

วรรณคดีที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการศึกษาความลำเอียงของข้อกระทงและแบบทดสอบ ตามที่ผู้วิจัยได้ศึกษามาพบว่า ส่วนใหญ่จะเป็นงานวิจัยที่ใช้วิธีการวิเคราะห์ความลำเอียงโดยไม่มีเกณฑ์ภายนอก (Absence of Criterion) เป็นเครื่องตัดสินความลำเอียงของแบบทดสอบ ส่วนการวิเคราะห์ความลำเอียงของแบบทดสอบโดยมีเกณฑ์ภายนอกนั้น มีปัญหาหลักของการวิเคราะห์คือการไม่สามารถหาเกณฑ์ภายนอกที่มีความตรงตามทำนาย มีความยุติธรรม เทียบตรงได้เท่าที่ควร ทำให้ไม่ค่อยมีการศึกษาความลำเอียงโดยใช้เกณฑ์ภายนอกเท่าที่ควร ผู้วิจัยจึงได้นำผลงานวิจัยที่ใช้วิธีการวิเคราะห์ความลำเอียงโดยไม่มีเกณฑ์ภายนอกเท่าที่ศึกษามาได้มารายงานดังนี้

ซัสซีย์ เผ่าพงษ์ (2526) ได้ศึกษาความลำเอียงของแบบทดสอบมาตรฐานวัดความถนัดทางการเรียนด้านคณิตศาสตร์และภาษาในระดับมัธยมต้นจำนวน 2 ฉบับซึ่งพัฒนาโดยสำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร ระหว่างกลุ่มผู้สอบเพศชายและเพศหญิง โดยใช้วิธีวิเคราะห์ด้วยโครงสร้างของข้อกระทงที่คำนวณค่าพารามิเตอร์ 3 ตัว เขามีความคิดว่าแบบทดสอบวัดความถนัดทางการเรียน ที่บอกว่าได้รับการพัฒนาเป็นแบบทดสอบมาตรฐานแล้วนั้น มุ่งที่จะตรวจสอบคุณภาพของแบบทดสอบว่าวัดในสิ่งที่ต้องการวัดได้หรือไม่ ด้วยการศึกษาเฉพาะค่าความเที่ยงของแบบทดสอบเท่านั้น ไม่มีการตรวจสอบเกี่ยวกับโครงสร้างของพฤติกรรมที่ต้องการวัดเมื่อแบ่งกลุ่มผู้สอบออกเป็นกลุ่มซึ่งมีความแตกต่างกันว่าเนื้อหาของข้อกระทงเหล่านั้นยังสามารถวัดได้พฤติกรรมที่ตรงกันหรือไม่ เพราะว่าแบบทดสอบความถนัดหรือผลสัมฤทธิ์ทางการเรียน มักจะมีข้อตกลงว่าจำนวนข้อกระทงทั้งฉบับมีข้อกระทงที่ผู้สอบต่างก็มีความเคยชินและไม่เคยชินกับข้อกระทงอยู่เท่า ๆ กัน ซึ่งในความเป็นจริงแล้วมีสภาพการณ์หลายอย่าง เช่น เพศ สภาพทางค่านิยมศาสตร์และอื่น ๆ ที่มีอิทธิพลบางอย่างทางค่านิยมธุรกิจและสังคม ทำให้ผู้สอบมีความแตกต่างกันในค่านิยมการประเมิน จึงเกิดการเสียเปรียบทางค่านิยมเนื้อหาของข้อกระทงแตกต่างกัน เมื่อกลุ่มหนึ่งเรียนรู้เนื้อหาที่ถูกทดสอบมากกว่าอีกกลุ่มหนึ่ง โครงสร้างของพฤติกรรมทางเนื้อหา

