาพประกอบ 15 แสดงเกณฑ์ในการพิจารณาเพื่อยอมรับความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

จากภาพประกอบ 15 แสดงให้เห็นว่าในการพิจารณาผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยการพิจารณาความถูกต้องในการตรวจสอบและการระบุผิดพลาดในการตรวจสอบพร้อมกันนั้น เงื่อนไขที่ต้องการมากที่สุดคือกรณีที่มีความถูกต้องในการตรวจสอบสูงแต่มีการระบุผิดพลาดต่ำ เงื่อนไขที่ไม่ต้องการคือมีความถูกต้องในการตรวจสอบต่ำ แต่มีการระบุผิดพลาดสูง มีความถูกต้องในการตรวจสอบสูงและมีการระบุผิดพลาดสูง มีความถูกต้องในการตรวจสอบต่ำและมีการระบุผิดพลาดต่ำ

ผู้วิจัยใช้เกณฑ์ข้างต้นในการเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีเดียวกัน ได้ผลดังต่อไปนี้

3.1.1.1 เปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธี

ชิบเทสท์

การประยุกต์เทคนิคการวิเคราะห์ความแปรปรวนแบบตัวแปรพหุในการเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสท์เพื่อพิจารณาความถูกต้องในการตรวจสอบและการระบุผิดพลาดในการตรวจสอบพร้อมกันด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป SPSS-PC for windows โปรแกรมจะเริ่มต้นด้วยการทดสอบนัยสำคัญของตัวแปรตาม(PและE)เมื่อกลุ่มตัวอย่างมีขนาดแตกต่างกันด้วยสถิติทดสอบพหุตัวแปร ในการศึกษาค้นคว้าครั้งนี้พิจารณาค่าสถิติ Wilks (λ) ถ้าพบว่ามีความสำคัญก็จะทำการทดสอบความแตกต่างระหว่างเวกเตอร์ค่าเฉลี่ยของตัวแปรตาม(Pและ E)จากกลุ่มตัวอย่างขนาดต่างๆเป็นรายคู่ ถ้าพบว่าคู่ใดมีความสำคัญก็จะทำการทดสอบความแตกต่างของตัวแปรตาม(Pและ E)เป็นรายตัวต่อไป การเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบได้ผลดังต่อไปนี้

3.1.1.1.1 กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ กลุ่มตัวอย่างมีขนาด 50 100 200 500 และ 1,000 คน

การทดสอบนัยสำคัญของตัวแปรตาม(P และ E)กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ กลุ่มตัวอย่างมีขนาด 50 100 200 500 และ 1,000 คน เมื่อพิจารณาค่าสถิติ Wilks ซึ่งมีค่า .561พบว่ามีความสำคัญทางสถิติที่ระดับ .05 จึงทดสอบความแตกต่างระหว่างเวกเตอร์ค่าเฉลี่ยของ P และ E จากกลุ่มตัวอย่างขนาดต่างๆเป็นรายคู่ ได้ผลดังแสดงในตาราง 19

ตาราง 19 แสดงผลการทดสอบความแตกต่างระหว่างเวกเตอร์ค่าเฉลี่ยของความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิปเทสต์เป็นรายคู่ กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ

ขนาดกลุ่มตัวอย่าง (คน)	1 (50)	2 (100)	3 (200)	4 (500)	5 (1,000)
1 (50)	-	.069	.655	21.283	57.699
2 (100)			1.147	23.741	61.554
3 (200)			-	14.488	46.365
4 (500)				-	9.744
5 (1,000)					-

* มีนัยสำคัญที่ระดับ .05

ผลการวิเคราะห์ตามตาราง 19 พบว่ากรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ เมื่อตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิปเทสต์กลุ่มตัวอย่างขนาด 50 100 และ 200 คน มีผลต่อตัวแปรตาม (P และ E) โดยรวมไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่กลุ่มตัวอย่างขนาด 50 100 และ 200 คน มีผลต่อตัวแปรตาม (P และ E) โดยรวมแตกต่างจากกลุ่มตัวอย่างขนาด 500 และ 1,000 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และ กลุ่มตัวอย่างขนาด 500 คน มีผลต่อตัวแปรตาม (P และ E) โดยรวมแตกต่างจากกลุ่มตัวอย่างขนาด 1,000 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ผลการทดสอบความแตกต่างระหว่างเวกเตอร์ค่าเฉลี่ยของตัวแปรตาม (P และ E) จากกลุ่มตัวอย่างขนาดต่างๆเป็นรายคู่พบว่า มีนัยสำคัญทางสถิติหลายคู่จึงต้องทดสอบความแตกต่างของตัวแปรตาม(PและE)เป็นรายตัว ได้ผลดังแสดงในตาราง 20

ตาราง 20 แสดงผลการทดสอบความแตกต่างของความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิปเทสต์เป็นรายตัว กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ

ตัวแปร	$\bar{X}$	S.D.	t	ตัวแปร	$\bar{X}$	S.D.	t
(4) กลุ่มตัวอย่าง 500 คน – (1) กลุ่มตัวอย่าง 50 คน							
P				E			
(1)	3.856	5.043	2.892*	(1)	3.750	5.206	5.833*
(4)	7.571	7.555		(4)	10.000	5.504	
(5) กลุ่มตัวอย่าง 1,000 คน – (1) กลุ่มตัวอย่าง 50 คน							
P				E			
(1)	3.856	5.043	7.153*	(1)	3.750	5.206	8.450*
(5)	11.715	5.910		(5)	13.250	6.008	
(4) กลุ่มตัวอย่าง 500 คน – (2) กลุ่มตัวอย่าง 100 คน							
P				E			
(2)	3.714	5.634	2.894*	(2)	3.375	4.035	6.864*
(4)	7.571	7.555		(4)	10.000	5.504	
(5) กลุ่มตัวอย่าง 1,000 คน – (2) กลุ่มตัวอย่าง 100 คน							
P				E			
(2)	3.714	5.634	6.930*	(2)	3.375	4.035	9.648*
(5)	11.715	5.910		(5)	13.250	6.008	
(4) กลุ่มตัวอย่าง 500 คน – (3) กลุ่มตัวอย่าง 200 คน							
P				E			
(3)	4.422	6.262	2.269*	(3)	4.875	4.937	4.901*
(4)	7.571	7.555		(4)	10.000	5.504	
(5) กลุ่มตัวอย่าง 1,000 คน – (3) กลุ่มตัวอย่าง 200 คน							
P				E			
(3)	4.422	6.262	5.989*	(3)	4.875	4.937	7.615*
(5)	11.715	5.910		(5)	13.250	6.008	
(5) กลุ่มตัวอย่าง 1,000 คน – (4) กลุ่มตัวอย่าง 500 คน							
P				E			
(4)	7.571	7.555	3.055*	(4)	10.000	5.504	2.820*
(5)	11.715	5.909		(5)	13.250	6.008	

* มีนัยสำคัญที่ระดับ .05

ผลวิเคราะห์ตามตาราง 20 แสดงให้เห็นว่ากรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ กลุ่มตัวอย่างขนาด 500 และ 1,000 คน ส่งผลให้วิธีซิปเทสที่มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบสูง และมีการระบุผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 50 100 และ 200 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และกลุ่มตัวอย่างขนาด 1,000 คน ส่งผลให้วิธีซิปเทสที่มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบสูงและมีการระบุผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 500 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3.1.1.1.2 กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ กลุ่มตัวอย่างมีขนาด 50 100 200 500 และ 1,000 คน

การทดสอบนัยสำคัญของตัวแปรตาม(P และ E)กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ กลุ่มตัวอย่างมีขนาด 50 100 200 500 และ 1,000 คน เมื่อพิจารณาค่าสถิติ Wilks ซึ่งมีค่า .654 พบว่ามีนัยสำคัญทางสถิติที่ระดับ.05 จึงทดสอบความแตกต่างระหว่างเวกเตอร์ค่าเฉลี่ยของ P และ E จากกลุ่มตัวอย่างขนาดต่างๆเป็นรายคู่ ได้ผลดังแสดงในตาราง 21

ตาราง 21 แสดงผลการทดสอบความแตกต่างระหว่างเวกเตอร์ค่าเฉลี่ยของความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิปเทสเป็นรายคู่ กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ

ขนาดกลุ่มตัวอย่าง (คน)	1 (50)	2 (100)	3 (200)	4 (500)	5 (1,000)
1 (50)	-	1.657	4.031	17.168	19.853
2 (100)		-	1.609	19.514	24.023
3 (200)			-	32.245	37.699
4 (500)				-	.547
5 (1,000)					-

* มีนัยสำคัญที่ระดับ.05

ผลการวิเคราะห์ตามตาราง 21 พบว่ากรณีที่แบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ เมื่อตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิปเทส กลุ่มตัวอย่างขนาด 50 คนกับ 100 คน และ 100 คนกับ 200 คน มีผลต่อตัวแปรตาม (P และ E) โดยรวมไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่กลุ่มตัวอย่างขนาด 50 100 และ 200 คน มีผลต่อตัวแปรตาม (P และ E) โดยรวมแตกต่างจากกลุ่มตัวอย่างขนาด 500 และ 1,000 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 กลุ่มตัวอย่างขนาด 500 คน มีผลต่อตัวแปรตาม (P และ E) โดยรวมไม่แตกต่างจากกลุ่มตัวอย่างขนาด 1,000 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ผลการทดสอบความแตกต่างระหว่างเวกเตอร์ค่าเฉลี่ยของตัวแปรตาม(PและE)จากกลุ่มตัวอย่างขนาดต่างๆเป็นรายคู่พบว่ามีความสำคัญทางสถิติหลายคู่จึงต้องทดสอบความแตกต่างของตัวแปรตาม(PและE)เป็นรายตัว ได้ผลดังแสดงในตาราง 22

ตาราง 22 แสดงผลการทดสอบความแตกต่างของความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิปเทสท์เป็นรายตัวระหว่าง 2 กลุ่ม กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ

ตัวแปร	$\bar{X}$	S.D.	t	ตัวแปร	$\bar{X}$	S.D.	t
(3) กลุ่มตัวอย่าง 200 คน – (1) กลุ่มตัวอย่าง 50 คน							
P				E			
(1)	4.250	5.703	.467	(1)	3.334	5.324	2.236*
(3)	4.750	4.980		(3)	5.417	3.279	
(4) กลุ่มตัวอย่าง 500 คน – (1) กลุ่มตัวอย่าง 50 คน							
P				E			
(1)	4.250	5.709	5.049*	(1)	3.334	5.324	3.012*
(4)	.125	.884		(4)	.920	1.942	
(5) กลุ่มตัวอย่าง 1,000 คน – (1) กลุ่มตัวอย่าง 50 คน							
P				E			
(1)	4.250	5.709	4.733*	(1)	3.334	5.324	4.108*
(5)	.250	1.768		(5)	.167	1.178	
(4) กลุ่มตัวอย่าง 500 คน – (2) กลุ่มตัวอย่าง 100 คน							
P				E			
(2)	3.625	4.391	5.526*	(2)	4.501	4.441	5.223*
(4)	.125	.884		(4)	.920	1.942	
(5) กลุ่มตัวอย่าง 1,000 คน - (2) กลุ่มตัวอย่าง 100 คน							
P				E			
(2)	3.625	4.391	5.042*	(2)	4.501	4.441	6.670*
(5)	.250	1.768		(5)	.167	1.178	
(4) กลุ่มตัวอย่าง 500 คน – (3) กลุ่มตัวอย่าง 200 คน							
P				E			
(3)	4.750	4.981	6.465*	(3)	5.417	3.880	7.330*
(4)	.125	.884		(4)	.920	1.942	
(5) กลุ่มตัวอย่าง 1,000 คน – (3) กลุ่มตัวอย่าง 200 คน							
P				E			
(3)	4.750	4.981	6.021*	(3)	5.417	3.880	9.158*
(5)	.250	1.768		(5)	.166	1.178	

*มีนัยสำคัญที่ระดับ .05

ผลวิเคราะห์ตามตาราง 22 แสดงให้เห็นว่ากรณีที่แบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 1,000 และ 500 คน ส่งผลให้วิธีซิปเทสที่มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบสูง และมีการระบุผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 50 100 และ 200 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 กลุ่มตัวอย่างขนาด 200 คน ส่งผลให้วิธีซิปเทสที่มีการระบุผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 50 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 กลุ่มตัวอย่างขนาดอื่น ๆ ส่งผลให้วิธีซิปเทสที่มีความถูกต้องและมีการระบุผิดพลาดในการตรวจสอบไม่แตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3.1.1.1.3 กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน

การทดสอบนัยสำคัญของตัวแปรตาม(PและE)กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน เมื่อพิจารณาค่าสถิติ Wilks ซึ่งมีค่า .706 พบว่ามีนัยสำคัญทางสถิติที่ระดับ.05 จึงทดสอบความแตกต่างระหว่างเวกเตอร์ค่าเฉลี่ยของ PและE จากกลุ่มตัวอย่างขนาดต่างๆเป็นรายคู่ ได้ผลดังแสดงในตาราง 23

ตาราง 23 แสดงผลการทดสอบความแตกต่างระหว่างเวกเตอร์ค่าเฉลี่ยของความถูกต้องและการระบุผิดพลาด ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิปเทสเป็นรายคู่ กรณีที่แบบทดสอบ ประกอบด้วยข้อสอบ 50 ข้อ

ขนาดกลุ่มตัวอย่าง (คน)	1 (50)	2 (100)	3 (200)	4 (500)	5 (1,000)
1 (50)	-	3.634	6.659	23.426	32.876
2 (100)		-	2.465	9.333	20.018
3 (200)			-	6.787	10.006
4 (500)				-	5.955
5 (1,000)					-

* มีนัยสำคัญที่ระดับ .05

ผลการวิเคราะห์ตามตาราง 23 พบว่ากรณีที่แบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ เมื่อตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิปเทส กลุ่มตัวอย่างขนาด 50 คน มีผลต่อตัวแปรตาม (P และ E) โดยรวมแตกต่างจากกลุ่มตัวอย่างขนาด 100 200 500 และ 1,000 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 กลุ่มตัวอย่างขนาด 100 และ 200 คน มีผลต่อตัวแปรตาม (P และ E) โดยรวมแตกต่างจากกลุ่มตัวอย่างขนาด 500 และ 1,000 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และกลุ่มตัวอย่างขนาด 500 คน มีผลต่อตัวแปรตาม (P และ E) โดยรวมแตกต่างจากกลุ่มตัวอย่างขนาด 1,000 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ผลการทดสอบความแตกต่างระหว่างเวกเตอร์ค่าเฉลี่ยของตัวแปรตาม(PและE) จากกลุ่มตัวอย่างขนาดต่างๆเป็นรายคู่พบว่ามีนัยสำคัญทางสถิติหลายคู่จึงต้องทดสอบความแตกต่างของตัวแปรตาม(PและE)เป็นรายตัว ได้ผลดังแสดงในตาราง 24

ตาราง 24 แสดงผลการทดสอบความแตกต่างของความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิปเทสท์เป็นรายตัวระหว่าง 2 กลุ่ม กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ

ตัวแปร	$\bar{X}$	S.D.	t	ตัวแปร	$\bar{X}$	S.D.	t
(2) กลุ่มตัวอย่าง 100 คน – (1) กลุ่มตัวอย่าง 50 คน							
P				E			
(1)	5.375	6.440	.824	(1)	3.432	3.174	2.233*
(2)	4.250	7.192		(2)	5.112	4.270	
(3) กลุ่มตัวอย่าง 200 คน – (1) กลุ่มตัวอย่าง 50 คน							
P				E			
(1)	5.375	6.440	1.006	(1)	3.432	3.174	3.398*
(3)	6.625	5.978		(3)	6.400	5.298	
(4) กลุ่มตัวอย่าง 500 คน – (1) กลุ่มตัวอย่าง 50 คน							
P				E			
(1)	5.375	6.440	-.513	(1)	3.432	3.174	6.918*
(4)	4.750	5.725		(4)	8.527	4.128	
(5) กลุ่มตัวอย่าง 1,000 คน – (1) กลุ่มตัวอย่าง 50 คน							
P				E			
(1)	5.375	6.440	3.194*	(1)	3.432	3.174	10.660*
(5)	8.750	3.788		(5)	10.032	3.015	
(4) กลุ่มตัวอย่าง 500 คน - (2) กลุ่มตัวอย่าง 100 คน							
P				E			
(2)	4.250	7.192	.385	(2)	5.113	4.270	4.065*
(4)	4.750	5.725		(4)	8.527	4.128	
(5) กลุ่มตัวอย่าง 1,000 คน – (2) กลุ่มตัวอย่าง 100 คน							
P				E			
(2)	4.250	7.192	3.915*	(2)	5.113	4.270	6.655*
(5)	8.750	3.788		(5)	10.032	3.015	
(4) กลุ่มตัวอย่าง 500 คน – (3) กลุ่มตัวอย่าง 200 คน							
P				E			
(3)	6.625	5.978	-1.602	(3)	6.400	5.298	2.239*
(4)	4.750	5.725		(4)	8.527	4.128	

ตาราง 24 (ต่อ)

ตัวแปร	$\bar{X}$	S.D.	t	ตัวแปร	$\bar{X}$	S.D.	t
(5) กลุ่มตัวอย่าง 1,000 คน – (3) กลุ่มตัวอย่าง 200 คน							
P				E			
(3)	6.625	5.978	2.123*	(3)	6.340	5.298	4.214*
(5)	8.750	3.788		(5)	10.032	3.015	
(5) กลุ่มตัวอย่าง 1,000 คน – (4) กลุ่มตัวอย่าง 500 คน							
P				E			
(4)	4.750	5.725	4.120*	(4)	8.527	4.128	2.083*
(5)	8.750	3.788		(5)	10.032	3.015	

* มีนัยสำคัญที่ระดับ .05

ผลวิเคราะห์ตามตาราง 24 พบว่ากรณีที่แบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ กลุ่มตัวอย่างขนาด 100 200 และ 500 คน ส่งผลให้วิธีชิบเทสที่มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบไม่แตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 จากกลุ่มตัวอย่างขนาด 50 คน แต่ส่งผลให้มีการระบุผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 50 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 กลุ่มตัวอย่างขนาด 500 คน ส่งผลให้มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบไม่แตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีการระบุผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 100 และ 200 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 กลุ่มตัวอย่างขนาด 1,000 คน ส่งผลให้มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบและการระบุผิดพลาดสูงกว่ากลุ่มตัวอย่างขนาด 50 100 200 และ 500 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3.1.1.2 เปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีดีเอฟไอที

การประยุกต์เทคนิคการวิเคราะห์ความแปรปรวนแบบตัวแปรพหุในการเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีดีเอฟไอทีเพื่อพิจารณาความถูกต้องในการตรวจสอบและการระบุผิดพลาดในการตรวจสอบพร้อมกันด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป SPSS-PC for windows โปรแกรมจะเริ่มต้นด้วยการทดสอบนัยสำคัญของตัวแปรตาม(PและE)เมื่อกกลุ่มตัวอย่างมีขนาดแตกต่างกันด้วยสถิติทดสอบพหุตัวแปร ในการศึกษานี้พิจารณาค่าสถิติ Wilks (λ) ถ้าพบว่ามีความสำคัญก็จะทำการทดสอบความแตกต่างระหว่างเวกเตอร์ค่าเฉลี่ยของตัวแปรตาม (P และ E) จากกลุ่มตัวอย่างขนาดต่างๆเป็นรายคู่ ถ้าพบว่าคู่ใดมีความสำคัญก็จะทำการทดสอบความแตกต่างของตัวแปรตาม (P และ E) เป็นรายตัวต่อไป การเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบได้ผลดังต่อไปนี้

3.1.1.2.1 กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน

การทดสอบนัยสำคัญของตัวแปรตาม(PและE)กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน เมื่อพิจารณาค่าสถิติ Wilks ซึ่งมีค่า .879 พบว่ามีนัยสำคัญทางสถิติที่ระดับ.05 จึงทดสอบความแตกต่างระหว่างเวกเตอร์ค่าเฉลี่ยของ P และ E จากกลุ่มตัวอย่างขนาดต่าง ๆ เป็นรายคู่ ได้ผลดังแสดงในตาราง 25

ตาราง 25 แสดงผลการทดสอบความแตกต่างระหว่างเวกเตอร์ค่าเฉลี่ยของความถูกต้องและการระบุผิดพลาด ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีดีเอฟไอทีเป็นรายคู่ กรณีที่แบบทดสอบ ประกอบด้วยข้อสอบ 30 ข้อ

ขนาดกลุ่มตัวอย่าง (คน)	1 (50)	2 (100)	3 (200)	4 (500)	5 (1,000)
1 (50)	-	2.476	2.077	8.498	6.506
2 (100)		-	.912	2.358	2.825
3 (200)			-	5.991	6.972
4 (500)				-	1.902
5 (1,000)					-

*มีนัยสำคัญที่ระดับ.05

ผลการวิเคราะห์ตามตาราง 25 พบว่ากรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ เมื่อตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีดีเอฟไอที กลุ่มตัวอย่างขนาด 50 และ 200 คน มีผลต่อตัวแปรตาม (P และ E) โดยรวมแตกต่างจากกลุ่มตัวอย่างขนาด 500 และ 1,000 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 กลุ่มตัวอย่างขนาดอื่น ๆ ส่งผลต่อตัวแปรตาม (P และ E) ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ผลการทดสอบความแตกต่างระหว่างเวกเตอร์ค่าเฉลี่ยของตัวแปรตาม (P และ E) จากกลุ่มตัวอย่างขนาดต่าง ๆ เป็นรายคู่พบว่า มีนัยสำคัญทางสถิติหลายคู่จึงต้องทดสอบความแตกต่างของตัวแปรตาม (P และ E) เป็นรายตัว ได้ผลดังแสดงในตาราง 26

ตาราง 26 แสดงผลการทดสอบความแตกต่างของความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีดีเอฟไอที่เป็นรายตัวระหว่าง 2 กลุ่ม กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ

ตัวแปร	$\bar{X}$	S.D.	t	ตัวแปร	$\bar{X}$	S.D.	t
(4) กลุ่มตัวอย่าง 500 คน – (1) กลุ่มตัวอย่าง 50 คน							
P				E			
(1)	39.518	25.773	-.853	(1)	50.139	25.509	2.062*
(4)	35.750	15.244		(4)	59.311	16.144	
(5) กลุ่มตัวอย่าง 1,000 คน – (1) กลุ่มตัวอย่าง 50 คน							
P				E			
(1)	39.518	25.773	.640	(1)	50.139	25.509	2.753*
(5)	42.855	25.036		(5)	62.600	18.346	
(4) กลุ่มตัวอย่าง 500 คน – (3) กลุ่มตัวอย่าง 200 คน							
P				E			
(3)	29.736	31.655	1.198	(3)	46.028	25.762	3.058*
(4)	35.750	15.244		(4)	59.311	16.144	
(5) กลุ่มตัวอย่าง 1,000 คน – (3) กลุ่มตัวอย่าง 200 คน							
P				E			
(3)	29.736	31.655	2.290*	(3)	46.028	25.762	3.680*
(5)	42.855	25.036		(5)	62.600	18.347	

*มีนัยสำคัญที่ระดับ.05

ผลวิเคราะห์ตามตาราง 26 แสดงให้เห็นว่ากรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ กลุ่มตัวอย่างขนาด 500 คน ส่งผลให้วิธีดีเอฟไอที่มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบไม่แตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 จากกลุ่มตัวอย่างขนาด 50 และ 200 คน แต่มีการระบุผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 50 และ 200 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 กลุ่มตัวอย่างขนาด 1,000 คน ส่งผลให้วิธีดีเอฟไอที่มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบไม่แตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 จากกลุ่มตัวอย่างขนาด 50 คน แต่มีการระบุผิดพลาดในการตรวจสอบสูงกว่าอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 กลุ่มตัวอย่างขนาด 1,000 คน ส่งผลให้วิธีดีเอฟไอที่มีความถูกต้องและมีการระบุผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 200 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3.1.1.2.2 กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน

การทดสอบนัยสำคัญของตัวแปรตาม(PและE)กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน เมื่อพิจารณาค่าสถิติ Wilks ซึ่งมีค่า .956 พบว่าไม่มีนัยสำคัญทางสถิติที่ระดับ.05 จึงไม่ทดสอบความแตกต่างระหว่างเวกเตอร์ค่าเฉลี่ยของ P และ E จากกลุ่มตัวอย่างขนาดต่าง ๆ เป็นรายคู่ แสดงว่ากรณีที่แบบทดสอบประกอบด้วยข้อสอบ 40 ข้อเมื่อตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีดีเอฟไอทีกลุ่มตัวอย่างขนาดต่าง ๆ ส่งผลต่อตัวแปรตาม (P และ E) โดยรวมไม่แตกต่างกัน

3.1.1.2.3 กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน

การทดสอบนัยสำคัญของตัวแปรตาม (P และ E) กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน เมื่อพิจารณาค่าสถิติ Wilks ซึ่งมีค่า .951 พบว่าไม่มีนัยสำคัญทางสถิติที่ระดับ.05 จึงไม่ทดสอบความแตกต่างระหว่างเวกเตอร์ค่าเฉลี่ยของ P และ E จากกลุ่มตัวอย่างขนาดต่าง ๆ เป็นรายคู่ แสดงว่ากรณีที่แบบทดสอบประกอบด้วยข้อสอบ 50 ข้อเมื่อตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีดีเอฟไอทีกลุ่มตัวอย่างขนาดต่าง ๆ ส่งผลต่อตัวแปรตาม (P และ E) โดยรวมไม่แตกต่างกัน

การศึกษาผลของขนาดกลุ่มตัวอย่างที่มีต่อความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ เมื่อแบบทดสอบมีความยาวแตกต่างกันโดยแยกพิจารณาเฉพาะวิธีซิปเทสต์แล้วจึงพิจารณาเฉพาะวิธีดีเอฟไอที ผลสรุปดังแสดงในตาราง 27

ตาราง 27 สรุปผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีเดียวกัน

ความยาว แบบทดสอบ (ข้อ)	ผลการตรวจสอบ	
	วิธีซิปเทสต์	วิธีดีเอฟไอที
30 ข้อ	กลุ่มตัวอย่างขนาด 1,000 คน ส่งผลให้ P และ E สูงกว่ากลุ่มตัวอย่างขนาด 50 100 200 และ 500 คน กลุ่มตัวอย่างขนาด 500 คน ส่งผลให้ P และ E สูงกว่ากลุ่มตัวอย่างขนาด 50 100 และ 200 คน	กลุ่มตัวอย่างขนาด 500 คน ส่งผลให้ P ไม่แตกต่างจากกลุ่มตัวอย่างขนาด 50 และ 200 คน แต่ E สูงกว่า
	กลุ่มตัวอย่างขนาด 50 100 และ 200 คน ส่งผลต่อ P และ E ไม่แตกต่างกัน	กลุ่มตัวอย่างขนาด 1,000 คน ส่งผลให้ P ไม่แตกต่างจากกลุ่มตัวอย่างขนาด 50 คน แต่ E สูงกว่าและส่งผลให้ P และ E สูงกว่ากลุ่มตัวอย่างขนาด 200 คน
		กลุ่มตัวอย่างขนาดอื่น ๆ ส่งผลต่อ P และ E ไม่แตกต่างกัน

ตาราง 27 (ต่อ)

ความยาว แบบทดสอบ (ข้อ)	ผลการตรวจสอบ	
	วิธีชิบเทสต์	วิธีดีเอฟไอที
40 ข้อ	กลุ่มตัวอย่างขนาด 1,000 และ 500 คน ส่งผลให้ P และ E สูงกว่ากลุ่มตัวอย่างขนาด 50 100 และ 200 คน	กลุ่มตัวอย่างขนาดต่าง ๆ ส่งผลต่อ P และ E ไม่แตกต่างกัน
	กลุ่มตัวอย่างขนาด 500 และ 1,000 คนส่งผลต่อ P และ E ไม่แตกต่างกัน กลุ่มตัวอย่างขนาด 50 100 และ 200 คน ส่งผลต่อ P และ E ไม่แตกต่างกัน	
50 ข้อ	กลุ่มตัวอย่างขนาด 1,000 คน ส่งผลให้ P และ E สูงกว่ากลุ่มตัวอย่างขนาด 50 100 200 และ 500 คน	กลุ่มตัวอย่างขนาดต่าง ๆ ส่งผลต่อ P และ E ไม่แตกต่างกัน
	กลุ่มตัวอย่างขนาด 500 คน ส่งผลให้ E สูงกว่า แต่ P ไม่แตกต่างจากกลุ่มตัวอย่างขนาด 50 100 และ 200 คน	
	กลุ่มตัวอย่างขนาด 100 และ 200 คน ส่งผลให้ E สูงกว่า แต่ P ไม่แตกต่างจากกลุ่มตัวอย่างขนาด 50 คน	

ผลวิเคราะห์ตามตาราง 27 แสดงให้เห็นว่าเมื่อตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์พบว่า กรณีที่เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 และ 50 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน ส่งผลให้ความถูกต้องและการระบุผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 50 100 200 และ 500 คน เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน ส่งผลให้ความถูกต้องและการระบุผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 50 100 และ 200 คน แต่ส่งผลต่อความถูกต้องและการระบุผิดพลาดในการตรวจสอบไม่แตกต่างจากกลุ่มตัวอย่างขนาด 500 คน เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ กลุ่มตัวอย่างขนาด 50 100 และ 200 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบไม่แตกต่างกัน

เมื่อตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีดีเอฟไอทีพบว่า กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ กลุ่มตัวอย่างขนาด 1,000 และ 500 คน ส่งผลให้ความถูกต้องในการตรวจสอบไม่แตกต่างจากกลุ่มตัวอย่างขนาด 50 และ 100 คน แต่ส่งผลให้การระบุผิดพลาดในการตรวจสอบมีค่าสูงกว่ากลุ่มตัวอย่างขนาด 1,000 คน ส่งผลให้ความถูกต้องและการระบุผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 200 คน กลุ่มตัวอย่างขนาดอื่น ๆ ส่งผลต่อความถูกต้องและการระบุผิดพลาดในการตรวจสอบไม่แตกต่างกัน เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 และ 50 ข้อ กลุ่มตัวอย่างขนาดต่าง ๆ ส่งผลต่อความถูกต้องและการระบุผิดพลาดในการตรวจสอบไม่แตกต่างกัน

สรุปได้ว่าการเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีเดียวกัน เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ กลุ่มตัวอย่างขนาด 50 100 และ 200 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิปเทสต์ที่ไม่แตกต่างกัน กลุ่มตัวอย่างขนาด 500 และ 1,000 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิปเทสต์สูงกว่ากลุ่มตัวอย่างขนาด 50 100 และ 200 คน แต่การระบุผิดพลาดในการตรวจสอบก็สูงกว่าด้วย เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีดีเอฟไอทีที่ไม่แตกต่างกัน

3.1.2 การเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิปเทสต์และวิธีดีเอฟไอที

การเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิปเทสต์และวิธีดีเอฟไอที ผู้วิจัยประยุกต์เทคนิคการวิเคราะห์ความแปรปรวนแบบตัวแปรพหุ กรณีกลุ่มที่ศึกษามี 2 กลุ่ม (Hotelling's T^2) เพื่อศึกษาผลของวิธีซิปเทสต์และวิธีดีเอฟไอทีที่มีต่อตัวแปรตาม 2 ตัวพร้อมกัน คือ ความถูกต้องในการตรวจสอบ(P)และการระบุผิดพลาดในการตรวจสอบ(E)ในเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน กรณีในการเปรียบเทียบที่ต้องการคือ กรณีที่มีความถูกต้องในการตรวจสอบสูง แต่มีการระบุผิดพลาดในการตรวจสอบต่ำ ทำการเปรียบเทียบในเงื่อนไขที่แบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน จากการพิจารณาค่าสถิติทดสอบพหุตัวแปร Wilks พบว่าทุกเงื่อนไขมีนัยสำคัญ จึงทดสอบความแตกต่างของตัวแปรตาม (P และ E) เป็นรายตัว ผลการทดสอบดังแสดงในตาราง 28

