

เชปอาร์ค และคนอื่น ๆ (Shepard and Others, 1985 : 81) กล่าวว่า เมื่อพบว่าโค้งลักษณะข้อสอบของกลุ่มอ้างอิง และกลุ่มเปรียบเทียบแตกต่างกันมากพอที่จะเป็นหลักฐานได้ว่าข้อสอบที่ศึกษาอยู่ไม่ได้วัดลักษณะแฝงเดียวกันสำหรับกลุ่มผู้สอบทั้งสองแล้ว จะใช้ดัชนีที่บ่งชี้ถึงระดับลำเอียง ดังนี้

1. พื้นที่ชนิดไม่มีเครื่องหมาย (Unsigned Area) เป็นค่าสัมบูรณ์ของพื้นที่ระหว่างโค้งลักษณะข้อสอบทั้งสอง
2. ผลบวกกำลังสองแบบที่ 1 (SOS 1) เป็นค่าผลบวกของความแตกต่างระหว่างกำลังสองทุกค่าของความสามารถที่เป็นไปได้
3. ผลบวกกำลังสองแบบที่ 2 (SOS 2) คล้าย SOS 1 แต่ให้นักหนักความแตกต่างกำลังสอง ด้วยส่วนกลับของความคลาดเคลื่อนของความแปรปรวนในการประมาณค่าความเป็นไปได้ ดังนั้น ถ้ามีการประมาณโค้งอย่างไม่ดีในส่วนตัวส่วนหนึ่งแล้ว จะทำให้ลดประโยชน์ต่อความแตกต่างระหว่างสองกลุ่ม
4. ไคสแควร์ ($IRT - \chi^2$) เป็นการทดสอบนัยสำคัญทางสถิติ โดยการเปรียบเทียบความแตกต่างของค่าความยาก และค่าอำนาจจำแนก ระหว่างกลุ่มพร้อม ๆ กัน โดย ลอร์ด (Lord, 1980)
5. พื้นที่ชนิดมีเครื่องหมาย เหมือนพื้นที่ชนิดไม่มีเครื่องหมายแต่มีการให้เครื่องหมายไว้เพื่อแสดงให้เห็นว่ากลุ่มใดเสียประโยชน์ และกลุ่มใดได้ประโยชน์ ถ้าโค้งลักษณะการตอบข้อสอบตัดกัน แสดงว่า มีความลำเอียงต่างกันสำหรับผู้มีความสามารถต่างกัน และค่าสัมบูรณ์ของพื้นที่ชนิดมีเครื่องหมายจะน้อยกว่าชนิดไม่มีเครื่องหมาย
6. ผลบวกกำลังสองแบบที่ 3 (SOS 3) เป็นผลบวกกำลังสองของความแตกต่างชนิดมีเครื่องหมายของ SOS 1 โดยการใช้เครื่องหมายของความแตกต่างแต่ละตัวที่ยังไม่ยกกำลังสอง
7. ผลบวกกำลังสองแบบที่ 4 (SOS 4) เป็นดัชนีคู่ขนานกับ SOS 2 แต่เป็นผลบวกที่มีการให้น้ำหนักและให้เครื่องหมาย

การวัดพื้นที่ของความแตกต่างระหว่างโค้งลักษณะข้อสอบของกลุ่มตัวอย่างเมื่อพิจารณาถึงขนาดของความลำเอียงของข้อสอบนั้นมีอยู่ 2 ลักษณะ คือ การวัดแบบช่วงเปิดและแบบช่วงปิด (Open and Closed Intervals) กรณีของการวัดแบบช่วงปิดจะกำหนดขอบเขตของการวัดให้อยู่ในช่วงความสามารถที่กำหนด ในขณะที่การวัดแบบช่วงเปิดจะวัดภายในช่วงความสามารถทั้งหมด เพื่อให้ได้พื้นที่ที่แน่นอน (Exact Area) ระหว่างโค้งลักษณะข้อสอบ 2 โค้ง ภายใต้การวัดแบบช่วงปิดนั้นสามารถวัดพื้นที่ได้ 2 ลักษณะ คือ พื้นที่ชนิดมีเครื่องหมายและไม่มีเครื่องหมาย (Signed and Unsigned Areas) โดยมีสูตรการคำนวณทั่ว ๆ ไป ดังนี้

จากฟังก์ชันการตอบข้อสอบแบบ 3 พารามิเตอร์

$$P(\theta) = c + (1-c)P^*(\theta)$$

$$\text{โดยที่ } P^*(\theta) = (1 + e[Da(\theta - b)])^{-1}$$

เมื่อ a, b และ c เป็นพารามิเตอร์ที่แสดงลักษณะของข้อสอบ และ D เป็นค่าคงที่ กำหนดให้ มีค่าเท่ากับ 1.7

พื้นที่ระหว่างความสามารถที่กำหนดช่วงไว้ 2 จุดบนสเกลความสามารถคำนวณได้จาก

$$S(\theta_1, \theta_2) = \int_{\theta_1}^{\theta_2} P(\theta) d\theta = c(\theta_2 - \theta_1) + (1-c)(Da)^{-1} \ln \frac{1 + e[Da(\theta_2 - b)]}{1 + e[Da(\theta_1 - b)]}$$

ในการศึกษาความลำเอียงของข้อสอบนั้น จะมีพารามิเตอร์ของ 2 กลุ่ม คือ กลุ่มอ้างอิง ได้แก่ a_R, b_R และ c_R และกลุ่มเปรียบเทียบ คือ a_F, b_F และ c_F พื้นที่ชนิดที่มีเครื่องหมาย (Closed-Interval Signed Area - CSA) สามารถคำนวณได้จาก

$$CSA = \int_{\theta_1}^{\theta_2} [P_R(\theta) - P_F(\theta)] d\theta = s_R(\theta_1, \theta_2) - s_F(\theta_1, \theta_2)$$

และพื้นที่ชนิดไม่มีเครื่องหมาย (Closed-Interval Unsigned Area - CUA) หรือ ϕ สามารถคำนวณได้จาก

$$CUA = \int_{\theta_1}^{\theta_2} |P_R(\theta) - P_F(\theta)| d\theta$$

สงบ ลักษณะ และคอฟแมน (สุรศักดิ์ อมรรัตนศักดิ์, 2531 : 54 ; อ้างจาก สงบ ลักษณะ และคอฟแมน, 1980 : 11) เสนอให้ใช้ค่า $CUA > .20$ เป็นเกณฑ์ชี้ความลำเอียงของข้อสอบ

สำหรับการวิจัยครั้งนี้ผู้วิจัยได้เลือกการเปรียบเทียบโค้งลักษณะข้อสอบแต่ละข้อสำหรับ กลุ่มผู้สอบ 2 กลุ่ม ว่าข้อสอบข้อใดบ้างเป็นข้อสอบที่มีความลำเอียง ซึ่งคำนวณโดยใช้โปรแกรม IRTDIF ใช้ค่า $CUA > .40$ เป็นค่าในการกำหนดว่าข้อสอบมีความลำเอียง (ทัศนีย์ พิรมนตรี, 2530 : 9 ; อ้างจาก Rudner, 1977)

งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบ

1. งานวิจัยภายในประเทศ

งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบเป็นการเปรียบเทียบผลการวิเคราะห์ความลำเอียงที่ใช้วิธีวิเคราะห์ต่างกัน โดยพิจารณาความลำเอียงที่เกิดขึ้นระหว่างประชากรกลุ่มย่อยที่มีเพศต่างกัน สภาพภูมิศาสตร์ต่างกัน และการพัฒนาเกณฑ์การตัดสินความลำเอียงทางเพศ ผู้วิจัยรวบรวมและสรุปได้ดังนี้

ซัชชัย เผ่าพงศ์ (2527 : 73 - 73) ได้ศึกษาความลำเอียงของข้อสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์และภาษา ระหว่างกลุ่มเพศชายและเพศหญิง ด้วยวิธีโค้งลักษณะข้อสอบแบบจำลองโลจิสติก 3 พารามิเตอร์ โดยใช้แบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์และภาษา ในระดับมัธยมศึกษา ซึ่งพัฒนาโดยสำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2524 ซึ่งประกอบด้วยทุกภาคภูมิศาสตร์ทั่วประเทศ 5 ภาค คือ ภาคเหนือ ภาคกลาง ภาคใต้ ภาคตะวันออก ภาคตะวันออกเฉียงเหนือ จำนวน 5,248 คน แบ่งเป็นกลุ่มที่ใช้วิเคราะห์ความลำเอียงของแบบทดสอบวัดความถนัดทางการเรียนด้านวิทยาศาสตร์ จำนวน 2,974 คน เป็นนักเรียนชาย 1,610 คน และนักเรียนหญิง 1,337 คน อีกกลุ่มหนึ่งใช้วิเคราะห์ความลำเอียงของแบบทดสอบความลำเอียงด้านภาษา จำนวน 2,301 คน เป็นนักเรียนชาย 1,316 คน นักเรียนหญิง 985 คน

ผลการศึกษาพบว่า แบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์ จำนวน 20 ข้อ มีข้อสอบที่มีความลำเอียงต่อกลุ่มใดกลุ่มหนึ่งโดยเฉพาะ 9 ข้อ คือ ลำเอียงต่อกลุ่มนักเรียนชาย 7 ข้อ และลำเอียงต่อกลุ่มนักเรียนหญิง 2 ข้อ เมื่อพิจารณาแยกตามระดับความสามารถมีข้อสอบที่ลำเอียงต่อกลุ่มนักเรียนชายและหญิงในระดับความสามารถปานกลางขึ้นไปมี จำนวน 5 ข้อ ซึ่งวัดในเรื่องร้อยละ การหาปริมาตร และการหาความยาวของเส้นในด้านของรูปสามเหลี่ยม จำนวนเรื่องละ 1 ข้อ อีกจำนวน 2 ข้อ เป็นเรื่องเกี่ยวกับโจทย์ปัญหา จากข้อสอบจำนวน 5 ข้อ ดังกล่าวนี้ เป็นข้อสอบที่มีความลำเอียงต่อกลุ่มนักเรียนหญิงในระดับความสามารถต่ำ และลำเอียงต่อกลุ่มนักเรียนชายในระดับความสามารถปานกลาง จำนวน 1 ข้อ ส่วนแบบทดสอบวัดความถนัดทางการเรียนด้านภาษา จำนวน 40 ข้อ มีข้อสอบที่ลำเอียงต่อกลุ่มใดกลุ่มหนึ่ง โดยเฉพาะ 11 ข้อ เป็นข้อสอบที่ลำเอียงต่อนักเรียนชาย 3 ข้อ และลำเอียงต่อนักเรียน 8 ข้อ เมื่อพิจารณาแยกตามระดับความสามารถมีข้อสอบที่มีความลำเอียงในระดับความสามารถปานกลางขึ้นไปมี

จำนวน 9 ข้อ ซึ่งวัดความเข้าใจเกี่ยวกับการอ่านคำประพันธ์ บทหรือกลอน อย่างละ 1 ข้อและวัดความเข้าใจเกี่ยวกับการอ่านข้อความ จำนวน 7 ข้อจากข้อสอบจำนวน 9 ข้อนี้เป็นข้อที่มีความลำเอียงต่อกลุ่มนักเรียนชาย จำนวน 1 ข้อ และเป็นข้อสอบที่มีความลำเอียงต่อกลุ่มนักเรียนหญิง จำนวน 6 ข้อ และมีข้อสอบที่มีความลำเอียงมากกว่าหนึ่งช่วงความสามารถ จำนวน 2 ข้อ จากผลการวิจัยครั้งนี้ เขาสรุปว่ามีแบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์และด้านภาษาฉบับนี้มีความลำเอียงต่อเพศ ดังนั้นเขาเสนอว่าในการที่จะนำแบบทดสอบฉบับนี้ไปทดสอบกับกลุ่มเพศชายและเพศหญิง ควรจะพิจารณาข้อที่มีค่าอำนาจจำแนกของทั้ง 2 กลุ่ม ที่แตกต่างกัน ไม่มาก ไม่ว่าข้อสอบข้อนั้นๆ จะมีค่าอำนาจจำแนกสูงหรือต่ำก็ตาม

ทัศนีย์ พิรมนตรี (2530 :76-78) ได้ศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของแบบทดสอบวิชาคณิตศาสตร์ โครงการตรวจสอบคุณภาพทางการศึกษาของสำนักงานทดสอบทางการศึกษาระดับชาตินิยมวิชาการ กระทรวงศึกษาธิการ โดยเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างกลุ่มนักเรียน ในกรุงเทพมหานครกับกลุ่มนักเรียนในภาคภูมิศาสตร์ทั้ง 5 ภาค คือภาคกลาง ภาคใต้ ภาคเหนือ ภาคตะวันออกเฉียงเหนือ และภาคตะวันออก ด้วยวิธีวิเคราะห์ 3 วิธี คือวิธีกำหนดจุดค่าเฉลี่ย วิธีทดสอบความแตกต่างระหว่างกลุ่มด้วยสถิติไคสแคว้ในโมเดลล็อกลิเนียร์ 2 โมเดล คือ โมเดลที่ไม่มีพารามิเตอร์ผลรวมระหว่างระดับคะแนนกับกลุ่ม โมเดลที่ไม่มีพารามิเตอร์ของผลหลักที่เกิดจากกลุ่ม และวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ กลุ่มตัวอย่างที่ใช้ศึกษาเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 จำนวน 7,036 คน

ผลการศึกษาพบว่า เมื่อแยกข้อสอบออกตามค่าความยากที่วิเคราะห์ด้วยวิธีดั้งเดิมมีข้อสอบที่มีความลำเอียง จำนวน 43 ข้อ และไม่ลำเอียง 17 ข้อ เหมือนกันทุกภาค เมื่อเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างกลุ่มนักเรียนในกรุงเทพมหานคร และกลุ่มนักเรียนภาคอื่นๆ พบว่าการวิเคราะห์ความลำเอียงด้วยวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ พบข้อสอบที่มีความลำเอียงมากที่สุด และมีข้อสอบที่มีความลำเอียงซ้ำกันระหว่างกลุ่มนักเรียนทุก ๆ ภาค กับนักเรียนในกรุงเทพมหานคร ในแต่ละวิธีมีจำนวนไม่เท่ากัน โดยที่วิธีกำหนดจุดเฉลี่ยและวิธีทดสอบความแตกต่างระหว่างกลุ่มด้วยสถิติไคสแคว้ในโมเดลล็อกลิเนียร์ 2 โมเดล มีจำนวนข้อสอบที่มีความลำเอียงมากที่สุดระหว่างกรุงเทพมหานคร และภาคตะวันออกเฉียงเหนือ ส่วนวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ มีจำนวนข้อสอบที่มีความลำเอียงมากที่สุด ระหว่างกรุงเทพมหานคร และภาคตะวันออก

การเปรียบเทียบผลการวิเคราะห์ความลำเอียง ทั้ง 3 วิธี กับความลำเอียงที่เกิดขึ้นในภาคเดียวกัน พบข้อสอบที่มีความลำเอียงซ้ำกัน ข้อสอบที่มีความลำเอียงส่วนมากเป็นข้อสอบที่ง่ายสำหรับนักเรียนในกรุงเทพมหานครมากกว่ากลุ่มนักเรียนภาคอื่นๆ จากการวิเคราะห์ด้วยวิธีกำหนดจุดเคลตา แต่เป็นข้อที่มีความลำเอียงในเกณฑ์ต่ำ สำหรับวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ และเป็นข้อที่ลำเอียงรูปแบบเดียวกันในการวิเคราะห์ด้วยวิธีทดสอบความแตกต่างระหว่างกลุ่มด้วยสถิติไคสแควร์ในโมเดลล็อกลิเนีย ทั้ง 2 โมเดล

พัชรี ปิยภักดิ์ (2531 ; 65-67) ได้ศึกษาเปรียบเทียบวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ด้วยวิธีแปลงค่าความยากวิธีไคสแควร์ และวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ ตัวแปรที่ศึกษาคือ เพศ และภาคภูมิศาสตร์ กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักเรียนชายและนักเรียนหญิง ชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2530 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติในกรุงเทพมหานครและจังหวัดสมุทรสาคร จำนวน 1,948 คน เพื่อวิเคราะห์หาดัชนีความลำเอียงของข้อสอบแล้วหาสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียงทั้ง 3 วิธีและศึกษาเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบก่อนและหลังการคัดเลือกข้อสอบที่ลำเอียงออกแล้ว

ผลการศึกษาพบว่า การวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีโค้งลักษณะข้อสอบพบจำนวนข้อสอบที่มีความลำเอียงมากที่สุด รองลงมาคือ วิธีไคสแควร์ และวิธีแปลงค่าความยากพบจำนวนข้อสอบที่ลำเอียงน้อยที่สุดและมีค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงต่อเพศระหว่างวิธีวิเคราะห์ความลำเอียง 3 วิธี มีค่าอยู่ระหว่าง .1713 ถึง .5618 ค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงต่อกลุ่มนักเรียนกรุงเทพมหานคร กับกลุ่มนักเรียนสมุทรสาครมีค่าอยู่ระหว่าง .1868 ถึง .6009 ส่วนค่าความเชื่อมั่นของแบบทดสอบก่อนและหลังคัดเลือกข้อสอบที่ลำเอียงแล้วแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

สุรศักดิ์ อมรรตณศักดิ์ (2531 : 200-207) ได้ศึกษาเปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบที่แตกต่างกัน 4 วิธี คือ วิธีวิเคราะห์ความแปรปรวน วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบที่มี 1 พารามิเตอร์ และวิธีโค้งลักษณะข้อสอบที่มี 3 พารามิเตอร์ จากแบบทดสอบที่ใช้สอบแข่งขันเพื่อบรรจุเข้ารับราชการเป็นข้าราชการครูในปี พ.ศ. 2529 ซึ่งมี 4 ฉบับ คือ ฉบับที่ 1 แบบทดสอบวิชาเอก (ความรู้เกี่ยวกับการประถมศึกษา) ฉบับที่ 2 แบบทดสอบวิชาการศึกษาและกฎหมาย ฉบับที่ 3 แบบทดสอบวิชาภาษาไทย ฉบับที่ 4 แบบทดสอบวิชาความสามารถทั่วไป เพื่อหาสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงระหว่างวิธีวิเคราะห์

ความลำเอียง 4 วิธีและเปรียบเทียบความแตกต่างของผลการคัดเลือกก่อนและหลังการศึกษาความลำเอียงของข้อสอบตามวิธีการคิดคะแนนรวมทั้งแตกต่างกัน 6 วิธี ในด้านจำนวนผู้ได้รับการคัดเลือก สักส่วนของชายต่อหญิงที่ได้รับการคัดเลือกและความเชื่อมั่นของแบบทดสอบ กลุ่มตัวอย่างที่ศึกษา คือ ผู้เข้าสอบแข่งขันเพื่อบรรจุเข้ารับราชการเป็นข้าราชการครู ในปี พ.ศ.2528 จำนวน 2,340 คน เป็นชาย 1,170 คน หญิง 1,170 คน

ผลการศึกษาพบว่า วิธีวิเคราะห์ความลำเอียงแต่ละวิธี พบจำนวนข้อสอบที่มีความลำเอียงต่างกัน โดยวิธีโค้งลักษณะข้อสอบที่มี 3 พารามิเตอร์ พบจำนวนข้อสอบที่มีความลำเอียงมากที่สุด รองลงมา ได้แก่ วิธีวิเคราะห์ความแปรปรวน และวิธีแปลงค่าความยากของข้อสอบ พบจำนวนข้อสอบที่มีความลำเอียงน้อยที่สุด ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงระหว่างวิธีวิเคราะห์ลำเอียง 4 วิธี พบว่า มีค่าอยู่ระหว่าง .7535 ถึง .9921 สำหรับจำนวนผู้ได้รับการคัดเลือกเมื่อเปรียบเทียบความแตกต่างของผลการคัดเลือกก่อนและหลังศึกษาความลำเอียงของข้อสอบตามวิธีคิดคะแนนคียบกับวิธีรวมคะแนนแบบอื่นๆ อีก 5 วิธี มีจำนวนผู้ได้รับการคัดเลือกแตกต่างกันประมาณ 4 เปอร์เซนต์ถึง 24 เปอร์เซนต์ ส่วนวิธีแปลงคะแนนคียบเป็นคะแนนมาตรฐานที่ปกติกับวิธีรวมคะแนนแบบอื่น ๆ อีก 4 วิธี มีจำนวนผู้ได้รับการคัดเลือกต่างกันประมาณ 6 เปอร์เซนต์ ถึง 23 เปอร์เซนต์ ส่วนความเชื่อมั่นของแบบทดสอบหลังจากคัดข้อสอบที่มีความลำเอียงพบว่าค่าความเชื่อมั่นของแบบทดสอบลดลงกว่าแบบทดสอบเดิมเล็กน้อย

สุวัฒน์ สุขมลสันต์ (2534) ได้วิเคราะห์ความลำเอียงของข้อสอบภาษาอังกฤษเพื่อคัดเลือกเข้าศึกษาต่อในมหาวิทยาลัย ด้วยวิธีวิเคราะห์ความลำเอียง 3 วิธี ได้แก่ วิธีกำหนดค่าเดลตา (Delta-Plot Method) วิธีโคสแควร์ชนิดที่แบ่งความสามารถของผู้สอบเป็น 3 ระดับ ได้แก่ กลุ่มที่มีระดับความสามารถต่ำ (ผู้ได้คะแนนรวม 0-40 คะแนน) กลุ่มมีระดับความสามารถปานกลาง (ผู้ได้คะแนนรวม 41-70 คะแนน) และกลุ่มที่มีระดับความสามารถสูง (ผู้ได้คะแนนรวม 71-100 คะแนน) วิธีการวัดพื้นที่ความแตกต่างระหว่างโค้งลักษณะข้อสอบที่วิเคราะห์ตามทฤษฎีการตอบข้อคำถามแบบ 3 พารามิเตอร์ จากแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข และชุด กขค ปี 2531-2533 ซึ่งมีข้อสอบชุดละ 100 ข้อ ตัวแปรที่ศึกษาความลำเอียง คือ เพศ และภาคภูมิศาสตร์ของผู้สอบ ซึ่งแยกออกเป็น 5 ภาค ตามภูมิสำเนาของผู้สอบ ได้แก่ ภาคกลาง ภาคตะวันออก ภาคใต้ ภาคตะวันออกเฉียงเหนือ และภาคเหนือ

เกณฑ์การตัดสินความลำเอียงของข้อสอบ สำหรับวิธีต่างๆ ที่ สุพัฒน์ สุขมลสันต์ กำหนดขึ้น คือ

1. วิธีกำหนดจุดค่าเฉลี่ย ข้อสอบที่ลำเอียง คือข้อที่มีระยะตั้งฉากจากคู่อันดับค่าเฉลี่ยไปยังเส้นแกนหลัก (d) มากกว่า 1.964 Sd. หรือน้อยกว่า -1.964 Sd.
2. วิธีไคสแควร์ ข้อสอบที่ลำเอียง คือ ข้อสอบที่ผู้สอบต่างกลุ่มกันมีอยู่ในระดับคะแนนเดียวกันมีสัดส่วนการตอบถูกหรือผิดแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ $\alpha = 0.017$
3. วิธีการวัดพื้นที่ความแตกต่างระหว่างโค้งลักษณะข้อสอบ ข้อสอบที่ลำเอียง คือ ข้อที่มีพื้นที่ความแตกต่างมากกว่า 0.40 โดยแบ่งเป็น 3 ระดับ
 - 3.1 พื้นที่ ที่มากกว่า 0.70 แสดงว่ามีความลำเอียงมาก
 - 3.2 พื้นที่ ที่มีค่าอยู่ระหว่าง 0.40-0.70 แสดงว่ามีความลำเอียงปานกลาง
 - 3.3 พื้นที่ ที่มีค่ามากกว่า 0.00 แต่น้อยกว่า 0.40 แสดงว่ามีความลำเอียงน้อย

ผลการศึกษาพบว่า แบบทดสอบภาษาอังกฤษ ฉบับ กข และ กขค ปี 2531-2533 มีความลำเอียงต่อเพศในช่วง 7-28 ข้อ และช่วง 4-41 ข้อ โดยมีแนวโน้มที่ลำเอียงต่อเพศชายมากกว่าเพศหญิง และลำเอียงต่อภาคภูมิศาสตร์ในช่วง 6-45 ข้อ และ 5-43 ข้อ โดยมีความลำเอียงต่อผู้สอบจากภาคอื่นมากกว่าภาคกลางประมาณ 2-3 เท่า

การเปรียบเทียบผลของการวิเคราะห์โดยใช้วิธีวิเคราะห์ต่างกัน พบว่า จำนวนข้อสอบที่ระบุว่ามีลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติ และแต่ละวิธีให้ผลที่มีความสัมพันธ์กันอย่างไม่มีนัยสำคัญทางสถิติ การวิเคราะห์ด้วยวิธีของทฤษฎีการตอบข้อคำถามเมื่อใช้เกณฑ์พื้นที่ที่แสดงความลำเอียงของข้อสอบ มากกว่า 0.40 พบว่า มีจำนวนข้อสอบที่มีความลำเอียงน้อยกว่า 2 วิธีแรกประมาณ 3-4 เท่า

นิรมล ชัยสวัสดิ์ (2537 : 57-58) ได้ศึกษาเปรียบเทียบผลการวิเคราะห์ความลำเอียงของแบบทดสอบวัดความเข้าใจในการอ่านภาษาไทยตามทฤษฎีคลาสสิกอลที่ใช้วิธีวิเคราะห์ต่างกัน 3 วิธี คือ วิธีแปลงค่าความยาก วิธีไคสแควร์ และวิธีวิเคราะห์ความแปรปรวน จากแบบทดสอบวัดความเข้าใจในการอ่านภาษาไทย จำนวน 5 ข้อ กลุ่มตัวอย่างที่ใช้ศึกษาเป็นนักเรียนชายหญิง ชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2536 ของโรงเรียนในสังกัดกรมสามัญศึกษา เขตพระโขนง กรุงเทพมหานคร จำนวน 1,006 คน โดยศึกษาเปรียบเทียบจำนวนข้อที่มีความลำเอียง ค่าความเชื่อมั่นที่ใช้สูตรคำนวณแบบ KR-20 และแบบแบ่งครึ่งฉบับ ของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว

ผลการศึกษาพบว่า การวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากกับวิธีโคสแควร์ และวิธีแปลงค่าความยาก กับวิธีวิเคราะห์ความแปรปรวน มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีโคสแควร์กับวิธีวิเคราะห์ความแปรปรวนพบจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติ

สำหรับค่าความเชื่อมั่นของแบบทดสอบที่ใช้สูตรคำนวณทั้งแบบ KR-20 และแบบแบ่งครึ่งฉบับหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว จากการวิเคราะห์ด้วยวิธีแปลงค่าความยาก วิธีโคสแควร์ และวิธีวิเคราะห์ความแปรปรวนแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

กาญจนา วิชาสุนทร (2538) ได้ศึกษาเกี่ยวกับการพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ โดยใช้ดัชนีการวิเคราะห์ 4 ตัว คือ พื้นที่ระหว่างโค้งการตอบข้อสอบชนิดที่มีเครื่องหมาย (SA) และ ไม่มีเครื่องหมาย (UA) จากวิธีใช้ทฤษฎีการตอบข้อคำถามตามแบบจำลอง 2 พารามิเตอร์ α_{MH} จากวิธีของเมนเทล-แฮนส์เซล และ B_{SIB} จากวิธี SIBTEST จากข้อมูลเชิงประจักษ์เพื่อพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ

ข้อมูลที่ใช้ศึกษาเป็นข้อมูลการตอบข้อสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของทบวงมหาวิทยาลัย ปีการศึกษา 2535 โดยใช้ความยาวของแบบทดสอบ 20, 30 และ 40 ข้อ สำหรับวิชาคณิตศาสตร์ และ 50, 60, 70 และ 80 ข้อ สำหรับวิชาภาษาอังกฤษ โดยใช้กลุ่มผู้สอบ 6 ขนาด คือ 100, 200, 400, 600, 800 และ 1,000 คน

การพัฒนาเกณฑ์กระทำโดยคำนวณค่าดัชนี ทั้ง 4 ตัว จากข้อมูลการตอบข้อสอบของผู้สอบเพศเดียวกัน เพศละ 50 ค่า สำหรับแต่ละความยาวของแบบทดสอบ และขนาดของกลุ่มผู้สอบ จากนั้นนำค่าดัชนีที่ได้ทั้งหมดมาวิเคราะห์หาค่าเฉลี่ย และกำหนดเกณฑ์จากค่าเฉลี่ย 2 ลักษณะคือ เกณฑ์ที่กำหนด โดยพิจารณาจากค่าเฉลี่ยรวม จากค่าดัชนีทุกข้อโดยไม่พิจารณาความแตกต่างในด้านความยาวของแบบทดสอบและขนาดของกลุ่มผู้สอบ และเกณฑ์ที่พิจารณาจากค่าเฉลี่ย ซึ่งพิจารณาถึงความยาวของแบบทดสอบและขนาดของกลุ่มผู้สอบ

ผลจากการศึกษาพบว่า เมื่อนำเกณฑ์ที่กำหนดไว้ไปตัดสินค่าดัชนีที่ได้จากการวิเคราะห์ความลำเอียงระหว่างผู้สอบเพศหญิงและชาย ไม่ว่าจะใช้เกณฑ์ลักษณะใดผลด้านจำนวนข้อสอบในรูปของร้อยละของข้อสอบที่มีความลำเอียง และความสอดคล้องของการตัดสินมีความไม่คงที่ทุก ๆ ความยาวของแบบทดสอบและทุก ๆ ขนาดกลุ่มผู้สอบ โดยที่เกณฑ์ในลักษณะหลังจะลดความไม่คงที่ลงได้ระดับหนึ่ง ทั้งนี้เนื่องจากค่าดัชนีแต่ละตัวที่ได้จากกลุ่มผู้สอบต่างเพศกันมีความไม่คงที่จะแปรเปลี่ยนไปเมื่อความยาวของแบบทดสอบที่ใช้ และขนาดของกลุ่มผู้สอบเปลี่ยนไป

นอกจากนั้นพบว่าที่ความยาวของแบบทดสอบ 40 ข้อ สำหรับวิชาคณิตศาสตร์ และ 80 ข้อ สำหรับวิชาภาษาอังกฤษมีจำนวนข้อสอบที่ได้รับการตัดสินว่ามีความลำเอียงมากกว่าในความยาวอื่นๆ มาก อีกทั้งยังพบว่าเกณฑ์ที่กำหนดได้จากการศึกษาครั้งนี้มีค่าสูงกว่าเกณฑ์ขั้นต่ำที่ควรใช้ : ในกรณีของค่า UA ในขณะที่ α_{MH} มีค่าใกล้เคียงกับเกณฑ์ขั้นต่ำที่ควรใช้ในเกณฑ์ตัดสินความลำเอียงของข้อสอบจากเกณฑ์ที่พัฒนาได้ครั้งนี้ คือ : SA : > .40 , UA > .50 เมื่อใช้ความยาวของแบบทดสอบ 50 ข้อ ขึ้นไป และ : SA : > .80 , UA > 1.20 กรณีความยาวของแบบทดสอบสั้นกว่า 50 ข้อ ค่า $.60 > \alpha_{MH} > 1.40$ และ : SA : > .06

จิตสุดา ธราพร (2539) ได้ศึกษาการเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบ โจทย์ปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษาที่ใช้วิธีวิเคราะห์ต่างกัน 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบและวิธีของแมนเทิล-แฮนส์เชล เพื่อเปรียบเทียบค่าความเชื่อมั่นที่คำนวณโดยใช้สูตร KR-20 และใช้สูตรแบบแบ่งครึ่งฉบับ ของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออก กลุ่มตัวอย่างที่ศึกษาเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษาในจังหวัดกระบี่ จำนวน 2,689 คน ซึ่งเลือกมาโดยการสุ่มอย่างง่าย

ผลการวิจัย พบว่า แต่ละวิธีที่พบจำนวนข้อสอบที่มีความลำเอียงแตกต่างกัน ซึ่งวิธีของแมนเทิล-แฮนส์เชล พบจำนวนข้อสอบที่มีความลำเอียงมากที่สุด และวิธีแปลงค่าความยากของข้อสอบพบจำนวนข้อสอบที่มีความลำเอียงน้อยที่สุด โดยที่การวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากของข้อสอบกับวิธีโค้งลักษณะข้อสอบ และวิธีแปลงค่าความยากของข้อสอบกับวิธีแมนเทิล-แฮนส์เชล มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนวิธีโค้งลักษณะข้อสอบกับวิธีของแมนเทิล-แฮนส์เชล มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

สำหรับค่าความเชื่อมั่นของแบบทดสอบ โจทย์ปัญหาทางคณิตศาสตร์ที่คำนวณโดยใช้สูตร KR-20 หลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกจากการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากของข้อสอบกับวิธีโค้งลักษณะข้อสอบ และวิธีแปลงค่าความยากของข้อสอบ กับวิธีของแมนเทิล-แฮนส์เชล มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และวิธีโค้งลักษณะข้อสอบกับวิธีของแมนเทิล-แฮนส์เชล มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ส่วนค่าความเชื่อมั่นของแบบทดสอบที่คำนวณโดยใช้สูตรแบบแบ่งครึ่งฉบับ หลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกจากการวิเคราะห์ความลำเอียง 3 วิธี มีค่าความแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ส่วนแบบทดสอบความเข้าใจทางภาษา ค่าความเชื่อมั่นโดยใช้สูตรแบบ KR-20 และ

แบบแบ่งครึ่งฉบับ หลังคัดเลือกข้อสอบที่มีความลำเอียงออกจากการวิเคราะห์ความลำเอียง มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

เรวดี อินทรสระ (2539:277) ได้ศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่วิเคราะห์ความลำเอียงต่อเพศด้วยวิธีใช้ทฤษฎีการตอบข้อคำถาม (IRT) วิธีแมนเทิล-แฮนส์เซล (MH) และวิธี SIBTEST การตัดสินใจผลการสอบที่คิดคะแนนมาตรฐานที่ปกติและคะแนนน้ำหนักความสามารถและสาเหตุความลำเอียงของข้อสอบ โดยศึกษาความลำเอียงของข้อสอบคัดเลือกเข้าศึกษาในชั้นปีที่ 1 ประเภทรับตรง ปีการศึกษา 2538 ของมหาวิทยาลัยสงขลานครินทร์ ในวิชาภาษาไทย ก วิชาสังคมศึกษา ก วิชาภาษาอังกฤษ กข วิชาละ 8,127 คน (ชาย 2,722 คน หญิง 5,405 คน) วิชาภาษาไทย กข วิชาสังคม กข และวิชาภาษาอังกฤษ กขก วิชาละ 5,415 คน (ชาย 1,454 คน หญิง 3,961 คน) ความเที่ยงตรงเชิงพยากรณ์ศึกษาจากคะแนนสอบคัดเลือกกับเกรดภาคเรียนที่ 1 ปีการศึกษา 2538 ของนักเรียนที่ได้รับการคัดเลือกจากประเภทรับตรง สายวิทยาศาสตร์ 763 คน และสายศิลปศาสตร์ 281 คน และสาเหตุความลำเอียงของข้อสอบจากการระบุสาเหตุของนักวัดผลการศึกษาหรืออาจารย์ผู้สอน จำนวน 50 คน และนิสิตนักศึกษาที่เรียนในสาขาวิชานั้นๆ วิชาละ 50 คน

ผลการศึกษาพบว่า วิธีการตรวจสอบความลำเอียงทั้ง 3 วิธี ตัดสินจำนวนข้อสอบที่ลำเอียงแตกต่างกันในวิชาภาษาไทย ก ฉบับที่ 2 และวิชาสังคมศึกษา ก ฉบับที่ 1 ที่ระดับนัยสำคัญทางสถิติที่ .05 นอกนั้นต่างกันที่ระดับนัยสำคัญทางสถิติที่ .01 โดยที่วิธีใช้ทฤษฎีการตอบข้อคำถามตัดสินจำนวนข้อสอบที่ลำเอียงมากที่สุด ความสำคัญของลำดับที่ของการสอบไม่ว่าจะคิดคะแนนมาตรฐานที่ปกติ หรือคิดคะแนนน้ำหนักความสามารถและใช้ข้อสอบจำนวนทั้งหมดหรือใช้เฉพาะข้อสอบที่ปราศจากความลำเอียงต่างมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ประสิทธิภาพในการทำนายผลสัมฤทธิ์ทางการเรียนในสายวิทยาศาสตร์ การคิดคะแนนน้ำหนักความสามารถที่ใช้ข้อสอบจำนวนทั้งหมดและใช้เฉพาะข้อสอบที่ปราศจากความลำเอียงมีประสิทธิภาพในการทำนายสูงกว่าการคิดคะแนนมาตรฐานที่ปกติ ส่วนสายศิลปศาสตร์การคิดคะแนนมาตรฐานที่ - ปกติทั้งที่ใช้ข้อสอบจำนวนทั้งหมดและใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง มีประสิทธิภาพในการทำนายสูงกว่าการคิดคะแนนน้ำหนักความสามารถและสาเหตุของความลำเอียงของข้อสอบต่อเพศทั้งชายและหญิง เกิดจากข้อสอบเป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะ จะมีโอกาสตอบถูกมากกว่า เป็นเรื่องราวที่กลุ่มนั้นๆ สนใจและเป็นข้อสอบที่ถามความจำ

2. งานวิจัยต่างประเทศ

สำนักงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบในต่างประเทศนั้นมีผู้วิจัยทำวิจัยไว้หลายท่าน ซึ่งผู้วิจัยขอนำเสนอ 2 ประเด็น คือ

1. การวิเคราะห์ความลำเอียงของข้อสอบจากข้อมูลจำลอง และ
2. การวิเคราะห์ความลำเอียงของข้อสอบจากข้อมูลจริง

2.1 การวิเคราะห์ความลำเอียงของข้อสอบจากข้อมูลจำลอง

เบอร์ก (Berk, 1982 : 164; quoting Merz and Grossen, 1979) ได้ศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อสอบ โดยวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยาก วิธีพอยท์ไบซีเรียล วิธีโคสแควร์ การวิเคราะห์ห่อองค์ประกอบ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ และวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ โดยใช้ข้อมูลสมมติ

ผลการศึกษาพบว่า การวิเคราะห์ความลำเอียงด้วยวิธีวิเคราะห์ด้วยพอยท์ไบซีเรียล เป็นวิธีที่ให้ผลการวิเคราะห์สอดคล้องกับวิธีอื่นน้อย ส่วนวิธีแปลงค่าความยากให้ผลการวิเคราะห์สอดคล้องกับวิธีอื่นมาก รวมทั้งวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ และ 1 พารามิเตอร์ ให้ผลการวิเคราะห์สอดคล้องกับวิธีอื่นมากเช่นกัน วิธีโคสแควร์และวิธีวิเคราะห์ห่อองค์ประกอบให้ผลการวิเคราะห์สอดคล้องกับวิธีอื่นปานกลาง แต่วิธีแปลงค่าความยากเป็นวิธีที่เหมาะสมเพราะไม่มีข้อจำกัดในเรื่องจำนวนกลุ่มตัวอย่าง และการคำนวณ

ชูเนอร์ (Schoener, 1984) ได้ศึกษาเปรียบเทียบวิธีการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีการทางสถิติที่ใช้ทฤษฎีการตอบข้อคำถามที่มี 3 พารามิเตอร์ และการตัดสินใจด้วยผู้ตัดสินจำนวน 24 คน เป็นผู้ตัดสินผิวดำ 8 คน ผิวดาว 8 คน และเชื้อสายสเปน 8 คน เป็นชายและหญิงจำนวนเท่า ๆ กัน โดยใช้แบบทดสอบอิงเกณฑ์ กลุ่มตัวอย่างที่นักเรียนชายและหญิงที่ระดับชั้นมัธยมศึกษาที่มีภูมิหลังทางวัฒนธรรมต่าง ๆ กันจำนวน 1,064 คน การตัดสินใจของผู้ตัดสินที่มีความรู้ทางด้านคณิตศาสตร์เป็นอย่างดี และใช้แบบฟอร์มการใช้คะแนนเป็นช่วง (Rating Form) สำหรับการวิเคราะห์ ได้พิจารณาถึงความสอดคล้องจากการตัดสินใจของกลุ่มผู้ตัดสินทั้ง 3 กลุ่ม และความสอดคล้องระหว่างวิธีการทางสถิติทางการตัดสินใจของผู้ตัดสิน โดยใช้ค่าสถิติของ แคปป่า (Kappa)

ผลการศึกษาพบว่า ความสอดคล้องระหว่างการตัดสินใจโดยผู้เชี่ยวชาญต่างกลุ่มไม่นับสำคัญทางสถิติสำหรับทั้งความลำเอียงทางด้านวัฒนธรรมและทางเพศ แต่พบความสอดคล้องกันอย่างมีนัยสำคัญทางสถิติในตัวบ่งชี้บางตัวที่คำนวณขึ้นเพื่อรวมการให้คะแนนของผู้ตัดสิน แต่

ไม่พบความสอดคล้องระหว่างการวิเคราะห์ความลำเอียงทั้ง 2 วิธี การตัดข้อสอบที่พบทางสถิติ ว่ามีความลำเอียงออกไป ไม่ทำให้ลำดับของคนในกลุ่มย่อยและรวมทั้งหมดเปลี่ยนแปลงไป นอกจากนี้ ทั้งแบบทดสอบฉบับเดิม และแบบทดสอบที่มีความลำเอียงออกแล้ว มีความสัมพันธ์อย่างมีนัยสำคัญทางสถิติกับแบบทดสอบวิชาคณิตศาสตร์อิงกลุ่มที่เป็นแบบทดสอบมาตรฐาน

รัดเนอร์ และคนอื่นๆ (Rudner and other, 1980) ได้เปรียบเทียบวิธีวิเคราะห์หาความลำเอียงของข้อสอบ โดยใช้วิธีวิเคราะห์ที่ต่างกัน 7 วิธี ได้แก่ วิธีแปลงค่าความยากของข้อสอบ 2 วิธี คือวิธีใช้เส้นแกนหลัก (TID-MA) และวิธีใช้เส้นแกน 45 องศา (TID-45) วิธีโคสแควร์ 2 วิธี คือวิธีโคสแควร์ที่แบ่งแบบ 5 ช่วง (CHI - 5) และวิธีโคสแควร์ที่แบ่งแบบหลายช่วง (CHI-N) และวิธีโค้งลักษณะข้อสอบ 3 วิธี คือ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ (ICC-3) วิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ ชนิดค่าสถิติเหมาะสม (ICC-1F) และชนิดความแตกต่างในความง่ายของข้อสอบ (ICC-1E) โดยใช้ข้อมูลสมมติ (Simulate Data) ที่ผู้วิจัยสร้างขึ้นเอง

ผลการศึกษาพบว่า วิธีโค้งลักษณะข้อสอบที่มี 3 พารามิเตอร์ เป็นวิธีค้นพบข้อสอบที่มีความลำเอียงจากข้อมูลสมมติที่ดีที่สุด รองลงมาได้แก่วิธีวิเคราะห์แบบโคสแควร์แบบ 5 ช่วง ส่วนวิธีแปลงค่าความยากของข้อสอบทั้ง 2 วิธี ผู้วิจัยแนะนำให้ใช้ได้ทางปฏิบัติ เมื่อพิจารณาความสัมพันธ์ของจำนวนข้อสอบที่มีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียงต่างกัน วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ วิธีโคสแควร์ ที่แบ่งแบบ 5 ช่วง และวิธีแปลงค่าความยากที่ใช้เส้นแกน 45 องศา เป็นวิธีที่ถูกต้องมากที่สุด คือ มีค่าสหสัมพันธ์เท่ากับ .80, .73 และ .63 ส่วนวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ ชนิดค่าสถิติเหมาะสม มีความถูกต้องน้อยที่สุด มีค่าสหสัมพันธ์เท่ากับ .55 วิธีวิเคราะห์ความลำเอียงทุกวิธี ยกเว้นวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ ชนิดสถิติเหมาะสมสามารถค้นพบข้อสอบที่มีความลำเอียงในกรณีที่มีความยากแตกต่างกันเท่านั้น แต่วิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ ชนิดค่าสถิติเหมาะสม มีความไวต่อความลำเอียงของข้อสอบที่มีทั้งความยากและค่าอำนาจจำแนกแตกต่างกัน นอกจากนี้ยังพบว่าวิธีโคสแควร์ที่แบ่งแบบ 5 ช่วง มีประสิทธิภาพเท่าๆ กับวิธีโค้งลักษณะข้อสอบที่มี 3 พารามิเตอร์ และถ้าได้คำนึงถึงความยาวของแบบทดสอบหรือขนาดของความลำเอียงของข้อสอบแล้ว ทั้งสองวิธีก็สามารถนำมาใช้แทนกันได้

ธิสเซน สไตน์เบิร์ก และไวเนอร์ (Thissen, Steinberg and Wainer, 1988) ได้ศึกษาความแตกต่างของโค้งการตอบข้อสอบระหว่างกลุ่มด้วยทฤษฎีการตอบข้อคำถาม เพื่อศึกษาถึงสภาพการณ์ในการทดสอบสมมติฐานความเท่ากันของค่าพารามิเตอร์ ด้วยค่าโคสแควร์แบบอัตราส่วน Likelihood (IRT-LR) โดยใช้ข้อมูลสมมติ (Simulata Data) แล้วจึงทำการเปรียบเทียบการทดสอบสมมติฐานความเท่ากันของค่าพารามิเตอร์ด้วยวิธี IRT-LR กับวิธีของแมนเทิล-แฮนส์เซล

ผลการศึกษาพบว่า วิธีไคสแควร์แบบอัตราส่วน Likelihood (IRT-LR) สามารถใช้วิเคราะห์ความลำเอียง ของข้อสอบได้เมื่อมีจำนวนกลุ่มตัวอย่างน้อย ๆ คือประมาณ 500 คน ความแตกต่างของค่าความยากของข้อสอบระหว่างกลุ่ม มีค่าเท่ากับ 0.3 และความแตกต่างในด้านค่าสัมประสิทธิ์การเดามีค่าเท่ากับ 0.1 โดยจำนวนร่วมจะมากหรือน้อยก็ไม่มีผลกระทบ ถ้าความแตกต่างมีค่าน้อยกว่านี้ไม่สามารถที่จะใช้วิธีไคสแควร์แบบอัตราส่วน Likelihood (IRT-LR) วิเคราะห์ได้ว่าข้อสอบลำเอียงหรือไม่ แต่ถ้าใช้กลุ่มตัวอย่างจำนวนมากก็อาจจะทำการวิเคราะห์ได้

การระบุความลำเอียงของข้อสอบ ใช้วิธีเปรียบเทียบความแตกต่างระหว่างกลุ่มในข้อสอบที่ต้องการวิเคราะห์ความลำเอียงกับข้อสอบร่วม ซึ่งมีข้อตกลงว่าข้อสอบร่วมนั้นไม่มีความลำเอียงอย่างไรก็ตามควรได้ศึกษาเพื่อหาวิธีเลือกข้อสอบร่วม และการทดสอบความลำเอียงของข้อสอบทั้งฉบับ สำหรับผลการวิเคราะห์ความลำเอียงของข้อสอบ ด้วยวิธีไคสแควร์แบบอัตราส่วน Likelihood (IRT-LR) กับวิธีของแมนเทิล-แฮนส์เซล ทั้งสองวิธีให้ผลการวิเคราะห์คล้ายคลึงกัน ข้อเสนอนี้จากการวิจัยครั้งนี้ คือ วิธีของแมนเทิล-แฮนส์เซล อาจจะใช้ในการวิเคราะห์ความลำเอียงของแบบทดสอบยาว ๆ หรืออาจใช้เครื่องมือในการกลั่นกรองข้อสอบก่อนที่จะใช้วิธีไคสแควร์แบบอัตราส่วน Likelihood (IRT-LR) ต่อไป

สวามินาธาน และโรเจอร์ส (Swaminathan and Rogers, 1990) ได้ตรวจสอบความลำเอียงของข้อสอบ ด้วยวิธีการถดถอยแบบโลจิสต์กับวิธีของแมนเทิล-แฮนส์เซล (MH) จากการจำลองข้อมูลการตอบข้อสอบด้วยโปรแกรม DATAGEN แบบ 3 พารามิเตอร์ และข้อมูลการจำลองความลำเอียงใช้แบบ 2 พารามิเตอร์ จำลองข้อมูลที่มีความลำเอียง 2 แบบ คือ ความลำเอียงแบบสม่ำเสมอ (Uniform) และการลำเอียงแบบไม่สม่ำเสมอ (Nonuniform) โดยใช้ค่าอำนาจจำแนกของ 2 กลุ่ม เท่ากันส่วนค่าความยากของข้อสอบแปรเปลี่ยนไปตามกลุ่มเพื่อให้เกิดระดับความลำเอียงต่าง ๆ กัน ตามต้องการ แต่ในลักษณะการลำเอียงไม่สม่ำเสมอ (Nonuniform) นั้นให้ค่าความยากของข้อสอบของ 2 กลุ่มเท่ากัน แต่ค่าอำนาจจำแนกแปรเปลี่ยนไปตามกลุ่ม โดยศึกษากับขนาดกลุ่มตัวอย่าง 2 ขนาด คือ 250 และ 500 คน และความยาวของแบบทดสอบ 3 ขนาด คือ 40, 60 และ 80 ข้อ

ผลการศึกษา พบว่า การตรวจสอบความลำเอียงของข้อสอบ ทั้ง 2 วิธี ให้ผลใกล้เคียงกัน ในกรณีความลำเอียงแบบสม่ำเสมอ (Uniform) วิธีของแมนเทิล-แฮนส์เซล ตรวจสอบความลำเอียงของข้อสอบได้ดีกว่าวิธีการถดถอยแบบโลจิสต์เล็กน้อย โดยมีการตรวจสอบความลำเอียงได้ถูกต้องร้อยละ 75 กรณีใช้กลุ่มตัวอย่าง 250 คน และตรวจสอบได้ถูกต้องร้อยละ 100 ในกรณีใช้กลุ่มตัวอย่าง 500 คน ในทุกความยาวของแบบทดสอบ แต่ในกรณีของความลำเอียงแบบไม่สม่ำเสมอ

(Nonuniform) นั้น วิธีของแมนเทล-แฮนส์เชล ไม่สามารถตรวจสอบความลำเอียงได้ ส่วนวิธีการถดถอยแบบโลจิสต์ ตรวจสอบความลำเอียงได้ถูกต้องประมาณร้อยละ 50 ในกรณีกลุ่มตัวอย่างน้อย และข้อสอบสั้น และตรวจสอบความลำเอียงได้ถูกต้องประมาณร้อยละ 75 ในกรณีกลุ่มตัวอย่างขนาดใหญ่และแบบทดสอบยาว แต่กรณีมีความลำเอียงแบบไม่สม่ำเสมอ โดยที่โค้งลักษณะข้อสอบตัดกันที่ปลายจุดของโค้งด้านใดด้านหนึ่งไม่ว่าจะเป็นด้านความสามารถสูงหรือต่ำก็ตาม วิธีของแมนเทล-แฮนส์เชล จะตรวจสอบความลำเอียงได้ดีกว่า

ในการตรวจสอบผิดพลาด คือ ระบุว่าข้อสอบมีความลำเอียง ทั้งๆ ที่ไม่ลำเอียงนั้น วิธีของแมนเทล-แฮนส์เชล ดีกว่า โดยตรวจสอบผิดพลาดประมาณร้อยละ 1 ส่วนวิธีถดถอยแบบโลจิสต์ ตรวจสอบผิดพลาดประมาณร้อยละ 1-6 และเสียค่าใช้จ่ายมากกว่าประมาณ 3-4 เท่า แม้ว่าจะมีความยืดหยุ่นและเป็นแบบทั่วไปมากกว่าวิธีของแมนเทล-แฮนส์เชล

ประโยชน์ของวิธีการถดถอยแบบโลจิสต์ อยู่ตรงที่เป็นวิธีการที่มีพื้นฐานเป็นแบบจำลองที่สามารถเพิ่มตัวแปรความสามารถเข้าไปในแบบจำลองได้ ซึ่งทำให้เข้าใจในธรรมชาติของความลำเอียงได้ดีกว่า

คลอเซอร์ และคนอื่นๆ (Clauser and Others, 1991) ได้ศึกษาถึงอิทธิพลต่างๆ ที่มีผลต่อวิธีการตรวจสอบความลำเอียงของข้อสอบ ด้วยวิธีของแมนเทล-แฮนส์เชล ข้อมูลที่ใช้เป็นข้อมูลที่จำลองขึ้นเพื่อใช้ในการศึกษาตรวจสอบความลำเอียงแบบสม่ำเสมอ ข้อมูลที่จำลองขึ้นเป็นแบบทดสอบจำนวน 75 ข้อ ให้มีความลำเอียงอยู่ 16 ข้อ และมีระดับความแตกต่างของค่าความยากหลายระดับ ระหว่างกลุ่มอ้างอิงและกลุ่มเป้าหมาย กลุ่มตัวอย่างที่ใช้ศึกษามีจำนวน 3,000 คน เป็นกลุ่มอ้างอิง 1,500 คน กลุ่มเป้าหมาย 1,500 คน มีการวิเคราะห์ระดับความยากของแบบทดสอบที่ต่างกันหลายระดับ กับระดับความยากของแบบทดสอบ 5 ระดับ อำนาจจำแนก 4 ระดับ และการกระจายของความสามารถ 2 ลักษณะ

ผลการศึกษาพบว่า สิ่งที่มีอิทธิพลต่อการตรวจสอบความลำเอียงของข้อสอบด้วยวิธีของแมนเทล-แฮนส์เชล คือ ความแตกต่างในระดับความยากระหว่างผู้สอบทั้ง 2 กลุ่ม รวมทั้งค่าอำนาจจำแนก และค่าความยากของข้อสอบด้วย จากการวิจัยนี้ทำให้ได้ข้อเสนอแนะว่า วิธีของแมนเทล-แฮนส์เชล นี้ใช้ได้ดี สำหรับข้อสอบที่มีค่าอำนาจจำแนกสูง และมีแนวโน้มจะไม่สามารถตรวจสอบความลำเอียงของข้อสอบ ในข้อสอบที่มีค่าความยากสูงได้

เมเซอร์ และคนอื่นๆ (Mazor and Other, 1991) ได้ศึกษาวิธีการตรวจค้นความลำเอียงของข้อสอบด้วยวิธีของแมนเทล-แฮนส์เชล การตรวจค้นความลำเอียงของข้อสอบจะใช้การทดสอบค่า $MH - \chi^2$ มี $df = 1$ จากข้อมูลที่จำลองขึ้นเป็นข้อสอบ จำนวน 5 ชุด และ 75 ชุด โดยศึกษา

กับกลุ่มตัวอย่างที่มีขนาดต่างกัน 5 ขนาด คือ 2,000, 1,000, 500, 200 และ 100 คน เพื่อเปรียบเทียบการตรวจค้นความลำเอียงของข้อสอบ

ผลการศึกษา พบว่า เมื่อใช้กลุ่มตัวอย่างขนาด 2,000 คน วิธีนี้วิเคราะห์ความลำเอียงของข้อสอบผิดพลาดร้อยละ 30 และเมื่อใช้จำนวนกลุ่มตัวอย่างตั้งแต่ 500 คนลงไป วิธีนี้จะวิเคราะห์ความลำเอียงของข้อสอบได้ผิดพลาดร้อยละ 50 ส่วนข้อสอบที่ไม่สามารถตรวจค้นความลำเอียงได้ ได้แก่ ข้อสอบที่ยากมากๆ ข้อสอบที่มีค่าความยากต่างกันเล็กน้อยสำหรับกลุ่มผู้สอบ 2 กลุ่ม และเป็นกลุ่มที่มีค่าอำนาจจำแนกต่ำ

สำหรับการวิเคราะห์ความลำเอียงของข้อสอบโดยใช้ข้อมูลจริง มีผู้ศึกษาดังนี้

ผองจิต อินทสุวรรณ (Intasuwat, 1979 :92) ได้ศึกษาวิเคราะห์ความลำเอียงของข้อสอบระหว่างกลุ่มผู้สอบที่มีเชื้อชาติต่างกัน ด้วยวิธีวิเคราะห์ 3 วิธี คือ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ วิธีโค้งลักษณะข้อสอบตามแบบจำลองของราสซ์ และวิธีไคสแควร์ การศึกษาครั้งนี้ใช้ข้อมูลจริง จากการใช้แบบทดสอบ ไอ อี เอ (IEA : The International Association for the Evaluation of Education Achievement) โดยศึกษากับกลุ่มนักเรียนที่มีอายุ 14 ปี ชาวอังกฤษ อเมริกัน และนิวซีแลนด์ จำนวน 8,601 คน

ผลการศึกษาพบว่า วิธีวิเคราะห์ความลำเอียงทั้ง 3 วิธี มีค่าสหสัมพันธ์อยู่ระหว่าง 0.51-0.98 โดยที่วิธีวิเคราะห์แบบไคสแควร์ และวิธีโค้งลักษณะข้อสอบตามแบบจำลองของราสซ์ มีค่าสหสัมพันธ์สูงถึง .98

เบอร์ก (Berk, 1982 : 196 ; quoting sheparrd, Camilli and Averill, 1981) ได้ศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อสอบ ระหว่างกลุ่มผู้สอบที่มีเชื้อชาติต่างกัน โดยวิธีวิเคราะห์ความลำเอียงด้วยวิธีไคสแควร์ วิธีกำหนดจุดเคลตา หรือวิธีแปลงค่าความยาก วิธีพอยท์ไบซีเรียล วิธีโค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์ และวิธีโค้งลักษณะข้อสอบแบบ 1 พารามิเตอร์ (Rasch Model) จากแบบทดสอบ Lorge-Thorndike ที่เป็นแบบทดสอบที่วัดความสามารถทางสมอง แบ่งเป็น 2 ฉบับ คือ ฉบับภาษา และไม่ใช่ภาษา การวิเคราะห์ความลำเอียงของข้อสอบผู้วิจัยใช้เกณฑ์ภายนอก และเกณฑ์ภายใน โดยที่เกณฑ์ภายใน คือ คะแนนจาก Raven's Colored Progressive Matrices ส่วนเกณฑ์ภายใน คือ คะแนนรวมจากแบบทดสอบ กลุ่มตัวอย่างที่ใช้มี 3 เชื้อชาติ จำนวน 1,500 คน เป็นนักเรียนผิวขาว ผิวดำ และซีกาโน (Cnicano) เชื้อชาติละ 500 คน

ผลการศึกษาพบว่า วิธีโค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์ เป็นวิธีที่เหมาะสมในทางทฤษฎีถูกต้อง และตรงที่สุด แต่มีข้อจำกัด คือ ต้องใช้กลุ่มตัวอย่างขนาดใหญ่ ส่วนวิธีอื่น ๆ ก็

สามารถใช้ได้ ถ้าผลการวิเคราะห์มีความสอดคล้องกับวิธีโค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์ ซึ่งเชปาร์ด และคณะได้เสนอว่า ในทางปฏิบัติสามารถใช้วิธีไคสแควร์ (Chi-square) แทนวิธีโค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์ และวิธีแปลงค่าความยาก (Angoff) สามารถนำมาใช้แทนวิธีโค้งลักษณะข้อสอบแบบ 1 พารามิเตอร์ (Rasch Model) ได้

แฮมเบิลตัน และคนอื่นๆ (Hambleton and Other, 1986) ได้ศึกษาเปรียบเทียบวิธีการวิเคราะห์ความลำเอียงของข้อสอบ 4 วิธี คือวิธีของแมนเทล-แฮนส์เชล วิธีการลงจุดค่าความยาก วิธีค่าความแตกต่างของค่าเฉลี่ยกำลังสอง (The Root Mean Squared Difference) และวิธีของพื้นที่รวม ซึ่ง 2 วิธีหลังนี้เป็นวิธีการที่ใช้ทฤษฎีการตอบข้อคำถาม แบบทดสอบที่ใช้ในการศึกษาเป็นแบบทดสอบที่ใช้ความสามารถในการอ่านของ คลิฟแลนด์ มีข้อสอบจำนวน 75 ข้อ โดยใช้กลุ่มตัวอย่างจำนวน 937 คน เป็นชาย 451 คนและเป็นหญิง 486 คน ใช้คะแนนจุดตัดในการแปลความหมายค่าสถิติที่แสดงความลำเอียงได้จากการจำลองข้อมูลขึ้น

ผลการศึกษาพบว่าวิธีวิเคราะห์ทั้ง 4 วิธี ให้ผลการวิเคราะห์ข้อสอบที่มีความลำเอียงได้ใกล้เคียงกัน แต่ปัญหาอยู่ที่ความไม่ชัดเจนในการกำหนดจุดตัด ความคลาดเคลื่อนชนิดที่หนึ่ง และการประมาณค่าพารามิเตอร์ที่ไม่ดี ข้อค้นพบที่เหตุได้อย่างเด่นชัด อยู่ที่ความสำคัญของการเลือกใช้ช่วงคะแนนความสามารถในการวัดความลำเอียง นอกจากนี้ พบว่า สามารถใช้วิธีของแมนเทล-แฮนส์เชล แทนวิธีที่ใช้ทฤษฎีการตอบข้อคำถาม เนื่องจากประหยัดกว่าทั้งเงินและเวลา

ดูลิทเทิล (Doolittle, 1987) ได้ศึกษาความแตกต่างในการตอบข้อสอบ ระหว่างผู้ที่มีเพศต่างกัน จากข้อสอบ 3 ประเภท ได้แก่ การคำนวณทางคณิตศาสตร์ เรขาคณิต และการใช้เหตุผลในการสอบ ACT (Assessment Mathematics Usual Test) ซึ่งมี 8 ฟอรัม ให้นักเรียนที่เป็นกลุ่มตัวอย่าง 8 กลุ่มทำ โดยจับคู่กลุ่มตัวอย่างชายและหญิง ซึ่งเป็นนักเรียนชั้นมัธยมศึกษาตอนปลายด้วย Profile ของวิชาวิทยาศาสตร์ที่ได้เรียนมาแล้ว และได้ออกแบบการวิเคราะห์ข้อมูลโดยใช้วิธี Split Plot Factorial Design ชนิด $2 \times 3 \times 8 \times 8$ (เพศ \times ประเภทของข้อสอบ \times ภูมิภาคด้านคณิตศาสตร์ \times ฟอรัมของแบบทดสอบ) จุดที่ต้องการเน้นในการศึกษาค้นคว้าครั้งนี้คือ ปฏิสัมพันธ์ระหว่างเพศกับประเภทของข้อสอบ

ผลการศึกษาแสดงให้เห็นว่า นักเรียนชายและนักเรียนหญิง ทำข้อสอบคำนวณได้ดีพอ ๆ กัน แต่นักเรียนหญิงจะทำข้อสอบเรขาคณิต และข้อสอบเชิงเหตุผลได้น้อยกว่านักเรียนชาย ซึ่งต้องใช้กระบวนการวางแผนก่อนการปัญหา ผลกระทบที่ใหญ่ที่สุดจากการศึกษาค้นคว้าครั้งนี้ คือ ผลกระทบที่เกิดจากภูมิภาคด้านคณิตศาสตร์ และเพศของกลุ่มตัวอย่าง

ดูลิทเทิล และเคลียร์ (Doolittle and Cleary, 1987) ได้ศึกษาวิธีวิเคราะห์ความลำเอียงของข้อสอบ โมเดลโลจิสติกแบบ 3 พารามิเตอร์ ที่วิเคราะห์ด้วยดัชนี Z ของลินน์ และฮาร์นิสซ์ (Linn and Harnisch, 1981) ระหว่างผู้สอบที่มีเพศต่างกัน โดยใช้ข้อสอบในแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ กลุ่มตัวอย่างที่ใช้มีประมาณ 1,300-1,400 คน เป็นหญิงประมาณร้อยละ 55 และแต่ละกลุ่มมีความเท่าเทียมกัน กลุ่มตัวอย่างจะแบ่งเป็นจำนวน 8 กลุ่มให้ทำแบบทดสอบ ACT (Assessment Mathematics Test) รวม 8 ฟอรัม ถ้าดัชนี Z ที่ได้เป็นบวก แสดงว่าข้อสอบข้อนั้นง่ายสำหรับกลุ่มเปรียบเทียบ ถ้าเป็นลบแสดงว่าข้อสอบข้อนั้นยากสำหรับกลุ่มเปรียบเทียบ

ผลการศึกษาพบว่า ดัชนี Z ที่คำนวณได้มีค่าเป็นลบ ในข้อสอบที่วัดด้านเรขาคณิตและการใช้เหตุผลเชิงพีชคณิต สำหรับข้อสอบที่วัดด้านเรขาคณิต การให้เหตุผลเชิงพีชคณิต และเรขาคณิตในทุกฟอรัม แสดงว่าง่ายสำหรับเพศชายมากกว่าเพศหญิง สำหรับดัชนี Z ที่เหลือส่วนมากมีค่าเป็นบวก

คิม และโคเฮน (Kim and Cohen, 1991) ได้ศึกษาความลำเอียงของข้อสอบที่สร้างขึ้นอย่างจงใจให้ลำเอียง โดยการเปรียบเทียบผลการคำนวณพื้นที่ความแตกต่างระหว่างโค้งที่ใช้จากทฤษฎีการตอบข้อคำถาม ของกลุ่มตัวอย่างผิวขาว และผิวดำ 2 วิธี ได้แก่ การคำนวณพื้นที่ภายใต้ความสามารถที่มีขอบเขตจำกัด และที่ไม่มีขอบเขตจำกัด

จำนวนกลุ่มตัวอย่างที่ใช้กลุ่มละ 1,000 คนซึ่งสุ่มมาจากเดิมที่มีผิวขาว 1,021 คน และผิวดำ 1,008 คน เพราะข้อจำกัด ในการใช้โปรแกรม BILOG ด้วยไมโคร-คอมพิวเตอร์ เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล คือ แบบทดสอบคำศัพท์เป็นแบบเลือกตอบ ชนิด 4 ตัวเลือก จำนวน 50 ข้อ เป็นคำศัพท์ภาษาอังกฤษมาตรฐาน 40 ข้อ อีก 10 ข้อ เป็นคำศัพท์ที่มีความง่ายสำหรับกลุ่มผิวดำ มากกว่ากลุ่มผิวขาว

การวิเคราะห์ความเป็นมิติเดียว (Unidimension) ด้วยการวิเคราะห์องค์ประกอบสำคัญ (Principal Componentes Analysis) พบว่า องค์ประกอบแรกที่ได้โดยไม่มีการหมุนแกน และใช้ค่าสหสัมพันธ์แบบเดตต้าคลอริด แสดงถึงความเป็นมิติเดียวได้พอเพียงในการแบบจำลองทฤษฎีการตอบข้อคำถาม โดยอธิบายความแปรปรวนได้ร้อยละ 25 ในกลุ่มผู้สอบผิวขาว ในขณะที่อธิบายได้ค่อนข้างน้อยในกลุ่มผู้สอบผิวดำ คืออธิบายได้ร้อยละ 16

การคำนวณพื้นที่ความแตกต่างระหว่างโค้งของกลุ่มตัวอย่าง 2 กลุ่ม โดยใช้โมเดลการตอบข้อคำถามแบบ 2 พารามิเตอร์และแบบ 3 พารามิเตอร์ ชนิดบังคับให้ค่าสัมประสิทธิ์การเดาของทั้งสองกลุ่มมีค่าเท่ากัน ($C = C_R = C_F = .23$) และใช้วิธีการวัดพื้นที่ใต้โค้ง 4 วิธี จากค่า

พารามิเตอร์ของข้อสอบที่ปรับแล้ว คือ พื้นที่แบบเปิดปิดเครื่องหมาย (Exact Unsigned Area: EUA) ได้จากการคำนวณพื้นที่ตลอดช่วงความสามารถโดยไม่จำกัดขอบเขตเพื่อให้ได้พื้นที่ที่แน่นอน พื้นที่แบบเปิดไม่คิดเครื่องหมาย (Exact Unsigned Area: EUA) พื้นที่แบบปิดคิดเครื่องหมาย (Closed-Interval Signed Area: CSA) และพื้นที่แบบปิดไม่คิดเครื่องหมาย (Closed-Interval Unsigned Area: CUA) กรณีที่เป็นพื้นที่ ที่คิดเครื่องหมายนั้นกลุ่มอ้างอิง (R) คือกลุ่มผิดค่า ดังนั้นถ้าค่า CSA หรือ ESA ที่คำนวณได้มีค่าเป็นบวก แสดงว่าข้อสอบข้อนั้นลำเอียงเข้าข้างกลุ่มผิดค่า

ข้อสอบที่สร้างเพื่อให้เกิดความลำเอียงเข้าข้างผิดค่า 10 ข้อ มีค่าเป็น 1 อีก 40 ข้อที่เหลือมีค่าเป็น 0 แล้วหาค่าสหสัมพันธ์แบบพอยท์ไบซีเรียล ระหว่างค่าดังกล่าวกับการวัดทั้ง 4 วิธี เพื่อดูผลการตรวจสอบแต่ละวิธี และหาค่าสหสัมพันธ์ระหว่างวิธีต่าง ๆ เพื่อดูความคล้ายคลึงกัน

การประมาณค่าอัตราความคลาดเคลื่อนของ การวิเคราะห์ความลำเอียง กระทำโดยมีข้อตกลงว่าการวัดพื้นที่แต่ละวิธีมีการกระจายเป็นแบบปกติ ค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานของการวัดแต่ละวิธีจะประมาณจากข้อสอบ 40 ข้อ กำหนดค่าวิกฤตแบบทางเดียวเป็น 2 ค่า คือ ที่ระดับ .05 และ .01 ข้อที่มีพื้นที่ใหญ่กว่าค่าวิกฤตจะเป็นข้อสอบที่มีความลำเอียง การจัดประเภทข้อสอบเช่นนี้ทำให้เกิดความคลาดเคลื่อน 2 ชนิด ได้แก่ ความคลาดเคลื่อนชนิดระบุว่าไม่ลำเอียง ทั้งๆ ที่ข้อสอบนั้นลำเอียง (FN = False Negatives) และระบุว่าข้อสอบลำเอียงทั้งๆ ที่ไม่ลำเอียง (FP = False Positives) และเมื่อใช้ค่าวิกฤตที่ $p = .05$ พบว่ามี FP มากกว่าที่ค่าวิกฤต $p = .01$

ผลการศึกษาพบว่า ในการวิเคราะห์ความลำเอียงของข้อสอบด้วยการคำนวณพื้นที่แบบช่วงเปิด และแบบช่วงปิดทั้งสองวิธีเกือบไม่มีความแตกต่าง เมื่อพิจารณาการใช้พื้นที่แบบไม่คิดเครื่องหมายไม่ว่าจะเป็นแบบช่วงเปิดหรือแบบช่วงปิด มีความไวต่อการตรวจค้นความลำเอียงของข้อสอบมากกว่าแบบที่คิดเครื่องหมาย นอกจากนั้นยังพบว่า การคำนวณพื้นที่แบบช่วงเปิด พื้นที่ชนิดคิดเครื่องหมายและไม่คิดเครื่องหมาย มีความสัมพันธ์กันมากกว่า พื้นที่ชนิดคิดเครื่องหมายและไม่คิดเครื่องหมายของการคำนวณพื้นที่แบบช่วงปิด

ในด้านของความคลาดเคลื่อนในการวิเคราะห์ความลำเอียง พบว่า วิธีคำนวณพื้นที่แบบ 3 พารามิเตอร์ โดยใช้พื้นที่แบบคิดเครื่องหมายมีอัตราความคลาดเคลื่อนต่ำกว่าวิธีอื่นๆ และไม่มี ความคลาดเคลื่อนเลยที่ระดับนัยสำคัญ .01 แต่มีความคลาดเคลื่อนในการระบุว่าข้อสอบลำเอียง โดยที่ไม่ลำเอียงจริง จำนวน 2 ข้อ ที่ระดับนัยสำคัญ .05 ส่วนวิธีคำนวณพื้นที่แบบ 3 พารามิเตอร์แบบไม่มีเครื่องหมายจะมีอัตราความคลาดเคลื่อนสูงกว่าวิธีอื่นๆ เล็กน้อย

ราชู และคนอื่นๆ (Raju and Others, 1993) ได้ศึกษาเปรียบเทียบผลการทดสอบ ไคสแควร์ ของลอร์ด การทดสอบ Z ของราชู (1990) ในการวัดพื้นที่ความแตกต่างระหว่างโค้ง ลักษณะข้อสอบทั้งพื้นที่ชนิดที่คิดเครื่องหมายและไม่คิดเครื่องหมาย โดยใช้โมเดล 2 พารามิเตอร์ และวิธีไคสแควร์ของแมนเทิล-แฮนส์เซล ในการวิเคราะห์ความลำเอียงของข้อสอบใช้เกณฑ์การ ตัดสินความลำเอียง ดังนี้คือ χ^2 ของ Lord มีค่าเท่ากับ 13.80 Z มีค่าอยู่ระหว่าง ± 8.30 และ ไคสแควร์ของแมนเทิล-แฮนส์เซล มีค่าเท่ากับ 10.83 โดยใช้ข้อมูลจากผลการตอบแบบทดสอบ Gater-Mac Ginitie Reading Tests (GMRT) เฉพาะแบบทดสอบย่อยที่วัดคำศัพท์ 45 ข้อ เป็นแบบ เลือกตอบชนิด 5 ตัวเลือก กลุ่มตัวอย่างที่ใช้เป็นนักเรียนระดับ 10 และ 12 ในปี 1987 จำนวน 839 คน เป็นนักเรียนผิวดำ 245 คน นักเรียนผิวขาว 436 คน นักเรียนหญิง 440 คนและนักเรียน ชาย 399 คน

ผลการศึกษาพบว่าในกรณีการวิเคราะห์ความลำเอียงระหว่างนักเรียนผิวดำ และนักเรียน ผิวขาว จากการทดสอบด้วย χ^2 และ Z ทั้งพื้นที่ชนิดคิดเครื่องหมายและไม่คิดเครื่องหมาย ทั้ง 3 วิธี พบข้อสอบที่มีความลำเอียง 1 ข้อ ได้แก่ ข้อ 41 ส่วนวิธีไคสแควร์ของแมนเทิล-แฮนส์เซล พบข้อสอบที่มีความลำเอียง 2 ข้อ คือ ข้อ 14 และ 27 ซึ่งข้อสอบทั้ง 3 ข้อ เป็น ข้อสอบที่มีค่าความยากต่างกันมากระหว่างกลุ่มผู้สอบผิวดำและผิวขาว

ส่วนผลการวิเคราะห์ความลำเอียงระหว่างนักเรียนชายและนักเรียนหญิง 3 วิธีแรก พบ ข้อสอบที่มีความลำเอียงซ้ำกัน 4 ข้อ ส่วนวิธีไคสแควร์ของแมนเทิล-แฮนส์เซล พบข้อสอบที่มี ความลำเอียง 5 ข้อ ซ้ำกับ 3 วิธีแรก 4 ข้อ คิดเป็นร้อยละ 80 ได้แก่ ข้อ 2, 18, 23, 33 และข้อ สอบที่ไม่มี ความลำเอียงซ้ำกับ 3 วิธีแรก คือข้อ 41 และข้อสอบทุกข้อมีความยากสูงกว่าข้อสอบ ที่ไม่มีความลำเอียง

จากการค้นพบว่า วิธีไคสแควร์ของแมนเทิล-แฮนส์เซล ค้นพบข้อสอบที่มีความลำเอียง ได้มากกว่าวิธีที่ใช้ทฤษฎีการตอบข้อคำถามเช่นนี้ คำนกับการค้นพบของสวามินาธานและโรเจอร์ส (1990) ที่พบว่าวิธีที่ใช้ทฤษฎีการตอบข้อคำถามสามารถค้นพบข้อสอบที่มีความลำเอียงได้มากกว่า วิธีของแมนเทิล-แฮนส์เซล

ข้อควรสังเกตจากการวิจัยครั้งนี้ คือ การทดสอบ Z ของราชู อาจให้ผลที่ถูกต้องกว่า การกำหนดเกณฑ์ตัดสินความแตกต่างระหว่างพื้นที่ได้โค้งลักษณะข้อสอบตามอำเภอใจ เพราะอาจ เกิดค่าความคลาดเคลื่อนมาตรฐานในการวัด ซึ่งเป็นตัวหนึ่งที่มีการมาเกี่ยวข้อง ทำให้ข้อสอบที่มี ความแตกต่างของพื้นที่น้อยกว่าเป็นข้อสอบที่มีความลำเอียง และข้อสอบที่มีค่าความแตกต่างของ พื้นที่มากกว่าเป็นข้อสอบที่ไม่มีความลำเอียงได้

โคเฮน และคิม (Cohen and Kim, 1993) ได้ศึกษาเปรียบเทียบการใช้ทฤษฎีการตอบข้อคำถามแบบ 2 พารามิเตอร์ โดยใช้การทดสอบ χ^2 ของลอร์ด และการทดสอบความแตกต่างของพื้นที่ระหว่างโค้งลักษณะข้อสอบชนิดคิดเครื่องหมาย โดยใช้สถิติ Z ของราชู จากข้อมูลจำลองขึ้นตามทฤษฎีการตอบข้อสอบ 2 พารามิเตอร์ ซึ่งจำลองข้อสอบ 2 ขนาดคือ 20 และ 60 ข้อ จำนวนผู้สอบ 2 ขนาด คือ 100 และ 500 คน และมีจำนวนข้อสอบลำเอียงในแบบทดสอบที่จำลอง 3 แบบ คือ มีข้อสอบที่ลำเอียงร้อยละ 0, 10 และ 20

ผลการศึกษา พบว่า แบบทดสอบที่ยาวการระบุข้อสอบลำเอียงผิดพลาด (FP) จะมากขึ้น และถ้ามีจำนวนข้อสอบลำเอียงอยู่ในแบบทดสอบน้อย จำนวนข้อสอบที่ถูกระบุลำเอียงผิดพลาดจะยิ่งเพิ่มมากขึ้นด้วย องค์ประกอบที่มีอิทธิพลต่อการระบุข้อสอบลำเอียงผิดพลาด (FP) อีกประการ คือ ระดับนัยสำคัญ (α) ถ้าค่าระดับนัยสำคัญยิ่งมากจะเกิด FP มาก นั่นคือที่ $\alpha = .05$ จะเกิด FP มากกว่าที่ $\alpha = .01$ สำหรับการทดสอบ Z ของราชูเกิดในทุกสถานการณ์ที่ศึกษา แต่การทดสอบ χ^2 ของลอร์ด จะเกิดในบางสถานการณ์เท่านั้น

และพบว่าแบบทดสอบที่ยาวกว่า มีร้อยละของข้อสอบลำเอียงมากกว่าจะมีข้อสอบที่ถูกระบุว่ามีข้อผิดพลาด (FN) มากกว่า แต่จะเกิดขึ้นกับการทดสอบด้วย Z ของราชู มากกว่าการทดสอบด้วย χ^2 ของลอร์ด

กรณีที่มีจำนวนผู้สอบเท่ากับ 100 คน พบว่า ความแตกต่างในการระบุข้อสอบไม่ลำเอียงผิดพลาด สำหรับ Z ของราชู และ χ^2 ของลอร์ด ไม่แตกต่างกันมากเท่ากับกรณีที่ใช้กลุ่มตัวอย่าง 500 คน ซึ่งมีการระบุข้อสอบไม่ลำเอียงผิดพลาด สำหรับ χ^2 ของลอร์ด น้อยกว่า Z ของราชู โดยเฉพาะอย่างยิ่งในกรณีที่มีข้อสอบลำเอียงในแบบทดสอบมากที่สุด คือร้อยละ 20

โคเฮน และคิม ได้ให้ความเห็นว่า การทดสอบด้วย χ^2 ของลอร์ดให้ผลการตรวจสอบข้อสอบลำเอียงได้ดีกว่าการทดสอบ Z ของพื้นที่ ที่คิดเครื่องหมายอาจเป็นเพราะ Z (ESA) ใช้การเปรียบเทียบเฉพาะค่าความยาก (b) ในขณะที่ χ^2 ใช้ทั้งค่าความยากและค่าอำนาจจำแนก (b และ a) แต่ไม่สามารถอธิบายความแตกต่างได้ในกรณีใช้ Z(H) ซึ่งทดสอบพื้นที่ที่ไม่คิดเครื่องหมายเพราะ Z(H) ก็ใช้ทดสอบค่าพารามิเตอร์ทั้ง 2 ตัวเช่นกัน

อย่างไรก็ตาม จากผลการวิจัยเปรียบเทียบวิเคราะห์ความลำเอียงของข้อสอบที่กล่าวมาปรากฏว่าวิธีวิเคราะห์ความลำเอียงของข้อสอบไม่ว่าจะศึกษาจากข้อมูลสมมติ (Simulate Data) หรือศึกษาจากข้อมูลจริง (Empirical Data) ทุกวิธีมีบางส่วนที่สัมพันธ์กัน และมีบางส่วนที่แตกต่างกัน และให้ผลสรุปที่คล้ายคลึงกัน