ที่เคยผ่านมาก็มีความแตกต่างกัน ทำให้กลุ่มที่มีประสบการณ์มากกว่าสามารถทำข้อกระทงได้มากกว่ากลุ่มอื่น ๆ โดยแสดงความสามารถของพฤติกรรมเพียงแต่ความรู้ความจำเท่านั้น ไม่ใช่ความถนัดหรือผลสัมฤทธิ์ที่ต้องการวัดที่แท้จริง ในงานวิจัยของเขามุ่งที่จะศึกษากับกลุ่มที่มีความแตกต่างกันทางด้านเพศ เนื่องจากเขาได้ศึกษาพบว่า เพศชายและหญิงมีการเรียนรู้และพัฒนาที่แตกต่างกัน โดยเฉพาะความสามารถทางคณิตศาสตร์และภาษา เพราะข้อกระทงวิชาคณิตศาสตร์จำเป็นต้องอาศัยความสามารถด้านการอ่านใจหยาบปัญหาซึ่งกลุ่มเพศชายและหญิงมีความสามารถทางการอ่านไม่เท่ากันหรือในข้อกระทงที่ใช้วัดเกี่ยวกับภาษา เพศชายและหญิงมีการได้เปรียบเสียเปรียบเกี่ยวกับภาษาที่ใช้แตกต่างกัน กลุ่มตัวอย่างที่ใช้เป็นกลุ่มนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2524 เพศชายและหญิงซึ่งประกอบด้วยทุกภาคภูมิศาสตร์ทั่วประเทศ 5 ภาค คือ ภาคเหนือ ภาคกลาง ภาคใต้ ภาคตะวันออก และภาคตะวันออกเฉียงเหนือ เป็นการควบคุมตัวแปรทางสภาพภูมิศาสตร์ซึ่งอาจจะเป็นตัวแปรที่จะทำให้เกิดความลำเอียงได้ ได้ใช้วิธีวิเคราะห์ความลำเอียงของข้อกระทงด้วยโด่งลักษณะของข้อกระทงที่คำนวณค่าพารามิเตอร์ 3 ตัว ที่ให้ผลการวิเคราะห์คงที่ และมีความสัมพันธ์สอดคล้องร่วมกับวิธีอื่น เห็นผลการวิเคราะห์ชัดเจนและเข้าใจเกี่ยวกับความลำเอียงของข้อกระทงมากกว่าวิธีอื่น ผลการวิจัยพบว่าแบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์มีข้อกระทงที่ลำเอียงต่อกลุ่มนักเรียนชายและหญิงในระดับปานกลางขึ้นไป จำนวน 5 ข้อ ซึ่งวัดในเรื่อง ร้อยละ การหาปริมาตรและการหาความยาวของเส้นในค่านของรูปสามเหลี่ยมจำนวนเรื่องละ 1 ข้อ และเรื่องโจทย์ปัญหาทางคณิตศาสตร์จำนวน 2 ข้อจากจำนวน 5 ข้อ เป็นข้อที่มีความลำเอียงต่อนักเรียนชายโดยเฉพาะ 4 ข้อ และเป็นข้อกระทงที่มีความลำเอียงต่อกลุ่มนักเรียนหญิงในช่วงความสามารถแรกและลำเอียงต่อกลุ่มนักเรียนชายในช่วงความสามารถต่อมาจำนวน 1 ข้อ ส่วนแบบทดสอบวัดความถนัดทางการเรียนด้านภาษาเกี่ยวกับการอ่านเข้าใจจำนวน 40 ข้อ ข้อกระทงที่มีความลำเอียงในระดับปานกลางขึ้นไป มีจำนวน 9 ข้อ ซึ่งวัดความเข้าใจเกี่ยวกับการอ่านคำประพันธ์ บท ร้อยกรองอย่างละ 1 ข้อ และวัดความเข้าใจเกี่ยวกับการอ่านข้อความจำนวน 7 ข้อ จากจำนวนข้อกระทงที่ลำเอียง 9 ข้อนี้ เป็นข้อที่มีความลำเอียงต่อกลุ่มนักเรียนชายโดยเฉพาะจำนวน 1 ข้อ และจำนวน 6 ข้อที่มีความลำเอียงต่อกลุ่มนักเรียนหญิงโดยเฉพาะ และมี 2 ข้อ ซึ่งมีความลำเอียงช่วงความสามารถแรกต่อนักเรียนชาย ช่วงความสามารถต่อมาลำเอียงต่อนักเรียนหญิง จากผลการวิจัยเขาได้สรุปว่า แบบทดสอบมาตรฐาน 2 ฉบับนี้