ตาราง 28 แสดงผลการทดสอบความแตกต่างของความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีซิปเทสต์และวิธีดีเอฟไอที ในเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

ตัวแปร	ผลต่างของค่าเฉลี่ย F	
แบบทดสอบมี 30 ข้อ กลุ่มตัวอย่าง 50 คน		
P (DFIT-SIB)	35.661	91.925*
E (DFIT-SIB)	46.389	158.207*
แบบทดสอบมี 30 ข้อ กลุ่มตัวอย่าง 100 คน		
P (DFIT-SIB)	30.067	56.596*
E (DFIT-SIB)	48.531	195.519*
แบบทดสอบมี 30 ข้อ กลุ่มตัวอย่าง 200 คน		
P (DFIT-SIB)	25.314	30.752*
E (DFIT-SIB)	41.153	123.005*

ตาราง 28 (ต่อ)

ตัวแปร	ผลต่างของค่าเฉลี่ย F	
แบบทดสอบมี 30 ข้อ กลุ่มตัวอย่าง 500 คน		
P (DFIT-SIB)	28.179	136.629*
E (DFIT-SIB)	49.311	417.113*
แบบทดสอบมี 30 ข้อ กลุ่มตัวอย่าง 1,000 คน		
P (DFIT-SIB)	31.140	73.272*
E (DFIT-SIB)	49.350	326.736*
แบบทดสอบมี 40 ข้อ กลุ่มตัวอย่าง 50 คน		
P (DFIT-SIB)	38.964	190.762*
E (DFIT-SIB)	57.050	406.258*
แบบทดสอบมี 40 ข้อ กลุ่มตัวอย่าง 100 คน		
P (DFIT-SIB)	34.656	241.415*
E (DFIT-SIB)	55.742	743.929*
แบบทดสอบมี 40 ข้อ กลุ่มตัวอย่าง 200 คน		
P (DFIT-SIB)	36.625	189.957*
E (DFIT-SIB)	56.041	576.696*
แบบทดสอบมี 40 ข้อ กลุ่มตัวอย่าง 500 คน		
P (DFIT-SIB)	33.421	169.747*
E (DFIT-SIB)	54.521	386.651*
แบบทดสอบมี 40 ข้อ กลุ่มตัวอย่าง 1,000 คน		
P (DFIT-SIB)	39.225	31.273*
E (DFIT-SIB)	53.480	419.214*
แบบทดสอบมี 50 ข้อ กลุ่มตัวอย่าง 50 คน		
P (DFIT-SIB)	38.324	232.078*
E (DFIT-SIB)	59.948	1,030.176*
แบบทดสอบมี 50 ข้อ กลุ่มตัวอย่าง 100 คน		
P (DFIT-SIB)	44.082	247.792*
E (DFIT-SIB)	58.846	438.346*
แบบทดสอบมี 50 ข้อ กลุ่มตัวอย่าง 200 คน		
P (DFIT-SIB)	40.582	216.389*
E (DFIT-SIB)	62.186	736.358*

ตาราง 28 (ต่อ)

ตัวแปร	ผลต่างของค่าเฉลี่ย F	
แบบทดสอบมี 50 ข้อ กลุ่มตัวอย่าง 500 คน		
P (DFIT-SIB)	37.402	186.803*
E (DFIT-SIB)	57.679	1,002.598*
แบบทดสอบมี 50 ข้อ กลุ่มตัวอย่าง 1,000 คน		
P (DFIT-SIB)	37.691	64.749*
E (DFIT-SIB)	51.143	704.498*

* มีนัยสำคัญที่ระดับ .05

ผลวิเคราะห์ตามตาราง 28 แสดงให้เห็นว่าการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที ในเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน ทั้ง 15 เงื่อนไข เมื่อพิจารณาความถูกต้องในการตรวจสอบและการระบุผิดพลาดในการตรวจสอบพร้อมกัน วิธีชิบเทสต์มีความถูกต้องในการตรวจสอบและมีการระบุผิดพลาดในการตรวจสอบน้อยกว่าวิธีดีเอฟไอที อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3.1.3 ผลการตรวจสอบความสอดคล้องในการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที

ความสอดคล้องในการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที คำนวณจากร้อยละของจำนวนข้อสอบที่ทั้งสองวิธีตรวจพบว่าทำหน้าที่เบี่ยงเบนซ้ำกันต่อจำนวนข้อสอบที่ทั้งสองวิธีตรวจพบว่าทำหน้าที่เบี่ยงเบนทั้งหมด ผลการการวิเคราะห์เมื่อแบบทดสอบมีความยาว 30 40 และ 50 ข้อ กลุ่มตัวอย่างมีขนาด 50 100 200 500 และ 1,000 คน สรุปได้ดังตาราง 29

ตาราง 29 สรุปความสอดคล้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที ในเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

ขนาดกลุ่ม ความยาว แบบทดสอบ (ข้อ)	50		100		200		500		1,000	
	ตัวอย่าง									
	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.
30	3.93	5.50	2.17	3.61	2.57	4.18	6.68	5.86	6.88	5.14
40	3.04	4.33	3.46	3.86	5.11	6.96	0.38	1.39	0.07	0.52
50	3.35	3.41	4.33	4.41	4.97	3.66	6.59	4.73	6.50	3.17

ผลวิเคราะห์ตามตาราง 29 แสดงให้เห็นร้อยละของความสอดคล้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอทีดังนี้ เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ กลุ่มตัวอย่างขนาด 50 คน มีค่าเฉลี่ย 3.93 ส่วนเบี่ยงเบนมาตรฐาน 5.50 กลุ่มตัวอย่างขนาด 100 คน มีค่าเฉลี่ย 2.17 ส่วนเบี่ยงเบนมาตรฐาน 3.61 กลุ่มตัวอย่างขนาด 200 คน มีค่าเฉลี่ย 2.57 ส่วนเบี่ยงเบนมาตรฐาน 4.18 กลุ่มตัวอย่างขนาด 500 คน มีค่าเฉลี่ย 6.68 ส่วนเบี่ยงเบนมาตรฐาน 5.86 กลุ่มตัวอย่างขนาด 1,000 คน มีค่าเฉลี่ย 6.88 ส่วนเบี่ยงเบนมาตรฐาน 5.14

เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 50 คน มีค่าเฉลี่ย 3.04 ส่วนเบี่ยงเบนมาตรฐาน 4.33 กลุ่มตัวอย่างขนาด 100 คน มีค่าเฉลี่ย 3.46 ส่วนเบี่ยงเบนมาตรฐาน 3.86 กลุ่มตัวอย่างขนาด 200 คน มีค่าเฉลี่ย 5.11 ส่วนเบี่ยงเบนมาตรฐาน 6.96 กลุ่มตัวอย่างขนาด 500 คน มีค่าเฉลี่ย 0.38 ส่วนเบี่ยงเบนมาตรฐาน 1.39 กลุ่มตัวอย่างขนาด 1,000 คน มีค่าเฉลี่ย 0.07 ส่วนเบี่ยงเบนมาตรฐาน 0.52

เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ กลุ่มตัวอย่างขนาด 50 คน มีค่าเฉลี่ย 3.35 ส่วนเบี่ยงเบนมาตรฐาน 3.41 กลุ่มตัวอย่างขนาด 100 คน มีค่าเฉลี่ย 4.33 ส่วนเบี่ยงเบนมาตรฐาน 4.41 กลุ่มตัวอย่างขนาด 200 คน มีค่าเฉลี่ย 4.97 ส่วนเบี่ยงเบนมาตรฐาน 3.66 กลุ่มตัวอย่างขนาด 500 คน มีค่าเฉลี่ย 6.59 ส่วนเบี่ยงเบนมาตรฐาน 4.73 กลุ่มตัวอย่างขนาด 1,000 คน มีค่าเฉลี่ย 6.50 ส่วนเบี่ยงเบนมาตรฐาน 3.17

3.2 การเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที

ความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบพิจารณาจากจำนวนหมวดข้อสอบที่แต่ละวิธีตรวจสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ การเปรียบเทียบผลการตรวจสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอทีในเงื่อนไขต่างๆ สรุปได้ดังตาราง 30

ตาราง 30 แสดงการเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที ในเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

ความยาวแบบ ทดสอบ ขนาด กลุ่มตัวอย่าง	30 ข้อ			40 ข้อ			50 ข้อ		
	P ₁	P ₂	z-test	P ₁	P ₂	z-test	P ₁	P ₂	z-test
50	0.7143	0.7143	0.0000	0.7143	0.5714	0.5578	0.8571	0.4286	1.6733
100	0.7143	0.7143	0.0000	0.7143	0.4286	1.0801	0.7143	0.5714	0.5578
200	0.5714	0.4286	0.5345	0.5714	0.4286	0.5345	0.7143	0.4286	1.0801
500	0.7143	0.5714	0.5578	0.8571	0.2857	2.1602	0.7143	0.5000	0.7917
1,000	0.8571	0.0000	3.2404	0.8571	0.3333	1.9353	0.8571	0.7143	0.6513

* มีนัยสำคัญที่ระดับ .05

หมายเหตุ P₁ หมายถึง สัดส่วนการตรวจสอบ DBF ถูกต้องด้วยวิธีชิบเทสต์

P₂ หมายถึง สัดส่วนการตรวจสอบ DBF ถูกต้องด้วยวิธีดีเอฟไอที

ผลวิเคราะห์ตามตาราง 30 แสดงให้เห็นว่าเมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ วิธีชิบเทสที่มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบมากกว่าวิธีดีเอฟไอที่อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อกลุ่มตัวอย่างมีขนาด 1,000 คน เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ วิธีชิบเทสที่มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบมากกว่าวิธีดีเอฟไอที่อย่างมีนัยสำคัญที่ระดับ .05 เมื่อกลุ่มตัวอย่างมีขนาด 500 คน เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ ไม่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ระหว่างวิธีการทั้งสอง

3.3 การเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบด้วยวิธีชิบเทสและวิธีดีเอฟไอที่

ความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบพิจารณาจากจำนวนครั้งของการตรวจสอบที่แต่ละวิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบได้ตรงกับความเห็นของผู้เชี่ยวชาญ การเปรียบเทียบผลการตรวจสอบ 50 ครั้งด้วยวิธีชิบเทสและวิธีดีเอฟไอที่ สรุปได้ตาราง 31

ตาราง 31 แสดงการเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบด้วยวิธีชิบเทสและวิธีดีเอฟไอที่ ในเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

ความยาวแบบ ทดสอบ ขนาด กลุ่มตัวอย่าง	30 ข้อ			40 ข้อ			50 ข้อ		
	P ₁	P ₂	z-test	P ₁	P ₂	z-test	P ₁	P ₂	z-test
50	0.7100	0.7778	-0.4292	0.8600	0.7000	1.9312	0.8600	0.8800	-0.2974
100	0.6600	0.8125	-1.7094	0.8000	0.7292	0.8271	0.8800	0.7200	2.0000
200	0.7800	0.8571	-0.9950	0.9400	0.9400	0.0000	0.9800	0.8298	2.5484
500	0.9200	0.9184	0.0298	0.6566	0.8980	-5.0072	1.0000	0.9388	1.7768
1,000	0.9800	0.9400	1.0206	0.0200	0.9200	-9.0162	1.0000	0.9200	2.0412

* มีนัยสำคัญที่ระดับ .05

หมายเหตุ P₁ หมายถึง สัดส่วนการตรวจสอบ DTF ถูกต้องด้วยวิธีชิบเทส

P₂ หมายถึง สัดส่วนการตรวจสอบ DTF ถูกต้องด้วยวิธีดีเอฟไอที่

ผลวิเคราะห์ตามตาราง 31 แสดงให้เห็นว่าเมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 และ 40 ข้อ ความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบด้วยวิธีชิบเทสและวิธีดีเอฟไอที่ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ วิธีชิบเทสที่มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบมากกว่าวิธีดีเอฟไอที่อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อกลุ่มตัวอย่างมีขนาด 100 200 และ 1,000 คน

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

การศึกษาค้นคว้าครั้งนี้มีจุดมุ่งหมายเพื่อศึกษาวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบที่เหมาะสมกับข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ โดยมีวัตถุประสงค์เฉพาะเพื่อเปรียบเทียบวิธีชิบเทสต์และวิธีดีเอฟไอที ในด้านต่าง ๆ ดังนี้

1. เปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบวิธีเดียวกันในเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

2. เปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ ด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที ในระดับต่าง ๆ ดังนี้

2.1 การทำหน้าที่เบี่ยงเบนของข้อสอบ

2.1.1 เปรียบเทียบความถูกต้องและการระบุผิดพลาดในการตรวจสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที ในเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

2.1.2 หาความสอดคล้องของการตรวจสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที ในเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

2.2 การทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ

เปรียบเทียบความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที ในเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

2.3 การทำหน้าที่เบี่ยงเบนของแบบทดสอบ

เปรียบเทียบความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที ในเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

สมมุติฐานในการวิจัยมีดังนี้

1. กลุ่มตัวอย่างขนาดต่าง ๆ ส่งผลต่อความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์แตกต่างกันทุกความยาวแบบทดสอบ

2. กลุ่มตัวอย่างขนาดต่าง ๆ ส่งผลต่อความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีดีเอฟไอทีแตกต่างกันทุกความยาวแบบทดสอบ

3. เมื่อพิจารณาความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบพร้อมกัน วิธีชิบเทสต์ให้ผลดีกว่าวิธีดีเอฟไอทีทุกเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

4. ความสอดคล้องของผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที มีค่าต่ำกว่าร้อยละ 50 ทุกเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

5. วิธีชิบเทสต์มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบมากกว่าวิธีดีเอฟไอที ทุกเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

6. วิธีชิบเทสต์มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบมากกว่าวิธีดีเอฟไอที ทุกเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน

กลุ่มตัวอย่างที่ใช้ในการศึกษาคั้งนี้เป็นนักเรียนเพศชายและเพศหญิง ชั้นมัธยมศึกษาปีที่ 1 ภาคการศึกษาที่ 2 ปีการศึกษา 2543 ของโรงเรียนสังกัดกรมสามัญศึกษา จังหวัดนนทบุรี ซึ่งได้มาจากการสุ่มแบบสองขั้นตอน มีขนาดโรงเรียนเป็นชั้นและมีห้องเรียนเป็นหน่วยการสุ่ม ได้โรงเรียนทั้งหมด 6 โรงเรียน จำนวน 50 ห้องเรียน นักเรียนทั้งหมด 2,259 คน เป็นเพศชาย 1,173 คน เพศหญิง 1,086 คน ใช้นักเรียนทั้งหมดของแต่ละเพศเป็นประชากรเทียม สุ่มกลุ่มตัวอย่างขนาด 1,000 500 200 100 และ 50 คน ขนาดละ 50 กลุ่ม เพื่อศึกษาความคงเส้นคงวาในการตรวจสอบการทำหน้าที่เบี่ยงเบน เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ และกลุ่มตัวอย่างมีขนาดแตกต่างกัน

เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 (ค 101) ที่ผู้วิจัยสร้างขึ้น ประกอบด้วยข้อสอบแบบเลือกตอบชนิด 5 ตัวเลือก จำนวน 50 ข้อ โดยมีข้อสอบที่ผู้เชี่ยวชาญพิจารณาว่าเป็นข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนต่อเพศชายจำนวน 16 ข้อ ซึ่งมีค่าความยากระหว่าง 0.15 ถึง 0.25 ($0.15 < P < 0.25$) อำนาจจำแนก 0.2 ขึ้นไป ($r \geq 0.2$) เมื่อเก็บรวบรวมข้อมูลแล้ว คัดเลือกข้อสอบตามสัดส่วนในตารางกำหนดข้อสอบจัดทำเป็นแบบทดสอบที่ประกอบด้วยข้อสอบ 40 ข้อและ 30 ข้อ

วิธีการดำเนินการวิจัย

1. ขอความร่วมมือในการเก็บรวบรวมข้อมูลจากอธิบดีกรมสามัญศึกษาและเก็บรวบรวมข้อมูลจากนักเรียนในโรงเรียนกลุ่มตัวอย่าง
2. บันทึกผลการตอบข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) จำนวน 50 ข้อของนักเรียนกลุ่มตัวอย่างทั้งหมด ตรวจสอบให้คะแนน พิจารณาคะแนนของนักเรียนทั้งสองเพศและแยกตามเพศ แล้วพิจารณาคัดเลือกข้อสอบตามสัดส่วนเนื้อหาในตารางกำหนดข้อสอบ จัดทำเป็นแบบทดสอบที่ประกอบด้วยข้อสอบ 40 และ 30 ข้อ พิจารณาคะแนนของนักเรียนทั้งสองเพศและแยกตามเพศเมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 และ 30 ข้อ
3. ตรวจสอบความเหมาะสมของข้อมูลการตอบข้อสอบของนักเรียนทั้งหมดและแยกตามเพศเมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ กับแบบจำลองที่ใช้ความสามารถหลายมิติ (M2PL)
4. จำลองข้อมูลเพื่อใช้ในการศึกษา โดยใช้นักเรียนแต่ละเพศที่ตอบข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 จำนวน 50 ข้อ เป็นประชากรเทียม สุ่มกลุ่มตัวอย่างแบบใส่คืนมีขนาด 50 100 200 500 และ 1,000 คน แต่ละขนาดสุ่ม 50 ครั้ง และดำเนินการเช่นเดียวกันในกรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 และ 40 ข้อ ดังนั้นต้องเตรียมแฟ้มข้อมูลการตอบข้อสอบของนักเรียนเพศชายและเพศหญิงในเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกันจำนวน $2 \times 3 \times 5 \times 50 = 1,500$ แฟ้ม
5. เตรียมแฟ้มข้อมูลสำหรับตรวจสอบการทำหน้าที่เบี่ยงเบนด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป SIBTEST โดยการนำแฟ้มข้อมูลการตอบข้อสอบของนักเรียนกลุ่มตัวอย่างมาจัดรูปแบบตามที่โปรแกรมต้องการและเตรียมแฟ้มข้อมูลสำหรับตรวจสอบการทำหน้าที่เบี่ยงเบนด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป DFIT คือ แฟ้มพารามิเตอร์ข้อสอบของเพศชายและเพศหญิง แฟ้มพารามิเตอร์ความสามารถของนักเรียนเพศหญิง ประเมินค่าสัมประสิทธิ์การเทียบมาตร (linking coefficient) เพื่อปรับเทียบพารามิเตอร์ข้อสอบของเพศชายและเพศหญิง
6. ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป SIBTEST และ DFIT

วิธีดีเอฟโอที เมื่อใช้กลุ่มตัวอย่าง 50 คน พบว่าความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 38.09 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 25.75 การระบุผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 50.14 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 25.51 เมื่อใช้กลุ่มตัวอย่าง 100 คน พบว่าความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 33.78 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 27.68 การระบุผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 51.28 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 23.95 เมื่อใช้กลุ่มตัวอย่าง 200 คน พบว่าความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 29.73 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 31.61 การระบุผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 41.03 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 25.76 เมื่อใช้กลุ่มตัวอย่าง 500 คน พบว่าความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 36.30 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 14.93 การระบุผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 59.94 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 16.09 เมื่อใช้กลุ่ม

วิธีชปเทศท์เมื่อใช้กลุ่มตัวอย่าง 50 คน พบว่าความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 5.38 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 6.44 การระบพิตพลาจใในการตรวจสอบมีค่าเฉลี่ย

ต่างกันทั้ง 5 กลุ่มมีเวกเตอร์ค่าเฉลี่ยของตัวแปรตาม (P และ E) ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นั่นคือ ขนาดกลุ่มตัวอย่างมีผลต่อตัวแปรตาม 2 ตัว (P และ E) โดยรวมไม่แตกต่างกัน

กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน พบว่าเมื่อตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีดีเอฟไอที กลุ่มตัวอย่างขนาดต่างกันทั้ง 5 กลุ่มมีเวกเตอร์ค่าเฉลี่ยของตัวแปรตาม (P และ E) ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นั่นคือ ขนาดกลุ่มตัวอย่างมีผลต่อตัวแปรตาม 2 ตัว (P และ E) โดยรวมไม่แตกต่างกัน

2.1.2 การเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที

เปรียบเทียบในเงื่อนไขที่แบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ กลุ่มตัวอย่างมีขนาด 50 100 200 500 และ 1,000 คน รวมทั้งหมด 15 เงื่อนไข พบว่าเมื่อพิจารณาความถูกต้องในการตรวจสอบและการระบุผิดพลาดในการตรวจสอบพร้อมกัน ทุกเงื่อนไขวิธีชิบเทสต์มีความถูกต้องในการตรวจสอบและมีการระบุผิดพลาดในการตรวจสอบน้อยกว่าวิธีดีเอฟไอทีอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

2.2 ความสอดคล้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที

กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ ความสอดคล้องในการตรวจสอบมีค่าเฉลี่ยต่ำกว่าร้อยละ 10 เมื่อกลุ่มตัวอย่างมีขนาด 50 100 200 500 และ 1,000 คน ความสอดคล้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอทีมีค่าเฉลี่ยเรียงตามลำดับดังนี้ 3.93 2.17 2.57 6.68 6.88

กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ ความสอดคล้องในการตรวจสอบมีค่าเฉลี่ยต่ำกว่าร้อยละ 10 เมื่อกลุ่มตัวอย่างมีขนาด 50 100 200 500 และ 1,000 คน ความสอดคล้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอทีมีค่าเฉลี่ยเรียงตามลำดับดังนี้ 3.04 3.46 5.11 0.38 0.07

กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ ความสอดคล้องในการตรวจสอบมีค่าเฉลี่ยต่ำกว่าร้อยละ 10 เมื่อกลุ่มตัวอย่างมีขนาด 50 100 200 500 และ 1,000 คน ความสอดคล้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอทีมีค่าเฉลี่ยเรียงตามลำดับดังนี้ 3.35 4.33 4.97 6.59 6.50

2.3 การเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที

พบว่าเมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ ความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบด้วยวิธีชิบเทสต์มากกว่าวิธีดีเอฟไอทีอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อกลุ่มตัวอย่างมีขนาด 1,000 คน เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ ความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบด้วยวิธีชิบเทสต์มากกว่าวิธีดีเอฟไอทีอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อกลุ่มตัวอย่างมีขนาด 500 คน เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ ไม่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ระหว่างวิธีการทั้งสอง

2.4 การเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที

พบว่าเมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 และ 40 ข้อ ความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบด้วยวิธีชิบเทสท์และวิธีดีเอฟไอทีไม่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ ความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบด้วยวิธีชิบเทสท์มากกว่าวิธีดีเอฟไอทีอย่างมีนัยสำคัญที่ระดับ .05 เมื่อกลุ่มตัวอย่างมีขนาด 100 200 และ 1,000 คน

โดยภาพรวมสรุปผลการวิจัยได้ดังนี้

1. เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ กลุ่มตัวอย่างมีขนาด 50 100 และ 200 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสท์ไม่แตกต่างกัน กลุ่มตัวอย่างขนาด 500 และ 1,000 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสท์สูงกว่ากลุ่มตัวอย่างขนาด 50 100 และ 200 คน แต่การระบุผิดพลาดในการตรวจสอบก็สูงกว่าด้วย ซึ่งไม่เป็นไปตามสมมุติฐานข้อที่ 1
2. เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีดีเอฟไอทีไม่แตกต่างกัน ซึ่งไม่เป็นไปตามสมมุติฐานข้อที่ 2
3. เมื่อพิจารณาความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบพร้อมกัน วิธีชิบเทสท์มีความถูกต้องและการระบุผิดพลาดในการตรวจสอบน้อยกว่าวิธีดีเอฟไอทีในทุกเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน ซึ่งไม่เป็นไปตามสมมุติฐานข้อที่ 3
4. ความสอดคล้องของการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสท์และวิธีดีเอฟไอทีมีค่าน้อยกว่าร้อยละ 10 ในทุกเงื่อนไขความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดแตกต่างกัน ซึ่งเป็นไปตามสมมุติฐานข้อที่ 4
5. การตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบพบว่าเมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ วิธีชิบเทสท์มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบมากกว่าวิธีดีเอฟไอที เมื่อกลุ่มตัวอย่างมีขนาด 1,000 คน เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ วิธีชิบเทสท์มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบมากกว่าวิธีดีเอฟไอที เมื่อกลุ่มตัวอย่างมีขนาด 500 คน เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ ไม่พบความแตกต่างระหว่างวิธีการทั้งสอง ซึ่งไม่เป็นไปตามสมมุติฐานข้อ 5
6. การตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบพบว่าเมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 และ 40 ข้อ ความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบด้วยวิธีชิบเทสท์และวิธีดีเอฟไอทีไม่แตกต่างกัน เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ วิธีชิบเทสท์มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบมากกว่าวิธีดีเอฟไอที เมื่อกลุ่มตัวอย่างมีขนาด 100 200 และ 1,000 คน ซึ่งไม่เป็นไปตามสมมุติฐานข้อที่ 6

อภิปรายผลการวิจัย

วัตถุประสงค์สำคัญในการศึกษาครั้งนี้เพื่อศึกษาวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนที่เหมาะสมกับข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ มีวัตถุประสงค์เฉพาะเพื่อเปรียบเทียบวิธีชิบเทสท์ และวิธีดีเอฟไอที ไม่ได้มุ่งนำผลการตรวจสอบไปใช้ในการปรับปรุงคุณภาพแบบทดสอบที่สร้างขึ้น และจากการค้นคว้าพบว่ายังไม่มีการศึกษาเพื่อเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของ

ข้อสอบ หมวดข้อสอบ และแบบทดสอบ จากข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที แต่มีการเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบจากแบบทดสอบที่วัดความสามารถมิติเดียวด้วยวิธีชิบเทสต์ กับวิธีแมนเทล-แฮนส์เซล เป็นต้น (Roussos & Stout. 1996 : 215-230) หรือเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีดีเอฟไอทีกับวิธีของลอร์ด การวัดพื้นที่แบบคิดเครื่องหมาย การวัดพื้นที่แบบไม่คิดเครื่องหมาย เป็นต้น (Raju, van der Linden & Fleer. 1995 : 353-368) จึงอภิปรายผลการวิจัยในประเด็นต่อไปนี้

1. การเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

1.1 ความถูกต้องในการตรวจสอบและการระบุผิดพลาดในการตรวจสอบในเงื่อนไขความยาวแบบทดสอบและขนาดกลุ่มตัวอย่างที่แตกต่างกัน ผลการศึกษาพบว่าเมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ กลุ่มตัวอย่างมีขนาด 50 100 200 500 และ 1,000 คน ผู้เชี่ยวชาญพิจารณาว่ามีข้อสอบแสดงการทำหน้าที่เบี่ยงเบนของข้อสอบ จำนวน 14 ข้อ วิธีชิบเทสต์ที่มีความถูกต้องในการตรวจสอบประมาณร้อยละ 4-12 มีแนวโน้มเพิ่มขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น ขณะเดียวกันการระบุผิดพลาดในการตรวจสอบมีค่าใกล้เคียงหรือสูงกว่าความถูกต้องในการตรวจสอบ และมีค่าประมาณร้อยละ 4-13 มีแนวโน้มเพิ่มขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น วิธีดีเอฟไอทีที่มีความถูกต้องในการตรวจสอบประมาณร้อยละ 30-43 ค่าที่ได้มีการเปลี่ยนแปลงไม่เป็นระบบ การระบุผิดพลาดในการตรวจสอบมีค่าสูงกว่าความถูกต้องในการตรวจสอบ มีค่าประมาณร้อยละ 41-62 ค่าที่ได้มีการเปลี่ยนแปลงไม่เป็นระบบ เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ ผู้เชี่ยวชาญพิจารณาว่ามีข้อสอบแสดงการทำหน้าที่เบี่ยงเบนของข้อสอบ จำนวน 16 ข้อ วิธีชิบเทสต์ที่มีความถูกต้องในการตรวจสอบประมาณร้อยละ 0.2-5 ค่าที่ได้มีการเปลี่ยนแปลงไม่เป็นระบบ การระบุผิดพลาดในการตรวจสอบมีค่าใกล้เคียงหรือสูงกว่าความถูกต้องในการตรวจสอบ มีค่าประมาณร้อยละ 1-5 ค่าที่ได้มีการเปลี่ยนแปลงไม่เป็นระบบ วิธีดีเอฟไอทีที่มีความถูกต้องในการตรวจสอบประมาณร้อยละ 33-44 ค่าที่ได้มีการเปลี่ยนแปลงไม่เป็นระบบ การระบุผิดพลาดในการตรวจสอบมีค่าสูงกว่าความถูกต้องในการตรวจสอบ มีค่าประมาณร้อยละ 53-62 ค่าที่ได้มีการเปลี่ยนแปลงไม่เป็นระบบ เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ ผู้เชี่ยวชาญพิจารณาว่ามีข้อสอบแสดงการทำหน้าที่เบี่ยงเบนของข้อสอบ จำนวน 16 ข้อ วิธีชิบเทสต์ที่มีความถูกต้องในการตรวจสอบประมาณร้อยละ 5-9 ค่าที่ได้มีการเปลี่ยนแปลงไม่เป็นระบบ การระบุผิดพลาดในการตรวจสอบมีค่าประมาณร้อยละ 3-10 ค่าที่ได้มีการเปลี่ยนแปลงไม่เป็นระบบ วิธีดีเอฟไอทีที่มีความถูกต้องในการตรวจสอบมีค่าประมาณร้อยละ 42-48 ค่าที่ได้มีการเปลี่ยนแปลงไปไม่เป็นระบบ การระบุผิดพลาดในการตรวจสอบมีค่าประมาณร้อยละ 61-69 ค่าที่ได้มีการเปลี่ยนแปลงไม่เป็นระบบ

จะเห็นว่าความถูกต้องในการตรวจสอบและการระบุผิดพลาดในการตรวจสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที มีค่าเปลี่ยนแปลงอย่างไม่เป็นระบบในเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดต่าง ๆ กัน นอกจากนั้นความถูกต้องในการตรวจสอบและการระบุผิดพลาดในการตรวจสอบด้วยวิธีชิบเทสต์มีค่าต่ำ แต่ความถูกต้องในการตรวจสอบด้วยวิธีดีเอฟไอทีมีค่าปานกลางและการระบุผิดพลาดในการตรวจสอบมีค่าสูงมาก แสดงให้เห็นความผิดปกติของผลการตรวจสอบการเบี่ยงเบนของข้อสอบ ซึ่งอาจมีสาเหตุดังนี้

1) การกำหนดเกณฑ์ในการตัดสินข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนโดยให้ผู้เชี่ยวชาญเป็นผู้พิจารณาตัดสินเพียงรอบเดียว อาจมีความไม่เหมาะสม เนื่องจากอาจเกิดความคลาดเคลื่อนในการพิจารณา ถ้าให้ผู้เชี่ยวชาญมีโอกาสพิจารณาการทำหน้าที่เบี่ยงเบนของข้อสอบทุกครั้งที่มีการทดลองใช้ข้อสอบและปรับปรุงข้อสอบ ทำให้ผู้เชี่ยวชาญมีโอกาสพิจารณาหลายรอบ เช่น การใช้เทคนิคเดลฟาย

(Delphi-Technique) ซึ่งมีการสอบถามผู้เชี่ยวชาญหลายรอบ หรืออาจให้คณะผู้เชี่ยวชาญกลุ่มเดิมหรือกลุ่มอื่นมีโอกาสร่วมกันพิจารณายืนยันอีกรอบ อาจให้ผลการตัดสินใจที่มีความคงเส้นคงวามากขึ้นและมีความคลาดเคลื่อนน้อยลง (Clauser, et al. 1997 : 345-358)