ข้อควรระวังส่วนใหญ่มีความลำเอียงทางการทดสอบ และมีข้อจำกัดเรื่องการเปรียบเทียบผลการสอบระหว่างกลุ่มนักเรียนชายและหญิง เพราะผลการสอบที่วัดได้ส่วนใหญ่มาจากความสามารถต่างชนิดกัน ไม่สามารถชี้ให้เห็นถึงความแตกต่างเกี่ยวกับระดับความสามารถในด้านเดียวกันได้อย่างชัดเจน ฉะนั้น ในการเปรียบเทียบผลการสอบระหว่างกลุ่มโดยตรงจึงอาจบอกได้เพียงแต่ว่าผลสอบของกลุ่มหนึ่งสามารถตอบข้อกระทงได้ถูกมากกว่าหรือน้อยกว่าอีกกลุ่มหนึ่ง เขาเสนอว่าในการที่จะนำข้อกระทงทั้ง 2 ฉบับนี้ไปใช้กับกลุ่มผู้สอบเพศชายและหญิงควรพิจารณาพารามิเตอร์ คือ ให้พิจารณาข้อที่มีค่าอำนาจจำแนกของทั้ง 2 กลุ่มไม่มีความแตกต่างกันมาก ไม่ว่าข้อนั้นจะมีค่าอำนาจจำแนกสูงหรือต่ำก็ตาม

Sinnott (1980, quoted in Carton & Marco, 1982: 287)

ได้ทำการวิเคราะห์ข้อมูลที่ได้จากแบบทดสอบ GMAT เพื่อจะแยกปัจจัยที่ทำให้ข้อกระทงบางข้อง่ายหรือยากเกินไปสำหรับกลุ่มผู้สอบกลุ่มน้อย โดยใช้วิธีการกำหนดจุดค่าเฉลี่ย (Delta-Plot Method) แบบทดสอบ GMAT ที่ใช้ศึกษาทำการทดสอบในเดือนมกราคม ปี 1978 ประกอบด้วย 5 ส่วน คือ Reading Comprehension, Problem Solving, Practical Business Judgement, Data Sufficiency และ Usage เธอมุ่งที่จะศึกษาตัวแปรทางด้านสังคมของผู้สอบที่จะเป็นตัวทำให้เกิดความลำเอียง (Sociological Dimension) คือ เพศ, เชื้อชาติ (Caucasion, Oriental, Hispanic, Blacks) เชื้อชาติและเพศ (เชื้อชาติดังกล่าวที่แบ่งเพศหญิงและชาย), อายุ (ต่ำกว่า 20, 20-22, 23-25, 26-29, 30-34, 35-39 และ 40-65) และ ความคล่องทางภาษา (U.S. Citizen, Chinese, English, French, Indo-Iranian, Japanese and Spanish) ขนาดกลุ่มตัวอย่างกลุ่มละ 400 คน ยกเว้น เชื้อชาติ Hispanic ที่มีจำนวนแค่ 309 คน เธอได้ทำการพัฒนาเกณฑ์ตัดสินความลำเอียงโดยการหาการกระจายของระยะห่างจากจุดค่าโคออดิเนตของเคลต้า 2 กลุ่มที่เปรียบเทียบถึงเส้นแกนหลัก ระหว่างกลุ่มตัวอย่างเดียวกันที่ถือว่าไม่ลำเอียง จุดตัดที่ใช้คือ 0.7 หลังจากพบข้อลำเอียง เธอได้อภิปรายผลโดยมุ่งไปที่การวิเคราะห์เนื้อหาของข้อที่ลำเอียง

Stricker (1981 quoted in Carton & Marco, 1982: 289)

ได้ศึกษาเปรียบเทียบดัชนีการตัดสิน ความลำเอียงในวิธีวิเคราะห์ความลำเอียง 3 วิธี คือ ระยะห่างซึ่งคำนวณโดยวิธีการกำหนดจุดค่าเฉลี่ย ดัชนีไคสแควร์ที่ใช้ตัดสินความแตกต่าง

ของค่าพารามิเตอร์ a, b จากวิธีตอบสนองข้อกระทง 3 พารามิเตอร์และใช้การหา Partial Correlation ของคะแนนข้อกระทงและกลุ่ม ด้วยคะแนนรวมทั้งหมคเป็นค่าคงที่โดยใช้แบบทดสอบ GRE Aptitude Test ซึ่งทดสอบในเดือนตุลาคม ปี 1977 และเดือนมกราคม ปี 1978 โดยศึกษาตัวแปรเชื้อชาติ (ขาว, ดำ) และเพศ กลุ่มตัวอย่างประกอบด้วยกลุ่มผู้สอบชายผิวขาว (1,122 คน), หญิงผิวขาว (1,471 คน), ชายผิวดำ (284 คน) และหญิงผิวดำ (626 คน) โดยจับคู่วิเคราะห์ระหว่างกลุ่มหญิงผิวดำและชายกับกลุ่มผู้สอบชายและหญิงผิวขาว การวิเคราะห์ข้อมูลจะวิเคราะห์แยกระหว่างผู้ชายและผู้หญิงในการวิเคราะห์ทางด้านเชื้อชาติ และวิเคราะห์แยกผู้สอบผิวดำและขาว ในการวิเคราะห์เพศเพื่อควบคุมตัวแปรทั้ง 3 วิธี ทดสอบที่ระดับนัยสำคัญ 0.05 และใช้สัมประสิทธิ์เคปป่า (Kappa Coefficient) และสถิติไคสแควร์ในการพิจารณาความสอดคล้องของทั้ง 3 วิธี เขาพบว่าดัชนีที่ใช้สถิติไคสแควร์ตัดสินความแตกต่างระหว่างโครงสร้างของข้อกระทงได้เปรียบวิธีที่ต้องตั้งเกณฑ์ เนื่องจากมีความคงที่และไม่เปลี่ยนแปลงตามเกณฑ์

Alderman และ Holland (1981 quoted in Carton and Marco 1982: 298) ได้ทำการวิจัยแบบทดสอบ TOFEL โดยมุ่งศึกษาตัวแปรทางด้านกลุ่มภาษา 7 กลุ่ม (Germanic, Spanish, African, Chinese I, Chinese II, Japanese, Arabic) ขนาดกลุ่มตัวอย่าง 500 ถึง 1,000 คน กลุ่มภาษาจีน 2 กลุ่มใช้ในการตั้งเกณฑ์ในการประเมินผลความล่าเอียง ใช้วิธีวิเคราะห์โดย Log-Linear Model โดยใช้คู่กับวิธีการกำหนดจุดค่าเคลต้า (Delta-Plot Method) ซึ่งเขาใช้ในการตรวจสอบผลและการแปลความในแต่ละกลุ่มเขาได้จำแนกผลการตอบเป็นตารางการเฉลจร เป็นกลุ่ม \times ระดับคะแนน 9 ช่วง \times ผลการตอบ วิธีของ Log-Linear Model คือ การหาโมเดลที่เหมาะสมกับข้อมูลเพื่อตรวจสอบว่าโอกาสในการตอบถูกของข้อกระทงขึ้นอยู่กับระดับคะแนนและกลุ่มหรือไม่ โดยใช้สถิติไคสแควร์มาทดสอบความเหมาะสมของโมเดลกับข้อมูล ผลการวิจัยพบว่าในแบบทดสอบ TOFEL ชุดหนึ่งที่พบว่าไคสแควร์สูงมากในแต่ละส่วน ซึ่งแสดงว่าข้อกระทงมีการเบี่ยงเบนซึ่งต้องได้รับการนำไปพิจารณาอย่างลึกซึ้งโดยผู้เชี่ยวชาญทางภาษาและทำการวิเคราะห์ซ้ำเพื่อยืนยันอีกครั้งหนึ่ง