2) ข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) ที่ผู้วิจัยสร้างขึ้นมีความยากและอำนาจจำแนกค่อนข้างสูง นักเรียนกลุ่มตัวอย่างเรียนวิชาดังกล่าวในภาคการศึกษาที่ 1 แต่ผู้วิจัยเก็บรวบรวมในภาคการศึกษาที่ 2 นักเรียนส่วนใหญ่ตอบแบบสอบถามว่าข้อสอบยากมากและลืมนเนื้อหาที่เรียนไปแล้ว จึงทำให้นักเรียนส่วนใหญ่ทั้งกลุ่มอ้างอิงและกลุ่มสนใจตอบข้อสอบโดยการเดาคำตอบ จึงทำให้การแจกแจงของคะแนนและการแจกแจงพารามิเตอร์ความสามารถหลักของกลุ่มตัวอย่างในเงื่อนไขของความยาวแบบทดสอบ 30 40 และ 50 ข้อ มีการแจกแจงแบบเบ้ขวาแทนที่จะมีการแจกแจงแบบปกติ ซึ่งจะทำให้ลักษณะการแจกแจงของข้อมูลสอดคล้องกับข้อตกลงเบื้องต้นตามกรอบแนวคิดของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ และจากการศึกษาของโมนาโค (Monaco, 1997.) โดยใช้ข้อมูลจำลองพบว่า เมื่อการแจกแจงพารามิเตอร์ความสามารถมีความเบ้ ทำให้ความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบลดลง

3) วิธีการที่ใช้ในการตรวจสอบความเหมาะสมของพหุคูณของข้อมูลการตอบข้อสอบกับแบบจำลองการตอบข้อสอบที่ใช้ความสามารถหลายมิติชนิด 2 พารามิเตอร์ (M2PL) อาจมีความไม่เหมาะสมเนื่องจากใช้วิธีการทางอ้อม คือ ตรวจสอบว่าชุดของข้อมูลการตอบข้อสอบละเมิดข้อตกลงของความเป็นมิติเดียวของแบบทดสอบ และตรวจสอบว่าข้อมูลการตอบข้อสอบสามารถวิเคราะห์ค่าพารามิเตอร์ข้อสอบด้วยแบบจำลองการตอบข้อสอบที่ใช้ความสามารถหลายมิติชนิด 2 พารามิเตอร์ (M2PL) ได้ทุกข้อ อาจเป็นข้อสรุปที่มีข้อโต้แย้งได้ ถ้าใช้วิธีการทางสถิติตรวจสอบทางตรงอาจทำให้ผลการตรวจสอบมีความคลาดเคลื่อนน้อยลง วิธีการทางสถิติทางตรงที่ใช้ในการตรวจสอบมิติของข้อมูลการตอบข้อสอบได้แก่ การใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป DIMTEST, DETECT เป็นต้น (Douglas, et al. 1996 : 465-484, Zhang & Stout. 1998)

4) ปริมาณข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนในแบบทดสอบมีมากเกินไป องค์ประกอบหนึ่งที่มีผลต่อการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบคือ สัดส่วนของข้อสอบแสดงการทำหน้าที่เบี่ยงเบนในแบบทดสอบ ในการศึกษาครั้งนี้มีข้อสอบแสดงการทำหน้าที่เบี่ยงเบนในแบบทดสอบคิดเป็นร้อยละ 47 40 และ 32 ในแบบทดสอบที่ประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ งานวิจัยเกี่ยวกับการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบส่วนใหญ่กำหนดสัดส่วนของข้อสอบแสดงการทำหน้าที่เบี่ยงเบนต่ำกว่าการศึกษาในครั้งนี้ เช่น สวามินาธานและโรเจอร์ (Swaminathan & Rogers. 1990 : 361-370) กำหนดสัดส่วนข้อสอบแสดงการทำหน้าที่เบี่ยงเบนในแบบทดสอบร้อยละ 20 ราชู แวนเดอร์ลินเด็นและฟลีร์ (Raju, van der Linden & Fleer. 1995 : 353-368) กำหนดสัดส่วนข้อสอบแสดงการทำหน้าที่เบี่ยงเบนในแบบทดสอบร้อยละ 0 5 10 และ 20 ญานัทธ สี่หะมงคล (2539) กำหนดสัดส่วนข้อสอบแสดงการทำหน้าที่เบี่ยงเบนในแบบทดสอบร้อยละ 0 10 และ 20 นอกจากนั้น นารายานานและสวามินาธาน (Narayanan & Sawaminathan. 1996 : 263) กล่าวว่าสัดส่วนของข้อสอบแสดงการทำหน้าที่เบี่ยงเบนของข้อสอบ ในแบบทดสอบมีผลต่อการตรวจสอบ DIF ถ้ามีข้อสอบแสดงการทำหน้าที่เบี่ยงเบนของข้อสอบปริมาณมากทำให้ความถูกต้องในการตรวจสอบลดลง

5) การใช้ทางเลือกในโปรแกรมคอมพิวเตอร์สำเร็จรูปที่ไม่เหมาะสม (choice of computer algorithm) ส่งผลให้ดัชนี DIF ที่คำนวณได้มีความคลาดเคลื่อนได้ (Camilli & Shappard. 1994 : 139) ใน

การวิเคราะห์ด้วยโปรแกรม SIBTEST และ DFIT ครั้งนี้ ใช้ทางเลือกที่เครื่องกำหนดโดยอัตโนมัติ (default) และทำการวิเคราะห์ครั้งเดียว หากใช้ทางเลือกอื่นอาจทำให้ผลการวิเคราะห์มีความคลาดเคลื่อนลดลง ในกรณีของวิธีซิปเทสท์อาจใช้การวิเคราะห์แบบ 2 ขั้นตอน กล่าวคือในตอนแรกวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบ โดยใช้คะแนนรวมของผู้สอบเป็นเกณฑ์การจับคู่ของกลุ่มผู้สอบ และในขั้นตอนที่สองวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบซ้ำอีกครั้ง โดยตัดข้อสอบที่ตรวจสอบพบการทำหน้าที่เบี่ยงเบนในขั้นตอนแรกออกก่อนที่จะคำนวณคะแนนรวมของผู้สอบ การวิเคราะห์ดังกล่าวจะทำให้เกณฑ์การจับคู่ของกลุ่มผู้สอบมีความบริสุทธิ์ ซึ่งจะส่งผลให้การระบุผิดพลาดในการตรวจสอบมีค่าลดลง (วลีมาส แซ่อึ้ง, 2543 : 282) หรือเลือกใช้การวิเคราะห์แบบยืนยัน (Confirmatory analysis) ซึ่งเป็นทางเลือกหนึ่งของโปรแกรมคอมพิวเตอร์สำเร็จรูป SIBTEST ที่ใช้ในการวิเคราะห์ครั้งนี้ (Stout. 1992) หรือใช้โปรแกรมคอมพิวเตอร์สำเร็จรูปรุ่นใหม่ที่มีการปรับปรุงให้เหมาะสมกับข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติมากขึ้น คือ MULTISIB (Stout, et al. 1997 : 195-213) เมื่อใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป DFIT อาจใช้วิธีเทียบมาตรฐานซ้ำ (two-stage iterative linking procedure) โดยการตัดข้อสอบที่ตรวจสอบพบการทำหน้าที่เบี่ยงเบนครั้งแรกออกก่อนการคำนวณสัมประสิทธิ์การเทียบมาตรฐาน (linking coefficient) ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป IPLINK และนำค่าที่ได้ไปใช้ในการวิเคราะห์ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป DFIT (Oshima, et al. 1997 : 265) อาจทำให้มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบสูงขึ้น

6) ขนาดของกลุ่มตัวอย่างเป็นตัวแปรหนึ่งที่มีผลต่อการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในกรณีของโปรแกรม SIBTEST ซึ่งใช้วิธีการของสถิตินอนพาราเมตริก อาจใช้กลุ่มตัวอย่างขนาดเล็กได้ กลุ่มตัวอย่างขนาดเล็กสุดที่ควรใช้คือ 100 คน (Stout, et al. 1997 : 199) รุสโซและสเตาต์ (Roussos & Stout. 1996 : 215-230) ศึกษาผลของการใช้กลุ่มตัวอย่างขนาดเล็กที่มีต่ออัตราความคลาดเคลื่อนชนิดที่ 1 วิเคราะห์ด้วยวิธี SIBTEST และ MH โดยใช้กลุ่มตัวอย่างขนาดละ 100 200 500 และ 1,000 คน ในการศึกษาที่ 1 และใช้กลุ่มตัวอย่างขนาดละ 500 1,000 3,000 คนในการศึกษาที่ 2 นอกจากนี้ กาญจนา วัธนสุนทร (2538) ศึกษาเพื่อพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ พบว่า วิธี SIBTEST และ MH ควรใช้กลุ่มตัวอย่าง 600 คนขึ้นไป จิตติมา วรรณศรี (2539) ได้เปรียบเทียบประสิทธิภาพของวิธีแมนเทิล-แฮนส์เซล และวิธีซิปเทสท์พบว่า เมื่อกลุ่มตัวอย่างมีขนาด 200 และ 500 คน ทั้งสองวิธีสามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบได้ถูกต้อง 50% แต่ถ้ากลุ่มตัวอย่างขนาด 1,000 คน สามารถตรวจสอบได้ถูกต้อง 100% กรณีของวิธี DFIT ใช้วิธีการของสถิติพาราเมตริกย่อมต้องใช้กลุ่มตัวอย่างขนาดใหญ่ในการศึกษา ดังที่แอกเคอร์แมน (Ackerman. 1994 : 136) กล่าวว่าแบบจำลองการตอบข้อสอบที่ใช้ความสามารถหลายมิติต้องการกลุ่มตัวอย่างขนาดใหญ่มาก (requires extremely large sample sizes) ซึ่งอาจมีข้อจำกัดสำหรับสถานการณ์การทดสอบทั่ว ๆ ไป และงานวิจัยส่วนใหญ่ที่ใช้วิธีดีเอฟไอที ใช้กลุ่มตัวอย่าง 1,000 คนขึ้นไป (Raju, et al. 1995 ; Oshima, Raju & Flowers. 1997 ; Oshima, et al. 1998 ; Larry & Oshima. 1998 ; Flowers, Oshima & Raju. 1999 ; Mulgreen. 2001) ยกเว้นงานวิจัยของราจูในปี ค.ศ. 1995 (Raju, et al. 1995) ที่ใช้กลุ่มตัวอย่าง 500 และ 1,000 คน ศึกษาข้อมูลการตอบข้อสอบมิติเดียว ในการศึกษาครั้งนี้ใช้กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน ศึกษาข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติจึงอาจไม่เหมาะสม ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิปเทสท์ เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ พบว่ากลุ่มตัวอย่างขนาด 50 100 และ 200 คน ส่งผลให้ความถูกต้องในการตรวจสอบไม่แตกต่างกัน แต่กลุ่มตัวอย่างขนาด 500 และ 1,000 คน ส่งผลให้ความถูกต้องและการระบุผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่าง

ขนาด 50 100 และ 200 คน เมื่อใช้วิธีดีเอฟไอที่พบว่ากลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน ส่งผลให้ความถูกต้องในการตรวจสอบไม่แตกต่างกัน แสดงให้เห็นว่ากลุ่มตัวอย่างขนาดต่าง ๆ ที่ใช้ในการศึกษานี้ อาจไม่มีผลต่อความไวในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบอย่างเพียงพอ ดังผลการวิจัยของสวามินาธานและโรเจอร์ส (Swaminathan & Rogers. 1990 : 361-370) ที่กล่าวว่าวิธีซิปเทสที่ใช้วิธีการของสถิตินอนพาราเมตริก เมื่อทดสอบกับสถิติ Z ซึ่งมีการแจกแจงแบบปกติไม่จำเป็นต้องใช้กลุ่มตัวอย่างขนาดใหญ่มาก แต่วิธีการถดถอยโลจิสติกใช้วิธีการของสถิติพาราเมตริก เมื่อทดสอบกับสถิติไคสแควร์ซึ่งมีการแจกแจงแบบเชิงเส้นกำกับ (asymptotic) จำเป็นต้องใช้กลุ่มตัวอย่างขนาดใหญ่ ซึ่งวิธีการดีเอฟไอที่ใช้ในการศึกษานี้ใช้วิธีการของสถิติพาราเมตริกที่ทดสอบกับสถิติไคสแควร์ที่มีการแจกแจงแบบเชิงเส้นกำกับเช่นกัน กรณีที่ศึกษากับข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติจึงควรใช้กลุ่มตัวอย่างขนาดใหญ่มากกว่า 1,000 คน ซึ่งเป็นกลุ่มตัวอย่างขนาดใหญ่สุดที่ใช้ในการศึกษานี้

7) ความยาวของแบบทดสอบเป็นตัวแปรหนึ่งที่มีอิทธิพลต่อความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ เพราะแบบทดสอบที่มีความยาวทำให้คะแนนมีความเชื่อมั่นและทำให้การประมาณค่าพารามิเตอร์ความสามารถมีความถูกต้องด้วย ดังนั้นความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบจะเพิ่มขึ้นเมื่อความยาวแบบทดสอบเพิ่มขึ้น (Narayanan & Sawaminathan. 1996 : 263) แต่ในการศึกษานี้พบว่าเมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ ความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ เมื่อกลุ่มตัวอย่างมีขนาด 50 100 200 500 และ 1,000 คน มีลักษณะเดียวกัน คือ วิธีซิปเทส กลุ่มตัวอย่างขนาด 50 100 และ 200 คน ส่งผลต่อความถูกต้องไม่แตกต่างกัน แต่กลุ่มตัวอย่างขนาด 500 และ 1,000 คน ส่งผลต่อความถูกต้องและการระบุผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 50 100 และ 200 คน วิธีดีเอฟไอที่ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน ส่งผลต่อความถูกต้องในการตรวจสอบไม่แตกต่างกัน แสดงว่าความยาวของแบบทดสอบไม่ส่งผลต่อความถูกต้องในการตรวจสอบ ซึ่งสอดคล้องกับการศึกษาของกาญจนา (2538) ที่พบว่า ความยาวของแบบทดสอบไม่มีผลกระทบต่อความถูกต้องในการตรวจสอบด้วยวิธีซิปเทสและวิธีแมนเทิล-แฮนส์เซล นอกจากนี้ในการศึกษานี้พบความผิดปกติของการตรวจสอบด้วยวิธีซิปเทส เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ กลุ่มตัวอย่างมีขนาด 1,000 คน จากการทดลองซ้ำ 50 กลุ่มตรวจสอบพบข้อสอบแสดงการทำหน้าที่เบี่ยงเบนของข้อสอบเพียงกลุ่มเดียว นอกนั้นตรวจสอบไม่พบ แต่ข้อมูลชุดเดียวกันเมื่อใช้ตรวจสอบด้วยวิธีดีเอฟไอที่ ปรากฏว่าตรวจสอบพบการทำหน้าที่ของกลุ่มต่าง ๆ ตามปกติ การศึกษาโดยใช้แบบทดสอบที่ประกอบด้วยข้อสอบ 40 ข้อ เคยมีการศึกษามาก่อนแล้ว เช่น การศึกษาโดยใช้ข้อมูลจำลองที่นารายานานและสวามินาธาน (Narayanan & Swaminathan. 1994 : 315-328) ใช้แบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ วิเคราะห์ด้วยวิธี SIBTEST และ MH ราชู แวนเดอร์ลินเด็นและฟลีร์ (Raju, Van der Linden & Fleer. 1995 : 353-368) ใช้ข้อมูลจำลองจากแบบทดสอบที่ประกอบด้วยข้อสอบ 40 ข้อ วิเคราะห์ด้วยวิธีการของลอร์ด การวัดพื้นที่แบบมีเครื่องหมาย การวัดพื้นที่โดยไม่มีเครื่องหมาย และวิธีดีเอฟไอที่ การศึกษาโดยใช้ข้อมูลจำลองและการใช้ข้อมูลเชิงประจักษ์อาจให้ผลไม่สอดคล้องกันและการตรวจสอบด้วยวิธีการที่แตกต่างกันก็อาจให้ผลแตกต่างกัน เพราะเกิดความคลาดเคลื่อนของการวัดที่แตกต่างกัน (Camilli & Shappard. 1994 : 139)

8) การประมาณค่าพารามิเตอร์ความสามารถของผู้สอบกลุ่มสนใจ (เพศหญิง) เพื่อนำไปใช้ในการคำนวณด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป DFIT โดยใช้โปรแกรม BILOG Version 3.04 อาจไม่เหมาะสม เพราะค่าที่คำนวณได้เป็นค่าประมาณของพารามิเตอร์ความสามารถที่ข้อสอบต้องการวัด (θ)

เท่านั้น ไม่ได้ประมาณค่าความสามารถอื่นที่ข้อสอบไม่ต้องการวัด (11) ด้วย จึงควรใช้โปรแกรมคอมพิวเตอร์ที่เหมาะสมในการคำนวณพารามิเตอร์ความสามารถหลายมิติ ซึ่งมีการแจกแจงแบบปกติหลายมิติ (multivariate normal distribution) เช่น วิธีการของนิวตัน-ราฟสัน (Newton-Raphson iteration ; Ackerman. 1994 : 133) หรือใช้โปรแกรมคอมพิวเตอร์สำเร็จรูปที่ประมาณค่าได้ทั้งพารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถตามกรอบแนวคิดทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติได้ ซึ่งในอนาคตอาจมีผู้พัฒนาโปรแกรมคอมพิวเตอร์สำเร็จรูปที่มีความสามารถดังกล่าว (Oshima Raju & Flowers. 1997 : 269) ถ้าใช้วิธีการประมาณค่าพารามิเตอร์ความสามารถที่เหมาะสมอาจทำได้ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนข้อสอบด้วยโปรแกรม DFIT มีประสิทธิภาพสูงขึ้น

1.2 ในการศึกษาครั้งนี้ความสอดคล้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอทีที่มีค่าค่อนข้างต่ำและมีการเปลี่ยนแปลงไม่เป็นระบบในเงื่อนไขของความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดต่าง ๆ อาจเป็นผลมาจากความไม่คงที่ของผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบซึ่งเกิดจากสาเหตุต่าง ๆ ตามที่ได้อภิปรายมาแล้วในข้อ 1.1 จึงส่งผลกระทบต่อความสอดคล้องของการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วย

2. การเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ

ในการศึกษาครั้งนี้ความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที ไม่มีเงื่อนไขใดที่ตรวจสอบได้ถูกต้องร้อยละ 100 อาจเป็นเพราะความคลาดเคลื่อนในการพิจารณาตัดสินการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบที่ผู้เชี่ยวชาญมีโอกาสพิจารณาครั้งเดียวก่อนการทดลองใช้ข้อสอบ ถ้าเปิดโอกาสให้ผู้เชี่ยวชาญมีโอกาสพิจารณาซ้ำ เมื่อมีการปรับปรุงคุณภาพข้อสอบหลังจากทดลองใช้ และตัดข้อสอบออกจากแบบทดสอบแล้ว หรือให้ผู้เชี่ยวชาญกลุ่มเดิมหรือกลุ่มอื่นมีโอกาสร่วมกันพิจารณาซ้ำเพื่อตรวจสอบความถูกต้องอีกรอบ ดังเช่นที่กล่าวในข้อ 1.1 (1) อาจทำให้ความคลาดเคลื่อนลดลง หรืออาจเป็นเพราะการแบ่งหมวดข้อสอบไม่เหมาะสม เนื่องจากใช้การแบ่งหมวดข้อสอบตามเนื้อหาเพียงอย่างเดียว ในการแบ่งหมวดข้อสอบอาจใช้วิธีการอื่น เช่น ดักกลาส รุสโซ และสเตาต์ (Douglas, Roussos & Stout. 1996 : 465-488) ได้พิจารณาแบ่งหมวดข้อสอบโดยใช้ผู้เชี่ยวชาญเป็นผู้พิจารณา ใช้วิธีการทางสถิติ และใช้วิธีการทางสถิติร่วมกับผู้เชี่ยวชาญ โอชิม่าและคณะ (Oshima, et al. 1998 : 353-369) แบ่งหมวดข้อสอบโดยจัดตามลำดับขั้นของความคิด (cognitive classification) ตามวัตถุประสงค์การสอนและตามเนื้อหา นอกจากนั้น ขนาดกลุ่มตัวอย่างที่ใช้ในการศึกษาอาจไม่เหมาะสม เช่น กรณีนีที่แวนเด็นเบอร์กและเกอร์ล (Vandenberghe & Gierl. 2001) ได้ตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบวิชาคณิตศาสตร์ วิทยาศาสตร์ และสังคมศาสตร์ โดยใช้นักศึกษาชาวอะบอริจินและไม่ใช่ชาวอะบอริจิน มีสัดส่วนดังนี้ 956:2,000 480:2,000 971:2,000 โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป SIBTEST ปรากฏว่าผลไม่เป็นไปตามสมมติฐานที่กำหนดไว้และสรุปว่าข้อจำกัดเกิดจากจำนวนกลุ่มตัวอย่างของนักศึกษาชาวอะบอริจินซึ่งมีจำนวนจำกัด โอชิม่าและคณะ (Oshima, et al. 1998) ตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบจากข้อมูลการตอบข้อสอบที่ใช้ความสามารถมิติเดียว โดยใช้ข้อมูลเชิงประจักษ์ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป DFIT ใช้กลุ่มตัวอย่างขนาดละ 1,000 คน ทั้งกลุ่มเปรียบเทียบและกลุ่มสนใจ ดังนั้นการศึกษานี้ใช้กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน อาจไม่เหมาะสมกับการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบจากข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ

3. การเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบ

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที ใน การศึกษาครั้งนี้ไม่มีเงื่อนไขใดที่ตรวจสอบได้ถูกต้องร้อยละ 100 แต่ความถูกต้องในการตรวจสอบในเงื่อนไข ต่าง ๆ มีค่าสูง อาจเป็นเพราะในการศึกษานี้เลือกใช้การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ ลำเอียงเข้าข้างเพศชาย และกำหนดให้ข้อสอบข้ออื่น ๆ ในแบบทดสอบเป็นข้อสอบที่ไม่แสดงการทำหน้าที่ เบี่ยงเบนของข้อสอบ จึงทำให้ปริมาณ DIF ของแต่ละข้อมีโอกาสสะสมทำให้แสดงการทำหน้าที่เบี่ยงเบน ระดับแบบทดสอบมากกว่าการเลือกใช้การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบแบบที่ไม่กำหนดให้ ข้อสอบข้ออื่นในแบบทดสอบแสดงการทำหน้าที่เบี่ยงเบนของข้อสอบ อาจทำให้มีการหักล้างกันของปริมาณ DIF ในแต่ละข้อ และทำให้มีโอกาสสะสมทำให้เกิด DTF น้อยลง (Nandakumar, 1993 ; Oshima, et al. 1997) และผลการศึกษาในครั้งนี้สอดคล้องกับการศึกษาของ นนดาคุมาร์ (Nandakumar, 1993 : 293-311) ที่ตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบด้วยวิธีชิบเทสต์โดยใช้ข้อมูลจำลองทำการศึกษาซ้ำ 100 ครั้ง พบว่าการตรวจสอบถูกต้องไม่ถึงร้อยละ 100

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำไปใช้

1.1 การศึกษาครั้งนี้พบว่าวิธีชิบเทสต์และวิธีดีเอฟไอที ตรวจสอบการทำหน้าที่เบี่ยงเบนของ ข้อมูลการตอบข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 40 และ 50 ข้อ ให้ผลแตกต่างกันในเงื่อนไขขนาดกลุ่มตัวอย่างแตกต่างกัน คือ วิธีชิบเทสต์พบว่ากลุ่มตัวอย่าง ขนาด 50 100 และ 200 คน ส่งผลต่อความถูกต้องในการตรวจสอบไม่แตกต่างกัน แต่เมื่อกลุ่มตัวอย่าง เพิ่มขึ้นเป็น 500 และ 1,000 คน ส่งผลให้มีความถูกต้องและการระบุผิดพลาดในการตรวจสอบการทำหน้าที่ เบี่ยงเบนของข้อสอบสูงกว่ากลุ่มตัวอย่างขนาด 50 100 และ 200 คน แต่วิธีดีเอฟไอทีกลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบไม่ แตกต่างกัน อาจเป็นเพราะกลุ่มตัวอย่างขนาดดังกล่าวส่งผลต่อความไวในการตรวจสอบการทำหน้าที่ เบี่ยงเบนของข้อสอบด้วยวิธีการทั้งสองไม่มากพอ และตามกรอบความคิดเชิงทฤษฎีของแบบจำลองการตอบ ข้อสอบที่ใช้ความสามารถหลายมิติ ในการศึกษาควรใช้กลุ่มตัวอย่างขนาดใหญ่ จะทำให้ผลการตรวจสอบการ ทำหน้าที่เบี่ยงเบนมีความถูกต้องสูงและการระบุผิดพลาดต่ำ จึงควรใช้วิธีดังกล่าวในสถานการณ์การสอบที่มีผู้ เข้าสอบเป็นจำนวนมาก มุ่งสร้างแบบทดสอบมาตรฐานหรือพัฒนาคลังข้อสอบ และควรเลือกใช้วิธีการชิบเท สต์มากกว่าวิธีดีเอฟไอที แม้ว่าวิธีชิบเทสต์มีความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ น้อยกว่าวิธีดีเอฟไอที แต่มีการระบุผิดพลาดในการตรวจสอบน้อยกว่าด้วย และวิธีชิบเทสต์มีความถูกต้องใน การตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบและแบบทดสอบมากกว่าวิธีดีเอฟไอที เมื่อกลุ่ม ตัวอย่างมีขนาดใหญ่ นอกจากนั้นยังใช้ได้สะดวก ประหยัดทั้งเวลาและแรงงานในการวิเคราะห์

1.2 การศึกษาครั้งนี้พบว่ามีการระบุผิดพลาดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของ ข้อสอบสูง สาเหตุหนึ่งอาจเนื่องจากมีสัดส่วนของข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนในแบบทดสอบ มากกว่าร้อยละ 20 ดังผลการวิจัยหลายเรื่อง แต่ในการศึกษานี้มีวัตถุประสงค์ในการสร้างแบบทดสอบให้มี ข้อสอบส่วนหนึ่งแสดงการทำหน้าที่เบี่ยงเบน และไม่ได้มุ่งพัฒนาคุณภาพของแบบทดสอบ จึงไม่ได้คัดเลือก ข้อสอบดังกล่าวออกจากแบบทดสอบ ในการสร้างและพัฒนาคุณภาพแบบทดสอบเพื่อนำไปใช้จริง ควรทำการ ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบทุกขั้นตอน เริ่มตั้งแต่การสร้างตารางกำหนดข้อสอบ จนกระทั่ง ภายหลังการนำแบบทดสอบไปใช้แล้ว โดยใช้ทั้งวิธีการทางสถิติร่วมกับการใช้ดุลยพินิจของผู้เชี่ยวชาญ จะ

ช่วยให้สามารถสร้างและพัฒนาแบบทดสอบให้มีคุณภาพสูงตรงตามจุดมุ่งหมายยิ่งขึ้น เช่น การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในกระบวนการสร้างและพัฒนาแบบทดสอบที่ดำเนินการโดยศูนย์บริการการทดสอบทางการศึกษา (Educational Testing Service : ETS) ในสหรัฐอเมริกา เป็นต้น

1.3 การศึกษาครั้งนี้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ แต่ไม่ได้นำผลการตรวจสอบไปใช้ ในทางปฏิบัติการนำผลการตรวจสอบการทำหน้าที่เบี่ยงเบนไปใช้มีวัตถุประสงค์แตกต่างกัน เป็นต้นว่า ในการปรับปรุงคุณภาพข้อสอบรายข้อควรพิจารณาผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ การพิจารณาผลกระทบจากการทำหน้าที่เบี่ยงเบนของข้อสอบหลายข้อที่รวมกันด้วยเงื่อนไขใดเงื่อนไขหนึ่งพร้อมกัน ควรพิจารณาผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ ส่วนผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของแบบทดสอบเป็นการพิจารณาผลกระทบจากการทำหน้าที่ของข้อสอบทุกข้อในแบบทดสอบพร้อมกัน ใช้ในกรณีที่ต้องการพิจารณาคะแนนรวมจากการทำข้อสอบ เช่น การคัดเลือก การจัดตำแหน่ง เป็นต้น ในการนำผลการตรวจสอบการทำหน้าที่เบี่ยงเบนไปใช้ นอกจากต้องคำนึงถึงวัตถุประสงค์แล้ว ควรใช้ผลการตรวจสอบด้วยวิธีการทางสถิติรวมกับการใช้ดุลยพินิจของผู้เชี่ยวชาญจะทำให้ได้ผลดียิ่งขึ้น

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 ควรทำการศึกษาซ้ำด้วยวิธีการเดิม คือ เปรียบเทียบวิธีชิบเทสต์และวิธีดีเอฟไอที ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบจากข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ อาจใช้เนื้อหาวิชาอื่น ระดับชั้นอื่น กำหนดตัวแปรให้กว้างขวางขึ้น และแก้ไขข้อบกพร่องต่าง ๆ ในการวิจัยครั้งนี้ตามที่ได้กล่าวมาแล้วในการอภิปรายผล พร้อมทั้งศึกษาที่ระดับนัยสำคัญหลาย ๆ ระดับ เช่น .01, .05, .10

2.2 ควรมุ่งศึกษาการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ เพื่อปรับปรุงคุณภาพข้อสอบและแบบทดสอบ โดยการพิจารณาค่าความเที่ยงตรงเชิงโครงสร้าง ค่าความเชื่อมั่นของแบบทดสอบ หรือสารสนเทศของแบบทดสอบ ภายหลังการตัดข้อสอบที่ตรวจสอบพบการทำหน้าที่เบี่ยงเบนออกไปแล้ว พร้อมทั้งศึกษาการแปลความหมายคะแนนที่ได้จากการตอบข้อสอบที่ใช้ความสามารถหลายมิติด้วย

2.3 ควรศึกษาในสถานการณ์การสอบที่ผู้สอบจำนวนมาก เช่น การสอบคัดเลือก การสร้างแบบทดสอบมาตรฐานของหน่วยงานที่รับผิดชอบ การสอบระดับภูมิภาค หรือระดับประเทศ เพื่อจะได้มีกลุ่มตัวอย่างขนาดใหญ่สอดคล้องกับแนวคิดของแบบจำลองการตอบข้อสอบที่ใช้ความสามารถหลายมิติ ซึ่งส่งผลให้มีความถูกต้องในการตรวจสอบมากขึ้นและการระบุผิดพลาดในการตรวจสอบลดลง หรือศึกษาตัวแปรตัวอื่น ๆ ได้อย่างกว้างขวาง เพื่อจะได้สารสนเทศประกอบการตัดสินใจนำไปใช้ในทางปฏิบัติจริงได้อย่างมีประสิทธิภาพ

2.4 ควรศึกษาเพื่อเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ ด้วยวิธีชิบเทสต์และวิธีดีเอฟไอที โดยใช้ข้อมูลจำลอง ทั้งการให้คะแนนรายข้อแบบสองค่าและการให้คะแนนรายข้อแบบหลายค่า โดยการจำลองข้อมูลให้สอดคล้องตามกรอบความคิดของทฤษฎีการตอบข้อสอบที่ใช้ความสามารถหลายมิติ เป็นต้นว่าลักษณะการแจกแจงความสามารถของผู้สอบในมิติต่าง ๆ และตรวจสอบการทำหน้าที่เบี่ยงเบนในเงื่อนไขของตัวแปรต่าง ๆ เช่น ความยาวแบบทดสอบ ลักษณะของข้อสอบ ขนาดกลุ่มตัวอย่าง สัดส่วนของข้อสอบแสดงการทำหน้าที่เบี่ยงเบนของข้อสอบ อัตราส่วนของกลุ่มเปรียบเทียบต่อกลุ่มสนใจ เพื่อเปรียบเทียบประสิทธิภาพของการตรวจสอบด้วยวิธีทั้งสอง

ในด้านต่าง ๆ เช่น ความถูกต้องในการตรวจสอบ การระบุผิดพลาดในการตรวจสอบทั้งประเภทที่ 1 และประเภทที่ 2 ความสอดคล้องของการตรวจสอบ สหสัมพันธ์ของการตรวจสอบ เพื่อจะได้เป็นแนวทางในการศึกษา โดยใช้ข้อมูลเชิงประจักษ์ได้อย่างมีประสิทธิภาพต่อไป

บรรณานุกรม

ภาษาไทย

- กาญจนา วันนสุนทร. (2538). การพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- จิตติมา วรณศรี. (2539). การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทิล-แฮนส์เชล กับวิธีซิปเทสท์ เมื่อความยาวแบบทดสอบ ขนาดกลุ่มตัวอย่าง และอัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกัน. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- ญาณภัทร สีหะมงคล. (2540). การเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี $Lord's \chi^2$ วิธี $Raju's Area Measures$ และวิธี $Closed Interval Area$. วิทยานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร. ถ่ายเอกสาร.
- ทองอยู่ สาระ. (2543). การเปรียบเทียบอำนาจการตรวจสอบและการจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบสมมาตรและไม่สมมาตรระหว่างวิธีแมนเทิล-แฮนส์เชล และวิธีการถดถอยโลจิสติก โดยใช้ความยาวแบบทดสอบและขนาดกลุ่มตัวอย่างต่างกัน. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- นพมาศ พิพัฒน์สุข. (2541). การเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทิล – แฮนส์เชล กับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเมื่อใช้เกณฑ์จับคู่เปรียบเทียบแตกต่างกันในแบบทดสอบชนิดพหุมิติ. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- บุญเชิด ภิญโญนนตพงษ์. (2527). การทดสอบอิงเกณฑ์. เทคนิคและวิธีการ. กรุงเทพมหานคร : โอเดียนสโตร์.
- บุญเชิด ภิญโญนนตพงษ์. (2542). สมประสิทธิ์ r_B : การประมาณค่าความเชื่อมั่นสำหรับแบบทดสอบผลสัมฤทธิ์ชนิดเลือกตอบที่ประกอบด้วยความยากรายข้อต่างกัน. รายงานการวิจัย. ภาควิชาการวัดผลและวิจัยการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- พรรณี จิตมาศ. (2540). การเปรียบเทียบผลการวิเคราะห์ความลำเอียงข้อสอบโดยใช้ขนาดกลุ่มผู้สอบและวิธีวิเคราะห์ต่างกัน. วิทยานิพนธ์ กศ.ม. (วัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร. ถ่ายเอกสาร.
- เรวดี อินทะสระ. (2539). ผลการตรวจสอบความลำเอียงของข้อสอบต่อการศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่คิดคะแนนต่างกัน. วิทยานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร. ถ่ายเอกสาร.
- วลีมาส แซ่อิง. (2543). การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบบอเนกรูป ระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์

- วิธีแมนเทิล-แฮนส์เซล และวิธีถดถอยโลจิสติก. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผล การศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- ศิริชัย กาญจนวาสี. (2535). “ทฤษฎีการทดสอบ,” ใน *การพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน*. หน้า 84-135. นนทบุรี. มหาวิทยาลัยสุโขทัยธรรมาธิราช.
- เสรี ชัดเข้ม. (2539). *การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันแบบไม่สม่ำเสมอของข้อสอบ ระหว่างวิธีแมนเทิล-แฮนส์เซลแบบปกติ กับวิธีแมนเทิล-แฮนส์เซลแบบแบ่งกลุ่มความสามารถ ผู้สอบและความยาวของข้อสอบ*. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- อารี วัชรโสติกุล. (2543). *การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยใช้ รูปแบบและวิธีต่างกัน*. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- อุทุมพร จามรมาน. (2540). *ทฤษฎีการวัดทางจิตวิทยา*. กรุงเทพมหานคร : ฟีนีქซ์บุ๊คซิง.