การศึกษาเปรียบเทียบเทคนิคทางสถิติที่ใช้วิเคราะห์ความลำเอียงของข้อกระทง โดยสร้างข้อมูลเชิงทฤษฎี มีผู้ศึกษา เช่น Rudner, Getson และ Knight (1979 quoted in Burrill, 1982:162) ได้ทำการศึกษาวิธีตรวจสอบความลำเอียง 5 วิธี คือ TID (Transformed Item Difficulties หรือ Delta-Plot Method), ICC-3 (Three Parameter Logistic Model), ICC-1 (Rasch Model), x^2 ที่แบ่งช่วงความสามารถเป็น 5 ช่วง และ x^2 ที่แบ่งช่วงความสามารถหลายช่วงโดยสร้างข้อกระทงที่มีความลำเอียงในพารามิเตอร์ a และ b โดยมีส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์เหล่านั้นอย่างละ 4 ค่า และใช้แบบทดสอบมีความยาว 4 ค่า คือ 20, 40, 60, 80 ข้อ ใช้ผู้สอบ 2 กลุ่ม 1,200 คน พบว่าทุก ๆ เทคนิค ยกเว้น x^2 ที่แบ่งช่วงความสามารถหลายช่วงเท่านั้นจะแสดงความสัมพันธ์ที่เพิ่มขึ้นระหว่างการตรวจสอบความลำเอียงและการแยกความลำเอียง เมื่อความยาวของแบบทดสอบเพิ่มขึ้นและเพิ่มมากที่สุดเมื่ออยู่ระหว่าง 20 และ 40 ข้อสำหรับเทคนิค ICC-3 รองลงมา คือ x^2 ที่แบ่ง 5 ช่วงความสามารถและต่ำสุด คือ ICC-1

ถ้าความลำเอียงไม่ได้เกิดที่พารามิเตอร์ b นั่นคือ ความแตกต่างของกลุ่มขึ้นอยู่กับค่าอำนาจจำแนกของข้อกระทง (พารามิเตอร์ a) วิธี x^2 (5 ช่วง) จะมีประสิทธิภาพมากที่สุดในการตรวจสอบ รองลงมาคือ ICC-1 และต่ำสุดคือ TID

ถ้าสร้างความลำเอียงในพารามิเตอร์ b เพียงตัวเดียว TID และ x^2 (5 ช่วง) จะมีประสิทธิภาพมากที่สุด วิธี x^2 และ TID จะตรวจสอบได้มากหรือน้อยขึ้นอยู่กับพารามิเตอร์ a

ค่าความสัมพันธ์ระหว่างกันทั้งใน 5 วิธี มีความสัมพันธ์ตามกันสูง ระหว่าง x^2 (5 ช่วง) และ TID และระหว่าง x^2 (5 ช่วง) กับ x^2 (หลายช่วง) ความสัมพันธ์ระหว่าง ICC-1 กับวิธีอื่นค่อนข้างต่ำ

Rudner สรุปผลว่าวิธีตรวจสอบ 3 วิธีที่ให้การประมาณที่ค่อนข้างตรงและแม่นยำ คือ ICC-3, x^2 (5 ช่วง) และ TID (Delta-Plot Method) โดยที่วิธี ICC-3 จะขึ้นอยู่กับค่าพารามิเตอร์ได้เที่ยงตรงหรือไม่ วิธี x^2 จะมีประสิทธิภาพขึ้นกับการแบ่งช่วงการประมาณความสามารถ ส่วนวิธี TID จะไม่ไวต่อความลำเอียงที่เกิดจากการจำแนกของกลุ่มต่างกัน

นอกจากนี้ Laksana (1979 quoted in Burrill, 1982: 168) ได้เปรียบเทียบเทคนิควิธี 2 วิธีในการวิเคราะห์ความลำเอียง คือ วิธี ANOVA และวิธีตอบสนองข้อกระทง 3 พารามิเตอร์ (ICC-3) กับแบบทดสอบ ITBS ชุดการแก้ปัญหาทางคณิตศาสตร์และคำศัพท์ กลุ่มตัวอย่างคือ กลุ่มเพศชายผิวดำ, เพศหญิงผิวดำ, เพศชายผิวขาว และเพศหญิงผิวขาว สำหรับวิธี ANOVA ได้ใช้แบบการวิเคราะห์ 3 องค์ประกอบ คือ ผิว, เพศ และข้อกระทงโดยพิจารณาเป็นผลคงที่ คำนวณความลำเอียงที่ใช้ตัดสินคือความมีนัยสำคัญในผลรวม (Interaction) ระหว่างข้อกระทงกับผิวหรือข้อกระทงกับเพศ ส่วนวิธี ICC-3 ได้ใช้โปรแกรม LOGIST ในการประมาณค่าพารามิเตอร์ และใช้การเปรียบเทียบ ICC ของ 2 กลุ่ม เป็นตัวตัดสินความลำเอียงโดยคำนวณหาพื้นที่ระหว่าง ICC 2 กลุ่ม และได้ตั้งเกณฑ์ความลำเอียงโดยไม่มีแสดงทิศทางว่าข้อที่ลำเอียงจะมีพื้นที่ระหว่าง ICC มากกว่าหรือเท่ากับ 0.30 เขาสรุปว่าวิธี ANOVA ใช้ง่ายแต่ให้ความหมายไม่กว้างนัก วิธี ICC-3 ให้ความหมายกว้างกว่า แต่ปฏิบัติได้ยากกว่าและต้องการการประมาณค่าพารามิเตอร์ที่เที่ยงตรง