ภาษาอังกฤษ

- Ackerman, Terry A. (1992). “A Didactic Explanation of Item Bias, Item Impact, and Item Validity From a Multidimensional Perspective,” *Journal of Educational Measurement*. 29(1) : 67-91.
- Ackerman, Terry A. (1994) . “A Discussion of Measurement Direction in a Multidimensional Latent Space and the Role It Plays in Bias Detection,” in *Modern Theories of Measurement: Problems and Issues*. P. 105-139. Ottawa : University of Ottawa.
- Ackerman, Terry A. (1996, December). Graphical Representation of Multidimensional Item Response Theory Analysis,” *Applied Psychological Measurement*. 20(4) : 311-329.
- Allen, Mary J. & Yen, Wendy, M. (1979). *Introduction to Measurement Theory*. California : Brooks/Cole Publishing Company.
- Ansley, Timothy N. & Forsyth, Robert A. (1985, March). “An Examination of the Characteristics of Unidimensional IRT Parameter Estimates Derived From Two-Dimensional Data,” *Applied Psychological Measurement*. 9(1) : 37-48.
- Camilli, Gregory & Sheppard, Lorrie A. (1994). *Methods for Identifying Biased Test Items*. California : Sage Publications.
- Camilli, Gregory. (1992, June). “A Conceptual Analysis of Differential Item Functioning in Terms of a Multidimensional Item Response Model,” *Applied Psychological Measurement*. 16(2) : 129-147.
- Carlson, James E. (1987). *Multidimensional Item Response Theory Estimation: A Computer Program*. (ACT Research Report Series; 87-19)
- Clauser, E. et al. (1996). “A Comparison of Alternative Matching Strategies for DIF Detection in Tests That Are Multidimensional,” *Journal of Educational Measurement*. 33(2) : 202-214.

- Clauser, Brain E. et al. (1997). "Development of a Scoring Algorithm to Replace Expert Rating for Scoring a Complex Performance-Based Assessment," *Applied Measurement in Education*. 10(4) : 345-358.
- Douglas, A., Roussos, A. & Stout, F. (1996). "Item – Bundle DIF Hypothesis Testing: Identifying Suspect Bundles and Assessing Their Differential Functioning," *Journal of Educational Measurement*. 33(4) : 465-484.
- Flowers, P., Oshima, C. & Raju, S. (1999, December). "A Description and Demonstration of the Polytomous – DFIT Framework," *Applied Psychological Measurement*. 23(4) : 309-326.
- Gipps Caroline & Murphy Patricia. (1994). *A Fair Test ? Assessment, Achievement and equity*. Buckingham : Open University press.
- Goldstein, Harvey. (1996). "Statistical and Psychometric Models for Assessment," in *Assessment : Problems, Developments and Statistical Issues*. England : John Wiley & Sons Ltd.
- Guilford, J.P. & B. Fruchter. (1978) *Fundamental Statistics in Psychology and Education*. Singapore : McGraw-Hill.
- Hambleton, K., Sawaminathan, H. & Rogers, H. (1991). *Fundamentals of Item Response Theory*. California : Sage Publications.
- Holland, Paul W. & Wainer, Howard. (1993). *Differential Item Functioning*. Hillsdale, New Jersey : Lawrence Erlbaum Associates.
- Joseph, Hair F. et al. (1998). *Multivariate Data Analysis*. New Jersey : Lawrence Erlbaum Associates.
- Larry R. & Oshima T.C. (1998). "Differential Functioning and Language Translation : A Cross-National Study with a Test Development for Certification," (online). Available : <http://thailis-web.uni.net.th/ERIC/>
- Laveault, Dany et al. (1994). *Modern Theories of Measurement: Problems and Issues*. Ottawa : University of Ottawa.
- Linn, Robert L. (1993). "The Use of Differential Item Functioning Statistics: A Discussion of Current Practice and Future Implications," in *Differential Item Functioning*. P. 349-363. Hillsdale, New Jersey : Lawrence Erlbaum Associates.
- McCauley, Cynthia D. & Mendoza Jorge. (1985, December). "Simulation Study of Item bias Using a Two – Parameter Item Response Model," *Applied Psychological Measurement*. 9(4) : 389-400.
- Miller, Timothy R. (1991). *Empirical Estimation of Standard Errors of Compensatory MIRT Model Parameters Obtained from the NOHARM Estimation Program*. (ACT Research Report Series; 91-2)
- Monaco, Malina. (1997). "A Monte Carlo Study of Skewed Theta Distribution on DIF Indices," (online). Available <http://ericir.syr.edu/plweb.../>

- Mulqueen, Matthew C. (2001). "Identification of latent constructs and assessment of measurement equivalence across rating sources on a 360-degree performance appraisal instrument," (online). available : <http://www.lib.umi.com/dissertations/fullcit/3010021>
- Nandakumar Ratana. (1993). "Simultaneous DIF Amplification and Cancellation: Shealy – Stout' s Test for DIF," *Journal of Educational Measurement*. 30(4) : 293-311.
- Nandakumar Ratana. (1994). "Assessing Dimensionality of a Set of Item Responses – Comparison of Different Approaches," *Journal of educational Measurement*. 31(1) : 17-35.
- Narayanan, P. & Swaminathan, H. (1994, December). "Performance of the Mantel-Haenszel and Simultaneous Item Bias Procedures for Detecting Differential Item Functioning" *Applied Psychological Measurement*. 18(4) : 315-328.
- Narayanan, Pankaja & Sawaminathan, H. (1996). "Identification of Items that Show Nonuniform DIF," *Applied Psychological Measurement*. 20(3) : 257-274.
- Oshima, C. & Miller, David. (1990). "Multidimensionality and IRT – Based Item Invariance Indexes: The Effect of Between – Group Variation in Trait Correlation," *Journal of Educational Measurement*. 27(3) : 273-283.
- Oshima, C., Raju, S. & Flowers, P. (1997). "Development and Demonstration of Multidimensional IRT – Based Internal Measures of Differential Functioning of Items and Tests," *Journal of Educational Measurement*. 34(3) : 253-272.
- Oshima, C., Raju, S. & flowers P. (1998). "Differential Bundle Functioning Using the DFIT Framework: Procedures for Identifying Possible Sources of Differential Functioning," *Applied Measurement in Education*. 11(4) : 353-369.
- Oshima, T. C. & Miller, David D. (1992, September). "Multidimensionality and Item Bias in Item Response Theory," *Applied Psychological Measurement*. 16(3) : 237-248.
- Raju, S. et al. (1993). "An Empirical Comparison of the Area Methods, Lord' s Chi – Square Test, and the Mantel – Haenszel Technique for Assessing Differential Item Functioning," *Educational and Psychological Measurement*. 53 : 301-314.
- Raju, S., van der Linden, J. & Fleer F. (1995, December). "IRT-based Internal Measures of Differential Functioning of Items and Tests," *Applied Psychological Measurement*. 19(4) : 353-368.
- Reckase, Mark D. & McKinley, Robert L. (1991, December). "The Discriminating Power of Items That Measure More than One Dimension," *Applied psychological Measurement*. 15(4) : 361-373.
- Reckase, Mark D. (1985, December). "The Difficulty of Test Items That Measure More Than One Ability," *Applied Psychological Measurement*. 9(4) : 401-412.
- Reckase, Mark D. (1994). "What is the 'Correct' Dimensionality for a Set of Item Response Data?, in *Modern Theories of Measurement: Problems and Issues*. P. 87-92. Ottawa : University of Ottawa.

- Reckase, Mark D. (1997). "A Linear Logistic Multidimensional Model for Dichotomous Item Response Data," in *Handbook of Modern Item Response Theory*. P. 271-286. New York : Springer – Verlag.
- Reckase, Mark D. (1997, March). "The Past and Future of Multidimensional Item Response Theory," *Applied Psychological Measurement*. 21(1) : 25-36.
- Roussos, Louis A. & Stout, William F. (1996). "Simulation Studies of the Effects of Small Sample Size and Studied Item Parameters of SIBTEST and Mantel – Haenszel Type I Error Performance," *Journal of Educational Measurement*. 33(2) : 215-230.
- Roussos, Louis A. & Stout, William F. (1996, December). "A Multidimensionality – Based DIF Analysis Paradigm," *Applied Psychological Measurement*. 20(4) : 355-371.
- Ryan, Katherine E. & Chiu, Shuwan. (2001). "An Examination of Item Context Effects, DIF and Gender DIF," *Applied Measurement in Education*. 14(1) : 73-90.
- Spray, A. et al. (1990). *Comparison of Two Logistic Multidimensional Item Response Theory Models*. (ACT Research Report Series; 90-8)
- Stevens, James. (1996). *Applied Multivariate Statistics for the Social Sciences*. Mahwah, N.J. : Lawrence Erlbaum Associates.
- Stout, William & Roussos, Louis. (1992). *SIBTEST User Manual*. Illinois : University of Illinois.
- Stout, William. et al. (1997). "MULTISIB : A procedure to Investigate DIF When a Test Is Intensionally Two-Dimensional," *Applied Psychological Measurement*. 21(3) : 195-213.
- Swaminathan, Hariharan & Roger, Jane H. (1990). "Detecting Differential Item Functioning Using Logistic Regression Procedures," *Journal of Educational Measurement*. 27(4) : 361-370.
- Vandenberghe, Christine N. & Gierl, Mark J. (2001). "Differential Bundle Functioning on Three Achievement Tests : A comparison of Aboriginal and Non-Aboriginal Examinees," (online). Available : <http://www.education.ualberta.ca/educ/psych/crame>.
- Whitmore, Marjorie L. & Schumacker, Randall E. (1999). "A Comparison of Logistic Regression and Analysis of Variance Differential Item Functioning Detection Methods," *Educational and Psychological Measurement*. 59(6) : 910-927.
- Wilson, Jame W. (1971). "An Evaluation of Learning in Secondary School Mathematics," in *Handbook on Formative Evaluation of Student Learning*. USA : McGraw-Hill.
- Yamane, Taro. (1974). *Statistics : an introductory analysis*. 2 nd ed. New York : Harper.
- Zhang, Jinming & Stout William. (1998). "The Theoretical DETECT Index of Dimensionality and Its Application to Approximate Simple Structure" (online). Available <http://Chatsiri.hypermart.net/Elearning/stat7.htm>
- Zieky, Micheal. (1993). "Practical Questions in the use of DIF Statistics in Test Development," in *Differential Item Functioning*. P. 337-347. Hillsdale, New Jersey : Lawrence Erlbaum Associates.

ภาคผนวก ก

รายชื่อผู้เชี่ยวชาญที่ตรวจสอบคุณภาพของแบบทดสอบ

1. อาจารย์สุชาติ สิริมีนนท์
อาจารย์ 2 ระดับ 7 โรงเรียนบางกะปิ กรุงเทพมหานคร
2. อาจารย์พรณี เจียมสุบุตร
อาจารย์ 2 ระดับ 6 โรงเรียนสวนกุหลาบวิทยาลัย กรุงเทพมหานคร
3. อาจารย์ลัดดา เสือสีพันท์
อาจารย์ 2 ระดับ 6 โรงเรียนสุวรรณารามวิทยาคม กรุงเทพมหานคร
4. ว่าที่ ร.ต.อรรคพล คำภู
อาจารย์ 2 ระดับ 6 โรงเรียนลาดปลาเค้าพิทยาคม กรุงเทพมหานคร
5. อาจารย์อนันต์ โพธิกุล
ผู้ช่วยหัวหน้าหมวดคณิตศาสตร์ โรงเรียนอัสสัมชัญ กรุงเทพมหานคร
6. อาจารย์อารี วัชรโสติกุล
อาจารย์หมวดวิชาคณิตศาสตร์ระดับมัธยมศึกษา โรงเรียนพระหฤทัยคอนแวนต์ กรุงเทพมหานคร

รายชื่อโรงเรียนกลุ่มตัวอย่าง

ทดลองใช้เครื่องมือครั้งที่ 1

โรงเรียนเตรียมอุดมศึกษาน้อมเกล้า นนทบุรี

ทดลองใช้เครื่องมือครั้งที่ 2

โรงเรียนนวมินทราชูทิศหอวัง นนทบุรี

โรงเรียนศรีบุญยานนท์

ทดลองใช้เครื่องมือครั้งที่ 3

โรงเรียนวัดเขมาภิรตาราม

โรงเรียนปากเกร็ด

โรงเรียนสวนกุหลาบวิทยาลัย

โรงเรียนบางบัวทอง

โรงเรียนโพธิ์นิมิตรวิทยา

โรงเรียนไทรน้อย

ภาคผนวก ง

ตาราง 32 สรุปผลการพิจารณา DIF DBF DTF ผลการวิเคราะห์ข้อสอบรายข้อและผลการคัดเลือกข้อสอบ
สำหรับแบบทดสอบฉบับต่าง ๆ

เนื้อหา	ลำดับข้อ	รหัสข้อสอบ	IOC				DIF	ฉ.1 ทดลอง		ฉ.2 ทดลอง		ฉ.3		ฉ.4	ฉ.5	
			ช.	ญ.	รวม	65 ข้อ		(1) p	r	60 ข้อ	(2) p	r	50 ข้อ	40 ข้อ	30 ข้อ	
บทที่ 1																
1.2 ระบบเลขฐานสิบ	1	1.2.1	0.3	0.0	0.2	X	1	0.55	0.22	1	0.74	0.37	1	1	1	
	2	1.2.2	0.3	-0.3	0.0	X										
1.3 เลขยกกำลัง	3	1.3.1	-0.3	-0.6	-0.5	X	2	0.33	0.22	2	0.61	0.17	2			
	4	1.3.2	0.3	0.3	0.3	X										
	5	1.3.3	0.3	0.3	0.3	X	3	0.25	0.28	3	0.40	0.40	3	2	2	
	6	1.3.4	0.3	0.3	0.3	X	4	0.25	0.28							
	DBF(1)		-0.3	-0.3	-0.3	X										
บทที่ 2																
2.2 จำนวนเฉพาะ	7	2.2.1	0.0	-0.3	-0.2	X	5	0.24	0.20	4	0.20	0.18	4	3	3	
2.3 การแยกตัวประกอบ	8	2.3.1	0.0	-0.3	-0.2	X	6	0.30	0.19	5	0.40	0.44	5	4	4	
2.4 ท.ร.ม.	9	2.4.1	0.3	-0.3	0.0	X	7	0.33	0.23	6	0.31	0.38	6			
	10	2.4.2	0.0	-0.3	-0.3	X										
	11	2.4.3	0.3	0.6	0.5	✓	8	0.15	0.11	7	0.13	0.18				
2.5 ค.ร.น	12	2.4.4	0.3	1.0	0.7	✓	9	0.17	0.20	8	0.27	0.35	7	5	5	
	13	2.5.1	0.3	-0.3	0.0	X										
	14	2.5.2	0.0	-0.3	-0.3	X	10	0.31	0.28	9	0.41	0.31	8	6		
	15	2.5.3	0.3	0.6	0.5	✓	11	0.17	0.15	10	0.16	0.19				
	16	2.5.4	0.6	0.6	0.7	✓	12	0.25	0.28	11	0.15	0.23	9	7	6	
	DBF(2)		0.3	-0.6	-0.3	X										
บทที่ 3																
3.1 ค่าประจำหลักของทศนิยม	17	3.1.1	0.3	-0.6	-0.3	X	13	0.47	0.05	12	0.60	0.42				
	18	3.1.2	-0.3	0.3	0.0	X	14	0.26	0.20	13	0.23	0.15	10	8	7	
3.2 การเปรียบเทียบทศนิยม	19	3.2.1	0.0	-0.6	-0.3	X	15	0.22	0.17	14	0.27	0.50	11	9	8	
3.3 การบวกและการลบทศนิยม	20	3.3.1	-0.3	-0.3	-0.3	X	16	0.33	0.33	15	0.50	0.38	12			
	21	3.3.2	0.3	0.6	0.5	✓	17	0.22	0.15	16	0.42	0.35				
	22	3.3.3	0.3	0.6	0.5	✓	18	0.19	0.18	17	0.19	0.15	13	10	9	

ตาราง 32 (ต่อ)

เนื้อหา	ลำดับ ข้อ	รหัส ข้อ	IOC				DIF	ฉ.1 ทดลอง			ฉ.2 ทดลอง			ฉ.3		ฉ.4	ฉ.5
			สอบ	ช.	ญ.	รวม		ข้อ	(1)		ข้อ	(2)		ข้อ	ข้อ	ข้อ	
									p	r		p	r				
3.4 การคูณและการหาร ทศนิยม	23	3.4.1	-0.3	-0.6	-0.5	X	19	0.28	0.44	18	0.47	0.43	14	11			
	24	3.4.2	-0.3	0.0	-0.2	X	20	0.11	0.11	19	0.22	0.29	15				
	25	3.4.3	0.3	0.6	0.5	✓	21	0.18	0.15	20	0.23	0.17	16	12			
	26	3.4.4	0.6	0.6	0.7	✓	22	0.25	0.20	21	0.36	0.19	17	13	10		
	27	3.4.5	0.3	0.6	0.5	✓	23	0.11	0.02								
	28	3.4.6	0.3	0.6	0.5	✓											
3.5 การแทนเศษส่วนด้วย ทศนิยม	29	3.5.1	0.3	-0.3	0.0	X	24	0.28	0.20	22	0.44	0.25					
	30	3.5.2	0.0	-0.3	-0.2	X	25	0.27	0.22	23	0.47	0.29	18	14	11		
	DBF(3)		-0.3	-0.3	-0.5	X											
บทที่ 4																	
4.1 ค่าประมาณจาก การวัด	31	4.1.1	-0.6	0.3	-0.2	X	26	0.31	0.17	24	0.51	0.33	19	15	12		
4.2 การบิดเบือน	32	4.2.1	0.3	0.0	0.2	X	27	0.31	0.05								
	33	4.2.2	0.3	0.0	0.2	X	28	0.31	0.28	25	0.46	0.23	20				
	34	4.2.3	0.3	0.6	0.5	✓	29	0.19	0.15	26	0.23	0.21	21	16	13		
4.3 การประมาณค่า	35	4.3.1	0.3	0.6	0.5	✓	30	0.15	0.16	27	0.24	0.18	22	17			
	36	4.3.2	0.6	0.6	0.7	✓	31	0.21	0.17	28	0.25	0.15	23	18	14		
	37	4.3.3	0.3	0.6	0.5	✓											
DBF(3)		0.3	0.6	0.5	✓												
บทที่ 5																	
5.2 ส่วนของเส้นตรงและ ความยาวส่วนของ เส้นตรง	38	5.2.1	0.3	0.3	0.3	X	32	0.21	0.19	29	0.22	0.16					
	39	5.2.2	0.3	0.0	0.2	X	33	0.31	0.15	30	0.78	0.21	24	19	15		
5.3 ขนาดของมุม	40	5.3.1	0.3	0.3	0.3	X	34	0.27	0.22	31	0.26	0.17	25	20	16		
	41	5.3.2	0.0	0.3	0.2	X	35	0.44	0.56	32	0.24	0.19	26				
	42	5.3.3	0.0	0.0	0.0	X	36	0.36	0.17	33	0.43	0.30	27	21	17		
5.4 มุมฉากและมุมตรง	43	5.4.1	0.0	0.0	0.0	X	37	0.39	0.44	34	0.47	0.19	28	22			
	44	5.4.2	0.3	0.0	0.2	X	38	0.25	0.16	35	0.30	0.22	29	23	18		
	45	5.4.3	0.0	0.0	0.0	X	39	0.29	0.19	36	0.32	0.21	30	24			
DBF(5)		0.3	0.0	0.2	X												
บทที่ 6																	
6.1 เศษส่วน	46	6.1.1	0.0	0.3	0.2	X											
	47	6.1.2	0.0	0.3	0.2	X	40	0.33	0.50	37	0.41	0.23	31	25	19		

ภาคผนวก จ

ตาราง 33 สรุปจำนวนข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 (ค 101) ในแบบทดสอบฉบับต่าง ๆ

เนื้อหา	ฉบับที่มีข้อสอบ 50 ข้อ		ฉบับที่มีข้อสอบ 40 ข้อ		ฉบับที่มีข้อสอบ 30 ข้อ	
	จำนวน	จำนวนข้อที่	จำนวน	จำนวนข้อที่	จำนวน	จำนวนข้อที่
	ข้อสอบ	แสดง DIF	ข้อสอบ	แสดง DIF	ข้อสอบ	แสดง DIF
บทที่ 1						
1.2 ระบบเลขฐานสิบ	1	0	1	0	1	0
1.3 เลขยกกำลัง	2	0	1	0	1	0
รวม	3	0	2	0	2	0
บทที่ 2						
2.2 จำนวนเฉพาะ	1	0	1	0	1	0
2.3 การแยกตัวประกอบ	1	0	1	0	1	0
2.4 ห.ร.ม.	2	1	1	1	1	1
2.5 ค.ร.น.	2	1	2	1	1	1
รวม	6	2	5	2	4	2
บทที่ 3						
3.1 ค่าประจำหลักของทศนิยม	1	0	1	0	1	0
3.2 การเปรียบเทียบทศนิยม	1	0	1	0	1	0
3.3 การบวกและการลบทศนิยม	2	1	1	1	1	1
3.4 การคูณและการหารทศนิยม	4	2	3	2	1	1
3.5 การแทนเศษส่วนด้วยทศนิยม	1	0	1	0	1	0
รวม	9	3	7	3	5	2
บทที่ 4						
4.1 ค่าประมาณจากการวัด	1	0	1	0	1	0
4.2 การปัดเศษ	2	1	1	1	1	1
4.3 การประมาณค่า	2	2	2	2	1	1
รวม	5	3	4	3	3	2
บทที่ 5						
5.2 ส่วนของเส้นตรง	1	0	1	0	1	0
5.3 ขนาดของมุม	3	0	2	0	2	0
5.4 มุมฉากและมุมตรง	3	0	3	0	1	0
รวม	7	0	6	0	4	0

ตาราง 33 (ต่อ)

เนื้อหา	ฉบับที่มีข้อสอบ 50 ข้อ		ฉบับที่มีข้อสอบ 40 ข้อ		ฉบับที่มีข้อสอบ 30 ข้อ	
	จำนวน	จำนวนข้อที่	จำนวน	จำนวนข้อที่	จำนวน	จำนวนข้อที่
	ข้อสอบ	แสดง DIF	ข้อสอบ	แสดง DIF	ข้อสอบ	แสดง DIF
บทที่ 6						
6.1 เศษส่วน	1	0	1	0	1	0
6.2 การเปรียบเทียบเศษส่วน	2	0	1	0	1	0
6.3 การบวกและการลบเศษส่วน	2	1	1	1	1	1
6.4 การคูณและการหารเศษส่วน	3	1	3	1	1	1
6.5 โจทย์ปัญหาเศษส่วน	4	3	4	3	3	3
รวม	12	5	10	5	7	5
บทที่ 7						
7.1 ความยาว	1	0	1	0	1	0
7.2 มาตรฐาน	2	0	1	0	1	0
7.3 พื้นที่	3	2	3	2	2	2
7.4 ปริมาตรของทรงสี่เหลี่ยมมุมฉาก	2	1	1	1	1	1
รวม	8	3	6	3	5	3
รวมทั้งฉบับ	50	16	40	16	30	14

ภาคผนวก ฅ

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธี SIBTEST และวิธี DFIT ในเงื่อนไข
ความยาวแบบทดสอบและกลุ่มตัวอย่างขนาดต่าง ๆ

ตาราง 34 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) เมื่อแบบทดสอบประกอบด้วยข้อสอบ
30 ข้อ กลุ่มตัวอย่างขนาด 50 คน

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																		รวม	สรุปผล					
	ผู้เชี่ยวชาญ	5	6	9	10	13	14	21	22	23	24	25	28	29	30	14	P	E								
1	SIBTEST	-																0	0.00	0.00						
	DFIT	*****																-	-	-						
2	SIBTEST	7																1	0.00	6.25						
	DFIT	1	2	3	4	7	8	10	11	12	15	16	17	18	19	20	21	22	24	21	42.86	93.75				
		25	27	29																						
3	SIBTEST	13	27																2	7.14	6.25					
	DFIT	1	2	4	6	7	8	9	10	11	13	14	17	22	23	29	30	16	64.29	43.75						
4	SIBTEST	2	4	12	25	29																5	14.29	18.75		
	DFIT	3	4	7	8	11	12	13	17	19	22	25	26											12	21.43	56.25
5	SIBTEST	7	9																2	7.14	6.25					
	DFIT	1	4	6	7	9	10	15	16	17	19	23	24	28	30	14	42.86	50.00								
6	SIBTEST	6	19																2	7.14	6.25					
	DFIT	1	3	4	8	11	12	15	17	19	21	22	25	26	27	29	15	21.43	75.00							
7	SIBTEST	1	8																2	0.00	12.50					
	DFIT	1	8	10	13	17	19	22	27											8	21.43	31.25				
8	SIBTEST	24																2	7.14	0.00						
	DFIT	*****																-	-	-						
9	SIBTEST	-																0	0.00	0.00						
	DFIT	3	5	6	7	8	14	15	16	17	20	21	22	23	25	26	27	28	30	18	57.14	62.50				
10	SIBTEST	-																0	0.00	0.00						
	DFIT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	28	85.71	100.00				
		19	20	21	23	24	25	26	27	28	30															
11	SIBTEST	14																1	7.14	0.00						
	DFIT	*****																-	-	-						
12	SIBTEST	-																0	0.00	0.00						
	DFIT	1	4	8	10	11	14	22											7	21.43	25.00					
13	SIBTEST	-																0	0.00	0.00						
	DFIT	1	4	10	11	15	17	26											7	7.14	37.50					
14	SIBTEST	29																1	7.14	0.00						
	DFIT	4	11	12	15	16	19	20	24	25	26	30											11	21.43	50.00	
15	SIBTEST	1	10																2	7.14	6.25					
	DFIT	1	2	5	8	11	12	14	19	20	22	23	26	30											13	35.71
16	SIBTEST	13	16	27																3	7.14	12.50				
	DFIT	3	4	6	7	8	13	14	15	16	18	19	21	22	23	24	25	28	29	19	78.57	50.00				
		30																								
17	SIBTEST	11																1	0.00	6.25						
	DFIT	1	17																2	0.00	12.50					