จะเห็นว่าความลำเอียงของข้อกระทงจะหมายถึงข้อกระทงที่วัดความสามารถหรือคุณลักษณะทางจิตวิทยาของผู้สอบในแต่ละกลุ่มได้ไม่ตรงกัน (Allen และ Yen อ้างถึงในซีซีซี, 2526:4) และจากงานวิจัยดังกล่าวมาแล้วสรุปได้ว่าวิธีการหาค่าความลำเอียงของข้อกระทงและแบบทดสอบโดยไม่มีเกณฑ์ภายนอกที่นิยมใช้ คือ วิธีที่เปรียบเทียบความแตกต่างระหว่างกลุ่มผู้สอบโดยพิจารณาจาก

1. ความแปรปรวนระหว่างการเป็นสมาชิกในแต่ละกลุ่มกับการตอบข้อกระทง (ANOVA)
2. ค่าความยากของข้อสอบ (TID หรือ Delta-Plot Method)
3. สัดส่วนการตอบข้อกระทงถูกต้องตามระดับคะแนน (Chi-Square Method, χ^2)
4. ความสามารถที่ไม่ขึ้นกับค่าความยากของข้อกระทงกับโอกาสในการตอบข้อกระทงถูก (Rasch Model, ICC-1)
5. ความสามารถที่ไม่ขึ้นกับค่าความยาก ค่าอำนาจจำแนก และค่าการเดาของข้อกระทง กับโอกาสในการตอบข้อกระทงถูก (Three-Parameter Logistic Model, ICC-3)

วิธีการดังกล่าวทั้ง 5 วิธี มีความแตกต่างและเหมือนกันดังที่สรุปไว้ในตารางที่ 1 (Osterlind, 1983: 78-79)

ตารางที่ 1 แสดงความเหมือนและความแตกต่างของวิธีการวิเคราะห์ความลำเอียงโดยไม่มีเกณฑ์ภายนอกที่นิยมใช้

วิธี	สิ่งที่ทำ การทดสอบ	การวิเคราะห์	การตัดสิน ความลำเอียง	การคำนวณ และ การแปลผล	ข้อจำกัด
1. ANOVA	ค่าความชาก	ผลรวมระหว่างการเป็นสมาชิกในกลุ่มกับการตอบข้อกระทงถูก	ความมีนัยสำคัญทางสถิติของ F-test ระหว่างผลหลักและผลรวม	ชาก, ค่อนข้างชาก	1. ต้องใช้เครื่องคอมพิวเตอร์คำนวณ 2. ขนาดกลุ่มตัวอย่างขึ้นอยู่กับจำนวนข้อและผู้สอบต้องมีขนาด 100 คนต่อกลุ่ม
2. TID	ค่าความชาก	ผลรวมระหว่างการเป็นสมาชิกในกลุ่มกับการตอบข้อกระทงถูก	การกำหนดระยะห่างจากจุดที่กำหนดไปเส้นแกนหลักซึ่งไม่มีกฎเกณฑ์ตายตัว	ง่าย, ง่าย	1. ขนาดกลุ่มตัวอย่าง 30 ถึง 50 คนต่อกลุ่ม 2. คลาดเคลื่อนเมื่อข้อกระทงมีค่าอำนาจจำแนกสูง
3. χ^2	ค่าความชาก	ความแตกต่างของสัดส่วนการตอบข้อกระทงถูกตามระดับคะแนนรวม	ความมีนัยสำคัญทางสถิติของ χ^2	ค่อนข้างง่าย, ง่าย	1. ขนาดกลุ่มตัวอย่างจำนวนมากกว่า 100 คนต่อกลุ่ม 2. การแย่งระดับคะแนนรวมเป็นช่วง ซึ่งไม่มีกฎเกณฑ์ตายตัวและการใช้คะแนนรวมแทนค่าความสามารถ
4. ICC-1 (b)	ค่าความชาก	ความแตกต่างของโอกาสในการทำข้อกระทงถูก	พื้นที่ระหว่าง ICC หรือการ Fit ICC กับโมเดล	ชาก, ชาก	1. ขนาดตัวอย่างมีจำนวนมากกว่า 100 คนต่อกลุ่ม 2. ข้อตกลงเบื้องต้นในเรื่องไม่มีการเดาและมีอำนาจจำแนกเท่านั้น 3. ใช้เครื่องคอมพิวเตอร์คำนวณโดยโปรแกรม BICAL
5. ICC-3	ค่าความชาก ค่าอำนาจจำแนกและค่าการเดา (b, a, c)	ความแตกต่างของโอกาสในการทำข้อกระทงถูก	พื้นที่ระหว่าง ICC และอีกหลายวิธี	ชาก, ชาก	1. ขนาดตัวอย่างต้องมากกว่า 1,000 คนต่อกลุ่ม 2. ค่าใช้จ่ายสูง, ใช้เครื่องคอมพิวเตอร์คำนวณโดยโปรแกรม LOGIST 3. ปัญหาการประมาณค่า a และ c 4. การปรับเสถียร 5. สถิติที่ใช้ในการคำนวณต้องใช้ผู้เชี่ยวชาญเรื่องสถิติ