ตาราง 34 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																รวม	สรุปผล					
	ผู้เชี่ยวชาญ	5	6	9	10	13	14	21	22	23	24	25	28	29	30	14	P	E						
18	SIBTEST	-																	0	0.00	0.00			
	DFIT	8	11	12	13	16											5	7.14	25.00					
19	SIBTEST	13																	1	7.14	0.00			
	DFIT	10	12	14	19	22	26	27	28								8	28.57	25.00					
20	SIBTEST	14	21	25														3	21.43	0.00				
	DFIT	1	2	4	7	11	12	16	17	18	25	26	29	30					13	21.43	62.50			
21	SIBTEST	-																	0	0.00	0.00			
	DFIT	22	29															2	14.29	0.00				
22	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	3	4	5	7	8	9	10	11	12	13	14	15	17	18	19	22	22	64.29	81.25		
		24	25	27	29																			
23	SIBTEST	2	12															2	0.00	12.50				
	DFIT	1	2	8	10	11	12	19	22	23	24	26							11	28.57	43.75			
24	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	28	92.86	93.75		
		19	21	22	23	24	25	27	28	29	30													
25	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	4	11	12	14	16	17	18	20											9	7.14	50.00	
26	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	3	4	5	6	7	8	9											9	21.43	37.50	
27	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	4	5	8	10	11	12	13	14	19	21	23	24	27	29				16	57.14	50.00	
28	SIBTEST	-																	0	0.00	0.00			
	DFIT	4	12															2	0.00	12.50				
29	SIBTEST	30																	1	7.14	0.00			
	DFIT	2	8	10														3	7.14	12.50				
30	SIBTEST	8																	1	0.00	6.25			
	DFIT	1	2	4	6	7	8	9	10	12	13	15	17	18	21					14	35.71	56.25		
31	SIBTEST	7																	1	0.00	6.25			
	DFIT	5	8	10	12	19	21	22											7	21.43	25.00			
32	SIBTEST	9	19															2	7.14	6.25				
	DFIT	2	4	5	6	7	10	11	12	16	17	19	20	30					13	28.57	56.25			
33	SIBTEST	2																	1	0.00	6.25			
	DFIT	1	2	3	4	7	8	9	10	11	12	13	14	15	16	17	19	20	21	24	64.29	93.75		
		24	25	26	27	28	30																	
34	SIBTEST	5	8	16	25														4	14.29	12.50			
	DFIT	1	2	4	5	6	7	8	11	13	18	19	21	22	25	27	30				16	42.86	62.50	
35	SIBTEST	15																	1	0.00	6.25			
	DFIT	1	2	3	4	8	9	10	11	12	13	14	15	16	17	18	19	20	21	26	85.71	87.50		
		22	23	24	25	27	28	29	30															
36	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	3	4	5	8	9	11	12	13	15	17	19	20	21	22	23	25	27	19	57.14	68.75		
		30																						
37	SIBTEST	30																	1	7.14	0.00			
	DFIT	1	4	6	8	10	12	19	24	28	30								10	35.71	31.25			
38	SIBTEST	16																	1	0.00	6.25			
	DFIT	*****																	-	-				

ตาราง 35 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) เมื่อแบบทดสอบประกอบด้วยข้อสอบ
30 ข้อ กลุ่มตัวอย่างขนาด 100 คน

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)															รวม	สรุปผล	
		ผู้เข้าข่าย	5	6	9	10	13	14	21	22	23	24	25	28	29	30		P	E
1	SIBTEST	-															0	0.00	0.00
	DFIT	1	2	3	8	22											5	7.14	25.00
2	SIBTEST	16															1	0.00	6.25
	DFIT	1	2	8	11	12	13	17	18	19	21	26	30				13	28.57	56.25
3	SIBTEST	-															0	0.00	0.00
	DFIT	3	5	10	11	12	13	14	15	16	18	19	20	21	22	23	24	25	26
		27	28	30													21	78.57	62.50
4	SIBTEST	-															0	0.00	0.00
	DFIT	1	4	8	10	12	17	24									7	14.29	31.25
5	SIBTEST	16	24														2	7.14	6.25
	DFIT	1	2	3	4	7	8	9	10	11	12	13	15	16	18	22	26	30	
																	17	35.71	75.00
6	SIBTEST	25															1	0.00	6.25
	DFIT	4															1	0.00	6.25
7	SIBTEST	-															0	0.00	0.00
	DFIT	1	2	4	8	9	10	11	12	13	14	15	17	18	19	20	21	22	25
		26	27														20	50.00	81.25
8	SIBTEST	29															1	7.14	0.00
	DFIT	1	3	4	7	8	10	11	12	13	14	15	17	18	19	20	21	22	23
		24	25	28	29												22	71.43	75.00
9	SIBTEST	-															0	0.00	0.00
	DFIT	1	8	9	10	11	12	17	20								8	14.29	37.50
10	SIBTEST	-															0	0.00	0.00
	DFIT	1	2	4	5	8	11	21									7	14.29	31.25
11	SIBTEST	-															0	0.00	0.00
	DFIT	4	8	19	26												4	0.00	25.00
12	SIBTEST	16	21	29													3	14.29	6.25
	DFIT	10	12	13	28												4	21.43	6.25
13	SIBTEST	-															0	0.00	0.00
	DFIT	1	2	3	4	5	8	9	10	11	12	13	15	16	17	19	20	21	23
		25	26	29													21	57.14	81.25
14	SIBTEST	12															1	0.00	6.25
	DFIT	1	2	3	4	5	6	7	8	9	10	11	12	13			13	35.71	50.00
15	SIBTEST	12															1	0.00	6.25
	DFIT	*****															-	-	-
16	SIBTEST	3	9	13	30												4	21.43	6.25
	DFIT	1	2	4	9	10	12	13	15	16	17	19	27				12	21.43	56.25
17	SIBTEST	-															0	0.00	0.00
	DFIT	*****															-	-	-
18	SIBTEST	2															1	0.00	6.25
	DFIT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	17	18	19	20
		21	22	23	25	26	27	28	29	30							27	92.86	93.75
19	SIBTEST	7	13														2	7.14	6.25
	DFIT	1	4	8	12	13	14	15	17	19	21	25					11	28.57	43.75
20	SIBTEST	2															1	0.00	6.25
	DFIT	1	11	16													3	0.00	18.75
21	SIBTEST	13															1	7.14	0.00
	DFIT	1	5	7	12	13	17	19	22	23	29						10	35.71	31.25

ตาราง 35 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																												รวม	สรุปผล	
	ผู้เชี่ยวชาญ	5	6	9	10	13	14	21	22	23	24	25	28	29	30	14	P	E														
22	SIBTEST	2	12	30												3	7.14	12.50														
	DFIT	1	2	3	4	5	7	8	10	11	12	13	14	17	18	19	20	21	22	23	71.43	81.25										
		24	26	28	29	30																										
23	SIBTEST	-														0	0.00	0.00														
	DFIT	1	2	4	6	8	9	10	11	12	13	16	17	19	21	22	25	26	30	18	57.14	62.50										
24	SIBTEST	-														0	0.00	0.00														
	DFIT	1	11	12	13	17	26	27								7	7.14	37.50														
25	SIBTEST	8	16													2	0.00	12.50														
	DFIT	1	3	4	8	10	11	12	13	14	20	28				11	28.57	43.75														
26	SIBTEST	-														0	0.00	0.00														
	DFIT	2	4	10	13											4	14.29	12.50														
27	SIBTEST	-														0	0.00	0.00														
	DFIT	2	4	7	10	11	12	16	17	19	24	28				11	21.43	50.00														
28	SIBTEST	16	25													2	7.14	6.25														
	DFIT	1	4	7	8	9	11	15	17	18	19	21	26	27	29	14	21.43	68.75														
29	SIBTEST	8	21													2	7.14	6.25														
	DFIT	1	2	3	4	5	7	9	10	12	13	14	15	17	20	21	22	23	24	24	92.86	68.75										
		25	26	27	28	29	30																									
30	SIBTEST	24														1	7.14	0.00														
	DFIT	1	2	5	8	9	10	12	17	19	21	22	25	26	27	14	42.86	50.00														
31	SIBTEST	5	15													2	7.14	6.25														
	DFIT	1	8	10	12	19										5	7.14	25.00														
32	SIBTEST	24	29													2	14.29	0.00														
	DFIT	1	2	3	4	5	8	9	12	13	14	15	26	27		13	14.29	68.75														
33	SIBTEST	-														0	0.00	0.00														
	DFIT	1	2	4	8	10	11	12	13	14	15	16	17	18	19	21	22	24	26	21	57.14	81.25										
		27	28	30																												
34	SIBTEST	25														1	7.14	0.00														
	DFIT	1	2	4	8	11	14	18	20	22						9	7.14	50.00														
35	SIBTEST	27														1	0.00	6.25														
	DFIT	1	2	4	5	6	7	9	10	11	12	14	16	17	19	21	26	27	17	42.86	68.75											
36	SIBTEST	15														1	0.00	6.25														
	DFIT	1	2	4	10	11	12	13	17	18	19	20	24	27	28	14	28.57	62.50														
37	SIBTEST	3	27	30												3	7.14	12.50														
	DFIT	1	3	4	6	8	9	10	11	12	14	16	22	24	26	27	28	16	42.86	62.50												
38	SIBTEST	-														0	0.00	0.00														
	DFIT	1	2	3	4	6	7	8	10	11	12	15	16	17	19	20	21	22	24	23	64.29	87.50										
		25	26	28	29	30																										
39	SIBTEST	13														1	7.14	0.00														
	DFIT	1	2	3	4	5	6	8	9	10	11	12	13	14	15	16	17	18	19	29	100.00	93.75										
		20	21	22	23	24	25	26	27	28	29	30																				
40	SIBTEST	9														1	7.14	0.00														
	DFIT	1	2	4	5	8	9	10	11	12	13	17	19	20	22	24	26	27	29	18	50.00	68.75										
41	SIBTEST	-														0	0.00	0.00														
	DFIT	11	12	13	19											4	7.14	18.75														
42	SIBTEST	3														1	0.00	6.25														
	DFIT	7	11	17	19	26	30									6	7.14	31.25														
43	SIBTEST	3	16													2	0.00	12.50														
	DFIT	2	3	5	6	7	9	11	15	18	19	20	22	23	24	26	29	16	50.00	56.25												

ตาราง 36 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)															รวม	สรุปผล						
	ผู้เชี่ยวชาญ	5	6	9	10	13	14	21	22	23	24	25	28	29	30	14	P	E						
22	SIBTEST	2																1	0.00	6.25				
	DFIT	1	4	12	13	14	21	25	26	27	28						10	35.71	31.25					
23	SIBTEST	16	29																2	0.00	6.25			
	DFIT	1	2	3	4	5	6	7	8	10	11	12	13	14	15	16	18	19	20	27	85.71	93.75		
		21	22	232	25	26	27	29	30															
24	SIBTEST	12	16																2	0.00	12.50			
	DFIT	1	2	4	11	17	19	28											7	7.14	37.50			
25	SIBTEST	2	8	9	16																4	7.14	18.75	
	DFIT	1	2	3	4	10	11	20	27	29								9	14.28	43.75				
26	SIBTEST	16	25	29																3	7.14	6.25		
	DFIT	2	6	8	10	12	19	23											7	14.28	31.25			
27	SIBTEST	15	16																2	0.00	12.50			
	DFIT	1	4	8	11	12	13	15	25	26	27						10	14.28	50.00					
28	SIBTEST	16	30																2	7.14	6.25			
	DFIT	1	4	6	8	11	12	17	19	25								9	14.28	43.75				
29	SIBTEST	-																0	0.00	0.00				
	DFIT	1	2	4	11	13	212	26	27								8	14.28	42.86					
30	SIBTEST	9																1	7.14	0.00				
	DFIT	1	2	4	8	10	11	12	13	15	17	19	20	22	26	14	14.28	75.00						
31	SIBTEST	-																0	0.00	0.00				
	DFIT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	18	19	20	26	100.00	75.00		
		21	22	23	24	25	28	29	30															
32	SIBTEST	16																1	0.00	6.25				
	DFIT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	17	18	19	29	100.00	93.75		
		20	21	22	23	24	25	26	27	28	29	30												
33	SIBTEST	2																1	0.00	6.25				
	DFIT	8	14	17	19	21											5	14.28	18.75					
34	SIBTEST	-																0	0.00	0.00				
	DFIT	4	8	10	11	12	19											6	7.14	31.25				
35	SIBTEST	-																0	0.00	0.00				
	DFIT	1	4	12	17	19	26											6	0.00	37.50				
36	SIBTEST	9	16																2	0.00	6.25			
	DFIT	1	2	3	4	6	7	8	9	10	11	12	13	14	15	17	19	20	21	25	85.71	81.25		
		22	23	24	26	28	29	30																
37	SIBTEST	9	16																2	0.00	6.25			
	DFIT	1	2	11											3	0.00	18.75							
38	SIBTEST	25																1	7.14	0.00				
	DFIT	1	2	4	6	7	8	10	11	12	13	14	17	19	20	21	24	26	28	18	50.00	68.75		
39	SIBTEST	24																1	7.14	0.00				
	DFIT	1	2	4	19											4	0.00	25.00						
40	SIBTEST	25																1	7.14	0.00				
	DFIT	1	4	8	12	17	19	26											7	0.00	43.75			
41	SIBTEST	8																1	0.00	6.25				
	DFIT	1	2	10	11	12											5	7.14	25.00					
42	SIBTEST	22																1	7.14	0.00				
	DFIT	1	2	4	5	8	10	11	12	14	19	22						11	28.57	43.75				
43	SIBTEST	8																1	0.00	6.25				
	DFIT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	28	85.71	100.00		
		19	20	21	22	23	24	25	26	27	28													

ตาราง 37 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																รวม	สรุปผล		
		ผู้เชี่ยวชาญ	5	6	9	10	13	14	21	22	23	24	25	28	29	30	P		E		
23	SIBTEST	9	13	15	16	25											5	21.43	12.50		
	DFIT	1	2	4	8	10	11	12	13	14	16	17	19	20	22	27	29	16	35.71	68.75	
24	SIBTEST	16															1	0.00	6.25		
	DFIT	1	2	4	8	9	10	11	12	14	20	22	28				12	28.57	50.00		
25	SIBTEST	8	16														2	0.00	12.50		
	DFIT	1	4	19													3	0.00	18.75		
26	SIBTEST	15	16														2	0.00	12.50		
	DFIT	1	4	6	10	12											5	7.14	25.00		
27	SIBTEST	9	16														2	7.14	6.25		
	DFIT	1	2	4	5	7	8	10	11	12	13	14	16	17	19	20	21	22	26	20	57.14
28	SIBTEST	16	24														2	7.14	6.25		
	DFIT	1	2	4	5	7	8	10	11	12	13	14	17	18	19	22	23	24	26	18	50.00
29	SIBTEST	16	25														2	7.14	6.25		
	DFIT	1	2	4	10	12	14	19	20	21	22	27	29				12	35.71	43.75		
30	SIBTEST	13	16														2	7.14	6.25		
	DFIT	1	2	4	8	9	11	12	19	21	24	26	27	28			13	28.57	56.25		
31	SIBTEST	1	2	9	16												4	7.14	18.75		
	DFIT	1	2	4	5	7	8	9	10	11	12	13	14	16	17	20	26	27	17	35.71	75.00
32	SIBTEST	16															1	0.00	6.25		
	DFIT	1	2	3	4	6	8	10	11	12	19	20	25	28			13	28.57	56.25		
33	SIBTEST	9	13	16	29												4	21.43	6.25		
	DFIT	1	2	4	10	11	12	14	17	19	20	21	22	23	26	28	15	42.86	56.25		
34	SIBTEST	2	7	13	16												4	7.14	18.75		
	DFIT	1	4	8	11	17	19	20	25	26	29	30					11	21.43	50.00		
35	SIBTEST	9	16														2	7.14	6.25		
	DFIT	1	2	3	4	6	8	10	11	12	14	16	17	20	21	26	30	16	35.74	68.75	
36	SIBTEST	2	24														2	7.14	6.25		
	DFIT	1	2	4	8	10	12	13	14	17	19	20	27				12	21.43	56.25		
37	SIBTEST	7	15	16													4	0.00	18.75		
	DFIT	1	2	4	5	6	7	10	12	13	15	16	17	26	28		14	35.71	56.25		
38	SIBTEST	7	9	15	24												4	14.29	12.50		
	DFIT	1	2	4	8	10	12	14	16	17	19	21	28				12	28.57	50.00		
39	SIBTEST	6	8	9	15												4	14.29	12.50		
	DFIT	4	6	8	10	11	12	13	15	17	19	21	22	24	26	28	29	16	57.14	56.25	
40	SIBTEST	16															1	0.00	6.25		
	DFIT	1	2	4	5	7	8	9	10	11	12	13	14	15	17	21	22	26	27	20	64.29
41	SIBTEST	5	9	16	19	25											5	21.43	12.50		
	DFIT	*****															-	-	-		
42	SIBTEST	-															0	0.00	0.00		
	DFIT	1	2	4	8	9	12	13	15	17	18	19	20	21	22	23	24	26	28	18	50.00
43	SIBTEST	6	7	9	14	15	25										6	28.57	12.50		
	DFIT	1	2	3	4	8	10	11	15	19	20	24	28	30			13	28.57	56.25		
44	SIBTEST	13	16														2	7.14	6.25		
	DFIT	1	2	4	10	12	14	22	27	29							9	28.57	31.25		
45	SIBTEST	2	7	8	16												4	0.00	25.00		
	DFIT	1	2	4	8	11	12	13	14	17	18	19	20	21	26	27	29	16	28.57	75.00	
46	SIBTEST	2	9	30													3	14.29	6.25		
	DFIT	1	2	4	6	8	10	11	13	14	15	17	18	24	26	27	28	16	42.86	62.50	

ตาราง 37 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)														รวม	สรุปผล				
	ผู้เชี่ยวชาญ	5	6	9	10	13	14	21	22	23	24	25	28	29	30	14	P	E			
47	SIBTEST	15	16	24												3	7.14	12.50			
	DFIT	1	2	4	8	10	11	12	13	14	15	16	17	18	19	21	26	28	17	35.71	75.00
48	SIBTEST	15	16													2	0.00	12.50			
	DFIT	1	2	4	7	8	10	11	12	14	16	17	19	20	21	24	26		16	28.57	75.00
49	SIBTEST	8	15	16												3	0.00	18.75			
	DFIT	1	2	4	5	7	8	10	11	12	13	14	15	16	17	18	19	21	22	24	64.29
50	SIBTEST	15	16													2	0.00	12.50			
	DFIT	1	2	3	4	5	10	11	12	17	23	24	26	27		13	28.57	56.25			
		$\bar{X}$								SIBTEST								7.57	10.00		
										DFIT								36.30	59.44		
		S.D.								SIBTEST								7.56	5.50		
										DFIT								14.93	16.09		

ตาราง 38 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) เมื่อแบบทดสอบประกอบด้วยข้อสอบ
30 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน

กลุ่มที่	วิธีตรวจสอบ		ข้อสอบที่แสดงการหาหน้าที่เบี่ยงเบน (DIF)														รวม		สรุปผล		
	ผู้เชี่ยวชาญ	5	6	9	10	13	14	21	22	23	24	25	28	29	30	14	P	E			
1	SIBTEST	16														1	0.00	6.25			
	DFIT	1	2	4	5	8	10	11	12	13	14	15	17	19	28	14	35.71	56.25			
2	SIBTEST	8	16													2	0.00	12.50			
	DFIT	1	2	4	8	12	17	21	22	24	26					10	21.43	43.75			
3	SIBTEST	2	7	8	9	13	16	25								7	21.43	25.00			
	DFIT	2	3	4	5	7	8	10	11	13	14	15	17	18	19	20	22	23	24	24	78.57
4	SIBTEST	7	8	13	16											4	7.14	18.75			
	DFIT	1	2	4	8	9	10	11	12	14	15	17	21	23	26	14	35.71	56.25			
5	SIBTEST	9	13	16												3	14.29	6.25			
	DFIT	1	2	4	8	10	11	12	13	14	17	19	23	29		13	35.71	50.00			
6	SIBTEST	7	9	13	16											4	14.29	12.50			
	DFIT	1	2	4	8	10	11	12	13	14	15	17	19	28		13	28.57	56.25			
7	SIBTEST	9	13	16												3	14.29	6.25			
	DFIT	1	2	4	6	10	11	12	14	17	23	26	28			12	35.71	50.00			
8	SIBTEST	8	13	16	25											4	14.29	12.50			
	DFIT	1	2	4	12	14	19	26	27	28						9	14.29	43.75			
9	SIBTEST	9	13	16												3	14.29	6.25			
	DFIT	1	2	4	8	10	11	12	14	17	19	28				11	21.43	50.00			
10	SIBTEST	8	9	13	16											4	14.29	12.50			
	DFIT	1	2	4	8	10	11	12	13	14	17	19	23	26		13	28.57	56.25			
11	SIBTEST	2	8	9	13	16	25									6	21.43	18.75			
	DFIT	1	2	4	8	10	11	12	13	14	17	19	21	28		13	35.71	50.00			
12	SIBTEST	7	8	9	13	16										5	14.29	18.75			
	DFIT	1	2	4	8	10	12	14	17	21	23	27				11	28.57	43.75			
13	SIBTEST	16														1	0.00	6.25			
	DFIT	1	2	4	10	11	12	14	17	19	23	28				11	28.57	43.75			
14	SIBTEST	7	9	16												3	7.14	12.50			
	DFIT	1	2	4	10	12	14	21	26	28						9	28.57	31.25			
15	SIBTEST	13	16													2	7.14	6.25			
	DFIT	1	2	4	5	8	12	26								7	7.14	37.50			
16	SIBTEST	8	9	16												3	7.14	12.50			
	DFIT	1	2	4	8	10	11	12	13	14	17	18	19	20	22	24	26	16	35.71	68.75	
17	SIBTEST	2	8	13	16											4	7.14	18.75			
	DFIT	1	2	4	10	11	12	13	14	15	17	19	22	23	26	27	28	16	42.86	62.50	
18	SIBTEST	7	8	9	13	15	16	25								7	21.43	25.00			
	DFIT	1	2	4	8	10	11	12	14	19	20	21	22	23	26	14	35.71	56.25			
19	SIBTEST	9	16													2	7.14	6.25			
	DFIT	1	2	4	8	10	11	12	13	14	15	17	27	28		13	28.57	56.25			
20	SIBTEST	8	9	13	16	25										5	21.43	12.50			
	DFIT	1	2	4	10	11	12	17	21	22	26					10	21.42	43.75			
21	SIBTEST	8	9	13	16											4	14.29	12.50			
	DFIT	1	2	4	8	10	11	12	13	15	17	19	22	25	28	14	28.57	62.50			
22	SIBTEST	2	7	8	9	13	15	16	25							8	21.43	31.25			
	DFIT	1	2	4	8	10	11	12	13	14	17	19	21	24	26	28	15	42.86	56.25		

ตาราง 38 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)															รวม	สรุปผล			
		ผู้เชี่ยวชาญ	5	6	9	10	13	14	21	22	23	24	25	28	29	30		14	P	E	
23	SIBTEST	7	9	13	16												4	14.29	12.50		
	DFIT	1	2	4	7	8	10	11	12	13	14	17	19	20	26	27	28	16	28.57	75.00	
24	SIB11TEST	8	9	15	16												4	7.14	18.75		
	DFIT	1	4	8	12	21	22	24	26								8	21.43	31.25		
25	SIBTEST	7	8	9	15	16											5	7.14	25.00		
	DFIT	1	2	4	8	10	11	12	13	14	16	17	19	21	22	24	26	273	18	50.00	68.75
26	SIBTEST	2	8	9	13	16											5	14.29	18.75		
	DFIT	1	2	3	4	7	8	9	10	11	12	13	14	17	19	22	27	30	17	42.86	68.75
27	SIBTEST	9	16	24	25												4	21.43	6.25		
	DFIT	1	2	4	8	10	11	12	13	14	17	19	23	28			13	35.71	50.00		
28	SIBTEST	13	16														2	7.14	6.25		
	DFIT	1	2	4	8	10	11	12	13	14	15	16	17	19	23	27	28	16	35.71	68.75	
29	SIBTEST	8	9	13	16												4	14.29	12.50		
	DFIT	1	2	4	10	11	12	13	14	15	17	19					11	21.43	50.00		
30	SIBTEST	13	16														2	7.14	6.25		
	DFIT	1	2	3	4	5	7	8	9	10	11	12	13	14	15	16	17	18	19	25	78.57
31	SIBTEST	8	13	16													3	7.14	12.50		
	DFIT	1	2	3	4	6	7	8	9	10	11	12	13	14	15	16	17	18	19	26	71.43
32	SIBTEST	2	9	13	16												4	7.14	18.75		
	DFIT	1	2	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	29	100.00
33	SIBTEST	2	9	13	16												4	14.29	12.50		
	DFIT	2	3	4	5	7	8	9	10	11	12	13	14	15	16	17	19	20	21	26	85.71
34	SIBTEST	13	16														2	7.14	6.25		
	DFIT	2	3	4	6	8	9	10	11	12	13	14	15	16	17	18	19	21	23	25	85.71
35	SIBTEST	7	8	9	13	16											5	14.29	18.75		
	DFIT	2	3	4	6	8	9	10	11	12	13	14	15	16	17	18	19	20	22	26	85.71
36	SIBTEST	9	13	16													3	14.29	6.25		
	DFIT	1	2	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	28	92.85
37	SIBTEST	13	16														2	7.14	6.25		
	DFIT	1	2	3	4	7	8	9	10	11	12	13	14	15	16	17	18	19	20	28	85.71
38	SIBTEST	8	13	16	25												4	14.29	12.50		
	DFIT	1	2	3	4	5	6	7	8	10	11	12	13	14	15	16	17	18	19	26	78.57
39	SIBTEST	8	9	16													3	7.14	12.50		
	DFIT	2	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	28	100.00
40	SIBTEST	2	7	9	13	16											5	14.29	18.75		
	DFIT	1	2	4	10	11	12	13	142	24	26						11	35.71	37.50		
41	SIBTEST	8	13	16	25												4	14.29	12.50		
	DFIT	1	2	4	8	9	10	11	12	14	15	19	26	27			13	21.43	62.50		

ตาราง 38 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)														รวม	สรุปผล					
		ผู้เชี่ยวชาญ	5	6	9	10	13	14	21	22	23	24	25	28	29		30	14	P	E		
42	SIBTEST	8	16														2	0.00	12.50			
	DFIT	1	2	4	5	8	9	10	11	12	13	14	16	19	20	22	23	26	27	18	50.00	68.75
43	SIBTEST	9	13	15	16												4	14.29	12.50			
	DFIT	1	2	4	10	11	12	14	17	22	23	26	28				12	35.71	43.75			
44	SIBTEST	2	8	9	13	16											5	14.29	18.75			
	DFIT	1	2	4	8	10	11	12	14	17	19	23	26				12	21.43	56.25			
45	SIBTEST	8	9	13	16												4	14.29	12.50			
	DFIT	1	2	4	10	11	12	14	15	21	22	23	24	26	30		14	50.00	43.75			
46	SIBTEST	8	9	16													3	7.14	12.50			
	DFIT	1	2	4	8	10	11	12	13	14	17	19	20	21	26	27	15	28.57	68.75			
47	SIBTEST	2	7	13	16												4	7.14	18.75			
	DFIT	1	2	3	4	8	10	11	12	13	14	15	17	19	21	23	27	16	35.71	68.75		
48	SIBTEST	8	9	13	16	25											5	21.43	12.50			
	DFIT	1	2	4	10	12	14	19	26								8	7.14	43.75			
49	SIBTEST	9	16														2	14.28	6.25			
	DFIT	1	2	4	8	10	11	12	13	14	15	17	19	20	27		14	21.43	68.75			
50	SIBTEST	8	13	16	24												4	14.29	12.50			
	DFIT	1	2	4	8	9	10	12	13	14	19	21	22	23	26		14	35.71	56.25			
		$\bar{X}$									SIBTEST					11.57		13.13				
											DFIT					42.86		62.00				
		S.D.									SIBTEST					5.93		5.96				
											DFIT					25.04		18.46				

ตาราง 39 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																		รวม	สรุปผล	
	ผู้เชี่ยวชาญ	5	7	10	12	13	16	17	18	27	30	31	32	33	37	39	40	16	P	E		
17	SIBTEST	16	27	37														3	18.75	0.00		
	DFIT	1	2	4	5	6	9	11	13	15	16	17	18	20	21	22	23	24	25	25	56.25	70.83
		29	30	32	33	34	38	39														
18	SIBTEST	34																1	0.00	4.17		
	DFIT	1	4	9	11	15	16	17	22	25	26	27	30	31	36	38		15	35.71	41.67		
19	SIBTEST	-																0	0.00	0.00		
	DFIT	1	3	6	7	8	9	10	11	13	14	15	16	17	18	20	21	22	24	28	56.25	79.17
		25	27	28	29	32	33	34	35	36	38											
20	SIBTEST	-																0	0.00	0.00		
	DFIT	8	9	13	15	17	18	25	26	28	29	32	33	34	39			14	37.50	37.50		
21	SIBTEST	29																1	0.00	4.17		
	DFIT	1	2	3	4	6	7	9	10	11	13	14	15	16	17	18	19	21	24	28	62.50	75.00
		26	29	30	31	32	34	35	36	38	39											
22	SIBTEST	10	20															2	6.25	4.17		
	DFIT	1	2	4	6	8	9	10	11	13	14	16	18	20	21	22	24	25	26	25	37.50	83.33
		28	29	32	35	36	38	39														
23	SIBTEST	17																1	6.25	0.00		
	DFIT	2	3	4	8	9	12	13	16	20	21	22	24	25	27	28	29	30	32	24	50.00	70.83
		34	35	36	37	38	39															
24	SIBTEST	6	40															2	6.25	4.17		
	DFIT	1	2	3	4	5	6	11	12	13	14	15	17	19	20	21	22	25	26	29	56.25	83.33
		27	28	29	30	31	32	34	35	36	38	39										
25	SIBTEST	18	33	39														3	18.75	0.00		
	DFIT	1	2	3	9	13	20	22	36									8	6.25	29.17		
26	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	7	8	9	11	13	16	17	18	21	22	24	25	27	28	23	43.75	62.50
		29	33	35	36	40																
27	SIBTEST	-																0	0.00	0.00		
	DFIT	1	3	4	6	8	9	10	13	14	15	16	21	27	29	30	34	35	36	21	43.75	58.33
		37	38	39																		
28	SIBTEST	33																1	6.25	0.00		
	DFIT	1	2	3	4	6	7	9	11	14	15	16	18	21	22	23	24	26	28	27	50.00	79.17
		29	30	32	33	34	35	36	37	39												
29	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	11	13	14	17	18	27	29	32	33	34	35	37	16	43.75	41.67		
30	SIBTEST	2																1	0.00	4.17		
	DFIT	1	2	4	6	9	11	13	15	16	17	18	20	21	22	24	25	26	30	26	62.50	66.67
		32	33	34	35	37	38	39	40													
31	SIBTEST	19	20	33														3	6.25	8.33		
	DFIT	1	2	3	4	7	8	9	11	12	15	17	18	19	22	29	31	33	36	20	43.75	54.17
		38	39																			
32	SIBTEST	34																1	0.00	4.17		
	DFIT	2	4	6	9	11	14	15	16	22	25	29	32	33	34	38	39	16	25.00	50.00		
33	SIBTEST	-																0	0.00	0.00		
	DFIT	2	4	5	9	11	14	16	18	19	20	21	22	23	28	30	34	36	38	19	25.00	62.50
		39																				
34	SIBTEST	34																1	0.00	4.17		
	DFIT	1	2	5	6	7	10	11	12	13	14	15	18	22	23	26	27	28	29	23	56.25	54.17
		32	33	35	37	38																

ตาราง 39 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																		รวม	สรุปผล	
	ผู้เชี่ยวชาญ	5	7	10	12	13	16	17	18	27	30	31	32	33	37	39	40	16	P	E		
35	SIBTEST	-																0	0.00	0.00		
	DFIT	9	11	15	29	36	38											6	0.00	25.00		
36	SIBTEST	2	3	4	16	21	24	25	28	30	35							10	12.50	33.33		
	DFIT	1	2	3	4	9	10	11	12	13	14	15	16	21	25	27	28	29	30			
		32	33	38																		
37	SIBTEST	34	38															2	0.00	8.33		
	DFIT	1	3	4	8	9	13	14	15	19	21	24	26	29				13	6.25	50.00		
38	SIBTEST	20																1	0.00	4.17		
	DFIT	1	2	4	6	9	10	11	12	13	15	16	18	22	23	24	26	27	28			
		29	30	32	33	35	36	38	39													
39	SIBTEST	1	10															2	6.25	4.17		
	DFIT	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	18	19			
		20	22	23	24	25	26	27	28	29	30	35	37	38	39	40						
40	SIBTEST	5	16	39														3	18.75	0.00		
	DFIT	2	3	4	5	6	7	8	9	11	12	13	14	15	16	18	19	20	21			
		22	23	25	26	28	29	30	31	32	33	34	35	36	38	40						
41	SIBTEST	9	16															2	6.25	0.00		
	DFIT	1	2	3	6	9	11	13	15	16	19	20	24	25	28	36	37	39	40			
42	SIBTEST	3																1	0.00	4.17		
	DFIT	1	2	4	5	6	17	18	25	26	29	33	34	38				13	25.00	37.50		
43	SIBTEST	5	10	22	40													4	18.75	4.17		
	DFIT	1	4	6	9	11	15	32	35									8	6.25	29.17		
44	SIBTEST	26	30															2	6.25	4.17		
	DFIT	1	2	3	12	13	14	29	32	35								9	18.75	29.17		
45	SIBTEST	30																1	6.25	0.00		
	DFIT	1	4	6	7	8	9	13	15	16	17	18	19	20	21	24	29	30	35			
		37	38																			
46	SIBTEST	-																0	0.00	0.00		
	DFIT	1	3	4	6	7	8	9	11	14	15	16	17	18	19	21	22	23	24			
		25	28	29	30	32	35	36														
47	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	3	4	6	8	9	11	14	15	16	17	19	20	21	25	26	27			
		28	30	32	33	35	36	38	39													
48	SIBTEST	33																1	6.25	0.00		
	DFIT	1	2	4	6	7	10	11	15	16	20	21	22	25	26	27	29	33	34			
		36	37	38	39	40																
49	SIBTEST	7																1	6.25	0.00		
	DFIT	1	5	6	8	10	13	14	15	16	21	23	24	25	26	27	29	30	35			
		36																				
50	SIBTEST	14																1	0.00	4.17		
	DFIT	4	6	9	12	13	14	15	16	19	22	24	25	29	30	31	32	33	34			
		39																				
		$\bar{X}$									SIBTEST									4.25		4.17
											DFIT									44.01		61.59
		$S.D.$									SIBTEST									5.71		3.74
											DFIT									18.81		17.69

ตาราง 40 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) เมื่อแบบทดสอบประกอบด้วยข้อสอบ
40 ข้อ กลุ่มตัวอย่างขนาด 100 คน

กลุ่มที่	วิธีตรวจสอบ ผู้เชี่ยวชาญ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																		รวม	สรุปผล	
		5	7	10	12	13	16	17	18	27	30	31	32	33	37	39	40	16	P		E	
1	SIBTEST	9	20	25														3	0.00	12.50		
	DFIT	1	3	4	6	7	11	12	14	15	18	20	21	22	24	28	30	31	33	21	37.50	62.50
		34	35	38																		
2	SIBTEST	8	20															2	0.00	8.33		
	DFIT	1	2	4	5	6	7	8	9	11	13	14	15	16	18	19	21	22	23	26	50.00	75.00
		25	26	27	28	30	31	35	36													
3	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	3	4	6	7	12	13	14	15	16	17	18	21	22	23	25	26	25	6.25	66.67
		29	30	32	33	36	38	40														
4	SIBTEST	16	19	32														3	12.50	4.17		
	DFIT	1	2	4	6	8	9	11	12	13	14	15	16	18	21	22	24	25	26	27	50.00	79.17
		28	29	30	32	34	35	38	39	40												
5	SIBTEST	9	20	32														3	6.25	8.33		
	DFIT	1	2	3	4	6	7	9	11	13	14	15	16	17	19	22	23	26	27	22	43.75	62.50
		32	35	36	40																	
6	SIBTEST	33																1	6.25	0.00		
	DFIT	2	5	6	9	11	13	16	17	19	20	21	23	25	27	28	29	30	32	22	43.75	62.50
		34	35	36	38																	
7	SIBTEST	19																1	0.00	4.17		
	DFIT	1	2	9	15	20	22	27	29	31	35							10	12.50	33.33		
8	SIBTEST	9	33	39														3	12.50	4.17		
	DFIT	*****																-	-	-		
9	SIBTEST	-																0	0.00	0.00		
	DFIT	2	5	6	7	8	9	10	11	12	13	14	15	16	17	19	21	22	25	27	68.75	66.67
		26	27	28	29	30	32	33	35	36												
10	SIBTEST	33																1	6.25	0.00		
	DFIT	1	2	3	4	6	7	8	9	11	12	14	15	18	24	27	28	29	32	23	43.75	66.67
		33	34	36	38	39																
11	SIBTEST	9	14															2	0.00	8.33		
	DFIT	1	2	4	6	9	11	12	13	14	15	17	21	22	23	25	29	35	36	19	25.00	62.50
		37																				
12	SIBTEST	3	39															2	6.25	4.17		
	DFIT	1	2	3	4	5	6	9	10	11	12	13	14	15	16	17	19	20	21	27	50.00	79.17
		22	25	26	27	33	35	36	38	39												
13	SIBTEST	2																1	0.00	4.17		
	DFIT	1	3	4	5	6	7	9	13	14	18	19	21	22	23	24	25	26	28	25	37.50	79.17
		29	32	34	35	36	38	39														
14	SIBTEST	12	15	25	38													4	6.25	12.50		
	DFIT	1	2	3	4	5	6	8	9	10	11	12	14	15	16	19	21	22	25	25	43.75	75.00
		26	28	30	35	37	38	39														
15	SIBTEST	9	16	20	33													4	12.50	8.33		
	DFIT	1	2	4	6	10	13	14	15	16	19	20	21	22	25	27	29	30	31	22	37.50	66.67
		32	34	36	38																	
16	SIBTEST	3	40															2	6.25	4.17		
	DFIT	1	3	4	6	7	10	11	13	14	15	16	17	18	20	21	22	23	24	25	50.00	70.83
		25	30	31	35	36	37	38														
17	SIBTEST	3																1	0.00	4.17		
	DFIT	6	14	15	18	22	25	26	29	30								9	12.50	29.17		

ตาราง 40 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																		รวม	สรุปผล	
	ผู้เชี่ยวชาญ	5	7	10	12	13	16	17	18	27	30	31	32	33	37	39	40	16	P	E		
18	SIBTEST	8	15	16	25													4	6.25	12.50		
	DFIT	1	2	4	5	8	10	11	12	13	15	16	17	18	19	21	22	24	25	24	62.50	58.33
		27	29	32	33	38																
19	SIBTEST	31																1	6.25	0.00		
	DFIT	1	2	4	6	9	11	13	14	15	16	17	18	21	22	25	26	27	28	22	31.25	70.83
		34	35	36	38																	
20	SIBTEST	16	36															2	6.25	4.17		
	DFIT	1	4	8	11	12	14	15	16	17	19	21	22	25	28	31	32	34	35	20	37.50	58.33
		38	39																			
21	SIBTEST	2	15															2	0.00	8.33		
	DFIT	1	3	4	5	11	14	15	16	21	22	24	25	27	29	30	31	35	36	20	37.50	58.33
		38	40																			
22	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	9	10	11	13	14	15	16	20	28	30	33		15	31.25	41.67		
23	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	7	8	9	10	11	12	13	14	15	16	17	18	21	22	27	62.50	70.83
		24	25	27	28	30	35	36	37	38												
24	SIBTEST	6	9	20														3	0.00	12.50		
	DFIT	1	3	4	6	11	12	13	16	18	20	22	24	25	28	34	37	16	31.25	50.00		
25	SIBTEST	17	39															2	6.25	0.00		
	DFIT	2	4	8	9	11	14	17	20	25	29	34	35	38				13	6.25	50.00		
26	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	5	7	9	11	12	14	15	16	22	24	25	27	28		15	25.00	45.83		
27	SIBTEST	16																1	6.25	0.00		
	DFIT	1	3	4	6	8	9	10	13	14	15	16	21	27	29	30	34	35	36	21	43.75	58.33
		37	38	39																		
28	SIBTEST	33																1	6.25	0.00		
	DFIT	1	2	3	4	5	6	7	8	9	10	11	13	16	17	18	19	21	22	26	68.75	62.50
		25	27	28	31	33	36	38	39													
29	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	5	6	9	10	11	13	16	17	21	22	25	27	28	36		17	37.50	45.83
30	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	3	4	5	6	9	10	11	14	17	22	25	27	28	29	33	34	21	37.50	62.50
		35	36	37																		
31	SIBTEST	9																1	0.00	4.17		
	DFIT	1	2	4	5	11	17	25	28	31	35	36	37					12	25.00	33.33		
32	SIBTEST	2																1	0.00	4.17		
	DFIT	1	4	8	13	14	16	17	18	19	20	21	22	23	24	25	2	30	32	34	23	43.75
		35	36	37	38																	
33	SIBTEST	20	33															2	6.25	4.17		
	DFIT	1	2	4	6	7	9	11	13	14	15	16	17	18	20	21	22	25	27	25	50.00	75.00
		28	29	31	35	36	38	39														
34	SIBTEST	34																1	0.00	4.17		
	DFIT	1	2	4	5	6	9	11	12	13	15	17	18	19	20	21	22	23	24	28	50.00	70.83
		25	26	27	29	30	33	35	38													
35	SIBTEST	19	28	2														2	0.00	8.33		
	DFIT	1	2	3	4	6	8	9	11	12	13	14	15	16	17	18	21	22	26	27	62.50	70.83
		27	28	29	30	33	36	37	38	40												

ตาราง 40 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																		รวม	สรุปผล	
	ผู้เชี่ยวชาญ	5	7	10	12	13	16	17	18	27	30	31	32	33	37	39	40	16	P	E		
36	SIBTEST	3	40															2	6.25	4.17		
	DFIT	1	3	4	9	11	14	17	21	22	29	32						11	12.50	37.50		
37	SIBTEST	20																1	0.00	4.17		
	DFIT	1	2	3	4	6	7	8	11	12	13	14	15	17	19	20	21	22	25	24	37.50	75.00
38	SIBTEST	10	16															2	12.50	0.00		
	DFIT	1	2	3	4	5	6	9	10	11	13	14	17	18	19	23	25	26	27	28	56.25	79.17
39	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	3	4	6	11	131	14	16	19	21	22	24	26	27	29	30	35	20	31.25	66.67
40	SIBTEST	3	12	39														3	12.50	4.17		
	DFIT	1	2	4	9	13	17	21	23	24	26	27	36	39				13	25.00	37.50		
41	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	8	9	10	11	12	13	15	17	18	19	21	24	25	31	32	21	43.75	58.33
42	SIBTEST	20	28															2	0.00	8.33		
	DFIT	1	3	4	5	8	9	11	13	141	16	18	192	21	22	23	25	26	27	26	50.00	75.00
43	SIBTEST	8	21															2	0.00	8.33		
	DFIT	1	3	10	11	13	17	21	22	252	27	29	33	35	36	38	40	16	37.50	41.67		
44	SIBTEST	1	2															2	0.00	8.33		
	DFIT	1	4	5	6	8	9	10	13	14	18	19	21	22	25	26	27	29	30	21	37.50	62.50
45	SIBTEST	2	16	19	20	28	40											6	12.50	16.67		
	DFIT	2	4	9	11	13	15	16	21	22	23	24	27	29	32	35		15	25.00	45.83		
46	SIBTEST	20																1	0.00	4.17		
	DFIT	1	2	4	6	8	9	10	11	13	14	15	18	19	21	22	25	26	27	26	43.75	70.83
47	SIBTEST	10																1	6.25	0.00		
	DFIT	1	2	4	5	9	11	15	16	18	19	20	21	22	27	28	35	36	39	18	31.25	54.17
48	SIBTEST	20																1	0.00	4.17		
	DFIT	*****																-	-	-		
49	SIBTEST	10	19	24	36													4	6.25	12.50		
	DFIT	1	2	4	6	11	13	14	16	17	21	22	28	38				13	18.75	41.67		
50	SIBTEST	5																1	6.25	0.00		
	DFIT	1	2	4	6	7	9	10	13	14	15	17	21	22	25	26	28	30	35	19	31.25	58.33
		38																				
		$\bar{X}$										SIBTEST								3.84	4.50	
												DFIT								38.28	60.25	
		$S.D.$										SIBTEST								4.56	4.44	
												DFIT								15.12	13.72	

ตาราง 41 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 200 คน

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																		รวม	สรุปผล	
	ผู้เชี่ยวชาญ	5	7	10	12	13	16	17	18	27	30	31	32	33	37	39	40	16	P	E		
1	SIBTEST	2	9	16	20													4	6.25	12.50		
	DFIT	1	4	6	8	9	11	13	14	15	17	19	21	22	24	25	27	28	30	24	31.25	79.17
		32	34	35	36	39																
2	SIBTEST	2	9	20														3	0.00	12.50		
	DFIT	7	16	17	19	21	22	27	28	34	37	38	39					12	37.50	25.00		
3	SIBTEST	38																1	0.00	4.17		
	DFIT	1	2	4	6	9	11	12	15	16	17	21	25	27	28	30	36	38	40	18	37.50	50.00
4	SIBTEST	9	20	33														3	6.25	8.33		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	26	38	19	37.50	54.17
		39																				
5	SIBTEST	12	20	33														3	12.50	4.17		
	DFIT	1	2	4	5	6	7	11	12	13	14	15	16	17	20	21	23	25	26	25	56.25	66.67
		28	35	36	37	38	39	40														
6	SIBTEST	20	25	32	38													4	6.25	12.50		
	DFIT	1	2	3	4	5	6	9	11	14	15	18	19	20	21	22	25	27	28	24	37.50	79.17
		30	32	33	34	35	38															
7	SIBTEST	24	38															2	0.00	8.33		
	DFIT	1	2	3	4	5	6	9	11	12	13	14	15	16	17	18	21	22	23	27	56.25	75.00
		25	26	27	29	30	32	35	36	38												
8	SIBTEST	16	39															2	12.50	0.00		
	DFIT	1	4	6	13	14	15	17	29	38								9	6.25	33.33		
9	SIBTEST	20	32															2	6.25	4.17		
	DFIT	1	4	6	7	8	11	13	14	15	16	17	18	21	22	24	25	26	28	23	43.75	66.67
		29	32	34	36	39																
10	SIBTEST	10	19	39														3	12.50	4.17		
	DFIT	1	2	3	4	6	8	9	11	12	14	15	16	18	20	21	22	27	28	25	43.75	79.17
		29	30	32	33	34	35	38														
11	SIBTEST	16																1	6.25	0.00		
	DFIT	1	2	4	6	9	11	13	14	15	17	18	26	30	34	35	37	40		17	37.50	45.83
12	SIBTEST	10	20	39														3	12.50	4.17		
	DFIT	1	2	3	4	5	6	9	10	11	12	13	14	15	17	18	19	20	21	33	81.25	87.50
		22	23	24	25	26	30	31	32	33	35	36	37	38	39	40						
13	SIBTEST	9																1	0.00	4.17		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	18	19	20	21	22	25	26	23	43.75	70.83
		29	30	35	37	39																
14	SIBTEST	20																1	0.00	4.17		
	DFIT	1	2	4	5	6	7	9	12	13	14	15	16	17	18	21	22	25	26	25	56.25	66.67
		27	28	30	34	35	36	38														
15	SIBTEST	15	20															2	0.00	8.33		
	DFIT	1	2	4	9	13	14	27	35	39								9	18.75	25.00		
16	SIBTEST	2	39															2	6.25	4.17		
	DFIT	1	2	3	4	6	7	8	9	11	12	13	15	16	17	18	19	20	21	28	62.50	75.00
		22	25	27	28	29	30	35	36	37	39											
17	SIBTEST	7	9	20	23													4	6.25	12.50		
	DFIT	1	2	4	6	7	9	11	13	14	15	17	19	20	23	24	27	29	30	22	43.75	62.50
		33	35	38	39																	

ตาราง 41 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																	รวม	สรุปผล		
	ผู้เชี่ยวชาญ	5	7	10	12	13	16	17	18	27	30	31	32	33	37	39	40	16	P	E		
18	SIBTEST	16	20	32														3	12.50	4.17		
	DFIT	2	4	5	6	9	11	12	13	14	15	16	17	18	20	21	22	25	26	25	56.25	66.67
		28	29	30	31	35	36	37														
19	SIBTEST	9																1	0.00	4.17		
	DFIT	1	2	9	11	14	16	18	22	28	29	30	32	35	36	38		15	25.00	45.83		
20	SIBTEST	2	19															2	0.00	8.33		
	DFIT	1	2	4	6	7	8	11	12	13	14	15	16	17	21	22	23	25	26	28	56.25	79.17
		27	28	29	31	32	34	35	36	38	39											
21	SIBTEST	15																1	0.00	4.17		
	DFIT	1	4	10	11	13	14	15	16	17	21	22	25	29	30	31	35	36		17	37.50	45.83
22	SIBTEST	2	20															2	0.00	8.33		
	DFIT	1	2	4	6	8	9	13	14	15	16	17	18	21	22	25	26	28	32	21	43.75	58.33
		33	36	39																		
23	SIBTEST	9	16															2	6.25	4.17		
	DFIT	1	6	13	15	21	22	28	29	38								9	6.25	33.33		
24	SIBTEST	9	15															2	0.00	8.33		
	DFIT	1	2	4	6	11	12	13	14	15	16	21	22	25	27	29	35	16	25.00	54.17		
25	SIBTEST	9	20	39														3	6.25	8.33		
	DFIT	1	2	3	4	5	6	9	11	12	13	14	15	16	18	21	22	25	26	23	37.50	66.67
		29	30	36	38	39																
26	SIBTEST	20	39															2	6.25	4.17		
	DFIT	1	2	4	6	7	8	9	11	12	13	14	15	16	17	20	21	22	24	26	50.00	75.00
		25	27	28	30	34	36	38	39													
27	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	8	9	12	13	14	15	16	17	21	22	24	25	28	29	30	26	56.25	70.83
		31	32	33	34	35	36	37	38													
28	SIBTEST	32																1	6.25	0.00		
	DFIT	1	4	6	7	9	10	11	14	15	17	20	21	25	26	29	30	33	35	20	37.50	58.33
		37	38																			
29	SIBTEST	9	20	40														3	6.25	8.33		
	DFIT	1	2	3	4	5	6	9	10	11	13	14	15	17	21	22	232	29	30	24	37.50	75.00
		34	35	36	38	40																
30	SIBTEST	10	15															2	6.25	4.17		
	DFIT	2	3	4	5	9	11	13	14	15	17	18	20	21	25	26	27	28	29	21	43.75	58.33
		30	32	36																		
31	SIBTEST	10																1	6.25	0.00		
	DFIT	1	2	4	6	9	11	13	14	15	16	17	18	21	22	23	25	28	29	23	31.25	75.00
		34	35	36	37	38																
32	SIBTEST	-																0	0.00	0.00		
	DFIT	1	3	4	5	6	8	9	11	13	15	18	19	22	23	28	29	32	33	23	37.50	70.83
		34	35	36	38	40																
33	SIBTEST	1	20															2	0.00	8.33		
	DFIT	1	2	3	4	5	6	10	11	13	15	16	17	18	21	23	24	25	26	26	56.25	75.00
		28	29	30	32	34	36	37	38													
34	SIBTEST	33																1	6.25	0.00		
	DFIT	1	2	4	6	9	11	12	13	14	16	17	20	21	22	25	27	28	30	21	37.50	62.50
		35	36	38																		
35	SIBTEST	10	15	16	39													4	18.75	4.17		
	DFIT	1	2	4	5	6	9	10	11	12	13	14	15	16	17	18	19	20	21	25	68.75	58.33
		26	28	32	33	34	37	40														

ตาราง 41 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ ผู้เชี่ยวชาญ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																รวม	สรุปผล			
		5	7	10	12	13	16	17	18	27	30	31	32	33	37	39	40		16	P	E	
36	SIBTEST	20	32															2	6.25	4.17		
	DFIT	1	2	3	4	5	6	8	9	10	11	12	14	16	17	18	21	22	25	25	62.50	62.50
		26	30	31	34	37	38	40														
37	SIBTEST	20																1	0.00	4.17		
	DFIT	1	2	4	6	7	9	11	12	13	15	17	18	20	21	22	24	25	28	23	43.75	66.67
		29	35	38	39	40																
38	SIBTEST	9	28															2	0.00	8.33		
	DFIT	1	2	4	6	11	12	13	14	15	16	21	22	24	25	27	29	35	38	18	25.00	58.33
39	SIBTEST	6	9	10	13													4	12.50	8.33		
	DFIT	1	2	3	4	5	6	9	10	11	13	14	15	17	21	22	23	25	29	25	37.50	79.17
		30	32	34	35	36	38	40														
40	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	7	8	9	11	12	13	14	15	16	17	19	27	32	33	20	43.75	54.16
		35	36																			
41	SIBTEST	38																1	0.00	4.17		
	DFIT	1	2	4	6	9	13	14	15	18	21	25	26	29	30	31	32	36	17	31.25	50.00	
42	SIBTEST	3	32															2	6.25	4.17		
	DFIT	1	2	4	6	8	9	11	13	14	15	17	18	21	22	25	28	30	17	25.00	54.17	
43	SIBTEST	16	20															2	6.25	4.17		
	DFIT	1	2	4	5	6	8	9	10	13	14	16	17	19	21	22	23	25	26	25	50.00	70.83
		27	29	30	33	35	36	38														
44	SIBTEST	2	8															2	0.00	8.33		
	DFIT	1	6	14	15	25	29	30	35	36								9	6.25	33.33		
45	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	5	6	7	9	10	11	12	13	14	15	16	17	18	19	20	35	93.75	83.33
		21	22	23	24	25	26	27	28	29	30	31	32	33	35	38	39	40				
46	SIBTEST	9	10	20	33	38												5	12.50	8.33		
	DFIT	1	2	4	5	6	9	10	11	12	13	14	15	16	17	21	22	23	25	25	62.50	62.50
		26	31	32	36	38	39	40														
47	SIBTEST	20																1	0.00	4.17		
	DFIT	1	2	4	6	7	9	12	13	14	16	17	20	21	26	29	30	33	34	20	43.75	81.25
		36	38																			
48	SIBTEST	9	10	32	34	36	38											6	12.50	16.67		
	DFIT	1	2	4	5	6	7	11	14	15	16	20	25	27	28	30	31	32	35	21	50.00	54.17
		36	37	38																		
49	SIBTEST	20																1	0.00	4.17		
	DFIT	1	2	4	6	8	9	11	14	15	17	21	22	24	26	29	32	35	36	19	12.50	70.83
		38																				
50	SIBTEST	9	32															2	6.25	4.17		
	DFIT	1	2	4	9	11	32											6	6.25	20.83		
		$\bar{X}$										SIBTEST		4.31		5.42						
												DFIT		41.38		61.46						
		$S.D.$										SIBTEST		4.75		3.88						
												DFIT		18.12		16.17						

ตาราง 42 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 500 คน

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																รวม	สรุปผล									
	ผู้เชี่ยวชาญ	5	7	10	12	13	16	17	18	27	30	31	32	33	37	39	40	16	P	E								
1	SIBTEST	-																	0	0.00	0.00							
	DFIT	*****																	-	-	-							
2	SIBTEST	-																	0	0.00	0.00							
	DFIT	1	2	4	6	11	12	13	14	15	16	17	21	22	24	25	26	30	32	19	37.50	54.17						
		34																										
3	SIBTEST	2																	1	0.00	4.17							
	DFIT	6	12	14	21	29	30															6	12.50	16.67				
4	SIBTEST	-																	0	0.00	0.00							
	DFIT	1	2	4	6	8	9	11	12	13	15	16	17	19	21	23	29	31	32	21	43.75	58.33						
		36	38	39																								
5	SIBTEST	-																	0	0.00	0.00							
	DFIT	1	2	4	6	8	9	10	11	14	15	16	17	18	19	21	22	24	25	26	50.00	75.00						
		26	27	31	32	34	36	38	39																			
6	SIBTEST	-																	0	0.00	0.00							
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	24	25	27	23	43.75	66.67						
		28	29	30	36	38																						
7	SIBTEST	-																	0	0.00	0.00							
	DFIT	1	3	4	6	9	13	14	15	16	21	22	26	27	29	36	37	38	17	25.00	66.67							
8	SIBTEST	-																	0	0.00	0.00							
	DFIT	1	2	4	6	9	11	13	15	16	18	22	23	26	28	30	31	36	38	18	25.00	58.33						
9	SIBTEST	-																	0	0.00	0.00							
	DFIT	1	2	4	7	9	13	14	15	16	18	20	21	22	25	26	27	28	29	23	50.00	62.50						
		30	34	35	37	39																						
10	SIBTEST	-																	0	0.00	0.00							
	DFIT	1	2	4	9	12	13	14	15	17	18	21	22									12	25.00	33.33				
11	SIBTEST	-																	0	0.00	0.00							
	DFIT	1	4	14	25																4	0.00	16.67					
12	SIBTEST	-																	0	0.00	0.00							
	DFIT	1	4	11	15	21	25															6	0.00	25.00				
13	SIBTEST	-																	0	0.00	0.00							
	DFIT	1	2	4	6	9	13	14	15	17	19	21	25	36	38					14	12.50	50.00						
14	SIBTEST	-																	0	0.00	0.00							
	DFIT	1	3	4	6	9	11	12	14	15	16	17	18	21	22	23	24	25	27	23	43.75	66.67						
		29	30	32	36	38																						
15	SIBTEST	6																	1	0.00	4.17							
	DFIT	1	6	15	18	25	27	29	30	32												9	25.00	20.83				
16	SIBTEST	-																	0	0.00	0.00							
	DFIT	1	4	6	9	11	12	13	14	15	16	17	20	21	22	23	25	30	36	21	43.75	58.33						
		37	38	39																								
17	SIBTEST	-																	0	0.00	0.00							
	DFIT	1	2	4	6	9	11	13	15	16	17	21	22	23	25	26	27	29	30	21	37.50	62.50						
		32	35	38																								
18	SIBTEST	-																	0	0.00	0.00							
	DFIT	1	2	4	5	8	10	11	12	13	15	16	17	18	19	20	21	22	24	23	56.25	58.33						
		25	27	29	32	33	38																					

ตาราง 42 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																		รวม	สรุปผล		
	ผู้เชี่ยวชาญ	5	7	10	12	13	16	17	18	27	30	31	32	33	37	39	40	16	P	E			
19	SIBTEST	-																			0	0.00	0.00
	DFIT	1	2	4	6	9	11	13	14	15	16	17	18	21	22	25	26	27	28	22	31.25	75.00	
		34	35	36	38																		
20	SIBTEST	15																			1	0.00	4.17
	DFIT	1	2	4	6	9	11	12	13	14	15	17	21	22	25	26	27	28	29	23	31.25	75.00	
		30	34	35	36	38																	
21	SIBTEST	34																			1	0.00	4.17
	DFIT	1	2	4	6	9	11	13	15	16	17	19	21	22	25	28	34	35	36	19	18.75	66.67	
		38																					
22	SIBTEST	-																			0	0.00	0.00
	DFIT	1	4	5	6	7	8	9	10	11	13	15	16	17	18	19	21	22	26	24	68.75	54.17	
		27	28	30	32	34	39																
23	SIBTEST	-																			0	0.00	0.00
	DFIT	1	4	6	9	14	15	16	17	18	22	28									11	18.75	33.33
24	SIBTEST	-																			0	0.00	0.00
	DFIT	1	2	4	6	7	9	11	12	13	15	16	17	21	22	25	26	32	36	20	37.50	54.17	
		38	39																				
25	SIBTEST	-																			0	0.00	0.00
	DFIT	1	2	5	6	7	8	9	11	13	14	15	17	20	22	25	28	29	32	22	43.75	66.67	
		34	36	38	39																		
26	SIBTEST	-																			0	0.00	0.00
	DFIT	1	2	4	5	6	8	9	11	12	13	14	15	16	17	18	21	22	25	24	37.50	75.00	
		26	28	29	34	36	38																
27	SIBTEST	32																			1	0.00	4.17
	DFIT	1	2	4	6	8	9	11	12	13	15	16	17	18	20	22	25	27	28	25	50.00	70.83	
		29	30	34	35	36	37	38															
28	SIBTEST	-																			0	0.00	0.00
	DFIT	1	2	3	4	6	8	9	11	12	13	14	15	16	17	18	19	22	23	27	43.75	83.33	
		24	25	26	29	32	35	36	37	38													
29	SIBTEST	-																			0	0.00	0.00
	DFIT	1	6	13	14	16	19	21	22	30	35	36	38									12	18.75
30	SIBTEST	-																			0	0.00	0.00
	DFIT	1	2	4	5	6	7	11	13	15	16	17	21	25	26	29	30	31	33	21	50.00	58.33	
		34	35	38																			
31	SIBTEST	-																			0	0.00	0.00
	DFIT	2	4	5	6	7	8	9	11	13	14	15	17	18	20	21	22	25	28	22	37.50	66.67	
		30	34	35	38																		
32	SIBTEST	1																			1	0.00	4.17
	DFIT	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	18	19	21	35	81.25	91.66	
		22	23	24	25	26	27	28	29	30	31	32	35	36	37	38	39	40					
33	SIBTEST	40																			1	6.25	0.00
	DFIT	1	2	4	6	13	15	16	17	21	22	25	26	27	29	30	36	37	38	19	43.75	50.00	
		39																					
34	SIBTEST	-																			0	0.00	0.00
	DFIT	1	2	4	6	7	11	12	13	14	15	16	17	18	19	20	21	22	25	24	43.75	70.83	
		27	28	34	35	36	39																
35	SIBTEST	14																			1	0.00	4.17
	DFIT	1	2	4	5	6	9	11	12	13	15	16	17	18	22	23	26	27	28	20	43.75	58.33	
		30	35																				

ตาราง 42 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																		รวม	สรุปผล	
	ผู้เชี่ยวชาญ	5	7	10	12	13	16	17	18	27	30	31	32	33	37	39	40	16	P	E		
36	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	9	11	12	13	14	15	17	18	20	21	22	25	29	30	19	37.50	54.17
		32																				
37	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	21	22	25	27	29	30	22	37.50	70.83
		34	35	36	38																	
38	SIBTEST	15	22															2	0.00	8.33		
	DFIT	1	2	4	6	8	9	11	12	13	14	15	16	17	18	21	22	25	26	24	50.00	66.67
		27	31	34	36	37	38															
39	SIBTEST	-																0	0.00	0.00		
	DFIT	1	6	25	36													4	0.00	16.67		
40	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	7	8	9	11	12	13	14	16	17	21	22	25	27	28	26	50.00	75.00
		29	30	32	34	35	36	38														
41	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	7	8	11	13	15	16	20	21	22	25	26	29	32	33	22	43.75	62.50
		36	37	38	39																	
42	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	9	11	13	14	15	17	22	25	34				13	12.50	45.83		
43	SIBTEST	2																1	0.00	4.17		
	DFIT	1	2	4	6	9	11	13	14	15	16	17	19	21	22	23	26	29	30	23	31.25	75.00
		34	35	36	38	39																
44	SIBTEST	-																0	0.00	0.00		
	DFIT	1	4	11	29	30	32	34										7	12.50	25.00		
45	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	21												5	0.00	20.83		
46	SIBTEST	-																0	0.00	0.00		
	DFIT	4	9	14	21	29	30	34										7	6.25	25.00		
47	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	6	9	15	21	30	34	36	38							10	6.25	37.50		
48	SIBTEST	20																1	0.00	4.17		
	DFIT	1	2	4	6	9	11	13	14	15	16	17	21	22	25	26	29	30	31	20	37.50	58.33
		32	34																			
49	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	7	9	11	12	13	14	16	19	20	21	22	25	28	29	25	43.75	75.00
		32	34	35	36	37	39															
50	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	3	4	7	8	12	13	14	16	21	22	23	24	25	26	27	34	23	43.75	70.83
		35	36	37	38	40																

ตาราง 43 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																		รวม	สรุปผล	
	ผู้เชี่ยวชาญ	5	7	10	12	13	16	17	18	27	30	31	32	33	37	39	40	16	P	E		
1	SIBTEST	9	10	16	20													4	12.50	8.33		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	27	29	23	50.00	65.50
		32	35	36	38	39																
2	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	3	4	6	9	11	12	13	14	15	16	17	18	21	22	25	26	24	43.75	70.83
		27	28	34	36	38	39															
3	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	11	12	13	16	17	18	21	22	25	29	32	36	38	17	37.50	50.00	
4	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	27	29	23	50.00	62.50
		32	35	36	38	39																
5	SIBTEST	-																0	0.00	0.00		
	DFIT	1	4	6	29													4	0.00	16.67		
6	SIBTEST	-																0	0.00	0.00		
	DFIT	1	4	6	9	11												5	0.00	20.83		
7	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	32	33	19	43.75	54.17
		36																				
8	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	27	30	21	43.75	58.33
		34	35	36																		
9	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	3	4	6	9	11	12	13	14	15	16	17	18	21	22	25	27	24	50.00	66.67
		29	32	35	36	37	38															
10	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	26	28	19	31.25	58.33
		38																				
11	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	28	35	20	31.25	62.50
		36	38																			
12	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	27	28	21	37.50	62.50
		35	36	38																		
13	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	21	22	25		15	25.00	45.83		
14	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	19	20	21	22	25	23	37.50	70.83
		27	28	35	36	38																
15	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	7	8	9	11	13	14	15	16	17	18	21	22	25	29	22	43.75	62.50
		32	36	37	38																	
16	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	30	33	21	43.75	58.33
		36	37	38																		
17	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	4	6	7	11	12	13	14	15	16	17	18	22	29	32	35	36	18	37.50	50.00

ตาราง 43 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ ผู้เชี่ยวชาญ	ข้อสอบที่แสดงการกำหนดที่เบี่ยงเบน (DIF)																				รวม	สรุปผล	
		5	7	10	12	13	16	17	18	27	30	31	32	33	37	39	40	16	P	E				
18	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	4	6	7	8	9	10	11	12	13	14	15	16	17	18	21	22	23	56.25	58.33		
		25	32	35	38	39																		
19	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	4	6	9	11	13	14	15	16	17	18	21	22	25	27	28	29	22	37.50	66.67		
		33	35	36	38																			
20	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	3	4	6	9	11	13	14	15	17	18	21	22	25	27	32	35	19	25.00	62.50		
		36																						
21	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	21	22	25	27	28	32	22	43.75	58.33		
		35	36	38	39																			
22	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	9	14	29	32														6	6.25	20.83	
23	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	34	36	19	31.25	58.33		
		38																						
24	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	26	29	22	50.00	58.33		
		30	32	36	37																			
25	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	29	30	21	37.50	62.50		
		34	36	38																				
26	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	3	4	6	9	11	12	13	14	15	16	17	18	21	22	25	29	22	37.50	66.67		
		32	34	36	38																			
27	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	27	28	21	37.50	62.50		
		35	36	38																				
28	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	34	36	19	31.25	58.33		
		38																						
29	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	27	28	22	31.25	70.83		
		34	35	36	38																			
30	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	29	32	19	37.50	54.17		
		36																						
31	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	35	36	20	37.50	58.33		
		37	38																					
32	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	4	6	9	11	13	14	15	16	17	18	21	22	25	34	36	39	18	31.25	54.17		
33	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	3	4	6	7	11	12	13	14	15	17	18	19	21	22	25	29	21	43.75	58.33		
		32	37	38																				
34	SIBTEST	-																	0	0.00	0.00			
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	22	25	27	28	34	22	37.50	66.67		
		35	36	38																				

ตาราง 43 (ต่อ)

กลุ่มที่	วิธีตรวจสอบผู้เชี่ยวชาญ	ข้อสอบที่แสดงการกำหนดน้ำหนักที่เบี่ยงเบน (DIF)																				รวม	สรุปผล			
		5	7	10	12	13	16	17	18	27	30	31	32	33	37	39	40	16	P	E						
35	SIBTEST	-																				0	0.00	0.00		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	28	32	21	37.50	62.50				
		35	36	38																						
36	SIBTEST	-																				0	0.00	0.00		
	DFIT	1	9	12	14																	4	6.25	12.50		
37	SIBTEST	-																				0	0.00	0.00		
	DFIT	1	2	4	6	9	10	11	12	13	14	15	16	17	18	21	22	25	27	24	43.75	70.83				
		29	30	34	35	36	38																			
38	SIBTEST	-																				0	0.00	0.00		
	DFIT	1	2	4	6	7	9	11	12	13	14	15	16	17	18	21	22	25	35	19	37.50	54.17				
		36																								
39	SIBTEST	-																				0	0.00	0.00		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	27	28	24	43.75	70.83				
		29	30	34	35	36	38																			
40	SIBTEST	-																				0	0.00	0.00		
	DFIT	1	4	9	12	14	29																	6	6.25	29.41
41	SIBTEST	-																				0	0.00	0.00		
	DFIT	1	2	4	32	35																	5	6.25	16.67	
42	SIBTEST	-																				0	0.00	0.00		
	DFIT	1	4	12	15																	4	6.25	12.50		
43	SIBTEST	-																				0	0.00	0.00		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	27	34	21	37.50	62.50				
		35	36	38																						
44	SIBTEST	-																				0	0.00	0.00		
	DFIT	1	2	4	6	8	9	11	12	13	14	15	16	17	18	20	21	22	25	25	37.50	79.17				
		27	28	29	34	35	36	38																		
45	SIBTEST	-																				0	0.00	0.00		
	DFIT	1	2	4	6	7	8	9	11	13	14	15	16	17	18	21	22	25	29	22	43.75	62.50				
		32	36	37	38																					
46	SIBTEST	-																				0	0.00	0.00		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	28	34	21	31.25	66.67				
		35	36	38																						
47	SIBTEST	-																				0	0.00	0.00		
	DFIT	1	9	12	14																	4	6.25	12.50		
48	SIBTEST	-																				0	0.00	0.00		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	27	28	23	43.75	66.67				
		32	34	35	36	38																				
49	SIBTEST	-																				0	0.00	0.00		
	DFIT	1	2	4	6	9	11	12	13	14	15	16	17	18	21	22	25	28	35	19	31.25	58.33				
		36																								
50	SIBTEST	-																				0	0.00	0.00		
	DFIT	1	12	15	29																	4	6.25	12.50		
		$\bar{X}$										SIBTEST										0.25		0.17		
												DFIT										33.12		53.58		
		S.D.										SIBTEST										1.77		1.18		
												DFIT										14.68		18.31		

ตาราง 44 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ กลุ่มตัวอย่างขนาด 50 คน

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการหาหน้าที่เบี่ยงเบน (DIF)																		รวม	สรุปผล									
		ผู้เข้าข่าย	7	9	13	16	17	21	22	23	35	38	39	40	41	46	48	49	P		E									
1	SIBTEST	50																			1	0.00	2.94							
	DFIT	1	2	3	5	8	9	10	11	14	15	16	17	18	19	20	22	23	24	28	43.75	61.76								
		27	31	34	36	37	38	40	43	47	50																			
2	SIBTEST	-																			0	0.00	0.00							
	DFIT	2	3	5	6	7	9	10	11	12	14	15	17	19	20	22	23	24	28	33	56.25	70.59								
		29	31	32	33	34	35	37	39	40	42	43	44	45	46	50														
3	SIBTEST	21	45																		2	6.25	2.94							
	DFIT	2	5	6	10	11	12	15	17	20	21	22	26	29	31	32	34	39	42	22	25.00	41.18								
		44	45	47	50																									
4	SIBTEST	19	40																		2	6.25	2.94							
	DFIT	1	4	5	6	8	11	14	15	17	19	20	21	27	28	31	32	37	40	21	31.25	47.06								
		43	46	49																										
5	SIBTEST	9	13	25	35	40	42																6	25.00	5.88					
	DFIT	5	8	10	11	13	15	18	19	20	21	25	26	27	28	34	36	42	44	19	12.50	50.00								
		50																												
6	SIBTEST	9																			1	0.00	2.94							
	DFIT	2	3	5	14	18	22	26	28	31	35	37	38	43	47						14	12.50	29.41							
7	SIBTEST	6	13	50																		3	6.25	5.88						
	DFIT	1	2	4	5	8	9	10	11	14	16	17	18	19	20	21	24	25	26	33	56.25	70.59								
		27	28	30	31	33	35	37	38	39	40	42	45	46	47	50														
8	SIBTEST	18	22																		2	6.25	2.94							
	DFIT	1	2	3	8	9	10	11	14	15	17	18	19	20	21	25	26	27	28	35	56.25	76.47								
		29	31	32	33	36	37	38	39	40	41	42	43	44	45	47	48	49												
9	SIBTEST	48																			1	6.25	0.00							
	DFIT	1	2	5	6	10	11	13	15	17	18	20	22	27	28	29	30	31	33	29	43.75	64.71								
		35	36	38	42	43	44	45	47	48	49	50																		
10	SIBTEST	18																			1	0.00	2.94							
	DFIT	2	3	5	6	7	8	9	10	11	14	15	17	18	19	20	21	22	25	34	62.50	70.59								
		26	28	29	31	33	34	35	36	37	40	42	43	45	46	48	50													
11	SIBTEST	1	19	23	46																	4	12.50	5.88						
	DFIT	1	2	3	5	8	9	10	11	12	14	15	16	18	19	20	21	22	23	35	62.50	70.59								
		24	26	27	28	31	33	34	36	37	38	40	41	42	43	46	49	50												
12	SIBTEST	11	35																		2	6.25	2.94							
	DFIT	1	2	3	4	5	8	14	15	17	18	20	21	22	26	28	29	32	42	20	25.00	66.67								
		45	46																											
13	SIBTEST	48																			1	6.25	0.00							
	DFIT	1	2	3	4	5	6	9	10	11	12	14	15	16	17	18	19	20	21	40	68.75	85.29								
		23	24	25	27	28	29	31	32	33	35	37	38	39	40	41	42	43	44											
	45	47	48	50																										
14	SIBTEST	35	48																		2	12.50	0.00							
	DFIT	1	2	4	6	8	9	10	11	12	14	15	17	21	23	25	26	28	30	29	43.75	64.71								
		31	32	33	35	36	40	41	42	43	44	47																		
15	SIBTEST	16	17																		2	12.50	0.00							
	DFIT	1	2	4	5	6	7	8	10	11	15	18	21	22	23	25	26	27	29	28	31.25	67.65								
		31	32	33	34	35	37	42	43	44	47																			

ตาราง 44 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																	รวม	สรุปผล		
	ผู้เชี่ยวชาญ	7	9	13	16	17	21	22	23	35	38	39	40	41	46	48	49	16	P	E		
16	SIBTEST	4	6	36	45													4	0.00	11.76		
	DFIT	1	2	4	11	16	18	21	25	29	30	31	32	33	36	38	42	44	45	19	18.75	47.06
		50																				
17	SIBTEST	1	6	18														3	0.00	8.82		
	DFIT	1	2	5	8	9	10	11	12	15	17	18	19	20	21	22	23	25	26	29	43.75	64.71
		27	28	29	30	31	36	38	43	47	48	50										
18	SIBTEST	38																1	6.25	0.00		
	DFIT	1	2	3	4	5	6	7	11	14	17	18	19	20	23	24	25	26	28	25	25.00	61.76
		30	31	32	33	37	40	42														
19	SIBTEST	21	31															2	0.00	5.88		
	DFIT	1	2	3	5	6	7	8	11	14	15	16	17	18	19	21	22	23	24	38	68.75	79.41
		25	26	27	28	29	31	32	34	36	37	38	39	40	41	42	43	45	47			
	48	50																				
20	SIBTEST	23	35	41														3	18.75	0.00		
	DFIT	1	3	4	5	7	8	10	11	13	14	17	18	19	21	24	25	27	29	29	43.75	64.71
		30	31	33	34	37	38	40	42	44	48	50										
21	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	3	5	6	8	15	16	17	18	20	21	22	26	27	28	29	31	29	50.00	61.76
		32	33	34	35	37	40	41	42	45	47	48										
22	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	5	6	8	10	14	15	16	18	19	20	22	26	28	32	34	37	26	17.65	58.82
		38	39	40	42	43	44	45	48													
23	SIBTEST	25	39															2	6.25	2.94		
	DFIT	2	3	4	5	6	9	10	11	13	17	20	21	22	24	25	26	27	29	32	62.50	64.71
		30	31	33	34	36	39	40	41	42	44	45	46	47	49							
24	SIBTEST	19																1	0.00	2.94		
	DFIT	4	5	6	7	8	10	11	12	13	14	15	16	18	19	20	21	23	25	33	23.53	73.53
		26	27	28	29	30	31	32	33	36	37	40	41	45	47	48						
25	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	3	5	8	10	11	13	14	15	19	20	24	26	34	35	36	39	20	25.00	47.06
		44	49																			
26	SIBTEST	47	49															2	6.25	2.94		
	DFIT	1	2	3	6	8	9	11	14	15	17	18	19	20	23	24	26	27	28	28	43.75	61.76
		31	32	34	36	37	38	40	44	46	48											
27	SIBTEST	11	21	42														3	6.25	5.88		
	DFIT	5	8	9	10	11	13	14	15	17	19	20	21	22	23	26	27	28	30	31	56.25	64.71
		31	32	34	37	38	39	40	41	44	45	48	50									
28	SIBTEST	9	50															2	0.00	5.88		
	DFIT	1	2	3	5	6	8	10	11	17	19	21	22	26	27	28	29	31	32	26	43.75	55.88
		35	37	39	40	41	42	44	45	50												
29	SIBTEST	9	13	14														3	12.50	2.94		
	DFIT	1	2	3	5	6	8	9	10	11	13	17	20	21	23	26	27	30	31	27	50.00	55.88
		32	35	36	37	38	40	43	44	48												
30	SIBTEST	41																1	6.25	0.00		
	DFIT	6	8	11	14	15	17	18	19	20	21	22	27	28	29	31	36	37	38	22	37.50	47.06
		40	42	46	47																	
31	SIBTEST	1																1	0.00	2.94		
	DFIT	2	3	4	5	6	7	8	9	10	11	13	14	15	17	18	19	20	21	38	62.50	82.35
		22	23	25	26	27	28	29	30	31	32	33	34	35	37	38	42	44	45			
	47	49																				

ตาราง 44 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																		รวม	สรุปผล	
		ผู้เชี่ยวชาญ	7	9	13	16	17	21	22	23	35	38	39	40	41	46	48	49	16		P	E
31	SIBTEST	1																	1	0.00	2.94	
	DFIT	2	3	4	5	6	7	8	9	10	11	13	14	15	17	18	19	20	21	38	62.50	82.35
		22	23	25	26	27	28	29	30	31	32	33	34	35	37	38	42	44	45			
		47	49																			
32	SIBTEST	1	3																2	0.00	5.88	
	DFIT	1	2	3	4	5	6	11	12	14	15	17	19	20	21	22	23	24	25	29	43.75	64.71
		27	28	29	33	40	41	44	45	46	47	49										
33	SIBTEST	-																	0	0.00	0.00	
	DFIT	1	2	3	5	6	8	9	10	11	13	14	17	18	19	20	22	23	25	36	62.50	76.47
		27	28	32	33	35	36	37	39	40	41	42	43	44	45	47	48	49	50			
34	SIBTEST	35																	1	0.00	2.94	
	DFIT	1	2	5	6	8	14	15	17	19	20	22	23	25	26	27	31	32	33	29	43.75	64.71
		34	35	36	40	41	42	43	44	45	48	49										
35	SIBTEST	37																	1	0.00	2.94	
	DFIT	1	2	3	5	6	8	11	12	14	18	20	21	22	26	27	31	32	33	24	25.00	58.82
		34	37	38	40	44	50															
36	SIBTEST	4	36	41															3	6.25	5.88	
	DFIT	1	2	4	5	6	7	8	9	15	16	17	18	21	24	25	26	29	32	31	62.50	61.76
		33	34	35	36	38	40	41	42	43	44	45	48	50								
37	SIBTEST	4	36	41															3	6.25	5.88	
	DFIT	1	2	3	4	5	6	8	9	10	11	13	14	17	18	19	20	21	22	37	56.25	82.35
		23	26	27	28	29	31	32	33	34	36	37	38	39	40	42	43	44	45			
		50																				
38	SIBTEST	10	13	21	28	46	50												6	18.75	8.82	
	DFIT	1	2	3	4	5	7	8	11	14	15	16	17	18	19	20	21	22	24	27	50.00	55.88
		26	29	37	38	40	43	44	46	50												
39	SIBTEST	36	41	42	47														4	6.25	8.82	
	DFIT	1	8	17	18	19	22	30	32	34	40	45							11	18.75	23.53	
40	SIBTEST	41																	1	6.25	0.00	
	DFIT	1	2	3	4	6	7	8	9	11	13	14	15	16	20	21	22	24	26	33	68.75	64.71
		27	28	31	32	33	35	36	37	38	40	41	44	45	47	49						
41	SIBTEST	35																	1	0.00	2.94	
	DFIT	1	2	3	5	6	8	10	11	14	17	18	19	20	21	22	23	24	25	33	43.75	76.47
		26	27	28	29	33	34	36	37	38	42	43	44	46	47	48						
42	SIBTEST	21	25	47															3	6.25	5.88	
	DFIT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	17	18	19	20	40	56.25	91.18
		21	23	24	25	26	27	28	29	30	31	33	35	37	38	39	42	43	44			
		45	46	47	50																	
43	SIBTEST	2	13	40	45														4	12.50	5.88	
	DFIT	1	3	8	9	11	15	17	20	23	24	25	27	28	30	31	32	34	36	27	50.00	64.71
		37	38	40	41	42	43	44	48	49												
44	SIBTEST	31																	1	0.00	2.94	
	DFIT	2	3	5	6	7	8	10	11	13	14	15	16	17	18	19	20	21	22	36	81.25	67.65
		23	25	27	28	31	32	34	35	37	38	39	42	43	45	46	48	49	50			
45	SIBTEST	20																	1	0.00	2.94	
	DFIT	1	2	4	5	6	7	8	10	11	15	18	21	22	25	27	29	31	32	27	31.25	64.71
		33	34	35	37	42	43	44	47	49												

ตาราง 45 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																	รวม	สรุปผล	
	ผู้เชี่ยวชาญ	7	9	13	16	17	21	22	23	35	38	39	40	41	46	48	49	16	P	E	
17	SIBTEST	4																1	0.00	2.94	
	DFIT	1	2	3	4	5	6	7	8	10	11	14	15	16	18	19	20	21	22		
		23	28	32	34	35	37	39	40	41	44	45	46	49	50				32	68.75	61.76
18	SIBTEST	-																0	0.00	0.00	
	DFIT	1	2	3	4	5	6	7	8	10	11	14	15	16	18	19	20	21	22		
		23	28	32	34	35	37	39	40	41	44	45	46	49	50				32	68.75	61.76
19	SIBTEST	7	25															2	6.25	2.94	
	DFIT	1	4	9	10	11	13	15	17	20	21	22	23	24	25	27	28	29	31		
		34	38	39	40	41	42	48	49	50									27	68.75	47.06
20	SIBTEST	25	40															2	6.25	2.94	
	DFIT	2	5	6	7	8	9	11	12	14	15	17	19	20	21	22	24	25	28		
		29	32	34	40	48	49												24	50.00	47.06
21	SIBTEST	-																0	0.00	0.00	
	DFIT	1	2	3	4	5	6	7	8	9	11	12	14	16	17	19	20	22	23		
		24	25	26	28	30	31	32	37	40	42	43	48	49					31	56.25	64.71
22	SIBTEST	11	21	25	38													4	12.50	5.88	
	DFIT	1	2	3	5	6	7	8	11	12	14	15	18	19	20	21	23	24	25		
		27	28	31	32	33	43	47	50										26	18.75	67.65
23	SIBTEST	36	48															2	6.25	2.94	
	DFIT	1	2	3	4	5	6	7	8	11	12	14	15	17	18	19	20	21	22		
		26	27	28	31	32	33	34	35	36	37	38	40	42	43	44	46	49	35	56.25	76.47
24	SIBTEST	11	23	50														3	6.25	5.88	
	DFIT	1	2	3	4	5	6	7	8	9	11	12	14	16	17	18	19	20	22		
		23	24	25	26	28	30	31	32	37	40	42	43	48	49				32	56.25	67.65
25	SIBTEST	3	25	28														3	0.00	8.82	
	DFIT	1	2	3	6	8	9	10	11	12	13	14	15	17	22	23	24	25	26		
		27	29	31	32	33	34	37	38	39	40	43	44	49	50				32	56.25	67.65
26	SIBTEST	11	19	36														3	0.00	8.82	
	DFIT	1	2	3	4	5	6	7	8	9	11	12	13	14	15	16	17	21	22		
		23	26	27	30	31	32	34	35	38	40	43	44	45	47	50			33	68.75	64.71
27	SIBTEST	25	50															2	0.00	5.88	
	DFIT	2	4	5	6	7	8	9	10	11	12	14	15	16	18	19	20	22	23		
		24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	42	43		
	44	45	46	47	48	50												42	62.50	94.12	
28	SIBTEST	25	45															2	0.00	5.88	
	DFIT	1	2	3	4	5	6	7	8	10	11	12	13	14	15	16	17	18	19		
		20	21	22	23	24	26	27	31	32	33	34	35	36	37	38	39	40	42		
	43	44	45	46	47	50												42	75.00	88.24	
29	SIBTEST	-																0	0.00	0.00	
	DFIT	1	2	3	5	6	7	8	9	10	11	14	15	17	19	21	22	23	24		
		25	26	28	30	31	32	33	34	35	36	37	38	42	44	45	50		34	50.00	76.47
30	SIBTEST	1	2	3	6	13	15	17	19	22	34	36	38	46	47	48	49		16	43.75	23.52
	DFIT	1	2	3	6	8	9	10	11	12	14	15	16	17	18	19	25	27	29		
		30	32	36	37	39	43	44	45	48	49	50							29	31.25	70.59
31	SIBTEST	34	47															2	0.00	5.88	
	DFIT	1	6	8	9	10	11	12	14	15	16	17	18	19	21	25	27	29	30		
		32	36	37	39	43	44	45	48	49	50								28	37.50	64.71

ตาราง 45 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																		รวม	สรุปผล	
	ผู้เชี่ยวชาญ	7	9	13	16	17	21	22	23	35	38	39	40	41	46	48	49	16	P	E		
32	SIBTEST	15	25	35	36	40												5	0.00	14.71		
	DFIT	1	2	5	8	9	11	12	13	14	15	18	19	22	23	24	25	26	28	31	37.50	73.53
		29	30	31	34	35	36	39	40	42	44	45	47	49								
33	SIBTEST	29																1	0.00	2.94		
	DFIT	1	2	3	6	8	11	12	13	14	15	16	17	18	19	20	23	24	29	27	50.00	55.88
		32	34	36	40	42	43	46	48	49												
34	SIBTEST	28																1	0.00	2.94		
	DFIT	1	2	3	4	5	7	8	9	11	15	17	19	20	21	22	23	24	29	34	68.25	67.65
		31	32	33	34	35	36	37	38	39	40	41	42	43	44	47	50					
35	SIBTEST	19																1	0.00	2.94		
	DFIT	1	2	5	6	8	11	12	13	14	15	16	17	18	19	20	22	23	24	30	43.75	67.65
		26	27	29	31	32	34	36	38	39	42	43	49									
36	SIBTEST	11	39	50														3	6.25	5.88		
	DFIT	1	4	8	11	12	14	16	17	19	21	23	25	26	27	28	31	32	35	25	43.75	52.94
		36	37	38	41	43	45	50														
37	SIBTEST	4	21	22														3	12.50	2.94		
	DFIT	5	6	8	11	15	17	19	20	21	24	27	28	31	38	45		15	18.75	35.29		
38	SIBTEST	12																1	0.00	2.94		
	DFIT	1	2	3	6	7	11	12	13	14	15	17	18	20	21	26	27	28	29	30	37.50	70.59
		31	32	33	35	36	39	43	44	45	46	47	49									
39	SIBTEST	25	45															2	0.00	5.88		
	DFIT	4	8	11	12	16	18	20	21	22	24	25	26	27	28	29	32	34	35	24	43.75	50.00
		39	40	41	42	44	45															
40	SIBTEST	25																1	0.00	2.94		
	DFIT	2	3	7	8	11	13	14	17	22	23	25	26	27	28	29	30	31	32	30	43.75	95.83
		33	34	35	36	37	39	43	44	45	47	49	50									
41	SIBTEST	11	48															2	6.25	2.94		
	DFIT	1	2	3	5	6	8	11	14	15	16	17	18	19	20	21	23	24	26	27	31.25	91.67
		27	28	30	31	32	34	41	43	47												
42	SIBTEST	15	25															2	0.00	5.88		
	DFIT	1	2	3	4	5	6	7	8	9	10	11	15	17	18	19	20	21	22	37	68.75	76.47
		23	26	27	28	29	31	33	34	35	36	38	39	40	42	43	44	45	49			
	50																					
43	SIBTEST	31	37	47														3	0.00	8.82		
	DFIT	1	2	6	7	8	11	12	14	16	17	20	21	22	26	27	28	31	32	27	43.75	58.82
		33	34	37	38	40	43	45	47	50												
44	SIBTEST	3	11	14	41													4	6.25	8.82		
	DFIT	*****																-	-	-		
45	SIBTEST	11	21															2	6.25	2.94		
	DFIT	1	2	3	5	6	7	8	11	14	17	18	19	20	21	22	23	24	25	36	56.25	79.41
		27	28	29	30	31	32	33	34	35	36	37	38	40	42	44	46	47	49			
46	SIBTEST	7	50															2	6.25	2.94		
	DFIT	1	2	3	5	6	8	14	15	17	18	19	20	21	22	23	24	25	26	35	56.25	76.47
		27	28	30	31	32	34	35	36	37	38	42	44	45	46	47	48	49				
47	SIBTEST	1	24															2	0.00	5.88		
	DFIT	1	2	3	4	5	6	7	8	9	11	12	13	14	15	16	17	18	19	47	93.75	94.12
		20	21	22	23	24	26	27	28	29	30	31	32	33	34	35	36	37	38			
		39	41	42	43	44	45	46	47	48	49	50										

ตาราง 45 (ต่อ)

[illegible]

ตาราง 46 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ กลุ่มตัวอย่างขนาด 200 คน

[illegible]

ตาราง 46 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																		รวม	สรุปผล	
	ผู้เชี่ยวชาญ	7	9	13	16	17	21	22	23	35	38	39	40	41	46	48	49	16	P	E		
16	SIBTEST	6	21	49														3	12.50	2.94		
	DFIT	1	2	5	23	24	35											6	12.50	11.76		
17	SIBTEST	23	30															2	6.25	2.94		
	DFIT	1	2	3	4	5	6	7	8	9	11	14	15	17	18	19	20	21	22	40	75.00	82.35
		23	24	25	26	27	28	29	31	32	33	34	35	36	38	39	40	41	43			
		44	45	47	48																	
18	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	3	5	7	9	10	11	15	17	18	20	21	22	23	27	28	32	26	56.25	50.00
		34	36	38	40	42	45	46	47													
19	SIBTEST	23	40	50														3	12.50	2.94		
	DFIT	1	2	3	5	6	9	10	12	14	15	17	18	19	20	22	23	26	27	33	56.25	70.59
		28	29	31	34	35	36	39	40	41	42	43	45	47	48	50						
20	SIBTEST	12	13	19	41													4	12.50	5.88		
	DFIT	1	2	3	5	8	11	12	14	15	18	19	22	23	25	26	27	28	30	28	31.25	67.65
		31	32	34	36	37	38	39	43	48	50											
21	SIBTEST	4	11	25														3	0.00	8.82		
	DFIT	1	3	5	6	8	11	14	15	19	20	28	32	38	43			14	6.25	38.24		
22	SIBTEST	19	25	48														3	6.25	5.88		
	DFIT	1	2	3	5	6	8	10	12	14	15	16	17	18	19	21	22	23	26	31	62.50	88.24
		27	31	32	34	38	40	42	43	44	45	46	47	48								
23	SIBTEST	16	31	40														3	12.50	2.94		
	DFIT	1	2	3	5	6	7	8	11	12	14	15	16	17	18	19	20	21	24	33	31.25	82.35
		26	27	28	29	30	31	32	33	34	36	37	39	42	43	50						
24	SIBTEST	1	21	22														3	12.50	2.94		
	DFIT	*****																-	-	-	-	-
25	SIBTEST	20	24															2	0.00	5.88		
	DFIT	1	2	5	6	7	8	11	12	14	15	18	19	21	22	24	26	27	28	33	56.25	70.59
		30	31	32	33	35	38	39	40	41	42	43	44	45	47	49						
26	SIBTEST	11	24	31	39													4	6.25	8.82		
	DFIT	1	3	5	6	10	14	15	17	18	19	20	22	23	26	27	28	30	31	29	37.50	67.65
		32	33	35	37	38	42	43	44	45	47	48										
27	SIBTEST	25																1	0.00	2.94		
	DFIT	1	2	3	4	5	7	11	12	14	15	16	18	19	20	21	26	27	28	29	31.25	70.59
		31	32	33	34	35	36	37	38	43	45	47										
28	SIBTEST	11	46															2	6.25	2.94		
	DFIT	1	2	3	4	5	6	7	8	9	10	11	13	14	15	17	18	19	20	38	68.75	79.41
		21	23	24	26	27	28	31	32	34	35	36	37	38	39	40	43	45	47			
		49	50																			
29	SIBTEST	11	13	19	21	24	40	47	49									8	25.00	11.76		
	DFIT	1	2	4	5	6	11	14	15	17	18	20	21	22	24	25	26	27	28	30	37.50	70.59
		29	30	31	32	34	35	41	42	43	45	46	47									
30	SIBTEST	3	28															2	0.00	5.88		
	DFIT	1	2	3	4	5	6	7	11	12	14	15	17	18	22	26	27	28	29	31	37.50	73.53
		31	32	33	34	36	38	40	42	43	44	45	48	50								
31	SIBTEST	7	19	24														3	6.25	5.88		
	DFIT	1	2	3	5	6	7	8	9	11	12	13	14	17	18	19	20	21	22	34	56.25	97.06
		26	27	28	29	31	32	33	36	38	42	44	45	46	47	49	50					
32	SIBTEST	-																0	0.00	0.00		
	DFIT	*****																-	-	-	-	-

ตาราง 46 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																	รวม	สรุปผล		
	ผู้เชี่ยวชาญ	7	9	13	16	17	21	22	23	35	38	39	40	41	46	48	49	16	P	E		
33	SIBTEST	25	40															2	6.25	2.94		
	DFIT	1	2	3	5	6	7	8	10	11	12	13	14	15	16	17	18	19	20	41	81.25	82.35
		21	22	23	25	26	27	28	31	32	33	34	35	36	38	40	41	43	44			
		45	46	47	48																	
34	SIBTEST	1	11	19	20	21	24	40										7	12.50	14.71		
	DFIT	1	2	3	5	6	7	8	10	11	14	15	17	18	20	21	22	23	24	33	43.75	76.47
		25	26	28	31	32	33	37	39	40	42	43	44	45	47	48						
35	SIBTEST	13	19	49														3	12.50	2.94		
	DFIT	1	2	3	5	6	7	8	10	11	14	15	17	18	19	20	21	26	28	29	37.50	67.65
		29	31	32	36	37	38	41	42	45	49	50										
36	SIBTEST	6	25															2	0.00	5.88		
	DFIT	1	2	3	5	6	7	8	11	14	15	16	17	18	19	20	22	25	27	32	50.00	70.59
		28	29	31	33	34	37	38	41	42	43	44	45	48	49							
37	SIBTEST	13	24	25	31	35												5	12.50	8.82		
	DFIT	1	2	3	5	6	7	8	9	10	13	14	15	16	17	19	20	21	22	34	68.75	67.65
		23	24	26	27	28	32	36	37	38	39	40	43	44	45	47	50					
38	SIBTEST	41	50															2	6.25	2.94		
	DFIT	1	2	3	5	7	9	15	16	18	19	20	22	23	24	25	26	27	31	32	62.50	64.71
		32	33	34	37	38	39	41	42	44	45	46	47	48	49							
39	SIBTEST	-																0	0.00	0.00		
	DFIT	1	5	6	7	8	10	11	12	14	15	16	17	18	19	20	26	27	31	21	18.75	52.94
		32	36	42																		
40	SIBTEST	40																1	6.25	0.00		
	DFIT	1	2	3	4	5	6	7	8	11	13	14	15	17	18	19	20	21	22	39	62.50	85.29
		23	27	28	29	30	31	32	33	34	35	36	37	38	39	40	42	43	44			
		45	46	47																		
41	SIBTEST	11	25															2	0.00	5.88		
	DFIT	1	2	3	5	8	14	15	20	26	27	28	31	32	37	45	50	16	0.00	47.06		
42	SIBTEST	1	6	19	20	23	25	31										7	6.25	17.65		
	DFIT	1	2	5	6	7	8	9	10	11	14	15	16	17	18	20	21	27	28	31	37.50	64.71
		31	32	33	35	36	37	38	39	41	42	43	44	50								
43	SIBTEST	21	24	47														3	6.25	5.88		
	DFIT	1	2	5	6	8	9	11	13	17	18	19	20	21	22	26	27	28	31	29	62.50	55.88
		34	35	36	37	38	40	41	44	45	46	49										
44	SIBTEST	1	13	25	41													4	12.50	5.88		
	DFIT	1	2	3	5	6	13	14	15	17	18	19	20	21	22	23	25	26	27	31	43.75	70.59
		28	30	31	32	33	34	35	36	37	38	43	45	47								
45	SIBTEST	10																1	0.00	2.94		
	DFIT	1	2	3	5	6	8	11	12	14	15	16	19	20	22	24	25	26	27	32	43.75	73.53
		30	32	33	38	39	41	43	44	45	46	47	48	49	50							
46	SIBTEST	21	27															2	6.25	2.94		
	DFIT	1	2	3	8	9	11	12	14	16	19	20	21	22	24	26	27	29	30	32	43.75	73.53
		31	32	33	34	35	37	39	40	43	44	45	46	47	50							
47	SIBTEST	1	3	17	25	31												5	6.25	11.76		
	DFIT	1	2	3	5	6	8	10	11	13	14	15	17	18	20	21	22	23	26	31	43.75	70.59
		27	28	29	31	32	34	35	36	37	38	42	44	45								
48	SIBTEST	11	25															2	0.00	5.88		
	DFIT	1	2	3	4	5	6	11	14	16	17	18	19	20	21	22	24	26	27	33	50.00	73.53
		29	31	32	33	34	36	37	38	40	41	43	44	45	47	49						

ตาราง 47 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ กลุ่มตัวอย่างขนาด 500 คน

กลุ่มที่	วัดตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																	รวม	สรุปผล		
	ผู้เข้าข่าย	7	9	13	16	17	21	22	23	35	38	39	40	41	46	48	49	16	P	E		
1	SIBTEST	9	13	19	24	25												5	12.50	8.82		
	DFIT	1	2	3	5	6	7	9	11	13	14	15	17	18	20	21	22	23	27	33	68.75	64.71
		29	32	34	35	36	37	38	43	44	45	46	47	48	49	50						
2	SIBTEST	2	3	7	13	19	25	31	49	50								9	18.75	17.65		
	DFIT	1	2	3	5	7	8	11	12	14	15	17	18	19	20	21	22	23	24	30	37.50	70.59
		27	28	29	31	32	33	34	40	43	44	45	49									
3	SIBTEST	11	25															2	0.00	5.88		
	DFIT	1	2	3	5	6	8	9	14	15	17	18	19	20	21	22	23	26	27	31	50.00	67.65
		28	30	31	32	34	35	36	37	40	43	45	47	49								
4	SIBTEST	3	4	11	19	21	25	50										7	6.25	17.65		
	DFIT	1	2	3	5	6	8	10	11	12	14	15	16	17	18	19	20	21	22	28	31.25	67.65
		26	30	31	32	35	37	40	41	45	47											
5	SIBTEST	11	21	25														3	6.25	5.88		
	DFIT	1	2	3	5	6	8	10	11	12	14	15	16	17	18	19	20	21	22	34	50.00	61.76
		24	25	26	27	28	31	32	34	35	36	37	38	40	44	45	47					
6	SIBTEST	1	11	13	25	31												5	6.25	11.76		
	DFIT	1	2	3	5	6	7	8	13	14	15	17	19	20	22	23	25	26	28	29	43.75	79.41
		31	34	35	36	37	38	45	46	47	48	50										
7	SIBTEST	2	11	25	47													4	0.00	11.76		
	DFIT	1	2	3	5	6	8	10	11	12	13	14	15	16	17	18	19	20	21	34	50.00	76.47
		22	23	26	27	28	31	32	33	34	35	36	40	43	45	47	49					
8	SIBTEST	11	25	31														3	0.00	8.82		
	DFIT	2	6	8	11	12	14	15	17	19	20	22	25	29	31	32	36	37	38	24	25.00	58.82
		39	40	45	46	47	50															
9	SIBTEST	6	11	19														3	0.00	8.82		
	DFIT	1	2	3	5	6	8	11	16	17	18	19	20	21	22	23	26	27	28	23	37.50	58.82
		29	32	38	43	44																
10	SIBTEST	3	10	11	21	25												5	6.25	11.76		
	DFIT	1	2	6	11	13	15	18	19	20	26	28	30	31	32	34	35	36	37	23	25.00	55.88
		38	39	40	43	45																
11	SIBTEST	11	19	25	40	50												5	6.25	11.76		
	DFIT	1	2	3	5	6	8	11	14	15	17	18	19	20	21	25	26	27	28	24	18.75	61.76
		31	32	35	43	45	47															
12	SIBTEST	11	25															2	0.00	5.88		
	DFIT	*****																-	-	-		
13	SIBTEST	1	3	9	11	15	17	19	21	25	31	48						11	25	20.59		
	DFIT	1	2	3	6	8	15	18	19	27	29	31	34	35	38	47	50	16	12.50	41.18		
14	SIBTEST	19																1	0.00	2.94		
	DFIT	1	2	3	5	6	7	8	10	11	13	14	15	16	17	18	19	20	21	38	68.75	79.41
		22	23	26	27	28	29	31	32	33	34	35	36	37	38	39	43	44	45			
		46	47																			
15	SIBTEST	11	13	25														3	6.25	5.88		
	DFIT	1	2	3	4	5	6	8	10	11	14	15	16	17	18	19	20	21	23	29	43.75	64.71
		25	26	31	32	34	35	36	39	40	43	44										
16	SIBTEST	3	11	21	24	25												5	6.25	11.76		
	DFIT	1	3	6	7	8	10	11	12	14	15	16	17	18	19	20	21	22	25	34	50.00	76.47
		26	27	28	31	32	33	34	35	36	37	38	40	42	44	45	46					

ตาราง 47 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																	รวม	สรุปผล		
	ผู้เชี่ยวชาญ	7	9	13	16	17	21	22	23	35	38	39	40	41	46	48	49	16	P	E		
17	SIBTEST	3	4	11	13	25												5	6.25	11.76		
	DFIT	2	3	5	6	7	8	10	14	15	17	18	19	20	22	26	29	31	32	22	18.75	55.88
		33	35	37	43																	
18	SIBTEST	11	21	24	25													4	6.25	8.82		
	DFIT	1	2	3	6	7	8	11	12	15	16	17	18	19	20	21	27	28	31	22	18.75	55.88
		32	34	35	43																	
19	SIBTEST	25																1	0.00	2.94		
	DFIT	1	2	3	4	5	6	7	8	10	11	12	14	17	18	19	20	21	23	34	56.25	73.53
		26	27	28	31	32	35	36	37	38	39	43	46	47	48	49	50					
20	SIBTEST	3	25	31	41	50												5	6.25	11.76		
	DFIT	1	3	5	6	8	11	12	14	15	16	17	18	20	21	22	23	25	26	35	56.25	76.47
		27	28	30	31	32	33	34	35	36	37	38	40	44	45	46	47	49				
21	SIBTEST	11	21	25	50													4	6.25	8.82		
	DFIT	1	2	3	4	5	6	8	11	14	16	17	18	19	20	21	22	23	26	29	37.50	67.65
		27	28	31	32	34	40	44	45	47	49	50										
22	SIBTEST	11	13	21	25	50												5	6.25	11.76		
	DFIT	5	8	15	16	17	18	19	20	28	32	34	37	43	47			14	12.50	35.29		
23	SIBTEST	11	25															2	0.00	5.88		
	DFIT	1	2	3	5	6	7	8	11	12	13	14	16	17	18	19	20	21	22	30	56.25	61.76
		26	27	28	31	35	37	38	43	45	46	48	50									
24	SIBTEST	21	24	25	40													4	12.50	5.88		
	DFIT	1	2	3	5	6	8	9	11	14	15	16	17	18	19	20	21	26	27	35	62.50	73.53
		28	29	31	32	33	34	35	37	38	40	41	42	43	44	46	47	49				
25	SIBTEST	11	25															2	0.00	5.88		
	DFIT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	45	75.00	97.06
		19	20	21	22	23	24	25	26	27	28	30	31	32	33	34	35	36	37			
26	SIBTEST	19	21	50														3	6.25	5.88		
	DFIT	1	2	3	5	6	7	8	11	12	14	15	17	18	19	20	21	22	23	33	31.25	82.35
		26	27	28	30	31	32	34	35	36	37	42	43	44	45	47						
27	SIBTEST	25	41															2	0.00	5.88		
	DFIT	1	2	3	4	5	6	8	9	11	14	15	17	18	19	20	21	22	23	31	43.75	70.59
		26	27	31	32	34	36	37	38	42	44	45	46	47								
28	SIBTEST	11	24															2	0.00	5.88		
	DFIT	1	2	3	4	5	6	8	9	11	12	14	15	16	17	18	19	20	21	35	56.25	76.47
		22	23	26	27	28	31	32	33	34	35	37	42	43	45	46	47	48				
29	SIBTEST	1	2	25														3	0.00	8.82		
	DFIT	1	2	3	5	14	19	28	32	33	39							10	0.00	29.41		
30	SIBTEST	3	4	25														3	0.00	8.82		
	DFIT	1	2	3	4	5	6	11	14	15	17	18	19	22	23	26	28	31	34	22	25.005	52.94
		37	40	46	50																	
31	SIBTEST	3	25															2	0.00	5.88		
	DFIT	1	2	3	5	6	8	9	10	11	13	14	15	18	19	20	22	23	25	35	26.47	76.47
		26	28	31	32	33	34	35	36	37	38	39	41	42	43	45	48	49				
32	SIBTEST	11	13	25	35	40												5	18.75	5.88		
	DFIT	1	2	3	4	5	7	8	11	14	15	16	17	18	19	20	21	22	23	35	26.47	76.47
		24	26	27	28	29	33	35	36	37	38	40	42	43	44	45	47	48				
33	SIBTEST	25	48	50														3	6.25	5.88		
	DFIT	1	2	3	5	6	8	10	11	12	13	14	15	16	17	18	19	21	22	36	68.75	73.53
		23	26	27	31	32	33	34	35	36	38	40	42	43	44	45	46	47	49			

ตาราง 47 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																		รวม	สรุปผล	
	ผู้เชี่ยวชาญ	7	9	13	16	17	21	22	23	35	38	39	40	41	46	48	49	16	P	E		
34	SIBTEST	11	21	23	25	40												5	12.50	8.82		
	DFIT	1	2	3	5	6	8	11	12	14	17	18	19	20	21	22	26	27	31	28	50.00	58.82
		34	35	36	38	39	42	45	46	47	48											
35	SIBTEST	11	21	24	25													4	6.25	8.82		
	DFIT	1	2	3	5	6	8	9	11	12	14	15	17	18	19	20	21	22	23	36	56.25	79.41
		24	25	26	27	28	31	32	33	35	37	39	43	44	45	46	47	48	50			
36	SIBTEST	11	13	22	25													4	6.25	8.82		
	DFIT	1	2	3	5	6	9	14	15	16	17	18	20	21	22	23	26	27	31	28	56.25	55.88
		34	35	38	39	42	43	44	45	47	48											
37	SIBTEST	11	13	19	21	25	47	50										7	6.25	17.65		
	DFIT	1	2	3	5	6	8	9	11	12	13	14	15	16	17	18	19	20	21	44	93.75	85.29
		22	23	24	25	26	27	28	30	31	32	33	35	37	38	39	40	41	42			
		43	44	45	46	47	48	49	50													
38	SIBTEST	3	21	24	25	41	45											6	12.50	11.76		
	DFIT	1	2	3	5	6	11	13	14	17	18	19	20	22	23	24	26	27	28	29	31.25	70.59
		29	31	32	34	35	37	41	44	45	49											
39	SIBTEST	25																1	0.00	2.94		
	DFIT	1	2	3	4	5	6	8	11	13	14	15	16	17	18	19	20	21	22	33	56.25	70.59
		23	24	25	26	27	31	32	34	36	37	38	39	40	42	46						
40	SIBTEST	25																1	0.00	2.94		
	DFIT	1	2	3	5	6	8	14	15	16	17	18	19	21	22	25	26	27	28	26	31.25	61.76
		29	31	32	33	34	36	44	48													
41	SIBTEST	3	11	20	25													4	0.00	11.76		
	DFIT	1	2	3	5	6	8	11	14	17	19	20	22	25	26	27	28	30	32	22	25.00	52.94
		35	43	46	47																	
42	SIBTEST	-																0	0.00	0.00		
	DFIT	1	2	3	4	5	6	8	12	14	15	16	17	18	19	21	22	24	26	28	37.50	64.71
		28	30	31	34	35	37	41	45	47	49											
43	SIBTEST	19	25	31														3	0.00	8.82		
	DFIT	1	2	3	4	5	6	7	8	9	11	14	15	17	18	19	20	25	26	28	37.50	64.71
		28	29	32	34	36	37	44	46	48	49											
44	SIBTEST	10	11	21	24	25												5	6.25	11.76		
	DFIT	1	2	3	5	6	8	9	13	14	15	17	18	19	20	21	22	23	25	33	43.75	76.47
		26	27	28	30	31	32	33	34	35	36	37	38	43	45	47						
45	SIBTEST	11	24															2	0.00	5.88		
	DFIT	1	2	3	4	5	6	8	10	11	14	15	16	17	18	19	20	21	23	30	43.75	67.65
		24	25	26	31	32	34	35	36	39	40	43	44									
46	SIBTEST	24	25	40														3	6.25	5.88		
	DFIT	1	2	3	4	5	6	7	8	14	15	16	17	18	19	20	21	22	26	30	56.25	61.76
		27	29	31	34	36	37	38	40	45	46	47	48									
47	SIBTEST	3	25	50														3	0.00	8.82		
	DFIT	1	2	3	5	6	8	11	14	15	17	18	19	21	22	25	26	27	28	28	37.50	64.71
		29	31	32	33	35	37	44	46	48	50											
48	SIBTEST	1	11															2	0.00	5.88		
	DFIT	1	2	3	4	6	8	9	11	13	14	18	19	26	27	29	31	32	33	28	37.50	64.71
		34	35	36	37	38	40	42	43	44	49											
49	SIBTEST	25	50															2	0.00	5.88		
	DFIT	1	2	3	4	5	6	8	9	10	11	13	14	16	18	20	21	22	23	32	62.50	64.71
		24	26	27	28	31	33	35	37	38	40	43	45	47	49							

ตาราง 48 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน

กลุ่มที่	วิธีตรวจสอบ ผู้เชี่ยวชาญ	ข้อสอบที่แสดงการหาหน้าทีเบี่ยงเบน (DIF)																		รวม	สรุปผล	
		7	9	13	16	17	21	22	23	35	38	39	40	41	46	48	49	16	P		E	
1	SIBTEST	11	13	19	21	24	25											6	12.50	11.76		
	DFIT	1	2	3	4	5	6	8	11	18	21	24	32	37	38	40	42	46	47	18	25.00	41.18
2	SIBTEST	11	13	19	21	24	25											6	12.50	11.76		
	DFIT	1	2	3	5	6	7	18	22	27	28	31	32	34	37	38	40	42	43	18	25.00	41.18
3	SIBTEST	3	11	19	21	24	25	41										7	12.50	14.71		
	DFIT	1	2	3	4	5	6	8	9	14	15	18	20	21	22	23	26	27	29	26	43.75	55.88
4	SIBTEST	11	25	41														3	6.25	5.88		
	DFIT	1	2	3	5	6	8	9	11	12	14	15	16	17	18	19	20	21	22	33	50.00	73.53
5	SIBTEST	3	11	13	21	24	25											6	12.50	11.76		
	DFIT	1	2	3	4	5	6	7	8	9	12	14	15	17	18	19	20	21	22	34	62.50	70.59
6	SIBTEST	11	21	24	25													4	6.25	8.82		
	DFIT	1	2	3	4	5	6	8	9	11	14	15	16	17	18	19	20	21	22	31	18.75	44.12
7	SIBTEST	3	11	13	25													4	6.25	8.82		
	DFIT	2	3	4	5	6	7	8	11	12	14	15	17	18	19	20	21	22	26	29	37.50	67.65
8	SIBTEST	3	11	13	21	24	25											6	12.50	11.76		
	DFIT	1	2	3	4	5	6	7	8	11	12	14	15	17	18	19	20	21	23	30	56.25	61.76
9	SIBTEST	11	13	21	24	25												5	12.50	8.82		
	DFIT	1	2	3	4	5	6	7	8	11	14	15	17	18	19	20	21	26	27	31	37.50	23.53
10	SIBTEST	3	11	19	21	24	25											6	6.25	14.71		
	DFIT	1	2	3	5	6	7	8	9	12	14	15	17	18	19	20	21	22	26	28	50.00	58.82
11	SIBTEST	11	13	25														3	6.25	5.88		
	DFIT	1	2	3	5	6	7	8	9	14	15	17	18	19	20	21	22	26	31	26	37.50	58.82
12	SIBTEST	1	11	13	21	25												5	12.50	8.82		
	DFIT	1	2	3	4	5	6	7	8	9	10	12	13	14	16	17	18	19	20	40	81.25	79.41
13	SIBTEST	3	11	13	19	21	24	25										7	12.50	4.71		
	DFIT	1	2	3	5	6	7	8	11	12	14	15	16	17	18	19	20	21	22	35	56.25	76.47
14	SIBTEST	1	11	13	21	24	25											6	12.50	11.76		
	DFIT	1	2	3	4	5	6	7	8	11	14	15	16	17	18	20	21	22	26	33	56.25	70.59
15	SIBTEST	11	21	25														3	6.25	5.88		
	DFIT	1	2	3	5	6	7	8	9	13	14	15	17	19	20	21	22	23	26	34	75.00	64.71
16	SIBTEST3	3	11	19	21	24	25											6	6.25	14.71		
	DFIT	1	2	3	4	5	6	8	11	14	15	16	17	19	20	21	22	26	29	28	43.75	61.76

ตาราง 48 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																		รวม	สรุปผล	
	ผู้เชี่ยวชาญ	7	9	13	16	17	21	22	23	35	38	39	40	41	46	48	49	16	P	E		
17	SIBTEST	1	3	11	13	19	21	24	25									8	6.25	20.59		
	DFIT	1	2	3	4	5	6	7	8	11	12	14	15	17	18	19	20	21	22	36	43.75	85.29
		23	25	26	27	28	29	31	32	33	34	35	36	43	44	45	46	47	50			
18	SIBTEST	11	13	21	24	25												5	12.50	8.82		
	DFIT	1	2	3	4	5	6	7	8	9	11	12	14	15	17	18	19	20	21	36	56.25	79.41
		22	26	27	28	29	31	32	34	35	37	38	41	43	44	45	46	47	50			
19	SIBTEST	11	21	24	25													4	6.25	8.82		
	DFIT	1	2	3	5	6	8	11	12	14	15	16	17	18	19	20	21	22	26	31	43.75	70.59
		27	28	29	31	32	33	35	41	43	45	46	47	49								
20	SIBTEST	11	13	21	24	25												5	12.50	8.82		
	DFIT	1	2	3	5	8	17	19	31	32	34	37	38	40	42	43	49	16	25.00	35.29		
21	SIBTEST	3	11	21	25													4	6.25	8.82		
	DFIT	1	2	3	4	5	6	7	8	11	14	15	17	18	19	20	21	22	26	31	43.75	70.59
		27	28	29	31	32	34	35	36	37	38	45	46	50								
22	SIBTEST	11	13	21	24	25												5	6.25	11.76		
	DFIT	1	2	3	4	5	6	7	8	9	14	15	17	19	20	21	22	26	29	26	55.625	50.00
		37	38	40	41	42	43	46	50													
23	SIBTEST	1	11	21	24	25	41											6	12.50	11.76		
	DFIT	1	3	5	6	8	11	12	14	15	17	18	19	20	21	22	26	31	33	25	17.71	55.88
		35	37	38	42	45	46	47														
24	SIBTEST	11	21	24	25													4	6.25	8.82		
	DFIT	1	2	3	4	5	6	8	11	14	15	17	18	19	20	21	22	26	27	32	43.75	73.53
		28	29	31	32	35	36	37	38	41	43	45	46	47	50							
25	SIBTEST	11	25															2	0.00	5.88		
	DFIT	1	2	3	4	5	6	8	14	17	18	19	20	21	22	26	27	31	33	25	31.25	58.82
		37	38	44	45	46	47	50														
26	SIBTEST	11	13	21	24	25												5	6.25	11.76		
	DFIT	1	2	9	11	14	18	22	31	32	37	38	40	42	43			14	25.00	29.41		
27	SIBTEST	11	13	19	21	24	25	31										7	12.50	14.71		
	DFIT	1	2	3	4	5	6	7	8	12	14	15	18	19	21	22	26	31	32	24	37.50	52.94
		37	38	40	42	46	50															
28	SIBTEST	11	13	21	25	31												5	12.50	8.82		
	DFIT	1	2	3	4	5	6	7	8	12	14	15	16	17	18	19	21	26	29	25	50.00	50.00
		32	37	38	40	43	46	48														
29	SIBTEST	1	11	21	25													4	6.25	8.82		
	DFIT	1	2	3	4	5	6	7	8	14	15	17	18	19	20	21	22	26	37	23	37.50	50.00
		38	42	45	46	50																
30	SIBTEST	11	13	21	25	41												5	18.75	5.88		
	DFIT	1	2	3	5	6	8	9	11	12	14	15	16	17	18	19	20	21	22	31	50.00	67.65
		26	27	28	31	32	34	35	37	38	43	45	46	47								
31	SIBTEST	11	13	21	24	25												5	12.50	8.82		
	DFIT	1	2	3	5	6	8	11	14	15	17	18	19	20	21	22	26	27	28	24	31.25	55.88
		31	35	43	45	46	47															
32	SIBTEST	11	19	21	24	25												5	6.25	11.76		
	DFIT	1	2	3	4	5	6	8	11	14	18	19	20	21	22	26	27	28	35	24	31.25	55.88
		37	38	42	43	46	47															
33	SIBTEST	11	21	24	25	50												5	6.25	11.76		
	DFIT	1	2	3	5	6	8	9	11	14	15	16	17	18	19	20	21	22	26	29	43.75	64.71
		27	31	33	34	35	36	37	43	45	47	48										

ตาราง 48 (ต่อ)

กลุ่มที่	วิธีตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบน (DIF)																		รวม	สรุปผล	
	ผู้เชี่ยวชาญ	7	9	13	16	17	21	22	23	35	38	39	40	41	46	48	49	16	P	E		
34	SIBTEST	3	11	13	24	25												5	6.25	11.76		
	DFIT	1	2	3	4	5	6	7	8	11	14	15	17	18	19	20	21	22	26	34	43.75	79.41
		27	28	31	32	33	34	35	36	37	38	40	42	43	45	47	50					
35	SIBTEST	11	13	21	24	25												5	12.50	8.82		
	DFIT	1	2	3	4	5	6	8	13	14	15	17	18	19	20	21	22	26	27	27	31.25	64.71
		28	31	32	33	38	45	46	47	50												
36	SIBTEST	3	11	13	24	25												5	6.25	11.76		
	DFIT	1	2	3	4	5	6	7	8	9	11	14	15	17	19	20	21	22	26	29	43.75	64.71
		27	28	29	32	37	38	42	43	46	47	50										
37	SIBTEST	3	11	21	24	25												5	6.25	11.76		
	DFIT	1	2	3	4	5	6	9	15	16	17	19	20	21	22	23	26	27	35	22	37.50	47.06
		36	37	42	46																	
38	SIBTEST	11	21	25														3	6.25	5.88		
	DFIT	1	2	3	4	5	6	8	11	14	15	16	17	18	19	20	21	22	26	29	37.50	64.65
		27	28	31	32	34	37	38	42	45	46	47										
39	SIBTEST	11	21	24	25													4	6.25	11.76		
	DFIT	1	2	3	5	6	7	8	11	14	15	17	18	19	20	21	22	26	27	27	37.50	61.76
		28	31	32	35	36	37	45	46	47												
40	SIBTEST	11	21	24	25													4	6.25	8.82		
	DFIT	1	2	3	5	6	8	9	11	12	14	15	17	19	20	21	22	26	27	29	50.00	61.76
		28	31	32	35	37	38	39	40	43	45	46										
41	SIBTEST	3	11	25														3	0.00	8.82		
	DFIT	1	2	3	4	5	6	7	8	11	14	15	16	17	18	19	20	21	22	35	62.50	73.53
		26	27	28	29	31	32	35	36	37	38	39	40	45	46	47	48	50				
42	SIBTEST	3	11	19	21	25	41											6	12.50	11.76		
	DFIT	1	2	3	5	6	7	8	11	12	14	15	17	18	19	20	22	26	28	30	43.75	67.65
		29	31	32	33	34	35	38	44	45	46	47	48									
43	SIBTEST	11	13	21	24	25												5	6.25	11.76		
	DFIT	1	2	3	4	5	6	7	8	11	14	15	17	18	19	20	21	22	23	32	37.50	76.47
		26	27	28	29	31	32	34	35	36	37	38	42	45	47							
44	SIBTEST	11	21	25														3	6.25	5.88		
	DFIT	1	2	3	4	5	6	7	8	11	14	15	16	17	18	19	20	21	22	31	50.00	67.65
		26	27	28	31	35	36	37	38	43	45	46	47	48								
45	SIBTEST	11	13	21	25													4	6.25	8.82		
	DFIT	1	2	3	5	6	8	11	13	14	15	17	19	20	21	23	26	27	29	30	37.50	70.59
		31	32	34	35	36	37	38	40	42	43	45	47									
46	SIBTEST	11	21	25	41													4	12.50	5.88		
	DFIT	1	2	3	4	5	6	7	8	11	14	15	17	18	19	20	21	22	26	27	31.25	61.76
		27	28	29	31	32	37	45	46	47												
47	SIBTEST	11	21	24	25	50												5	6.25	11.76		
	DFIT	1	2	3	4	5	6	8	11	14	15	16	17	18	19	20	21	22	23	31	50.00	67.65
		26	27	28	31	35	36	37	38	43	45	46	47	50								
48	SIBTEST	11	13	24	25													4	6.25	8.82		
	DFIT	1	2	3	4	5	6	8	9	14	15	17	18	19	20	21	26	27	28	23	25.00	55.88
		29	31	33	40	42																
49	SIBTEST	3	11	13	21	24	25											6	12.50	11.76		
	DFIT	1	2	3	5	6	8	11	12	14	15	17	18	19	20	21	22	25	26	30	31.25	73.53
		27	28	29	31	32	33	35	36	43	44	45	46									

ภาคผนวก ญ

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบด้วยวิธี SIBTEST และ DFIT ในเงื่อนไข
ความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดต่าง ๆ

ตาราง 49 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ (DBF) ด้วยวิธี SIBTEST และ DFIT
เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน

ขนาด กลุ่มตัวอย่าง	ความยาวแบบทดสอบ (30 ข้อ)	วิธีการ	หมวดข้อสอบ						
			1	2	3	4	5	6	7
50		ผู้เชี่ยวชาญ	0	0	0	1	0	1	0
		SIBTEST	0	0	0	0	0	0	0
100		DFIT	0	1	0	1	1	1	0
		SIBTEST	0	0	0	0	0	0	0
200		DFIT	1	0	0	1	0	0	0
		SIBTEST	1	0	0	0	0	0	0
500		DFIT	1	1	0	0	0	1	1
		SIBTEST	0	0	0	0	0	0	0
1,000		DFIT	0	0	0	0	1	0	0
		SIBTEST	0	0	0	1	0	0	0
		DFIT	1	1	1	0	1	0	1
		SIBTEST	1	1	1	0	1	0	1

ตาราง 50 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ (DBF) ด้วยวิธี SIBTEST และ DFIT
เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน

ขนาด กลุ่มตัวอย่าง	ความยาวแบบทดสอบ (40 ข้อ)	วิธีการ	หมวดข้อสอบ						
			1	2	3	4	5	6	7
50		ผู้เชี่ยวชาญ	0	0	0	1	0	1	0
		SIBTEST	0	0	0	0	0	0	0
100		DFIT	0	1	0	0	0	1	1
		SIBTEST	0	0	0	0	0	0	0
200		DFIT	1	1	1	1	1	1	0
		SIBTEST	1	0	0	0	0	0	0
500		DFIT	1	1	1	1	1	1	1
		SIBTEST	0	0	0	1	0	0	0
1,000		DFIT	1	1	1	1	1	1	1
		SIBTEST	0	0	0	1	0	0	0
		DFIT	1	1	1	1	*****	1	1
		SIBTEST	1	1	1	1	*****	1	1

ตาราง 51 ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของหมวดข้อสอบ (DBF) ด้วยวิธี SIBTEST และ DFIT
เมื่อแบบทดสอบประกอบด้วยข้อสอบ 50 ข้อ กลุ่มตัวอย่างขนาด 50 100 200 500 และ 1,000 คน

ขนาด กลุ่มตัวอย่าง	ความยาวแบบทดสอบ (50 ข้อ)	วิธีการ	หมวดข้อสอบ						
			1	2	3	4	5	6	7
50		ผู้เชี่ยวชาญ	0	0	0	1	0	1	0
		SIBTEST	0	0	0	0	0	1	0
		DFIT	1	1	1	1	0	1	1
100		SIBTEST	0	0	0	0	0	0	0
		DFIT	1	1	0	1	1	1	0
200		SIBTEST	1	0	0	1	0	0	0
		DFIT	1	1	0	1	1	1	1
500		SIBTEST	1	0	0	1	0	0	0
		DFIT	1	1	0	1	*****	1	1
1,000		SIBTEST	0	0	0	1	0	0	0
		DFIT	1	0	0	1	0	1	1

ประวัติย่อผู้วิจัย

ชื่อ-ชื่อสกุล	นางสาวสิริรัตน์ วิภาสศิลป์
วัน เดือน ปีเกิด	10 มิถุนายน 2498
สถานที่อยู่ในปัจจุบัน	18/158 ถนนสุขาประชาสรร 2 ตำบลบางพูด อำเภอปากเกร็ด จังหวัดนนทบุรี 11120
ตำแหน่งหน้าที่การงานในปัจจุบัน	ผู้ช่วยศาสตราจารย์ ระดับ 8
สถานที่ทำงานปัจจุบัน	ศูนย์วิจัยและพัฒนาแบบทดสอบ สำนักทะเบียนและวัดผล มหาวิทยาลัยสุโขทัยธรรมาราช ถนนแจ้งวัฒนะ อำเภอปากเกร็ด จังหวัดนนทบุรี 11120
ประวัติการศึกษา	
พ.ศ. 2516	มัธยมศึกษาปีที่ 5 จากโรงเรียนสตรีศรีสุริโยทัย
พ.ศ. 2521	กศ.บ. (คณิตศาสตร์) จากมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร
พ.ศ. 2525	ค.ม. (การวัดและประเมินผลการศึกษา) จากจุฬาลงกรณ์มหาวิทยาลัย
พ.ศ. 2545	กศ.ด. (การทดสอบและวัดผลการศึกษา) จากมหาวิทยาลัยศรีนครินทรวิโรฒ