หมายเหตุ ค่าความชากในวิธีที่ 1-3 คือ สัดส่วนของผู้ตอบข้อกระทงถูกกับจำนวนผู้สอบทั้งหมด


จากงานวิจัยที่เปรียบเทียบเทคนิควิธีทางสถิติที่ใช้วิเคราะห์ความลำเอียงของ
ข้อกระทงไม่ว่าจะใช้ข้อมูลที่สร้างขึ้นมาให้มีความลำเอียง แล้วใช้วิธีวิเคราะห์เข้าไปตรวจ
สอบ เช่นงานของRudner หรือการใช้ข้อมูลจริงมาทำการตรวจสอบด้วยเทคนิคหลาย ๆ
เทคนิค เช่นงานของ Intasuwan, 1979; Rudner & Convey, 1978; Shepard,
1980; Laksana, 1979, Ironson & Subkoviad, 1979 etc. ให้ผลสรุปที่
คล้ายคลึงกันในการเลือกใช้เทคนิควิธีวิเคราะห์ 3 วิธีที่มีความสัมพันธ์กันในผลการตรวจ
สอบ และให้การประมาณที่ค่อนข้างตรงและแม่นยำ คือ วิธี ICC-3, χ^2 และ TID
ตารางที่ 2 เป็นการสรุปการเปรียบเทียบเทคนิควิธีวิเคราะห์ความลำเอียงทั้ง 3 วิธี

ตารางที่ 2 : การเปรียบเทียบเทคนิควิธีวิเคราะห์ความลำเอียงของข้อกระทง 3 วิธี คือ
วิธีการกำหนดจุดค่าเคลตา(TID) วิธีการทดสอบความแตกต่างระหว่างกลุ่ม
ด้วยสถิติไคสแควร์ (χ^2) และวิธีการตอบสนองของข้อกระทง 3 พารามิเตอร์
(ICC-3)

วิธีวิเคราะห์ หัวเรื่อง	TID	χ^2	ICC-3
1. ข้อตกลงเบื้องต้น	ผลรวมระหว่างกลุ่มกับ ข้อกระทงเป็นตัวบ่งชี้ ถึงความลำเอียงภายใน ข้อกระทง	คะแนนรวมที่ทำได้จาก แบบทดสอบเป็นตัวบ่งชี้ ของความสามารถของ ผู้สอบ	แบบทดสอบต้องมีมิติ เดียวและโค้งลักษณะ ของข้อกระทงสามารถ แสดงฟังก์ชันของค่า ความสามารถและโอกาส ในการทำข้อกระทงข้อ นั้นถูกต้อง
2. ตัวที่ทำการตรวจ สอบ	ค่าความยาก	ค่าความยากโดยควม คุมการจำแนกของกลุ่ม	ค่าความยาก, ค่าอำนาจ จำแนก, ค่าการเคา
3. จุดที่ทำการวิ- เคราะห์	ผลรวมระหว่างการ เป็นสมาชิกของกลุ่ม กับการตอบข้อกระทง ถูก	ความแตกต่างในสัดส่วน ของการตอบข้อ กระทงถูกต้องตามการจำ- แนกของคะแนนรวม	ความแตกต่างของโอกาส ในการตอบข้อกระทงถูก

วิธีวิเคราะห์ หัวข้อเรื่อง	TID	χ^2	ICC-3
4. เกณฑ์การตัดสินใจ ความลำเอียง	เกณฑ์ที่สร้างขึ้นไม่ตาย ตัวเพื่อตัดสินระยะห่าง ของจุดค่าเฉลี่ย	การทดสอบสารูปสนิห ด้วยความมีนัยสำคัญ ทางสถิติของสถิติ ไคสแควร์	เกณฑ์ที่สร้างขึ้นไม่ตายตัว เพื่อตัดสินพื้นที่ระหว่าง ICC
5. ข้อจำกัด	1. คลาดเคลื่อนเมื่อ ข้อกระทงมีค่า อำนาจจำแนกสูง 2. ใช้กับกลุ่มตัวอย่าง ขนาดเล็ก 3. ค่าความยากแปร เปลี่ยนตามกลุ่ม ผู้สอบ	1. ไม่มีกฎเกณฑ์ตาย ตัวในการแบ่ง ระดับคะแนน 2. การใช้คะแนนรวม เป็นตัวแทนของ ความสามารถ 3. ไวต่อการกระจาย ของคะแนนและ ความแตกต่างของ ขนาดกลุ่มตัวอย่าง 4. ให้รายละเอียดของ ความลำเอียงน้อย 5. ค่าความยากเปลี่ยน ตามผู้สอบ	1. จำกัดขนาดตัวอย่าง ต้องมาก 2. การประมาณค่า พารามิเตอร์ a และ c มีจุดอ่อนใน LOGIST 3. ปัญหาการปรับสเกล 4. ค่าใช้จ่าย, เวลาใน การใช้เครื่อง คอมพิวเตอร์วิเคราะห์ 5. ใช้ความรู้ในการ ดำเนินการและแปล ผลได้ยาก 6. สถิติที่ใช้ไม่สามารถ หาได้โดยทั่ว ๆ ไป
6. ข้อดี	1. คำนวณง่าย 2. แปลผลและแสดงให้ผู อื่นที่ไม่มีพื้นฐานทาง สถิติเข้าใจได้ง่าย	1. คำนวณค่อนข้างง่าย กว่า ICC-3 2. มีเกณฑ์ตายตัวใน การแปลผล	1. ให้รายละเอียดของ ความลำเอียงมาก 2. ตรงตามหลักการ ทฤษฎี

จะเห็นว่าทั้ง 3 เทคนิควิธีมีบางส่วนที่สัมพันธ์กันและมีบางส่วนที่แตกต่างกัน
วิธีที่ถูกต้องตามทฤษฎีเป็นวิธีที่มีข้อจำกัดในทางปฏิบัติสูง โดยเฉพาะอย่างยิ่งในการ
คำนวณและแปลความผลการวิจัยค่อนข้างยาก ในขณะที่วิธีที่ง่ายต่อการคำนวณและการ
แปลผลการวิเคราะห์ก็มีข้อจำกัดในหลักการทฤษฎี และจากรายงานที่ทำการศึกษาเปรียบเทียบ
เทียบวิธีทั้ง 3 วิธีที่ได้กล่าวมาแล้วว่าผลการวิเคราะห์มีความสัมพันธ์กันและมีความคงที่
ค่อนข้างตรง (Rudner, Getson and Knight quoted in Burrill, 1982:163)
ทำให้ผู้วิจัยใช้ทั้ง 3 เทคนิควิธีมาวิเคราะห์ความลำเอียงของแบบทดสอบคณิตศาสตร์ของ
โครงการตรวจสอบคุณภาพทางการศึกษานักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย