บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

เอกสารและงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบผู้วิจัยนำเสนอตามลำดับ ดังนี้

1. การประเมินคุณภาพการศึกษาขั้นพื้นฐาน
2. ความเป็นมาของการศึกษาความลำเอียงของข้อสอบ
3. ความหมายของความลำเอียงของข้อสอบ
4. สาเหตุที่ทำให้ข้อสอบมีความลำเอียง
5. การตรวจสอบความลำเอียงของข้อสอบ
   5.1 การใช้วิธีภายนอก
   5.2 การใช้วิธีภายใน
6. วิธีการตรวจสอบความลำเอียงของข้อสอบ
   6.1 การวิเคราะห์ความลำเอียงของข้อสอบตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิม
   6.2 การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้ตารางการณ์จร
   6.3 การวิเคราะห์ความลำเอียงของข้อสอบตามแนวคิดทฤษฎีการตอบสนองข้อสอบ
7. งานวิจัยที่เกี่ยวข้อง
   7.1 งานวิจัยภายในประเทศ
   7.2 งานวิจัยต่างประเทศ

1. การประเมินคุณภาพการศึกษาขั้นพื้นฐาน

สำนักงานเขตพื้นที่การศึกษาปัตตานี เขต 2 (2551 : 1-12) ได้ดำเนินงานตามโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อการประกันคุณภาพผู้เรียนซึ่งมีรายละเอียดในการดำเนินงาน พอสรุปได้ดังนี้

ความสำคัญของการประเมิน

หลักสูตรการศึกษาขั้นพื้นฐานเป็นหลักสูตรแกนกลางสำหรับการจัดการศึกษาขั้นพื้นฐานที่มุ่งพัฒนาคนไทยให้เป็นมนุษย์ที่สมบูรณ์ทั้งสติปัญญา ความรู้ ร่างกาย จิตใจ อารมณ์ สังคมและคุณธรรมจริยธรรมในการดำรงชีวิต มีค่านิยมที่ดี สามารถอยู่ร่วมกันกับผู้อื่นได้อย่างมีความสุขและเน้นการสร้างกระบวนการเรียนรู้ที่นำไปสู่การพัฒนาคนและการดำรงชีวิตในสังคมการวัดและ

Prince of Songkla University Pattani Campus

ประเมินผลถือว่าเป็นส่วนหนึ่งของการจัดกิจกรรมการเรียนการสอนของครูที่มีความสำคัญยิ่ง และมีบทบาทสำคัญต่อการพัฒนาการเรียนรู้ของผู้เรียนทั้งทางตรงและทางอ้อม ผลทางตรง คือ จะให้ข้อมูลย้อนกลับที่จะนำไปสู่การพัฒนาการเรียนรู้ที่มีประสิทธิภาพ ส่วนผลทางอ้อม คือ การประเมินผลจะเป็นตัวชี้นำการเรียนการสอน ดังนั้นผู้บริหารสถานศึกษาจึงสามารถนำผลการประเมินมาใช้ให้เกิดประโยชน์อย่างแท้จริงต่อการจัดการศึกษา

การวัดและการประเมินผลการเรียนรู้จึงเป็นกระบวนการหนึ่งที่ให้ผู้สอนใช้พัฒนาคุณภาพผู้เรียน เพราะจะได้ข้อมูลสารสนเทศที่แสดงพัฒนาการความก้าวหน้า และความสำเร็จทางการเรียนของผู้เรียนรวมทั้งข้อมูลที่จะเป็นประโยชน์ต่อการส่งเสริมให้ผู้เรียนเกิดการพัฒนาการเรียนรู้อย่างเต็มศักยภาพและเพื่อให้การจัดการศึกษาของสถานศึกษามีคุณภาพที่แท้จริง ซึ่งสถานศึกษาในฐานะผู้รับผิดชอบในการจัดการศึกษาจึงต้องมีผลการเรียนรู้ของผู้เรียนจากการวัดและประเมินทั้งในระดับชั้นเรียน ระดับสถานศึกษา ระดับเขตพื้นที่การศึกษาและระดับชาติ ตลอดจนการประเมินภายนอกเพื่อใช้เป็นข้อมูลสร้างความมั่นใจเกี่ยวกับคุณภาพผู้เรียนให้แก่ผู้เกี่ยวข้องทั้งภายในและภายนอกสถานศึกษาซึ่งโดยหลักการการประเมินคุณภาพระดับชาติ สถานศึกษาจะต้องจัดให้ผู้เรียนทุกคนที่เรียนในแต่ละชั้นปีและในปีสุดท้ายของแต่ละช่วงชั้นคือ ชั้นประถมศึกษาปีที่3 ชั้นประถมศึกษาปีที่ 6 ชั้นมัธยมศึกษาปีที่3 และชั้นมัธยมศึกษาปีที่6 เข้ารับการประเมินในระดับชาติในกลุ่มสาระการเรียนรู้ที่สำคัญคือ ภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ สังคมศึกษา ศาสนาและวัฒนธรรม และภาษาต่างประเทศ ข้อมูลที่ได้จากการประเมินจะนำไปใช้ในการพัฒนาคุณภาพผู้เรียนและคุณภาพการจัดการศึกษาของสถานศึกษาแต่ละแห่งต่อไป

หลักการดังกล่าวสอดคล้องกับนโยบายหลักของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานซึ่งได้กำหนดว่า ในปีการศึกษา2550 นี้ให้ดำเนินการสอบวัดเพื่อประเมินผลการจัดการศึกษาโดยมีเป้าหมายเพื่อการประกันคุณภาพผู้เรียนตามที่พระราชบัญญัติการศึกษาแห่งชาติได้กำหนดไว้ในมาตรา47 ว่า ให้มีระบบการประกันคุณภาพการศึกษาเพื่อพัฒนาคุณภาพและมาตรฐานการศึกษาทุกระดับและมาตรา48 ให้หน่วยงานต้นสังกัดและสถานศึกษาจัดให้มีระบบการประกันคุณภาพการศึกษาภายในสถานศึกษาและต้องดำเนินการอย่างต่อเนื่องโดยการจัดทำรายงานประจำปีเสนอต่อหน่วยงานต้นสังกัด หน่วยงานที่เกี่ยวข้องและเปิดเผยต่อสาธารณชนเพื่อนำไปสู่การพัฒนาคุณภาพและมาตรฐานการศึกษาและเพื่อรองรับการประกันคุณภาพภายนอกการประเมินคุณภาพการศึกษาขั้นพื้นฐานจึงเป็นกระบวนการ วิธีการเพื่อให้ได้ข้อมูลที่จะเป็นตัวบ่งชี้ถึงผลสำเร็จในการจัดการศึกษาซึ่งเป็นส่วนประกอบสำคัญในการประกันคุณภาพภายใน หลักสูตรสถานศึกษา พ.ศ. 2544 จึงกำหนดแนวการวัดและประเมินผลการเรียนรู้เพื่อให้ได้ข้อมูลสารสนเทศที่แสดงพัฒนาการความก้าวหน้าและสำเร็จทางการเรียนของผู้เรียน ซึ่งสถานศึกษาต้องจัดให้มีการ

Prince of Songkla University Pattani Campus

ประเมินผลการเรียนรู้ให้เป็นไปในมาตรฐานเดียวกันทั้งในระบบชั้นเรียนระดับสถานศึกษา ระดับเขตพื้นที่การศึกษา และระดับชาติ ข้อมูลที่ได้จากการประเมินจะนำไปใช้ในการพัฒนาคุณภาพของผู้เรียนและคุณภาพการจัดการศึกษาของสถานศึกษาแต่ละแห่งและเพื่อเป็นสารสนเทศรองรับบริบทของการประเมินภายนอกดังที่กล่าวไว้ข้างต้นสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) และทุกเขตพื้นที่การศึกษายังมีจุดมุ่งหมายเพื่อเป็นการตรวจสอบ ควบคุม กำกับ ดูแล และรักษาคุณภาพการศึกษาของสถานศึกษาจึงได้กำหนดโครงการให้มีการประเมินผลสัมฤทธิ์นักเรียนในทุกโรงเรียนทุกเขตพื้นที่การศึกษาทั่วประเทศ ให้ได้มาตรฐานในปีการศึกษา2550 โดยที่สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานจะประเมินในช่วงชั้นที่ 1 (ประถมศึกษาปีที่ 3) ช่วงชั้นที่ 3 (มัธยมศึกษาปีที่3) สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ประเมินในช่วงชั้นที่2 (ประถมศึกษาปีที่ 6) และช่วงชั้นที่ 4 (มัธยมศึกษาปีที่6) โดยมีเป้าหมายเพื่อประเมินนักเรียนทุกคนครอบคลุมทุกโรงเรียน สำหรับผลการประเมินทุกช่วงชั้นที่ได้จะเป็นตัวบ่งชี้คุณภาพการศึกษาขั้นพื้นฐานในภาพรวมและใช้เป็นข้อมูลสำคัญประกอบการตัดสินใจในการกำหนดนโยบายเพื่อวางแผนการพัฒนาคุณภาพการศึกษาของเขตพื้นที่การศึกษา ตลอดจนใช้ปรับปรุงพัฒนาตนเองของผู้เรียนและการจัดการเรียนการสอนของสถานศึกษา

สำนักทดสอบทางการศึกษาได้รับมอบหมายจากสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานให้จัดการประเมินคุณภาพการศึกษาขั้นพื้นฐานและสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ซึ่งดูแลการจัดสอบนักเรียนในช่วงชั้นที่2 ทั่วประเทศ จึงได้พิจารณารูปแบบและวิธีการประเมินที่เหมาะสมร่วมกับสำนักงานเขตพื้นที่การศึกษาทั่วประเทศ เพื่อให้ได้ผลการประเมินที่น่าเชื่อถือและสภาวะบ่งชี้คุณภาพการศึกษาที่ชัดเจนของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน เขตพื้นที่การศึกษาและสถานศึกษาได้อย่างสมเหตุสมผลสำนักงานเขตพื้นที่การศึกษาปัตตานี เขต 2 ตั้งอยู่ในพื้นที่3 จังหวัดชายแดนภาคใต้และได้รับผลกระทบจากสถานการณ์ความไม่สงบอยู่อย่างต่อเนื่องอันทำให้เกิดการสูญเสียชีวิตและทรัพย์สินตลอดจนความไม่ปลอดภัยในการเดินทาง การเสียขวัญและกำลังใจในการปฏิบัติงานอย่างต่อเนื่องตลอดระยะเวลา 5 ปีที่ผ่านมา ครูผู้สอนต้องเสี่ยงภัยในการเดินทางไปปฏิบัติราชการบ้างก็ขอย้ายออกนอกพื้นที่บ้างก็ขอย้ายกลับภูมิลำเนาเดิม อัตรากำลังครูในบางโรงเรียนจึงไม่เพียงพอ ขณะเดินทางและระหว่างการปฏิบัติงานอยู่ที่โรงเรียนก็ต้องคอยระวังภัย จึงส่งผลให้การจัดการเรียนการสอนไม่เป็นไปอย่างต่อเนื่อง อย่างไรก็ตามภารกิจหลักคือ การจัดการศึกษา ด้วยเหตุนี้สำนักงานเขตพื้นที่การศึกษาปัตตานี เขต 2 จึงได้มีความตระหนักเล็งเห็นความจำเป็นและความสำคัญในการดำเนินการประเมินผลสัมฤทธิ์นักเรียนทุกช่วงชั้นและนักเรียนทุกคนในปีการศึกษา 2550

Prince of Songkla University Pattani Campus

ดังนั้น ในปีการศึกษา2550 สำนักงานเขตพื้นที่การศึกษาปัตตานี เขต2 โดยกลุ่มนิเทศติดตามและประเมินผลการจัดการศึกษา กลุ่มงานวัดและประเมินผลการจัดการศึกษาจึงได้ดำเนินการประเมินผลสัมฤทธิ์ทางการเรียนของนักเรียนทุกคนที่เรียนในปีสุดท้ายของช่วงชั้นที่1 (ประถมศึกษาปีที่ 3) ช่วงชั้นที่ 2 (ประถมศึกษาปีที่ 6) และช่วงชั้นที่ 3 (มัธยมศึกษาปีที่3) ประเมินผลสัมฤทธิ์ทางการเรียนของนักเรียนทุกคนในชั้นประถมศึกษาปีที่2 ชั้นประถมศึกษาปีที่ 5 และมัธยมศึกษาปีที่ 2 ตามที่สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) และสำนักทดสอบทางการศึกษาได้กำหนดขึ้นและเมื่อดำเนินการเสร็จสิ้นให้สำนักงานเขตพื้นที่การศึกษาสรุปรายงานผลการประเมินรายบุคคลและโดยภาพรวมให้กับทุกสถานศึกษา เขตพื้นที่การศึกษาและสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานทราบ

วัตถุประสงค์

1. เพื่อรายงานสรุปผลการดำเนินงานการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อการประกันคุณภาพผู้เรียน
2. เพื่อรายงานผลสัมฤทธิ์นักเรียนในชั้นประถมศึกษาปีที่2 ชั้นประถมศึกษาปีที่ 5 มัธยมศึกษาปีที่2 ชั้นประถมศึกษาปีที่ 3และมัธยมศึกษาปีที่ 3
3. เพื่อรายงานผลสัมฤทธิ์นักเรียน (ผล O-NET) ในช่วงชั้นที่ 2 (ชั้นประถมศึกษาปีที่ 6)
4. เพื่อให้ได้ข้อมูลย้อนกลับสำหรับใช้ในกระบวนการตัดสินใจและกำหนดแผนพัฒนาคุณภาพการศึกษา ระดับเขตพื้นที่การศึกษาและระดับสถานศึกษา

เป้าหมาย

ในปีการศึกษา 2550 มีนักเรียนกลุ่มเป้าหมายในแต่ละระดับชั้นและช่วงชั้นดังนี้

ชั้นประถมศึกษาปีที่ 2	(LAS)	จำนวน 3,325 คน
ชั้นประถมศึกษาปีที่ 5	(LAS)	จำนวน 3,963 คน
ชั้นมัธยมศึกษาปีที่2	(LAS)	จำนวน 3,528 คน
ชั้นประถมศึกษาปีที่ 3	(NT)	จำนวน 3,374 คน
ชั้นประถมศึกษาปีที่ 6	(O-NET)	จำนวน 3,704 คน
ชั้นมัธยมศึกษาปีที่3	(NT)	จำนวน 2,698 คน

การประเมินผลสัมฤทธิ์นักเรียน ปีการศึกษา2550 สำนักทดสอบทางการศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานได้ดำเนินการจัดสอบวัดผลสัมฤทธิ์ทางการเรียน (General Achievement Test) ทุกคนในชั้นประถมศึกษาปีที่2 ชั้นประถมศึกษาปีที่ 5 และมัธยมศึกษาปีที่2 และช่วงชั้นที่ 3 (มัธยมศึกษาปีที่3) ดำเนินการสอบโดยสอบอย่างครอบคลุมทุกโรงเรียนโดยใช้

Prince of Songkla University Pattani Campus

เครื่องมือที่เป็นแบบทดสอบของสำนักทดสอบทางการศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ซึ่งมีรายละเอียดดังนี้

1. ชั้นประถมศึกษาปีที่ 2 จัดสอบวัดผลสัมฤทธิ์ทางการเรียนกลุ่มสาระการเรียนรู้ภาษาไทยและคณิตศาสตร์

2. ชั้นประถมศึกษาปีที่ 5 จัดสอบวัดผลสัมฤทธิ์ทางการเรียนกลุ่มสาระการเรียนรู้ภาษาไทย คณิตศาสตร์และภาษาต่างประเทศ (ภาษาอังกฤษ)

3. ชั้นมัธยมศึกษาปีที่2 จัดสอบวัดผลสัมฤทธิ์ทางการเรียนกลุ่มสาระการเรียนรู้ภาษาไทย คณิตศาสตร์และภาษาต่างประเทศ (ภาษาอังกฤษ)

4. ชั้นมัธยมศึกษาปีที่3 (ช่วงชั้นที่3) จัดสอบวัดผลสัมฤทธิ์ทางการเรียนกลุ่มสาระการเรียนรู้ภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ ภาษาอังกฤษ สังคมศึกษา ศาสนาและวัฒนธรรม

ส่วนสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ได้ดำเนินการจัดสอบประเมินผลสัมฤทธิ์นักเรียน (O-NET) ในช่วงชั้นที่ 2 (ประถมศึกษาปีที่ 6) ทุกโรงเรียนและทุกคน โดยจัดสอบรายวิชา ภาษาไทย คณิตศาสตร์ วิทยาศาสตร์และภาษาต่างประเทศ (ภาษาอังกฤษ)

ขอบเขตการประเมิน

การประเมินคุณภาพการศึกษาขั้นพื้นฐานเพื่อการประกันคุณภาพผู้เรียน ปีการศึกษา2550 จะเป็นการประเมินผลสัมฤทธิ์ของผู้เรียนตามจุดมุ่งหมายของหลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช2544 โดยการจัดการประเมินเป็น 2 ระดับ คือ

1. ส่วนกลาง ประเมินโดยวิธีสอบนักเรียนทุกคนในช่วงชั้นที่1 (ชั้นประถมศึกษาปีที่ 3) และช่วงชั้นที่ 3 (ชั้นมัธยมศึกษาปีที่3) ใช้แบบทดสอบชนิดเลือกตอบ (Multiple choices) โดยมีสาระการประเมินดังนี้

ช่วงชั้นที่ 1 (ชั้นประถมศึกษาปีที่ 3) สอบ NT วิชาภาษาไทยและคณิตศาสตร์

ช่วงชั้นที่ 2 (ชั้นประถมศึกษาปีที่ 6) สอบ O-NET วิชาภาษาไทย คณิตศาสตร์ วิทยาศาสตร์และภาษาต่างประเทศ (ภาษาอังกฤษ)

ช่วงชั้นที่ 3 (ชั้นมัธยมศึกษาปีที่3) สอบวิชาภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ ภาษาต่างประเทศ (ภาษาอังกฤษ) และสังคมศึกษา ศาสนาและวัฒนธรรม

2. สำนักงานเขตพื้นที่การศึกษาประเมินนักเรียนทุกคนในชั้นประถมศึกษาปีที่2 ชั้นประถมศึกษาปีที่ 5 และมัธยมศึกษาปีที่2 โดยประเมินในสาระสำคัญที่ต้องเร่งปรับปรุงเพื่อพัฒนาผู้เรียนเป็นรายบุคคล คือ กลุ่มสาระการเรียนรู้ภาษาไทย คณิตศาสตร์และภาษาต่างประเทศ (ภาษาอังกฤษ)

Prince of Songkla University Pattani Campus

ขอบเขตการรายงานผล

การรายงานผลการประเมิน รายงานการประมวลและวิเคราะห์ผลจากการประเมินทั้งในระดับส่วนกลางและระดับท้องถิ่น คือ

1. รายงานผลการประเมินชั้นประถมศึกษาปีที่2 ชั้นประถมศึกษาปีที่ 5 และมัธยมศึกษาปีที่ 2 โดยประมวลผลระดับเขตพื้นที่การศึกษา

2. รายงานผลการประเมินชั้นประถมศึกษาปีที่3 (NT) และมัธยมศึกษาปีที่3 (NT) โดยใช้ผลการประเมินและประมวลผลจากส่วนกลาง คือ สำนักทดสอบทางการศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน

3. รายงานผลการประเมินชั้นประถมศึกษาปีที่6 (O-NET) และมัธยมศึกษาปีที่6 (O-NET) โดยใช้ผลการประเมินและประมวลผลจากส่วนกลาง คือ สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)

ระยะเวลาดำเนินการ

การดำเนินการจัดสอบนักเรียน ได้ดำเนินการตามวัน เวลา ที่กำหนดพร้อมกันทั่วประเทศตามตารางสอบ ดังนี้

การจัดสอบ NT (National Test)

ชั้นประถมศึกษาปีที่ 3 และมัธยมศึกษาปีที่3 สอบวันที่12 กุมภาพันธ์2551

การจัดสอบ LAS (Local Area Assessment)

ชั้นประถมศึกษาปีที่2 ชั้นประถมศึกษาปีที่ 5และมัธยมศึกษาปีที่2 สอบวันที่19 กุมภาพันธ์2551

การจัดสอบ O-NET

ชั้นประถมศึกษาปีที่6 สอบวันที่4 มีนาคม 2551

แนวคิดในการประเมิน

การประเมินผลสัมฤทธิ์ทางการเรียน ปีการศึกษา2550 มีแนวคิดหลักในการประเมิน ดังนี้

1. การประเมินผลสัมฤทธิ์ทางการเรียนเป็นกระบวนการตรวจสอบว่า เมื่อจัดกระบวนการเรียนรู้ตามมาตรฐานการเรียนรู้และเนื้อหาสาระแล้วประสบผลสำเร็จมากน้อยเพียงใด

2. การวัดและประเมินผลเป็นกระบวนการที่ช่วยผู้สอนให้ได้ข้อมูลสารสนเทศที่แสดงถึงพัฒนาการความก้าวหน้าและความสำเร็จของการเรียนรู้ของผู้เรียนซึ่งเป็นประโยชน์ต่อการส่งเสริมให้ผู้เรียนเกิดการพัฒนาและเรียนรู้เต็มตามศักยภาพ

3. วิธีการวัดและประเมินผลต้องใช้วิธีการที่หลากหลายสอดคล้องกับสิ่งที่ต้องการวัดทั้งที่เป็นความรู้ ความคิดและการปฏิบัติ ส่วนมากแล้วมักจะใช้แบบทดสอบ อาจเป็นแบบปรนัยหรือไม่

Prince of Songkla University Pattani Campus

ก็อัตนัยนอกจากนี้ยังสามารถใช้แฟ้มสะสมงาน การสังเกต การสัมภาษณ์ ซึ่งเป็นวิธีที่เหมาะกับการวัดและการประเมินความดีงามและการปฏิบัติ

4. สถานศึกษาจะต้องจัดให้ผู้เรียนทุกคนที่เรียนในชั้นประถมศึกษาปีที่2 ชั้นประถมศึกษาปีที่ 5และมัธยมศึกษาปีที่2 รวมทั้งนักเรียนที่เรียนปีสุดท้ายของช่วงชั้นที่2 คือ ชั้นประถมศึกษาปีที่ 6 และช่วงชั้นที่ 3 คือ ชั้นมัธยมศึกษาปีที่3 เข้ารับการประเมินในระดับชาติ ในกลุ่มสาระการเรียนรู้ที่สำคัญตามที่กระทรวงศึกษาธิการกำหนด

นอกจากแนวคิดหลักดังกล่าวสำนักงานเขตพื้นที่การศึกษาปัตตานี เขต2 ได้ดำเนินการวัดและประเมินผลในทุกสถานศึกษา โดยเน้นการวัดด้านความรู้ ความคิดและด้านการปฏิบัติ เครื่องมือที่ใช้ประเมินเป็นแบบทดสอบปรนัย ในทุกระดับช่วงชั้น คือ ช่วงชั้นที่1 (ชั้นประถมศึกษาปีที่ 3) กลุ่มสาระการเรียนรู้หลัก คือภาษาไทย คณิตศาสตร์ ช่วงชั้นที่3 (มัธยมศึกษาปีที่3) คือ แบบทดสอบแบบเลือกตอบ กลุ่มสาระการเรียนรู้หลัก คือภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ ภาษาต่างประเทศ (ภาษาอังกฤษ) สังคมศึกษา ศาสนาและวัฒนธรรม ส่วนระดับชั้นประถมศึกษาปีที่ 2 กลุ่มสาระการเรียนรู้หลัก คือ ภาษาไทยและคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 5และชั้นมัธยมศึกษาปีที่2 ใช้แบบทดสอบลักษณะแบบปรนัยและสอบเฉพาะสาระการเรียนรู้ที่สำคัญ คือ ภาษาไทย คณิตศาสตร์และภาษาต่างประเทศ (ภาษาอังกฤษ)

แนวคิดในการสร้างแบบทดสอบ

แบบทดสอบวัดผลสัมฤทธิ์ทั่วไป(General Achievement Test) เป็นแบบทดสอบที่วัดความรู้ความสามารถในสาขากลุ่มสาระการเรียนรู้ต่าง ๆ ที่หลักสูตรกำหนดเนื้อหาสาระของข้อสอบ (Test Content) ได้จากการวิเคราะห์วัตถุประสงค์และคำอธิบายรายวิชา สื่อ ตำรา และเอกสารหลักสูตร แบบทดสอบทุกฉบับจะวัดความรู้ความเข้าใจ เกี่ยวกับสาระสำคัญและทักษะกระบวนการ (Process Skill) เช่นทักษะทางวิทยาศาสตร์ ทักษะการแสวงหาความรู้(Inquiry Skill) ทักษะการแก้ปัญหาและกระบวนการคิดอย่างมีวิจารณญาณ(Critical Thinking Process) เป็นต้น โดยไม่เน้นการท่องจำ ข้อมูลหรือข้อเท็จจริงที่เป็นรายละเอียดปลีกย่อยโดยทั่วไป ข้อสอบที่วัดทักษะเชิงกระบวนการจะมีจำนวนมากกว่าข้อสอบที่วัดความรู้ความเข้าใจ โดยได้กำหนดอัตราส่วน 60 : 40 และเป็นแบบทดสอบชนิดเลือกตอบ โดยที่ชั้นประถมศึกษาปีที่ 2 มี 3 ตัวเลือก ระดับสูงขึ้นไปมี 4 ตัวเลือก

ขอบข่ายของแบบทดสอบ

แบบทดสอบมาตรฐานปีการศึกษา2550 ตามหลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2544 แสดงดังตาราง 1

Prince of Songkla University Pattani Campus

ตาราง 1 แสดงจำนวนข้อของแบบทดสอบตามกลุ่มสาระการเรียนรู้ในแต่ละช่วงชั้น

ช่วงชั้นที่	กลุ่มสาระการเรียนรู้	จำนวนข้อ/คะแนนเต็ม
1 (ป.2) LAS	ภาษาไทย	30
	คณิตศาสตร์	30
2 (ป.5) LAS	ภาษาไทย	40
	คณิตศาสตร์	40
	ภาษาต่างประเทศ (ภาษาอังกฤษ)	40
3 (ม.2) LAS	ภาษาไทย	40
	คณิตศาสตร์	40
	ภาษาต่างประเทศ (ภาษาอังกฤษ)	40
1 (ป.3) NT	ภาษาไทย	30
	คณิตศาสตร์	30
3 (ม.3) NT	ภาษาไทย	40
	คณิตศาสตร์	40
3 (ม.3) (ต่อ) NT	วิทยาศาสตร์	40
	สังคมศึกษา ศาสนาและวัฒนธรรม	40
	ภาษาต่างประเทศ (ภาษาอังกฤษ)	40
2 (ป.6) O-NET	ภาษาไทย	40
	คณิตศาสตร์	40
	วิทยาศาสตร์	40
	ภาษาต่างประเทศ (ภาษาอังกฤษ)	40

Prince of Songkla University Pattani Campus

เกณฑ์การตรวจให้คะแนน

เกณฑ์การประเมินตามโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐานเพื่อการประกันคุณภาพผู้เรียน ประจำปีการศึกษา2550 ได้กำหนดเกณฑ์การตรวจให้คะแนนในทุกกลุ่มสาระการเรียนรู้และกำหนดระดับคุณภาพ ดังนี้

1. เกณฑ์การตรวจให้คะแนนคำตอบของแบบทดสอบทุกกลุ่มสาระการเรียนรู้แต่ละข้อ ตอบถูกให้ 1 คะแนน ตอบผิดให้ 0 คะแนน

2. เกณฑ์การประเมินผลรวมของคะแนนที่ได้โดยการจำแนกเป็นระดับคุณภาพ กำหนดเกณฑ์เป็น 3 ระดับ คือ

2.1 ระดับปรับปรุง

2.2 ระดับพอใช้

2.3 ระดับดี

ในแต่ละระดับชั้นกลุ่มสาระการเรียนรู้กำหนดคะแนนจุดตัด ช่วงของคะแนนระดับคุณภาพแสดงดังตาราง 2

ตาราง 2 แสดงคะแนนจุดตัดตามกลุ่มสาระการเรียนรู้ในแต่ละช่วงชั้น

ช่วงชั้น/ชั้น	สาระการเรียนรู้	คะแนนเต็ม	คะแนนจุดตัด		ระดับคุณภาพ		
			ขั้นผ่าน	ขั้นสูง	ปรับปรุง	พอใช้	ดี
1 (ป.2)	ภาษาไทย	30	12	19	1-11	12-18	19-30
LAS	คณิตศาสตร์	30	10	19	1-9	10-18	19-30
2 (ป.5)	ภาษาไทย	40	12	19	1-11	12-18	19-40
LAS	คณิตศาสตร์	40	10	19	1-9	10-18	19-40
	ภาษาอังกฤษ	40	14	20	1-13	14-19	20-40
3 (ม.2)	ภาษาไทย	40	12	19	1-11	12-18	19-40
LAS	คณิตศาสตร์	40	10	19	1-9	10-18	19-40
	ภาษาอังกฤษ	40	14	20	1-13	14-19	20-40

Prince of Songkla University Pattani Campus

2. ความเป็นมาของการศึกษาความลำเอียงของข้อสอบ

การศึกษาคุณภาพข้อสอบของผู้สอบกลุ่มต่างๆ ในประชากรมีมานานแล้ว แต่ด้านความยุติธรรมของข้อสอบหรือแบบสอบระหว่างผู้สอบกลุ่มต่างๆ เริ่มศึกษากันอย่างจริงจังในช่วงปลายทศวรรษของปี ค.ศ. 1960 มีการเสนอวิธีการต่างๆ เพื่อตรวจสอบความลำเอียงของข้อสอบ(Item bias) ความลำเอียงของแบบสอบ(Test bias) และความลำเอียงในการคัดเลือก(Selection bias) โดยนิยามความลำเอียงว่าเป็น ความคลาดเคลื่อนอย่างเป็นระบบ(Systematic error) ที่เกิดขึ้นจากการวัด ความพยายามของการตรวจสอบความลำเอียงดังกล่าว ดำเนินไปเพื่อจำแนกข้อสอบที่ทำหน้าที่ไม่เหมาะสมหรือไม่ยุติธรรมสำหรับปรับปรุง หรือตัดข้อสอบข้อนั้นออกจากแบบสอบเป็นการขจัดข้อสอบที่เกิดปัญหาความยุติธรรมระหว่างกลุ่มข้อสอบกลุ่มต่างๆ ที่มีลักษณะบางอย่างแตกต่างกัน เช่น เชื้อชาติ ศาสนา วัฒนธรรม ภูมิลำเนา สังคม เพศ ภาษา อายุ ประสบการณ์ เป็นต้น เพื่อพัฒนาแบบสอบให้มีคุณภาพเหมาะสมสำหรับนำไปใช้ทดสอบต่อไป(ศิริชัย กาญจนวาสี, 2550 : 115)

การศึกษาความลำเอียงของข้อสอบนั้น เริ่มมาจากการสังเกตผลการสอบคัดเลือก ซึ่งพบว่าไม่เป็นไปตามสัดส่วนสติปัญญาหรือโครงสร้างของประชากร กรณีที่ทำให้มีการตื่นตัวมากในเรื่องความลำเอียงของข้อสอบคือ ในปี ค.ศ. 1971 มาร์โคเดอฟูนิส (Marco Kefunis) และคนอื่นๆ ซึ่งถูกปฏิเสธจากโรงเรียนกฎหมายของมหาวิทยาลัยวอชิงตัน ได้ฟ้องร้องว่าเขาได้คะแนนสูงกว่าผู้ที่ได้รับการคัดเลือกบางคน และได้ยื่นฎีกาฟ้องร้อง ชาร์ล โอเดการ์ด(Charles Odegaurs) เพื่อให้พิจารณาทบทวนการคัดเลือกนักศึกษาใหม่ และจากนั้นเป็นต้นมาการพิจารณาตรวจสอบความลำเอียงของข้อสอบระหว่างผู้สอบกลุ่มย่อยก็ได้รับปฏิบัติกันจนถึงปัจจุบัน(Breland and Ironson, 1976 : 89) ซึ่งการตรวจสอบความลำเอียงจะทำการตรวจสอบก่อนและหลังจากการทดลองใช้ การสำรวจความลำเอียงของข้อสอบก่อนนำไปทดลองใช้จะตรวจสอบโดยผู้เชี่ยวชาญ โดยพิจารณาถึงรูปแบบของข้อสอบ เนื้อหา คำที่ใช้และอื่นๆ เพื่อไม่ให้เกิดความลำเอียง(รัชดาพร แก้วชาฎก, 2544 : 14)

ศิริชัย กาญจนวาสี (2550 : 116) ยังเสนอแนวคิดเกี่ยวกับการใช้คำและความหมายโดยมีประเด็นโต้แย้งว่าความลำเอียงของข้อสอบ เป็นผลการตัดสินว่าข้อสอบมีความยุติธรรมหรือไม่ อันส่งผลต่อการบรรลุจุดมุ่งหมายของการใช้แบบสอบหรือความลำเอียงของข้อสอบ เป็นสารสนเทศทางสถิติที่ได้จากข้อสอบเกี่ยวกับความสัมพันธ์ระหว่างคุณลักษณะที่ข้อสอบมุ่งวัดกับประสบการณ์ของผู้สอบกลุ่มต่างๆ ที่ทำการสอบเมื่อกลุ่มผู้สอบต่างกลุ่มกันตอบข้อสอบข้อเดียวกัน ความแตกต่างที่เกิดขึ้นอาจมาจากความไม่เหมาะสมของข้อคำถาม ซึ่งสามารถเกิดขึ้นได้หลายลักษณะหรือประสบการณ์ของผู้สอบซึ่งอาจมีลักษณะพื้นฐานเดิมแตกต่างกันในหลายสถานการณ์จึง

ไม่เหมาะสมที่จะใช้คำว่า ข้อสอบลำเอียง(Biased item) เนื่องจากเป็นภาษาที่มีความหมายในเชิงลบ ประกอบกับเกณฑ์ที่ใช้สำหรับตัดสินความลำเอียงยังมีความคลุมเครือและค่อนข้างสับสน ดั้งนั้นจึงควรเปลี่ยนมาใช้คำว่า กรทำหน้าที่ต่างกันของข้อสอบ(Differential Item Functioning ; DIF) ซึ่งเป็นคำที่มีความเป็นกลางและเหมาะสมกว่า

3. ความหมายของความลำเอียงของข้อสอบ

นักวิจัยทางการวัดผลหลายท่านได้ให้ความหมายของความลำเอียงของข้อสอบและการทำหน้าที่ต่างกันของข้อสอบ(ศิริชัย กาญจนวาสี, 2550 : 116-117) ไว้ดังนี้

ความลำเอียงของข้อสอบ หมายถึง สัดส่วนของผู้สอบที่ตอบข้อสอบได้ถูกต้องไม่เท่ากันในแต่ละกลุ่มประชากรที่ใช้ในการศึกษา เมื่อกลุ่มผู้สอบมีคะแนนเท่ากันและข้อสอบมีความเป็นเอกพันธ์(Scheuneman, 1979)

ความลำเอียงของข้อสอบ หมายถึง ข้อสอบที่มีค่าความยากสัมพัทธ์สำหรับสมาชิกของผู้สอบกลุ่มหนึ่งมากกว่าสมาชิกของผู้สอบอีกกลุ่มหนึ่ง(Rudner, Geston and Knight, 1980)

ความลำเอียงของข้อสอบ หมายถึง ความโน้มเอียงของข้อสอบที่เมื่อใช้คะแนนจากข้อสอบนั้นแล้ว ทำให้การตัดสินผลเป็นไปอย่างไม่ยุติธรรม(Popham, 1981)

ความลำเอียงของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกันสำหรับการวัดความสามารถหรือโอกาสในการตอบข้อสอบในทางบวกแตกต่างกันสำหรับการวัดเจตคติ เมื่อผู้สอบที่มีคุณลักษณะของการวัดในปริมาณเท่ากัน แต่มาจากกลุ่มประชากรย่อยที่แตกต่างกัน (Hulin, Drasgow and Parson, 1983)

ความลำเอียงของข้อสอบ หมายถึงโอกาสในการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มหนึ่งมีค่าต่ำกว่าหรือสูงกว่าผู้สอบอีกกลุ่มหนึ่งที่มีระดับความสามารถเดียวกัน(Dorans and Kulick, 1986)

ความลำเอียงของข้อสอบ หมายถึง คะแนนข้อสอบของกลุ่มผู้สอบที่มีความสามารถเท่ากันแต่มาจากต่างกลุ่มกัน มีความแตกต่างกันอย่างเป็นระบบ(Kederman, 1990)

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง สารสนเทศทางสถิติของข้อสอบที่ได้จากผลการตอบของผู้สอบต่างกลุ่มกันและมีความสามารถเท่ากัน แต่มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน (Holland and Wainer, 1993)

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ความเป็นพหุมิติในการวัดของข้อสอบซึ่งแสดงได้จากการแจกแจงความสามารถหลัก(Primary ability) ของกลุ่มผู้สอบตั้งแต่2 กลุ่มขึ้นไปมีความ

Prince of Songkla University Pattani Campus

เท่ากันแต่มีการแจกแจงความสามารถรอง (Secondary ability) แตกต่างกัน (Camili and Shepard,1994)

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ฟังก์ชันการตอบสนองข้อสอบซึ่งคำนวณจากกลุ่มผู้สอบกลุ่มย่อยที่ต่างกัน มีค่าไม่เท่ากัน(Narayanan and Swaminathan, 1996)

การทำหน้าที่ต่างกันของข้อสอบ(DIF) หมายถึง การที่ข้อสอบทำให้ผู้สอบจากต่งกลุ่มกันที่มีความสามารถหรือคุณลักษณะที่มุ่งวัดเท่ากัน มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน หรือมีฟังก์ชันตอบสนองข้อสอบแตกต่างกัน การทำหน้าที่ต่างกันของข้อสอบเกิดขึ้นเมื่อนำข้อสอบไปทดสอบกับผู้สอบกลุ่มย่อยต่างกัน ที่มีความสามารถหลัก(Primary ability) ระดับเดียวกันหรือมีคุณลักษณะแฝง(Latent trait) ที่ต้องการวัดเท่ากัน แต่มีความสามารถรอง(Secondary ability) แตกต่างกัน ทำให้ผู้สอบต่างกลุ่มที่นำมาจับคู่เปรียบเทียบมีโอกาสตอบข้อสอบถูกแตกต่างกัน (ศิริชัย กาญจนวาสี, 2550 : 117)

นอกจากนี้ก็ยังมีผู้ให้ความหมายไว้อีกมากมาย เช่น

เอลเลน และ เยน (Allen and Yen, 1979 : 129) ให้ความหมายของข้อสอบที่มีความลำเอียงว่าเป็นข้อสอบที่วัดความสามารถ หรือคุณลักษณะทางจิตวิทยาของผู้สอบแต่ละกลุ่มได้ไม่ตรงกัน เช่น เมื่อนำข้อสอบวัดความสามารถด้านคณิตศาสตร์ที่ใช้ภาษาสื่อความหมายของโจทย์ปัญหาไปทดสอบกับผู้สอบ2 กลุ่มที่มีความสามารถทางการอ่านไม่เท่ากัน ข้อสอบนั้นมีความลำเอียงต่อผู้สอบที่มีความสามารถทางการอ่านสูง

ลอร์ด (Lord, 1980 : 212) สับโคเวียค และคนอื่นๆ (Subkovaik and others, 1984 : 49) แมคคอลเลอร์ และเมนโดซา (McCanley and Mendoza, 1985 : 389) แฮมเบิลตัน และคณะ (Hambleton et. Al, 1991 : 110) ฮอล์แลนด์ และไวเนอร์ (Holland and Wainer, 1993 : 4, quoting Angoff, 1993) ให้ความหมายของข้อสอบที่มีความลำเอียงว่า เป็นข้อสอบที่ผู้สอบที่มีความสามารถเท่ากัน แต่อยู่ต่างกลุ่มกันมีโอกาสตอบข้อสอบข้อนั้นได้ถูกไม่เท่ากัน

ออสเตอร์ลายด์ (Osterlind, 1983 : 55) ให้ความหมายว่า ความลำเอียงเป็นความคลาดเคลื่อนอย่างมีระบบทางการวัด

เคลียรีและฮิลตัน (Cleary and Hilton, 1986 : 61) ให้ความหมายของข้อสอบที่มีความลำเอียงว่า เป็นข้อสอบที่เมื่อนำไปทดสอบกับบุคคลตั้งแต่สองกลุ่มขึ้นไปที่มีความสามารถเท่ากันแล้ว ปรากฏว่ามีคะแนนเฉลี่ยของข้อสอบข้อเดียวกันของแต่ละกลุ่มไม่เท่ากัน

คามิลลิ และเซพเพิร์ด (Camilli and Shepard, 1994 : 8) ได้ให้ความหมายของข้อสอบที่มีความลำเอียงว่า เป็นความคลาดเคลื่อนอย่างเป็นระบบหรือความไม่เที่ยงตรงของการวัด ทำให้

Prince of Songkla University Pattani Campus

สมาชิกบางกลุ่มได้ประโยชน์หรือเสียประโยชน์ เช่น แบบทดสอบหรือข้อสอบที่เข้าข้างกลุ่มนักเรียนชายมากกว่านักเรียนหญิง หรือคนผิวขาวกับคนผิวดำหรือระหว่างคนที่อยู่ในภูมิภาคต่างกัน

ถึงแม้ว่าปัจจุบัน นักวัดผลและนักวิจัยหันมาใช้คำว่าการทำหน้าที่ต่างกันของข้อสอบแทนคำว่า ความลำเอียงของข้อสอบ หรือคำอื่นๆ เนื่องจากเห็นว่าเป็นคำที่มีความหมายเป็นกลาง และเหมาะสมในเชิงวิชาการ แต่ก็ยังมีจุดเน้นที่ต่างกัน โดยความลำเอียงเป็นคำที่ใช้กันในทางสังคม เน้นสิ่งที่สังเกตได้ของกลุ่มผู้เข้าสอบและอาศัยเกณฑ์การตัดสินที่แสดงถึงความลำเอียง ส่วนการทำหน้าที่ต่างกันของข้อสอบเน้นลักษณะทางสถิติ ซึ่งเป็นส่วนหนึ่งของสิ่งที่แสดงถึงความลำเอียงของข้อสอบ

จากความเป็นมาและความหมายข้างต้น ผู้วิจัยใช้คำว่าข้อสอบที่มีความลำเอียงซึ่งผู้วิจัยมองถึงประเด็นการตรวจสอบความลำเอียงโดยการที่จะระบุว่าข้อสอบหรือแบบทดสอบนั้นมีความลำเอียงหรือไม่ต้องอาศัยเกณฑ์การตัดสินโดยใช้สถิติวิเคราะห์ข้อสอบ และการศึกษาครั้งนี้ ผู้วิจัยใช้สถิติวิเคราะห์ที่มีเกณฑ์การตัดสินว่าข้อสอบหรือแบบทดสอบนั้นมีความลำเอียงหรือไม่และยังบ่งบอกได้ว่าลำเอียงไปยังกลุ่มอ้างอิง (Reference Group) หรือกลุ่มเปรียบเทียบ (Focal Group) ดังนั้นพอสรุปได้ว่าแบบทดสอบที่มีความลำเอียงหมายถึง แบบทดสอบที่เมื่อนำไปทดสอบกับกลุ่มผู้สอบที่มีความสามารถเท่าเทียมกันจากประชากรเดียวกัน แต่มีลักษณะของกลุ่มย่อยแตกต่างกัน เช่น เพศ เชื้อชาติ ศาสนา วัฒนธรรม สภาพภูมิศาสตร์ เศรษฐกิจหรือแม้แต่ประสบการณ์เป็นต้น แล้วปรากฏว่าความน่าจะเป็นในการตอบข้อสอบข้อเดียวกันของกลุ่มผู้สอบถูกไม่เท่ากัน ทำให้มีการได้เปรียบเสียเปรียบจากการตอบข้อสอบเดียวกันและใช้เกณฑ์จากการวิเคราะห์ผลการตอบข้อสอบตามกลุ่มย่อยของผู้สอบเพื่อบ่งบอกว่ากลุ่มใดเป็นฝ่ายได้เปรียบหรือเสียเปรียบจากการตอบข้อสอบ

4. สาเหตุที่ทำให้ข้อสอบมีความลำเอียง

จากการวิเคราะห์ความลำเอียงด้วยค่าสถิติแบบต่างๆ ในแบบทดสอบหลายๆฉบับ เพื่อศึกษาถึงสาเหตุที่ทำให้เกิดความลำเอียงในตัวคำถามเหล่านั้น พบว่าสาเหตุที่น่าจะก่อให้เกิดความลำเอียงมีหลายสาเหตุด้วยกัน ซึ่งสามารถสรุปสาเหตุที่ก่อให้เกิดความลำเอียงของข้อสอบมากที่สุดได้ดังต่อไปนี้ (Scheunerman, 1982 : 180-195 อ้างถึงใน รัชดาพร แก้วชาฎา, 2544 : 13)

1. เดา (Guess) อาจจะเกิดจากข้อสอบนั้นยากเกินไป หรือเวลาไม่เพียงพอจะก่อให้เกิดความไม่เท่าเทียมกันในโอกาสการตอบข้อสอบถูกของผู้สอบแต่ละคน

2. ความเร็ว (Speed) หรือเวลาในการตอบจะทำให้เกิดการเดาหรือในกรณีเวลาน้อยอาจจะทำข้อสอบไม่ทันซึ่งจะมีผลกับข้อสอบข้อหลังๆ โดยเฉพาะในการศึกษาความลำเอียงของข้อสอบวัดความถนัด

3. ความกำกวมหรือความไม่ชัดเจน(Unclear) ของข้อคำถาม นั่นคือข้อคำถามขาดความเป็นปรนัย การใช้ภาษาถิ่นหรือใช้คำที่ไม่เป็นภาษากลางในการสื่อความหมายซึ่งจะก่อให้เกิดความลำเอียงกับกลุ่มภาษาใดภาษาหนึ่งขึ้นได้

4. ลำดับขั้นของคำถาม(Series) อาจจะเป็นสิ่งที่ก่อให้เกิดความสับสนหรือชี้แนะคำตอบข้อสอบบางข้อได้

5. สถานภาพทางสังคมหรือความเกี่ยวข้องทางสังคม(Social Implication) ก็เป็นสิ่งที่ก่อให้เกิดความแตกต่างระหว่างกลุ่มผู้สอบแต่ละกลุ่มได้

6. ประสบการณ์หรือการฝึกฝนของแต่ละกลุ่มย่อยเป็นสิ่งที่ก่อให้เกิดการได้เปรียบเสียเปรียบของแต่ละกลุ่มค่อนข้างจะชัดเจน

7. องค์ประกอบทางวัฒนธรรม ความเป็นอยู่ ขนบธรรมเนียมประเพณี เชื้อชาติ ศาสนา ก็จะเอื้อให้กับบางกลุ่มย่อย จึงทำให้เกิดการได้เปรียบเสียเปรียบในบางเนื้อหาวิชาได้

นอกจากนี้ความลำเอียงของข้อสอบอาจจะเกิดจากสาเหตุที่สำคัญ2 ประการ คือ

1. การเลือกเนื้อหา (Bias in Selection) คือผู้สร้างข้อสอบเลือกเนื้อหาเฉพาะส่วนใดส่วนหนึ่งมาสร้างข้อสอบทำให้ได้ข้อสอบที่มีเนื้อหาไม่ครอบคลุมและไม่ได้สัดส่วนที่สมดุลกัน

2. การสร้างข้อสอบ (Bias in Construction) คือการใช้ภาษาหรือข้อความบางอย่างที่เอื้อให้เกิดประโยชน์กับผู้สอบกลุ่มใดกลุ่มหนึ่ง

5. การตรวจสอบความลำเอียงของข้อสอบ

การศึกษาความลำเอียงของข้อสอบที่มีพื้นฐานจากการใช้หลักสถิติสามารถทำได้ 2 แนวทาง ได้แก่

1. การใช้วิธีภายนอก (External Method) การวิเคราะห์ความลำเอียงโดยใช้วิธีภายนอกหรือเกณฑ์ภายนอกนี้ สามารถนำไปใช้ได้ทั้งข้อสอบรายข้อและแบบทดสอบทั้งฉบับ วิธีการวิเคราะห์ จะเขียนกราฟแสดงความสัมพันธ์ระหว่างคะแนนจากตัวแปรเกณฑ์ภายนอกกับตัวแปรทำนาย วิธีนี้มีจุดมุ่งหมายที่จะวิเคราะห์ความถดถอยของตัวแปรทั้งสอง แล้วทำการเปรียบเทียบค่าความชัน (Slope) ค่าการตัดแกน (Intercept) ของเส้นกราฟ ในการวิเคราะห์ความลำเอียงของข้อสอบทั้งฉบับจะใช้คะแนนรวมเป็นตัวแปรทำนาย แต่ถ้าเป็นการวิเคราะห์รายข้อจะใช้ค่า

Prince of Songkla University Pattani Campus

ความยาก (p) ของข้อสอบแต่ละข้อเป็นตัวแปรทำนาย ส่วนตัวแปรเกณฑ์ภายนอกจะใช้คะแนนรวมหรือเกรดเกี่ยวกับผลสัมฤทธิ์ของงานบางอย่างที่ให้ทำ(Cronbach, 1970 ; Anastasi, 1976)

การใช้เกณฑ์ภายนอกมีข้อดี คือเกณฑ์ที่ใช้มีความเป็นอิสระจากข้อสอบและแบบสอบที่ต้องการตรวจสอบ แต่มีจุดอ่อนตรงที่ความเหมาะสมของเกณฑ์ที่จะนำมาใช้ในทางปฏิบัติเป็นการยากที่จะหาตัวแปรเกณฑ์ภายนอกจากแบบสอบฉบับอื่นที่มีความตรงเชิงทำนายและมีความยุติธรรมสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ถ้าตัวแปรเกณฑ์ภายนอกขาดคุณสมบัติดังกล่าวจะทำให้ผลการวิเคราะห์ขาดความแม่นยำและความสมบูรณ์(ศิริชัย กาญจนวาสี, 2550 : 121)

2. การใช้วิธีภายใน (Internal Method)) การวิเคราะห์ความลำเอียงโดยใช้วิธีภายในหรือเกณฑ์ภายใน เป็นวิธีที่นิยมใช้กันอย่างกว้างขวาง เป็นการนำวิธีการทางสถิติมาตรวจสอบโดยเน้นการตรวจสอบความตรงเชิงโครงสร้าง สามารถแบ่งเป็น 3 กลุ่มใหญ่ ๆ (Camilli and Shepard, 1994) ดังนี้

2.1 การตรวจสอบความลำเอียงของข้อสอบตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) เป็นวิธีการตรวจสอบความลำเอียงของข้อสอบที่ใช้กันมาเนิ่นนานแม้ว่าปัจจุบันวิธีนี้ไม่เป็นที่นิยมเนื่องจากนักวัดผลคิดค้นได้เทคนิคใหม่ๆ ที่ให้ผลดีกว่ารวมทั้งวิธีการตรวจสอบแนวนี้มีจุดอ่อนตรงที่ไม่สามารถแยกค่าความยากของข้อสอบออกจากค่าอำนาจจำแนกได้ ซึ่งวิธีการในกลุ่มนี้ได้แก่ วิธีการแปลงค่าความยาก วิธีGolden Rule วิธีวิเคราะห์ความแปรปรวนและวิธีพอยต์ไบซีเรียล เป็นต้น

2.2 การตรวจสอบความลำเอียงของข้อสอบตามแนวคิดทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) วิธีการตรวจสอบความลำเอียงของข้อสอบในแนวนี้ จะพิจารณาเปรียบเทียบฟังก์ชันการตอบสนองข้อสอบหรือเส้นโค้งคุณลักษณะระหว่างกลุ่มตั้งแต่2 กลุ่มขึ้นไปว่ามีความแตกต่างกันหรือไม่

การตรวจสอบความลำเอียงของข้อสอบตามแนวคิดของทฤษฎีการตอบสนองข้อสอบมีจุดเด่นกว่าวิธีตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิม ดังนี้

1) ทฤษฎีการตอบสนองข้อสอบสามารถประมาณค่าพารามิเตอร์ของข้อสอบ เช่น ค่าความยาก (b) ค่าอำนาจจำแนก(a) และค่าการเดา (c) ทำให้สามารถแยกค่าพารามิเตอร์ข้อสอบได้

2) คุณสมบัติทางสถิติของข้อสอบ สามารถอธิบายได้ละเอียดและชัดเจนกว่าเมื่อข้อสอบลำเอียงระหว่าง2 กลุ่ม

3) คุณสมบัติของข้อสอบสามารถแสดงด้วยแผนภาพที่เป็นกราฟทำให้เข้าใจง่ายและนำไปใช้ได้กว้างขวาง

Prince of Songkla University Pattani Campus

2.3 การตรวจสอบความลำเอียงของข้อสอบโดยใช้ตารางการณ์จร (Contingency Table Approach) วิธีการตรวจสอบความลำเอียงของข้อสอบในแนวนี้มีลักษณะดังนี้

1) ใช้สถิตินอนพาราเมตริกในการทดสอบ

2) ใช้คะแนนที่สังเกตได้

3) มีความสอดคล้องมากกับการเปรียบเทียบระหว่างเส้นกราฟคุณลักษณะของข้อสอบตามแนวทฤษฎีการตอบสนองข้อสอบ

4) ไม่จำเป็นต้องใช้สถิติขั้นสูง

5) ประหยัด

วิธีการตรวจสอบในกลุ่มนี้ได้แก่ ไคสแควร์ แมนเทล – แฮนส์เซล และลอกลิเนียร์ เป็นต้น

6. วิธีการตรวจสอบความลำเอียงของข้อสอบ

วิธีการตรวจสอบความลำเอียงของข้อสอบที่มักจะพบเจอในงานวิจัยพอสรุปได้ ดังนี้

1. การวิเคราะห์ความลำเอียงของข้อสอบตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิม

1.1 วิธีการแปลงค่าความยากของข้อสอบ (Transformed Item Difficulty)

เป็นการวิเคราะห์ความลำเอียงของข้อสอบตามวิธีการของ แองกอฟฟ์(Berk, 1982 : 97, quoting Angoff, 1972 ; also Angoff & Ford, 1973) คือจะต้องคำนวณหาค่าความยาก (P-Values) ของข้อสอบแต่ละข้อจากผู้สอบแต่ละกลุ่ม(คือกลุ่ม 1 กลุ่ม 2 ดังนั้นค่าความยากของข้อสอบแต่ละข้อจะมี 2 ค่า ตามจำนวนกลุ่มที่แบ่ง) ในการเปรียบเทียบค่าความยาก (P-Values) นี้จะใช้วิธีแปลงค่าความยาก (P-Values) เป็นคะแนนมาตรฐาน Z จากคะแนนมาตรฐาน Z แปลงให้เป็นความยากมาตรฐานเดลตา (Δ) ด้วยสมการ $\Delta = 4Z + 13$ แล้วนำค่าเดลตา(Delta) แต่ละคู่มาลงจุด คู่อันดับบนกราฟ ซึ่งแทนค่าเดลตาจะอยู่ในรูปวงรี แล้วคำนวณหาระยะทางที่จุดต่างๆ เหล่านี้ห่างจากเส้นแกนหลัก (Major Axis) ข้อสอบข้อใดที่ห่างจากเส้นแกนหลักน้อยกว่า 3Sd หรือมากกว่า 3Sd จะสรุปว่าข้อสอบข้อนั้นเป็นข้อสอบที่มีความลำเอียง ซึ่งคำนวณได้ดังนี้

จากสมการเส้นแกนหลัก

$$y = ax + b$$

เมื่อ a แทน ความชันของเส้นแกนหลัก

Prince of Songkla University Pattani Campus

b แทน ค่าคงที่ของการตัดแกน Y

โดยที่ค่า a และ b คำนวณได้จากสูตร

$$a = \frac{(S^2_y - S^2_x) + \sqrt{(S^2_y - S^2_x)^2 + 4r_{xy}{}^2 S^2_x S^2_y}}{2r_{xy} S_x S_y}$$

$$b = M_y - a M_x$$

เมื่อ x แทน ค่าเดลตาของกลุ่มที่ 1

y แทน ค่าเดลตาของกลุ่มที่ 2

M_x แทน ค่าเฉลี่ยของค่าเดลตา ของกลุ่มที่ 1

M_y แทน ค่าเฉลี่ยของค่าเดลตา ของกลุ่มที่ 2

S^2_x แทน ค่าแปรปรวนของค่าเดลตากลุ่มที่ 1

S^2_y แทน ค่าแปรปรวนของค่าเดลตากลุ่มที่ 2

S_x แทน ค่าความเบี่ยงเบนมาตรฐานของค่าเดลตากลุ่มที่ 1

S_y แทน ค่าความเบี่ยงเบนมาตรฐานของค่าเดลตากลุ่มที่ 2

r_{xy} แทน สหสัมพันธ์ของค่าเดลตา กลุ่มที่ 1 และกลุ่มที่ 2

แองกอฟฟ์ (Angoff) เสนอการคำนวณระยะทางตั้งฉากของจุดคู่อันดับค่าเดลตาของข้อสอบแต่ละข้อไปยังเส้นแกนหลัก จากสูตร

$$d_i = \frac{ax_i - y_i + b}{\sqrt{a^2 + 1}}$$

และ $S_d = \sqrt{1 - r_{xy}}$

เมื่อ d_i แทน ระยะทางตั้งฉากของจุดคู่อันดับค่าเดลตาของข้อสอบแต่ละข้อไปยังเส้นแกนหลัก

S_d แทน ค่าเบี่ยงเบนมาตรฐานของระยะ d

Prince of Songkla University Pattani Campus

2. การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้ตารางการณ์จร

2.1 วิธีของแมนเทล-แฮนส์เซล (Mantel-Haenszel)

วิธีแมนเทล-แฮนส์เซล (MH) เป็นเทคนิคที่แมนเทล-แฮนส์เซลได้เสนอขึ้นใช้ตั้งแต่ปี ค.ศ.1959 แต่ฮอลแลนด์และเทเยอร์ (Holland and Thayer) นำเสนอเพื่อใช้การตรวจสอบความลำเอียงของข้อสอบในปี1988 และก็มีการตรวจสอบด้วยวิธีแมนเทล-แฮนส์เซล หรือเรียกย่อๆ ว่า MH เป็นวิธีที่มีความคล้ายคลึงกับวิธีไคสแควร์ที่เสนอโดยชูเนแมน (Scheuneman, 1979) โดยการเปรียบเทียบผลการตอบข้อสอบของกลุ่มผู้สอบ 2 กลุ่มหรือที่ฮอลแลนด์ เรียกว่า กลุ่มอ้างอิงและกลุ่มเปรียบเทียบ (Reference and Focal Group) โดยกลุ่มใดกลุ่มหนึ่งเสียประโยชน์จากข้อสอบในกรณีที่ข้อสอบมีความลำเอียงการตรวจสอบด้วยวิธีการนี้ทำได้โดยแยกวิเคราะห์ข้อสอบเป็นรายข้อ พิจารณาผลการตอบข้อสอบของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีคะแนนรวมเท่ากันซึ่งตรวจสอบทุกๆ ระดับคะแนนรวมจากการสอบแล้วแสดงความถี่ของผู้เข้าสอบทั้งสองกลุ่มที่ตอบข้อนั้นถูก(1) ผิด (0) วิธีการในการวิเคราะห์มีดังนี้

ถ้าให้ N_{Rj} และ N_{Fj} แทนจำนวนผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีคะแนนรวมในช่วงคะแนน j และให้ T_j ซึ่งเท่ากับ $N_{Rj} + N_{Fj}$ เป็นจำนวนผู้สอบรวมทั้ง 2 กลุ่มที่ได้คะแนนรวมจากการสอบอยู่ในช่วงคะแนน j แล้วนำมาเขียนในรูปตาราง 2 ทาง แสดงผลการตอบถูก (1) และผิด (0) แสดงดังตาราง 3

ตาราง 3 แสดงข้อมูลการสอบของกลุ่มตัวอย่างกลุ่ม R และ F ที่มีคะแนนรวมอยู่ในช่วงคะแนน j

		คะแนนที่ได้จากข้อสอบที่ต้องการวิเคราะห์ความลำเอียง		
		1	0	รวม
กลุ่มผู้สอบ	R	A_j	B_j	N_{Rj}
	F	C_j	D_j	N_{Fj}
	รวม	N_{1j}	N_{0j}	T_j

จะได้ว่าในช่วงระดับคะแนน j

1. A_j และ B_j แทนจำนวนผู้สอบกลุ่มอ้างอิงที่ตอบข้อสอบที่ถูกต้องและไม่ถูกต้องตามลำดับในช่วงคะแนน j

2. C_j และ D_j แทนจำนวนผู้สอบกลุ่มเปรียบเทียบที่ตอบข้อสอบถูกต้องและไม่ถูกต้องตามลำดับในช่วงคะแนน j

Prince of Songkla University Pattani Campus

จากตาราง 3 นำมาคำนวณค่าความน่าจะเป็นในรูปของค่าสัดส่วนการตอบถูกผิดระหว่างกลุ่มโดยกำหนดให้เป็นค่า α_{MH} ดังนี้

$$\alpha_{MH} = \frac{\sum_{j=1}^{s} A_j D_j / T_j}{\sum_{j=1}^{s} B_j C_j / T_j}$$

เมื่อ α_{MH} แทน สัดส่วนการตอบถูกผิดระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

A_j แทน จำนวนผู้ตอบถูกในกลุ่มอ้างอิงที่ระดับคะแนน j

B_j แทน จำนวนผู้ตอบผิดในกลุ่มอ้างอิงที่ระดับคะแนน j

C_j แทน จำนวนผู้ตอบถูกในกลุ่มเปรียบเทียบที่ระดับคะแนน j

D_j แทน จำนวนผู้ตอบผิดในกลุ่มเปรียบเทียบที่ระดับคะแนน j

T_j แทน จำนวนผู้ตอบข้อสอบทั้งหมดที่ระดับคะแนน j

โดยที่ค่า $\alpha_{MH} = 1$ แสดงว่า ไม่มีความแตกต่างหรือไม่มีความลำเอียง ซึ่งหมายถึงกลุ่มตัวอย่างทั้ง2 กลุ่มทำข้อสอบได้ผลเช่นเดียวกันในแต่ละระดับคะแนนรวม ค่า $A_j D_j / T_j$ และ $B_j C_j / T_j$ มีค่าเท่ากันทั้ง2 กลุ่มทำให้ α_{MH} มีค่าเท่ากับ 1 ถ้าค่า α_{MH} มีค่ามากกว่า 1.00 นั่นคือ สมาชิกของกลุ่มอ้างอิงจะตอบข้อสอบข้อนั้นได้ถูกต้องมากกว่าสมาชิกในกลุ่มเปรียบเทียบและหากค่า α_{MH} มีค่าน้อยกว่า 1.00 นั่นหมายถึง สมาชิกของกลุ่มเปรียบเทียบจะตอบข้อสอบข้อนั้นได้ถูกต้องมากกว่าสมาชิกในกลุ่มอ้างอิงซึ่งแสดงถึงความแตกต่างหรือความลำเอียงที่เกิดขึ้นระหว่างกลุ่มตัวอย่าง2 กลุ่ม

ในทางปฏิบัติถึงแม้จะมีการควบคุมให้ผู้สอบ2 กลุ่ม มีความสามารถในสิ่งที่ต้องการวัดเท่ากัน แต่เป็นไปไม่ได้ที่ผลการตอบข้อสอบของผู้สอบทั้ง2 กลุ่มจะเป็นเช่นเดียวกัน แมนเทล-แฮนส์เซล (Mantel-Haenszel, 1959) จึงเสนอค่าสถิติไคสแควร์เพื่อทดสอบค่าที่ได้ว่าจะมีความแตกต่างจาก 1.00 อย่างมีนัยสำคัญทางสถิติหรือไม่ ที่ระดับชั้นแห่งความเป็นอิสระเท่ากับ1 ดังนี้

Prince of Songkla University Pattani Campus

สมมติฐาน

$$H_0 : \alpha_{MH} = 1$$

$$H_1 : \alpha_{MH} \neq 1$$

สถิติทดสอบ

$$\chi^2_{MH} = \frac{\left[\left|\sum_{j=1}^{s}\left[A_j - E(A_j)\right]\right| - 0.5\right]^2}{\sum_{j=1}^{s} Var(A_j)}$$

โดยที่ค่าผลรวมเป็นค่าที่ได้จากการรวมทุกชั้นของคะแนนรวม และค่าE(A_j) มีค่าดังนี้

$$\mathrm{E}\left(A_j\right) = \left(N_{Rj}\right)\left(N_{1j}\right)/T_j$$

$$\mathrm{Var}(A_j) = \frac{N_{Rj}\,N_{Fj}\,N_{1j}\,N_{0j}}{T_j^2\left(T_j - 1\right)}$$

เมื่อ χ^2_{MH} แทน ดัชนีความลำเอียงของข้อสอบ

$E\left(A_j\right)$ แทน ค่าคาดหวังในการตอบข้อสอบถูกของกลุ่มอ้างอิง

$Var\left(A_j\right)$ แทน ความแปรปรวนในการตอบข้อสอบถูกของกลุ่มอ้างอิง

N_{Rj} แทน จำนวนผู้สอบในกลุ่มอ้างอิงที่ช่วงคะแนนj

N_{Fj} แทน จำนวนผู้สอบในกลุ่มเปรียบเทียบที่ช่วงคะแนนj

N_{1j} แทน จำนวนผู้สอบที่ตอบข้อสอบถูกที่ช่วงคะแนนj

N_{0j} แทน จำนวนผู้สอบที่ตอบข้อสอบผิดที่ช่วงคะแนนj

T_j แทน จำนวนผู้สอบทั้งหมด

ข้อสอบที่มีความลำเอียง จะต้องมีค่า χ^2_{MH} ที่คำนวณได้มากกว่า χ^2_{MH} ที่ระดับชั้นแห่งความเป็นอิสระเท่ากับ 1 และนัยสำคัญทางสถิติที่ระดับ.05

การวิจัยครั้งนี้ ผู้วิจัยใช้การวิเคราะห์ความลำเอียงโดยวิธีแมนเทล-แฮนส์เซล โดยเปรียบเทียบผลการตอบของผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ หลังจากการจับคู่ผู้สอบตามระดับความสามารถ ซึ่งใช้คะแนนรวมรายบุคคลในแบบทดสอบเป็นเกณฑ์การจับคู่คำนวณโดยใช้

Prince of Songkla University Pattani Campus

โปรแกรม Microsoft Excel ข้อสอบที่มีความลำเอียง จะต้องมีค่า χ^2_{MH} ที่คำนวณมากกว่า χ^2_{MH} ที่ระดับนัยสำคัญทางสถิติ .05 และมีชั้นแห่งความเป็นอิสระเท่ากับ1 ส่วนค่า α_{MH} บ่งบอกถึงความน่าจะเป็นในการตอบข้อสอบของกลุ่มผู้สอบ

3. การวิเคราะห์ความลำเอียงของข้อสอบตามแนวคิดทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างความสามารถที่มีอยู่ภายในบุคคล(Latent trait or ability) กับผลการตอบข้อสอบหรือข้อคำถามโดยใช้โค้งลักษณะข้อสอบ(Item Characteristic Curve: ICC) ซึ่งมีการกำหนดลักษณะของข้อสอบด้วยพารามิเตอร์ทฤษฎีนี้ได้อธิบายความสัมพันธ์ดังกล่าวในรูปของฟังก์ชันคณิตศาสตร์ หรือโมเดลที่แสดงความสัมพันธ์ระหว่างระดับความสามารถ คุณลักษณะของข้อสอบและโอกาสของการตอบข้อสอบได้ถูกที่เรียกว่า ฟังก์ชันการตอบสนองข้อสอบซึ่งมีลักษณะความสัมพันธ์เป็นแบบฟังก์ชันโลจิสโค้งลักษณะข้อสอบมีหลายลักษณะขึ้นอยู่กับโมเดล (Model) หรือแบบจำลองที่ใช้อธิบายความสัมพันธ์ดังกล่าว โมเดลที่นิยมใช้กันคือ โมเดลแบบหนึ่งพารามิเตอร์ (One-Parameter Model) โมเดลแบบสองพารามิเตอร์ (Two-Parameter Model) และโมเดลแบบสามพารามิเตอร์ (Three-Parameter Model) (ศิริชัย กาญจนวาสี, 2550 : 53-54)

3.1 หลักการของทฤษฎีการตอบสนองข้อสอบ

ในการวิเคราะห์ข้อสอบโดยทั่วไปแล้วจะพิจารณาตามทฤษฎีดั้งเดิม (Classical Test Theory) ซึ่งเมื่อพิจารณารายข้อ (Item) จะดูจากค่าความยาก (p) และค่าอำนาจจำแนก(r) เมื่อพิจารณารวมทั้งฉบับ(Test) จะดูจากค่าความเชื่อมั่น (Reliability) และค่าความเที่ยงตรง (Validity) ซึ่งจากการพิจารณาโดยภาพรวมแล้ว พบว่ายังมีจุดอ่อนอยู่หลายประการ คือประการแรกค่าพารามิเตอร์ของข้อสอบแปรเปลี่ยนไปตามกลุ่มของผู้สอบที่แตกต่างกันในด้านความสามารถ (Ability) ประการที่สอง การเปรียบเทียบความสามารถของผู้สอบจำกัดอยู่ในสถานการณ์ที่ทดสอบและประการที่สามจะไม่สามารถบอกได้ว่าผู้เข้าสอบคนหนึ่งจะทำข้อสอบได้เพียงใดเมื่อได้เผชิญกับข้อคำถามหนึ่งยกเว้นเมื่อได้มีการใช้ข้อสอบนั้นแล้วกับกลุ่มตัวอย่างที่คล้ายคลึงกันกับบุคคลนั้น จากเหตุการณ์เหล่านี้ทำให้นักทดสอบทางจิตวิทยาสำรวจและพัฒนาทฤษฎีทฤษฎีที่พัฒนาจนถึงปัจจุบันคือทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) โดยใช้ข้อตกลงเบื้องต้นว่าความสามารถของบุคคล การทำข้อสอบที่วัดความสามารถนั้นเป็นอย่างไร ซึ่งทฤษฎีการตอบข้อสอบมีหลักการที่สำคัญอยู่ที่การใช้ผลการตอบแบบทดสอบมาอธิบายถึงความสามารถของผู้สอบในเรื่องที่ทำการทดสอบนั้นซึ่งจุดเด่นของการวิเคราะหทฤษฎีนี้ คือค่าพารามิเตอร์ของข้อสอบจะคงที่ โดยค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าสัมประสิทธิ์การเดา (c) จะเป็นค่าที่ไม่แปรเปลี่ยนไปตามกลุ่มผู้ตอบไม่ว่าจะนำไปสอบกับผู้ใดก็ตามเมื่อทราบลักษณะการตอบข้อสอบ

Prince of Songkla University Pattani Campus

ในแต่ละข้อคำถามของผู้ข้าสอบคนใด จะสามารถคำนวณหาค่าความสามารถที่แท้จริงของผู้สอบคนนั้นได้ โดยค่านี้จะสัมพันธ์โดยตรงกับคะแนนจริง ซึ่งถือเป็นลักษณะของความเป็นอิสระของข้อสอบ

3.2 ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อคำถาม

3.2.1 ความเป็นมิติเดียวของข้อสอบ (Unidimensionality Test) เป็นแบบทดสอบที่ประกอบด้วยข้อคำถามที่เป็นเอกพันธ์ นั่นคือแบบทดสอบนั้นจะต้องมุ่งวัด ความสามารถอย่างใดอย่างหนึ่งที่มีลักษณะเด่นๆเพียงความสามารถเดียว หากไม่กำหนดข้อตกลงเบื้องต้นเช่นนี้จะทำให้โมเดลที่ใช้มีความสลับซับซ้อนมาก(Hambleton and Swaminathan, 1985 : 38) สำหรับการตรวจสอบว่าแบบทดสอบนั้นวัดในมิติเดียวหรือไม่ สามารถทำได้โดยการวิเคราะห์องค์ประกอบ (Factor Analysis) แล้วสังเกตค่าไอเกน(Eigenvalue) ขององค์ประกอบสูงสุดว่าแตกต่างจากค่าไอเกนขององค์ประกอบที่เหลืออย่างชัดเจนหรือไม่ในขณะที่ค่าไอเกนขององค์ประกอบที่เหลือทั้งหมดมีค่าใกล้เคียงกัน(Warm, 1978 : 104 ; Lord, 1980 : 21)

3.2.2 ความเป็นอิสระในการตอบข้อสอบ (Local Independence) เป็นการกำหนดข้อตกลงเบื้องต้นเกี่ยวกับความเป็นอิสระในการตอบข้อสอบ กล่าวคือ การตอบข้อสอบข้อใดข้อหนึ่งถูกจะไม่มีผลต่อการตอบข้อสอบข้ออื่นๆ ในการตรวจสอบว่าข้อสอบแต่ละข้อเป็นไปตามข้อตกลงเกี่ยวกับความเป็นอิสระในการตอบข้อสอบหรือไม่ ให้สังเกตความเป็นมิติเดียวกันของแบบทดสอบถ้าแบบทดสอบมีความเป็นมิติเดียวแล้ว แบบทดสอบนั้นจะมีความเป็นอิสระในการตอบข้อสอบด้วย(Lord, 1980 : 19 ; Hambleton and Swaminathan, 1985 : 24)

3.2.3 โค้งลักษณะข้อสอบ(Item Characteristic Curve) เป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงถึงความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบข้อนั้นได้ถูกต้องตรงกับระดับความสามารถที่วัดได้ โดยใช้ชุดของข้อสอบหรือแบบทดสอบนั้นและความน่าจะเป็นที่ผู้สอบจะตอบข้อสอบได้ถูกต้องขึ้นอยู่กับโค้งลักษณะข้อสอบของแต่ละโมเดลที่ใช้แต่จะไม่ขึ้นกับการแจกแจงความสามารถของกลุ่มตัวอย่าง(Hambleton and Swaminathan, 1985 : 25) กล่าวคือคุณสมบัติของโค้งลักษณะของข้อสอบจะไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบฉะนั้นจึงทำให้ความน่าจะเป็นในการตอบข้อสอบถูกแต่ละข้อไม่แปรเปลี่ยนด้วย โดยผู้สอบที่มีความสามารถในระดับสูงจะมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องมากกว่าข้อสอบที่มีความสามารถในระดับต่ำ

Prince of Songkla University Pattani Campus

3.3 พารามิเตอร์ของทฤษฎีการตอบสนองข้อสอบ

3.3.1 พารามิเตอร์ของข้อสอบ (Item Parameter) ได้แก่

ค่าความยาก (b) หมายถึง สัดส่วนของคนที่ทำข้อสอบข้อนั้นถูก หรือหมายถึงค่าที่แสดงถึงระดับความสามารถของผู้สอบ (θ) ที่จุดโค้งลักษณะข้อสอบมีความชัน มากที่สุดมีค่าตั้งแต่ $-\infty$ ถึง ∞ แต่ในทางปฏิบัติจะอยู่ระหว่าง -3 ถึง +3 ค่า -3 บ่งบอกว่าข้อสอบนั้น ง่ายมาก และค่า +3 แสดงว่าข้อสอบนั้นยากมาก

ค่าอำนาจจำแนก(a) หมายถึง ความสามารถของข้อสอบที่แยกผู้สอบออกเป็น 2 กลุ่ม คือกลุ่มตอบถูกกับกลุ่มตอบผิด ในการวิจัยนี้หมายถึง ค่าที่เป็นสัดส่วนโดยตรงกับความชันของโค้งคุณลักษณะของข้อสอบ ณ จุดเปลี่ยนโค้งมีค่าตั้งแต่ $-\infty$ ถึง ∞ แต่ในทางปฏิบัติมีค่าตั้งแต่ 0 ถึง 2 เพราะค่า a ที่เป็นลบแสดงว่าข้อสอบไม่ดี ใช้ไม่ได้ต้องตัดทิ้ง ค่า 0 แสดงว่าข้อสอบไม่มีค่าอำนาจจำแนก ค่า +2 แสดงว่าข้อสอบมีค่าอำนาจจำแนกสูงในการคัดเลือกข้อสอบ ข้อสอบที่คัดไว้จะมีค่า a ตั้งแต่ 0.3 ขึ้นไป

ค่าสัมประสิทธิ์การเดา (c) หมายถึง ความน่าจะเป็นของบุคคลหนึ่งที่ไม่มีความสามารถในการตอบข้อสอบนั้นได้ถูกต้อง เป็นค่าที่แสดงถึงโอกาสการตอบข้อสอบถูกโดยไม่มีความรู้ในเรื่องนั้นๆ มีค่าจาก 0 ถึง 1 จะคัดเลือกเอาข้อสอบที่มีค่า c ต่ำกว่า 0.3

3.3.2 พารามิเตอร์ของผู้สอบ ได้แก่

ระดับความสามารถของผู้สอบ (θ) หมายถึง ศักยภาพของผู้สอบที่ประมาณได้จากการตอบข้อสอบตามทฤษฎีการตอบสนองข้อสอบ มีค่าอยู่ระหว่าง -3 ถึง +3 ค่า –3 แสดงว่ามีความสามารถต่ำและค่า +3 แสดงว่ามีความสามารถสูง

3.4 รูปแบบของโมเดลโลจิสติก

โลจิสติกโมเดล (Logistic Model) เป็นโมเดลที่ได้พัฒนาเพื่อให้เหมาะสมกับสถานการณ์การทดสอบ เช่น ในกรณีที่การตอบข้อสอบได้ด้วยการเดาปลายต่ำสุดของโค้งลักษณะข้อสอบไม่เป็นศูนย์ แสดงว่าการทำข้อสอบย่อมมีการเดาและในกรณีผู้สอบทำข้อสอบด้วยความสะเพร่า โมเดลนี้จะเพิ่มค่าความรอบคอบขึ้นอีกหนึ่งค่า ซึ่งจะเห็นได้จากปลายสูงสุดของโค้งลักษณะข้อสอบมีค่าน้อยกว่า 1 โลจิสติก โมเดล แบ่งเป็น 4 โมเดลย่อยๆ ตามพารามิเตอร์ ดังนี้ (Hambleton and Swaminathan, 1985 : 34-49)

โมเดลโลจิสติกถูกพัฒนาขึ้นเพื่อให้สะดวกต่อการนำไปใช้จึงพัฒนาขึ้นเป็น 3 รูปแบบ ดังนี้

3.4.1 โมเดลโลจิสติก 1 พารามิเตอร์ (One – Parameter Logistic Model) โมเดลนี้เบอร์นบอม พัฒนาขึ้นในปี 1968 ซึ่งบังเอิญตรงกับรูปแบบของราสช์ (Rasch, 1960) เป็น โมเดลที่

Prince of Songkla University Pattani Campus

อธิบายข้อสอบด้วยค่าพารามิเตอร์เพียงค่าเดียว คือ ค่าความยากซึ่งโอกาสที่ผู้สอบจะทำข้อสอบได้ถูกหรือไม่นั้น ขึ้นอยู่กับระดับความสามารถของตนเองกับระดับความยากของข้อสอบ ดังนั้นจึงถือว่าค่าการเดาเป็นศูนย์ ($c_i = 0$) และค่าอำนาจจำแนก($a_i = 1$) ของข้อสอบจะคงที่ทั้งฉบับ เขียนเป็นฟังก์ชันได้ดังนี้ (Birnbaum , 1968 : 402)

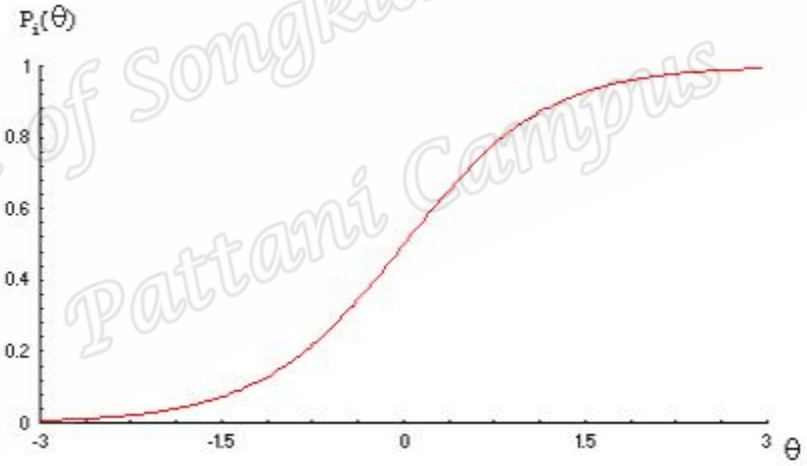
$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{\theta - b_i}} \ , \ i = 1,2,3,\ldots\ldots n$$

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

θ แทน ระดับความสามารถที่แท้จริงของผู้สอบ

b_i แทน ค่าความยากของข้อสอบข้อที่ i

e แทน ค่าคงที่มีค่าเท่ากับ 2.7182818



ภาพประกอบ 1 ค่าความหมายของค่า Item Parameter (1 พารามิเตอร์) ของข้อสอบ

ที่มา : สุพัฒน์ สุกมลสันต์, 2534 : 63

3.4.2 โมเดลโลจิสติก 2 พารามิเตอร์ (Two – Parameter Logistic Model)

เบอร์นบอร์ม (Birnbaum) ได้พัฒนาโมเดลนี้ขึ้นมาและกำหนดให้ทุกข้อไม่มีการเดาเกิดขึ้น คือ ค่า c_i มีค่าเป็นศูนย์ทุกข้อ กล่าวคือ ผู้สอบที่มีความสามารถต่ำสุด ไม่มีโอกาสที่จะทำข้อสอบถูกในข้อสอบที่มีค่าความยากสูง ซึ่ง เบอร์นบอร์ม ได้เสนอรูปแบบของสมการดังนี้ (Birnbaum, 1968 : 400)

Prince of Songkla University Pattani Campus

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

เมื่อ $P_i(\theta)$ แทน โอกาสที่ผู้มีความสามารถ จะทำข้อสอบข้อที่ i ได้ถูกต้อง

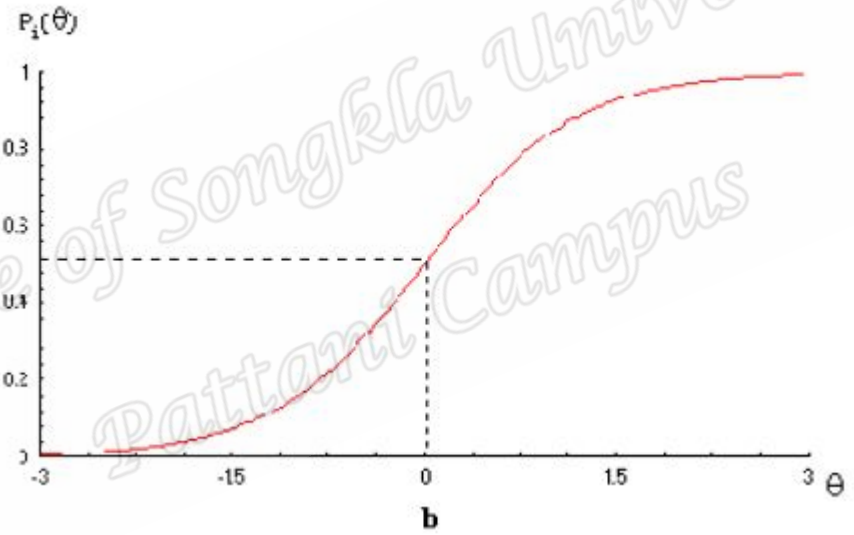
θ แทน ระดับความสามารถที่แท้จริงของผู้สอบ

a_i แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ i

b_i แทน ค่าความยากของข้อสอบข้อที่ i

e แทน ค่าคงที่มีค่าเท่ากับ 2.7182818

D แทน ค่าคงที่ ซึ่งมีค่า 1.7



ภาพประกอบ 2 ความหมายของค่า Item Parameter (2 พารามิเตอร์) ของข้อสอบ

ที่มา : สุพัฒน์ สุกมลสันต์, 2534 : 64

3.4.3 โมเดลโลจิสติก 3 พารามิเตอร์ (Three – parameter Logistic Model) เป็นโมเดลที่พัฒนามาจาก Two-Parameter Logistic Model เพื่อให้เหมาะกับแบบทดสอบที่มีอิทธิพลจากการเดาเข้ามาแฝงอยู่ด้วย และเป็นโค้งลักษณะข้อสอบที่แสดงถึงลักษณะข้อสอบที่มีค่าพารามิเตอร์ของข้อสอบ 3 ตัว ซึ่งเบอร์นบอร์ม ได้เสนอรูปแบบของสมการดังนี้ (Birnbaum , 1968 : 405)

Prince of Songkla University Pattani Campus

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

เมื่อ $P_i(\theta)$ แทน โอกาสที่ผู้มีความสามารถจะทำข้อสอบข้อที่ i ได้ถูกต้อง

θ แทน ระดับความสามารถที่แท้จริงของผู้สอบจะมีค่าอยู่ระหว่าง –3 ถึง +3 และค่า –3 แสดงว่ามีค่าความสามารถต่ำ1+3 แสดงว่ามีค่าความสามารถสูง

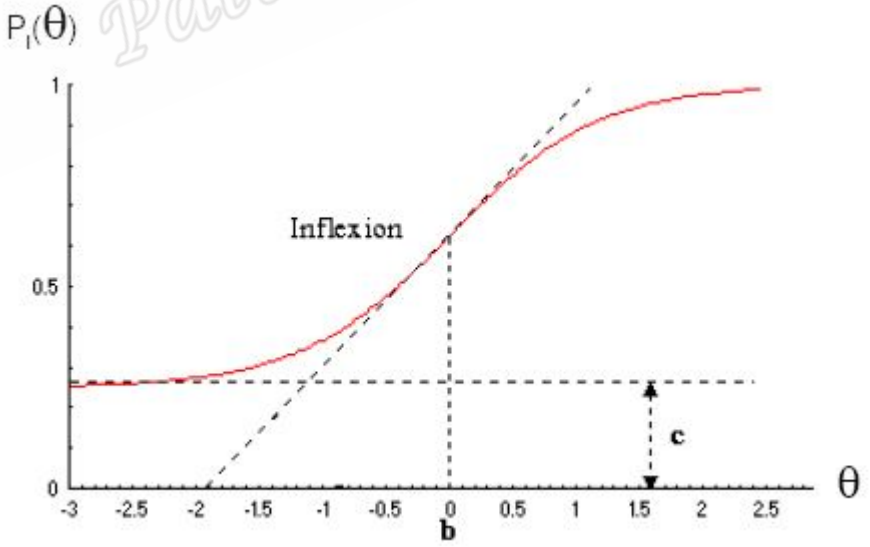
a_i แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ i ซึ่งมีค่าสัดส่วนโดยตรงกับค่าความชันของโค้ง ณ จุดเปลี่ยนโค้งที่จุด $\theta = b_i$ โดยทั่วไปมักจะเลือกใช้ช่วงของค่าอำนาจจำแนกของข้อสอบอยู่ระหว่าง 0 ถึง 2

b_i แทน ค่าความยากของข้อสอบข้อที่ i เป็นระดับความสามารถของผู้สอบ ณ จุดเปลี่ยนโค้งและมีค่าอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัติจะมีค่าอยู่ระหว่าง -2 ถึง +2 ค่า –2 แสดงว่าข้อสอบง่ายมาก และค่า +2 แสดงว่าข้อสอบยาก

c_i แทน ค่าการเดาของข้อสอบข้อที่ i เป็นความน่าจะเป็นหรือโอกาสของคนที่มีความสามารถต่ำจะตอบข้อสอบถูกมีค่าอยู่ระหว่าง 0 ถึง 1 โดยทั่วไปแล้ว ข้อสอบที่ดีจะต้องมีค่าการเดาต่ำกว่า 0.30

e แทน ค่าคงที่มีค่าเท่ากับ 2.7182818

D แทน ค่าคงที่มีค่าเท่ากับ 1.7

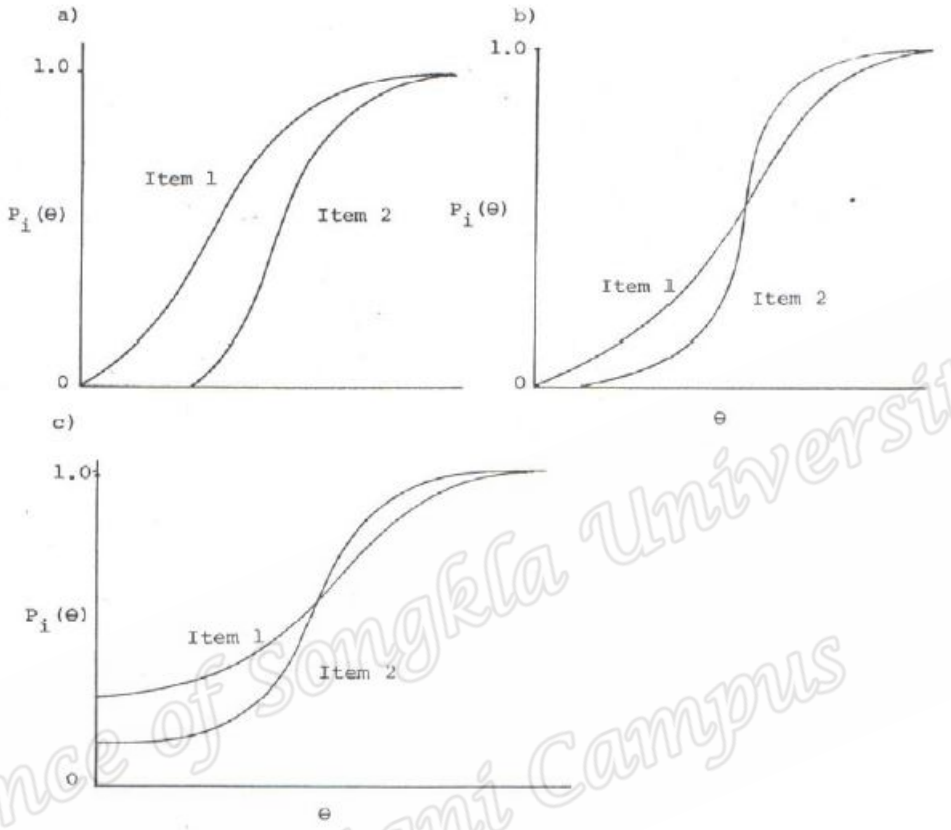


ภาพประกอบ 3 ความหมายของค่า Item Parameter (3 พารามิเตอร์) ของข้อสอบ

ที่มา : สุพัฒน์ สุกมลสันต์, 2534 : 65

Prince of Songkla University Pattani Campus

จากโมเดลโลจิสติกทั้ง 3 โมเดล สามารถเขียนแสดงได้ดังรูปต่อไปนี้



ภาพประกอบ 4 ICC ของโมเดลโลจิสติก 1, 2, 3 พารามิเตอร์

จากภาพประกอบ 4 แสดงถึงโอกาสที่ผู้ทำข้อสอบจะตอบข้อสอบได้ถูกต้องขึ้นอยู่กับค่าพารามิเตอร์ ได้แก่

รูป a) แสดงถึงโอกาสที่จะทำข้อสอบได้ถูกต้องขึ้นอยู่กับความสามารถของข้อสอบเพียงอย่างเดียว โดยถือว่าข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากันหมดคือเท่ากับหนึ่งและค่าการเดาเท่ากับศูนย์ในรูปแสดงให้เห็นว่าข้อสอบข้อที่ 2 ยากกว่าข้อที่ 1

รูป b) แสดงถึงโอกาสที่จะทำข้อสอบได้ถูกต้อง ขึ้นอยู่กับความยากและค่าอำนาจจำแนกในรูปแสดงให้เห็นว่าข้อสอบข้อที่ 2 ยากกว่าข้อที่ 1 และข้อสอบข้อที่ 2 จำแนกคนได้ดีกว่าข้อที่ 1

รูป c) แสดงถึงโอกาสที่จะทำข้อสอบได้ถูกต้อง ขึ้นอยู่กับค่าอำนาจจำแนก ค่าความยากและค่าการเดา ในรูปแสดงให้เห็นว่าข้อสอบข้อที่ 1 ยากกว่าข้อที่ 2 และข้อสอบข้อที่ 2 จำแนกคนได้ดีกว่าและมีค่าการเดาน้อยกว่าข้อที่ 1

Prince of Songkla University Pattani Campus

3.5 ฟังก์ชันสารสนเทศของข้อสอบตามแนวทฤษฎีการตอบสนองข้อสอบ

ในการประมาณค่าความสามารถของผู้สอบด้วยวิธี Maximum Likelihood นั้นความแน่นอนของการประมาณค่าความสามารถแสดงได้ในเทอมของ Information Function โดยที่ในโมเดลคลาสสิคอลนั้นเราศึกษาเรื่องความเที่ยง (Reliability) ของคะแนนและความคลาดเคลื่อนมาตรฐานของการวัด (Standard Error of Measurement) ซึ่งค่าที่ได้จะแปรเปลี่ยนไปตามกลุ่มผู้สอบ อันเป็นจุดอ่อนในการศึกษา แต่ในทฤษฎีการตอบสนองข้อสอบนั้นจะพิจารณาจากค่าฟังก์ชันสารสนเทศของแบบทดสอบ ซึ่งเป็นดัชนีบอกความแม่นยำในการประมาณค่าความสามารถที่แท้จริง (Lord, 1980 : 72) ได้จากผลรวมของฟังก์ชันสารสนเทศของข้อสอบโดยที่ค่าฟังก์ชันสารสนเทศของข้อสอบจะแตกต่างกันไปตามค่าพารามิเตอร์ของข้อสอบ ได้แก่ ค่าอำนาจจำแนก ค่าความยากและค่าการเดา เป็นต้น ค่าฟังก์ชันสารสนเทศของข้อสอบหาได้ดังนี้

$$I_i(\theta) = \frac{(P_i')^2}{P_i Q_i} \quad ; i = 1,2,3,.....n \qquad (1)$$

เมื่อ $I_i(\theta)$ แทน ค่า Item Information Functions

P_i' แทน ความชันของ ICC ที่ระดับความสามารถ θ

P_i แทน ความน่าจะเป็นที่ผู้สอบที่มีความสามารถ θ ตอบข้อสอบข้อที่ i ถูก

Q_i แทน 1 - P_i

สำหรับในโมเดลโลจิสติกแบบ 3 พารามิเตอร์จะให้ค่าอนุพันธ์และค่าสารสนเทศของข้อสอบ (Hambleton and Swaminathan, 1985 : 107) ดังนี้

$$P_i = \frac{c_i + e^{1.7a_i(\theta - b_i)}}{1 + e^{1.7a_i(\theta - b_i)}} \qquad (2)$$

$$Q_i = \frac{1 - c_i}{1 + e^{1.7a(\theta - b_i)_i}} \qquad (3)$$

$$P_i' = \frac{1.7a_i(1 - c_i)}{e^{1.7a_i(\theta - b)} + 2 + e^{-1.7a_i(\theta - b)}} \qquad (4)$$

Prince of Songkla University Pattani Campus

แทนค่า (2) , (3) และ (4) ลงใน (1) จะได้

$$I_i(\theta) = \frac{(1.7a_i)^2(1-c_i)}{\left[c_i + e^{1.7a_i(\theta-b)}\right]\left[1+e^{-1.7a_i(\theta-b_i)}\right]^2} \qquad (5)$$

จากสมการ (5) หมายถึง ค่าฟังก์ชันสารสนเทศของข้อสอบที่เขียนอยู่ในรูปของค่าพารามิเตอร์ข้อสอบ สำหรับค่าพารามิเตอร์ของแบบทดสอบหาได้จากผลรวมของค่าฟังก์ชันสารสนเทศข้อสอบทั้งหมดในแบบทดสอบ เขียนสมการได้ดังนี้

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

เมื่อ $I(\theta)$ แทน ค่าฟังก์ชันสารสนเทศของแบบทดสอบ Test Information Function (Birnburm in Lord and Novick, 1968 : 454)

จากสมการ(1) จะเห็นว่าข้อสอบแต่ละข้อมี Item Information Curve ซึ่งจะขึ้นอยู่กับความชันของ ICC และความแปรปรวนของการตอบข้อสอบถูกในแต่ละข้อในแต่ละระดับความสามารถ และยิ่งความชันของICC มีค่ามากๆ ประกอบกับความแปรปรวนมีค่าน้อยๆ Item Information Curve ที่ระดับความสามารถนั้นจะยิ่งสูงขึ้นสำหรับItem Information Curve ที่มีค่าสูงสุด ณ ระดับความสามารถใดก็จะจำแนกระดับความสามารถผู้สอบได้ดี ณ ระดับความสามารถนั้น (Hambleton and Cook, 1977 : 66) ดังนั้นประโยชน์ที่ได้จากประเด็นข้างต้นก็คือ ถ้ามีกลุ่มของ ข้อสอบอยู่ชุดหนึ่งที่สามารถทราบ Information Curve ของแต่ละข้อ เราก็จะสามารถสร้างแบบทดสอบฉบับหนึ่งให้มี Test Information Curve ณ ระดับหนึ่งของความสามารถตามที่เราต้องการได้ และนั่นหมายถึงว่าเราสามารถสร้างฉบับแบบทดสอบให้เป็นนไปตามจุดมุ่งหมายของการสอบได้ เช่น ถ้าต้องการได้แบบทดสอบคัดเลือก ก็ต้องเลือกใช้ข้อสอบที่มีความสูงสุดของโค้งที่ระดับความสามารถสูง ๆ ซึ่งก็คือ เลือกข้อสอบที่จะให้ได้ Test Information Curve สูงที่ระดับความสามารถสูงๆ เป็นต้น

3.6 การประยุกต์ทฤษฎีการตอบสนองข้อสอบเพื่อศึกษาความลำเอียงของข้อสอบ

จากข้อตกลงเบื้องต้น3 ประการของทฤษฎีการตอบข้อคำถาม คือ ความเป็นมิติเดียวของข้อสอบ ความเป็นอิสระในการตอบข้อสอบและโค้งลักษณะข้อสอบสามารถนำมาประยุกต์เพื่อศึกษาความลำเอียงของข้อสอบได้โดย ถ้ามีแบบทดสอบฉบับหนึ่งวัดความสามารถของข้อสอบ

Prince of Songkla University Pattani Campus

มากกว่า 1 กลุ่ม แล้วข้อสอบข้อเดียวกันวัดความสามารถของผู้สอบได้ไม่ตรงกัน เช่น กลุ่มหนึ่งอาจจะมีประสบการณ์เกี่ยวกับเนื้อหาของข้อสอบข้อนั้นมาก่อน หรือเมื่อตอบข้อสอบข้ออื่นถูกแล้วเป็นเหตุทำให้ความน่าจะเป็นของการตอบข้อสอบข้อนั้นถูกมีค่ามากกว่าอีกกลุ่ม ถึงแม้ว่าภายในแต่ละฉบับจะมีความเป็นเอกพันธ์กันทุกข้อ แต่เมื่อพิจารณาเปรียบเทียบข้อสอบเป็นรายข้อระหว่างกลุ่มผู้สอบทุกกลุ่ม โดยนำผลการสอบของแต่ละกลุ่มมาสร้างโค้งลักษณะข้อสอบบนกราฟเดียวกัน เพื่ออธิบายลักษณะของข้อสอบจะพบว่าโค้งลักษณะข้อสอบจากผู้สอบแต่ละกลุ่มแตกต่างกันถึงถือได้ว่าเป็นข้อสอบที่มีความลำเอียงแต่หากโค้งลักษณะข้อสอบของผู้สอบต่างกลุ่มกันมีลักษณะโค้งเหมือนกัน แสดงว่าข้อสอบข้อนั้นไม่ลำเอียง

3.6.1 คุณสมบัติของโค้งลักษณะข้อสอบ

ในการศึกษาความไม่แปรเปลี่ยนเกี่ยวกับลักษณะของข้อสอบนี้ ลอร์ด และโนวิค (Lord and Novick, 1968) ได้ศึกษาเกี่ยวกับค่าพารามิเตอร์ของข้อสอบข้อเดียวกัน ระหว่างกลุ่มผู้สอบที่มีความสามารถต่างกัน

โค้งลักษณะข้อสอบ ซึ่งมีลักษณะไม่เปลี่ยนแปลงตามจำนวนและระดับความสามารถของผู้สอบในกรณีที่กลุ่มผู้สอบมีระดับความสามารถต่ำค่าความน่าจะเป็นของการตอบข้อสอบถูกของผู้สอบสามารถจะอธิบายได้ด้วยโค้งลักษณะของข้อสอบในช่วงระยะความสามารถต่ำ และในทำนองเดียวกันเมื่อผู้สอบมีระดับความสามารถสูงความน่าจะเป็นของการตอบข้อสอบถูกก็ย่อมสามารถอธิบายได้ด้วยโค้งลักษณะของข้อสอบในช่วงระยะความสามารถสูง นั่นคือ ถ้าค่าพารามิเตอร์ของข้อสอบไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบแล้ว ข้อสอบที่นำไปทดสอบกับผู้ที่มีความสามารถแตกต่างกัน ถ้าข้อสอบข้อนี้วัดความสามารถอย่างเดียวกันของกลุ่มตัวอย่างทั้งสองกลุ่มแล้ว เส้นโค้งลักษณะข้อสอบก็ย่อมมีเส้นเดียว เส้นโค้งในช่วงความสามารถต่ำและความสามารถสูง ย่อมเป็นส่วนหนึ่งของโค้งลักษณะข้อสอบข้อนี้ เมื่อเส้นโค้งลักษณะข้อสอบมีได้เส้นเดียว ตัวพารามิเตอร์ซึ่งเป็นตัวกำหนดรูปร่างลักษณะของเส้นโค้งย่อมมีได้ค่าเดียวเช่นกัน หรืออาจกล่าวได้ว่า ถ้าค่าพารามิเตอร์ของข้อสอบไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบแล้ว ข้อสอบที่นำไปสอบกับผู้สอบที่มีความสามารถแตกต่างกัน เมื่อนำข้อสอบที่ได้จากแต่ละกลุ่มมาวิเคราะห์จะมีค่าพารามิเตอร์ของข้อสอบเป็นค่าเดียวกัน

การวิเคราะห์ความลำเอียงของข้อสอบด้วยโค้งลักษณะข้อสอบจะพิจารณาจากผลการตอบข้อสอบกับโอกาสในการตอบข้อสอบข้อนั้นๆถูก ระหว่างกลุ่มผู้สอบตั้งแต่2 กลุ่ม หากโค้งลักษณะข้อสอบมีความแตกต่างกันแสดงว่าข้อสอบข้อนั้นทำให้เกิดการได้เปรียบเสียเปรียบระหว่างกลุ่มผู้สอบที่มีลักษณะแตกต่างกันและค่าพารามิเตอร์ที่เป็นตัวกำหนดโค้งลักษณะข้อสอบตามทฤษฎีการตอบสนองข้อสอบนั้นมี 3 ค่า คือ ค่าความยาก (b_i) ค่าอำนาจจำแนก(a_i) และค่า

Prince of Songkla University Pattani Campus

สัมประสิทธิ์การเดา (c_i) ในการทดสอบว่าข้อสอบข้อใด มีความลำเอียงหรือไม่และจะใช้วิธีใดในการวิเคราะห์ความลำเอียงของข้อสอบให้มีความถูกต้องเหมาะสมเพียงใดนั้นขึ้นอยู่กับลักษณะความแตกต่างของโค้งลักษณะข้อสอบที่มีความลำเอียง ซึ่งสามารถแบ่งออกเป็น 2 แบบ คือ

1. แบบลำเอียงอย่างสม่ำเสมอหรือแบบลำเอียงคงที่ คือ โค้งลักษณะข้อสอบของกลุ่มต่างๆ จะมีค่าพารามิเตอร์ของข้อสอบเท่ากัน 2 ค่า คือ ค่าอำนาจจำแนกของข้อสอบ (a_i = ความชัน) ของแต่ละกลุ่มเท่ากันและมีค่าสัมประสิทธิ์การเดา (c_i) ที่ใกล้เคียงกันหรือเท่ากัน วิธีนี้จะมีเฉพาะค่าความยากของข้อสอบ (b_i) ตัวเดียวเท่านั้นที่แตกต่างกันและโค้งลักษณะข้อสอบของกลุ่มต่างๆ จะไม่ตัดกันตลอดต่อเนื่องของทุกระดับความสามารถในกลุ่มผู้สอบที่มีระดับความสามารถ (θ) เดียวกัน ผู้สอบที่อยู่ในกลุ่ม R มีโอกาสในการตอบข้อสอบข้อนี้ได้ถูกต้องมากกว่าผู้สอบที่อยู่ในกลุ่ม F ซึ่งในการวิเคราะห์ความลำเอียงของข้อสอบ จะใช้ค่าพารามิเตอร์เพียงตัวเดียว คือ ค่าความยากของข้อสอบ (b_i) มาคำนวณค่าดัชนีความลำเอียงของข้อสอบ นั่นคือ เป็นการวิเคราะห์ความลำเอียงของข้อสอบตามแบบจำลองโลจิสติกที่มี 1 พารามิเตอร์ หรือแบบจำลองของราส์ช (Rasch Model) นั่นเอง การคำนวณพื้นที่ระหว่างโค้งลักษณะข้อสอบวิธีนี้สามารถคำนวณได้ง่ายเพราะพื้นที่ที่คำนวณได้มีเฉพาะเครื่องหมายบวกเพียงอย่างเดียวเท่านั้น

2. แบบลำเอียงอย่างไม่สม่ำเสมอ หรือแบบลำเอียงไม่คงที่ คือ โค้งลักษณะข้อสอบของกลุ่มต่างๆ จะมีค่าพารามิเตอร์ของข้อสอบแตกต่างกัน 3 ค่า คือ ค่าความยาก (b_i) ค่าอำนาจจำแนก (a_i) และค่าสัมประสิทธิ์การเดา (c_i) ทำให้โค้งลักษณะข้อสอบของกลุ่มต่างๆ มีลักษณะตัดที่ระดับความสามารถสูงผู้สอบในกลุ่ม R มีโอกาสที่จะตอบข้อสอบนี้ได้ถูกต้องมากกว่าผู้สอบในกลุ่ม F ที่มีระดับความสามารถเดียวกันแต่สำหรับในกลุ่มผู้สอบที่มีระดับความสามารถต่ำแล้วผู้สอบในกลุ่ม F มีโอกาสที่จะตอบข้อสอบถูกมากกว่าผู้สอบในกลุ่ม R ในการวิเคราะห์ค่าความลำเอียงของข้อสอบจะนำค่าพารามิเตอร์ทั้ง 3 ค่ามาคำนวณค่าดัชนีความลำเอียงของข้อสอบ นั่นคือเป็นการวิเคราะห์ความลำเอียงของข้อสอบตามแบบจำลองโลจิสติก 3 พารามิเตอร์ การคำนวณค่าดัชนีความลำเอียงตามวิธีนี้จะมีความยุ่งยากมากและซับซ้อน เพราะว่าค่าพารามิเตอร์ของโค้งลักษณะข้อสอบแต่ละค่าจะแตกต่างกันทุกค่า ส่วนในกรณีที่ใช้พื้นที่ระหว่างโค้งลักษณะข้อสอบเป็นตัววัดดัชนีความลำเอียงของข้อสอบนั้นการคำนวณพื้นที่จะมีความยุ่งยากมาก เพราะพื้นที่ที่คำนวณได้จะต้องมีทั้งทางบวกและทางลบ เนื่องจากโค้งลักษณะข้อสอบของผู้สอบทั้งสองกลุ่มมีลักษณะตัดกัน

3.6.2 วิธีการวิเคราะห์ความลำเอียง

ตามทฤษฎีการตอบสนองข้อสอบนั้นข้อสอบจะลำเอียง ถ้าฟังก์ชันการตอบข้อสอบข้อนั้นๆของประชากรกลุ่มย่อยๆ 2 กลุ่ม ไม่เป็นเช่นเดียวกัน ซึ่งการวิเคราะห์ความลำเอียง

ของข้อสอบตามแนวทฤษฎีการตอบสนองข้อสอบสามารถทำได้โดยการเปรียบเทียบฟังก์ชันการตอบข้อสอบของผู้สอบกลุ่มย่อยที่ต้องการศึกษาและการเปรียบเทียบในทฤษฎีการตอบสนองข้อสอบมีวิธีการวิเคราะห์หลายวิธี ซึ่งพอสรุปได้ดังนี้ คือ

1) แบบจำลองโลจิสติกชนิด 1 พารามิเตอร์ ซึ่งกำหนดให้ค่าการเดา(c_i) เท่ากับ 0 และค่าอำนาจจำแนก(a_i) ของข้อสอบทุกข้อ เท่ากับ1 มีค่าความยาก (b_i) เพียงตัวเดียวที่แปรเปลี่ยนไปตามกลุ่มผู้สอบ วิธีนี้เป็นการวิเคราะห์ความแตกต่างของค่าพารามิเตอร์ความยาก (b_i) ของข้อสอบจากโค้งลักษณะข้อสอบของแต่ละกลุ่ม

เริ่มด้วยการทดสอบความเหมาะสมของข้อมูลทั้ง2 กลุ่มกับแบบจำลอง โดยทดสอบรายข้อด้วย $H_i(\theta)$ ซึ่งมีการกระจายแบบไคสแควร์ (χ^2) มี $df = N-1$ เพื่อตัดข้อที่ไม่เหมาะสมออกแล้วนำไปวิเคราะห์ความลำเอียงรายข้อ Z_i

$$H_i(\theta) = \sum_{j=1}^{N} \frac{\left[U_{ij} - \hat{P}_i\right]^2}{\left[\hat{P}_i . Q_i\right]}$$

เมื่อ U_{ij} แทน คะแนนจากการตอบข้อสอบ 0 - 1 ของคนที่ j ข้อที่ i

$$P_i(\theta) = \frac{1}{1+e\left[-D(\theta - b_i)\right]}$$

$$Z_i = \frac{b_{iA} - b_{iB}}{SE_{biA}^2 + SE_{biB}^2}$$

$$SE_{bi}^2 = \sqrt{I_{ai} / \left(I_{ai} I_{bi} I_{aibi}^2\right)}$$

ถ้าค่า Z_i ที่คำนวณได้ค่าสูง แสดงว่า ข้อสอบนั้นยากสำหรับกลุ่มหนึ่งมากกว่าอีกกลุ่มหนึ่ง หรืออาจจะวิเคราะห์ความลำเอียงเป็นรายข้อด้วย ซึ่งคำนวณได้จากสูตร

$$t = \frac{b_{iA} - b_{iB}}{\sqrt{SE_{biA}^2 + SE_{biB}^2}}$$

เมื่อ $b_{iA} - b_{iB}$ แทน ค่าความยากของข้อสอบที่ได้จากการวิเคราะห์ในกลุ่มที่

Prince of Songkla University Pattani Campus

1 และกลุ่มที่ 2 ตามลำดับ

$SE^2_{biA} + SE^2_{biB}$ แทน ค่าความคลาดเคลื่อนมาตรฐานของข้อสอบที่ได้จากการวิเคราะห์จากกลุ่มที่ 1 และกลุ่มที่ 2 ตามลำดับ

ค่า t ที่คำนวณได้ คือ ดัชนีความลำเอียงของข้อสอบ ซึ่งดราบา (Berk, 1982 : 138 ; quoting Draba, 1977) เสนอให้ใช้ค่า $t \geq 2.4$ เป็นเกณฑ์ในการตัดสินว่าข้อสอบมีความลำเอียง สำหรับวิธีการนี้เรียกอีกอย่างหนึ่งว่า เป็นการวิเคราะห์ความลำเอียงตามแบบจำลองของ ราส์ช (Rasch Model)

2) แบบจำลองโลจิสติกชนิด 2 พารามิเตอร์ กำหนดให้ค่าการเดา (c_i) เท่ากับ 0 ส่วนค่าความยาก (b_i) และค่าอำนาจจำแนก (a_i) แปรเปลี่ยนไปตามกลุ่มผู้สอบ วิธีนี้เริ่มด้วยการประมาณค่าพารามิเตอร์ของข้อสอบแยกตามกลุ่มผู้สอบ กลุ่มย่อย A และ B ด้วยวิธี Maximum Likelihood แล้วเปรียบเทียบโค้งลักษณะการตอบข้อสอบด้วยค่า F ซึ่งคำนวณได้จากสูตร

$$F = \frac{SSE(Pooled) - [SSE(A) - SSE(B)]}{SSE(A) + SSE(B)} \cdot \frac{J_A + J_B - 4}{2}$$

เมื่อ SSE(A) แทน ความคลาดเคลื่อนมาตรฐานของข้อสอบที่ได้จากการวิเคราะห์กลุ่มที่ 1

SSE(B) แทน ความคลาดเคลื่อนมาตรฐานของข้อสอบที่ได้จากการวิเคราะห์กลุ่มที่ 2

SSE(Pooled) แทน ความคลาดเคลื่อนมาตรฐานของข้อสอบที่ได้จากการวิเคราะห์กลุ่มที่ 1 และกลุ่มที่ 2

J_A, J_B แทน จำนวนช่วงความสามารถ

และมีค่าชั้นแห่งความเป็นอิสระ คือ $2, (J_A + J_B - 4)$ ซึ่งการทดสอบด้วยค่า F จะบอกว่าเส้นถดถอยรวมสามารถอธิบายความแปรปรวนของสัดส่วนการตอบของเส้นถดถอยจากแต่ละกลุ่มได้เท่ากันหรือไม่ ถ้าเท่ากันหรือไม่มีนัยสำคัญทางสถิติก็ไม่จำเป็นต้องใช้เส้นถดถอยแยกกัน เพราะความสัมพันธ์ระหว่างความสามารถและการตอบข้อสอบของผู้สอบทั้งสองกลุ่มเป็นเช่นเดียวกัน และข้อสอบก็ทำหน้าที่ในการวัดความสามารถของคน 2 กลุ่มนี้ได้เหมือนกัน

Prince of Songkla University Pattani Campus

3) แบบจำลองโลจิสติกชนิด 3 พารามิเตอร์ เป็นแบบที่ให้ค่าความยาก (b_i) ค่าอำนาจจำแนก(a_i) และค่าสัมประสิทธิ์การเดา (c_i) ของข้อสอบแปรเปลี่ยนไปตามกลุ่มย่อยของผู้สอบ

การคำนวณหาดัชนีความลำเอียงของข้อสอบวิธีนี้เริ่มจากการประมาณค่าความสามารถของผู้สอบและค่าพารามิเตอร์ของข้อสอบเป็นรายข้อรวมของทุกคนทุกกลุ่มแล้วแปลงค่าพารามิเตอร์ที่ได้จากผู้สอบ2 กลุ่ม ให้อยู่บนมาตรฐานเดียวกันบนค่าประมาณความยาก จากนั้นแยกประมาณค่าความยากและค่าอำนาจจำแนกตามกลุ่มย่อยโดยใช้ค่าการเดาที่ประมาณได้จากการรวมทุกกลุ่ม การทำค่าความยากให้เป็นมาตรฐานนั้นจะตัดข้อสอบที่มีค่าความยากสูงมากหรือต่ำมากออก

การวิเคราะห์ความลำเอียงของข้อสอบตามแบบจำลอง โลจิสติก 3 พารามิเตอร์ มีหลายวิธี เช่น ลินน์ ลีไวน์ ฮาสติ้งส และวอร์ดรอพ (Hulin, Drasgows and Parson, 1983 : 776 ; quoting Linn, Levine, Hastings and Wardrop, 1981) ได้พัฒนาการทดสอบความลำเอียงของข้อสอบจากความแตกต่างระหว่างโค้งลักษณะของข้อสอบของผู้สอบ2 กลุ่มย่อย เป็นเกณฑ์การตัดสินความลำเอียงของข้อสอบ ผู้ที่ศึกษาความลำเอียงโดยการเปรียบเทียบความแตกต่างของโค้งลักษณะข้อสอบเช่น รัดเนอร์ (Rudner, 1977 : quoting in Berk, 1982 : 118) รัดเนอร์ เกทสัน และไนท์ (Rudner, Getson and Knight, 1980) ไรออนสัน และสับโคเวียต (Ironson and Subkovial, 1979) ซึ่งเกณฑ์ที่ใช้พิจารณาความลำเอียง คือ มีพื้นที่ใต้โค้งต่างกันตั้งแต่ .20 ขึ้นไป ในขณะที่ทัศนีย์ พีรมนตรี ใช้ค่ามากกว่า.40 การใช้เกณฑ์ความลำเอียง.20 หรือ .40 แทนที่จะใช้ 0.00 นั้นเนื่องจากยอมให้มีความคลาดเคลื่อนที่อาจจะเกิดจากการสุ่ม ผู้ที่นำดัชนีนี้ไปใช้และใช้เกณฑ์ที่ต่างกัน ได้แก่ คิมและโคเฮน(Kim and Cohen, 1991) โดยมีข้อตกลงเบื้องต้นว่าการกระจายของความแตกต่างของพื้นที่ใต้โค้งที่วัดได้มีลักษณะเป็นโค้งปกติ เกณฑ์การตัดสินความลำเอียงจะพิจารณาค่าที่มากกว่าค่าวิกฤตชนิดทางเดียวที่ระดับ.05 และ .01 (รัชดาพร แก้วชาฎก, 2544)

นอกจากนี้ ลินน์ และคนอื่นๆ (Linn, et al. 1981) ได้อธิบายเกี่ยวกับการหาดัชนีความลำเอียงของข้อสอบ โดยใช้รากที่สองของค่าเฉลี่ยความแตกต่างกำลังสองของโค้งลักษณะข้อสอบของกลุ่มผู้สอบ2 กลุ่ม เป็นดัชนีความลำเอียงของข้อสอบในช่วงความสามารถ-3 ถึง +3 ที่ได้จากการคำนวณ จากสูตร ดังนี้

$$RMSD = \frac{1}{600}\left[\sum_{J=1}^{J}[P_{iA}(\theta_J) - P_{iB}(\theta_J)]^2\right]^{1/2}$$

Prince of Songkla University Pattani Campus

โดยได้จากการแบ่งช่วงความสามารถจาก -3 ถึง +3 เป็น 600 ช่วง ค่าสัมบูรณ์ของความแตกต่างของโค้งทั้ง2 ที่จุดกึ่งกลางของ 600 ช่วง จะเป็น $P_{iA}(\theta_j) - P_{iB}(\theta_j)$ คูณด้วยความยาวของช่วงแต่ละช่วง คือ .01 การบวกค่าพื้นที่ทั้ง 600 ช่วง จะได้ค่าประมาณของพื้นที่ทั้งหมดที่อยู่ระหว่างโค้งทั้งสอง

เชปาร์ด และคนอื่นๆ (Shepard, et al. 1985 : 81) กล่าวว่า เมื่อพบว่าโค้งลักษณะข้อสอบของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแตกต่างกันมากพอที่จะเป็นหลักฐานได้ว่าข้อสอบที่ศึกษาอยู่ไม่ได้วัดลักษณะแฝงเดียวกัน สำหรับกลุ่มผู้สอบทั้งสองแล้วจะใช้ดัชนีที่บ่งชี้ถึงระดับลำเอียง ดังนี้

1. พื้นที่ชนิดไม่มีเครื่องหมาย (Unsigned Area) เป็นค่าสัมบูรณ์ของพื้นที่ระหว่างโค้งลักษณะข้อสอบทั้งสอง

2. ผลบวกกำลังสองแบบที่1 (SOS 1) เป็นค่าผลบวกของความแตกต่างระหว่างกำลังสองทุกค่าของความสามารถที่เป็นไปได้

3. ผลบวกกำลังสองแบบที่2 (SOS 2) คล้าย SOS 1 แต่ให้น้ำหนักความแตกต่างกำลังสองด้วยส่วนกลับของความคลาดเคลื่อนของความแปรปรวนในการประมาณค่าความเป็นไปได้ ดังนั้นถ้ามีการประมาณโค้งอย่างไม่ดีในส่วนใดส่วนหนึ่งแล้วจะทำให้ลดประโยชน์ต่อความแตกต่างระหว่างสองกลุ่ม

4. ไคสแควส์ ($IRT - \chi^2$) เป็นการทดสอบนัยสำคัญทางสถิติ โดยการเปรียบเทียบความแตกต่างของค่าความยากและอำนาจจำแนกระหว่างกลุ่มพร้อมๆกัน โดย ลอร์ด (Lord, 1980)

5. พื้นที่ชนิดมีเครื่องหมาย เหมือนพื้นที่ชนิดไม่มีเครื่องหมายแต่มีการให้เครื่องหมายไว้เพื่อแสดงให้เห็นว่ากลุ่มใดเสียประโยชน์และกลุ่มใดได้ประโยชน์ ถ้าโค้งลักษณะการตอบข้อสอบตัดกัน แสดงว่า มีความลำเอียงต่างกันสำหรับผู้มีความสามารถต่างกันและค่าสัมบูรณ์ของพื้นที่ชนิดมีเครื่องหมายจะน้อยกว่าชนิดไม่มีเครื่องหมาย

6. ผลบวกกำลังสองแบบที่3 (SOS 3) เป็นผลบวกกำลังสองของความแตกต่างชนิดมีเครื่องหมายของ SOS 1 โดยการใช้ความหมายของความแตกต่างแต่ละตัวที่ยังไม่ยกกำลังสอง

7. ผลบวกกำลังสองแบบที่4 (SOS 4) เป็นดัชนีคู่ขนานกับ SOS 2 แต่เป็นผลบวกที่มีการให้น้ำหนักและให้เครื่องหมาย

Prince of Songkla University Pattani Campus

7. งานวิจัยที่เกี่ยวข้อง

7.1 งานวิจัยภายในประเทศ

งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบเป็นการเปรียบเทียบผลการวิเคราะห์ความลำเอียง ที่ใช้วิเคราะห์ต่างกัน โดยพิจารณาความลำเอียงที่เกิดขึ้นระหว่างประชากรกลุ่มย่อยที่มีเพศต่างกัน สภาพภูมิศาสตร์ต่างกัน และการพัฒนาเกณฑ์การตัดสินความลำเอียงทางเพศ ผู้วิจัยรวบรวมและสรุปได้ดังนี้

สุพัฒน์ สุกมลสันต์(2534 : 89) วิเคราะห์ความลำเอียงของข้อสอบภาษาอังกฤษเพื่อคัดเลือกเข้าศึกษาต่อในมหาวิทยาลัย ด้วยวิธีวิเคราะห์ความลำเอียง 3 วิธี ได้แก่ วิธีกำหนดค่าเดลตา (Delta-Plot Method) วิธีไคสแควร์ชนิดที่แบ่งความสามารถของผู้สอบเป็น 3 ระดับ ได้แก่ กลุ่มที่มีระดับความสามารถต่ำ(ผู้ได้คะแนนรวม0-40 คะแนน) กลุ่มมีระดับความสามารถปานกลาง(ผู้ได้คะแนนรวม 41-70 คะแนน) และกลุ่มที่มีระดับความสามารถสูง (ผู้ได้คะแนนรวม71-100 คะแนน) วิธีการวัดพื้นที่ความแตกต่างระหว่างโค้งลักษณะข้อสอบที่วิเคราะห์ตามทฤษฎีการตอบข้อคำถามแบบ 3 พารามิเตอร์ จากแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข และชุด กขค ปี2531-2533 ซึ่งมีข้อสอบชุดละ 100 ข้อ ตัวแปรที่ศึกษาความลำเอียง คือ เพศ และภาคภูมิศาสตร์ของผู้สอบ ซึ่งแยกออกเป็น 5 ภาค ตามภูมิลำเนาของผู้สอบ ได้แก่ ภาคกลาง ภาคตะวันออก ภาคใต้ ภาคตะวันออกเฉียงเหนือและภาคเหนือ

เกณฑ์การตัดสินความลำเอียงของข้อสอบ สำหรับวิธีต่างๆ ที่ สุพัฒน์ สุกมลสันต์ กำหนดขึ้น คือ

1. วิธีกำหนดจุดค่าเดลตา ข้อสอบที่ลำเอียง คือข้อที่มีระยะตั้งฉากจากคู่อันดับค่าเดลตาไปยังเส้นแกนหลัก(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 พื้นที่ที่มากกว่า 0.70 แสดงว่ามีความลำเอียงมาก

3.2 พื้นที่ที่มีค่าอยู่ระหว่าง 0.40-0.70 แสดงว่ามีความลำเอียงปานกลาง

3.3 พื้นที่ที่มีค่ามากกว่า 0.00 แต่น้อยกว่า 0.40 แสดงว่ามีความลำเอียงน้อย

ผลการศึกษาพบว่า แบบทดสอบภาษาอังกฤษ ฉบับ กข และ กขค ปี2531-2533 มีความลำเอียงต่อเพศในช่วง7-28 และช่วง 4-41 ข้อ โดยมีแนวโน้มที่ลำเอียงต่อเพศชายมากกว่าเพศหญิง

Prince of Songkla University Pattani Campus

และลำเอียงต่อภาคภูมิศาสตร์ในช่วง6-45 ข้อ และ 5-43 ข้อ โดยมีความลำเอียงต่อผู้สอบจากภาคอื่นมากกว่าภาคกลางประมาณ 2-3 เท่า

การเปรียบเทียบผลการวิเคราะห์โดยใช้วิธีวิเคราะห์ต่างกัน พบว่า จำนวนข้อสอบที่ระบุว่าลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติ และแต่ละวิธีให้ผลที่มีความสัมพันธ์กันอย่างไม่มีสัมพันธ์นัยสำคัญทางสถิติ การวิเคราะห์ด้วยวิธีของทฤษฎีการตอบข้อคำถามเมื่อใช้เกณฑ์พื้นที่แสดงความลำเอียงของข้อสอบ มากกว่า0.40 พบว่า มีจำนวนข้อสอบที่มีความลำเอียงน้อยกว่า 2 วิธีแรก ประมาณ 3-4 เท่า

เรวดี อินทสะระ (2539 : 277) ศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่วิเคราะห์ความลำเอียงต่อเพศด้วยวิธีใช้ทฤษฎีการตอบข้อคำถาม(TRT) วิธีแมนเทล-แฮนส์เซล (MH) และวิธี SIBTEST การตัดสินผลการสอบที่คิดคะแนนมาตรฐานที่ปกติและคะแนนน้ำหนักความสามารถและสาเหตุความลำเอียงของข้อสอบ โดยศึกษาความลำเอียงของข้อสอบคัดเลือกเข้าศึกษาในชั้นปีที่1 ประเภทรับตรง ปีการศึกษา 2538 ของมหาวิทยาลัยสงขลานครินทร์ ในวิชาภาษาไทย ก วิชาสังคมศึกษา ก วิชาภาษาอังกฤษ กข วิชาละ8,127 คน (ชาย 2,722 คน หญิง 5,405 คน) วิชาภาษาไทย กข วิชาสังคม กข และวิชาภาษาอังกฤษ กขค วิชาละ5,415 คน (ชาย 1,454 คน หญิง 3,961 คน) ความเที่ยงตรงเชิงพยากรณ์ศึกษาจากคะแนนสอบคัดเลือกกับเกรดภาคเรียนที่1 ปีการศึกษา 2538 ของนักเรียนที่ได้รับการคัดเลือกจากประเภทรับตรง สายวิทยาศาสตร์ 763 คน และสายศิลปศาสตร์ 281 คน และสาเหตุความลำเอียงของข้อสอบจากการระบุสาเหตุของนักวัดผลการศึกษาหรืออาจารย์ผู้สอน จำนวน50 คน และนิสิตนักศึกษาที่เรียนในสาขานั้นๆวิชาละ 50 คน

ผลการศึกษาพบว่า วิธีการตรวจสอบความลำเอียงทั้ง3 วิธี ตัดสินจำนวนข้อสอบที่ลำเอียงแตกต่างกันในวิชาภาษาไทย ก ฉบับที่2 และวิชาสังคมศึกษา ก ฉบับที่1 ที่ระดับนัยสำคัญทางสถิติที่ .05 นอกนั้นต่างกันที่ระดับนัยสำคัญทางสถิติที่01 โดยที่วิธีใช้ทฤษฎีตอบข้อคำถามตัดสินจำนวนข้อสอบที่ลำเอียงมากที่สุด ความสำคัญของลำดับที่ของการสอบไม่ว่าจะคิดคะแนนมาตรฐานที่ปกติ หรือคิดคะแนนน้ำหนักความสามารถและใช้ข้อสอบจำนวนทั้งหมดหรือใช้เฉพาะข้อสอบที่ปราศจากความลำเอียงต่างมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ.01

ประสิทธิภาพในการทำนายผลสัมฤทธิ์ทางการเรียนในสายวิทยาศาสตร์ การคิดคะแนนน้ำหนักความสามารถที่ใช้ข้อสอบจำนวนทั้งหมดและใช้เฉพาะข้อสอบที่ปราศจากความลำเอียงมประสิทธิภาพในการทำนายสูงกว่าการคิดคะแนนมาตรฐานที่ปกติ ส่วนสายศิลปศาสตร์การคิดคะแนนมาตรฐานที่ปกติทั้งที่ใช้ข้อสอบจำนวนทั้งหมดและใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง มีประสิทธิภาพในการทำนายสูงกว่าการคิดคะแนนน้ำหนักความสามารถและสาเหตุของความลำเอียงของข้อสอบต่อเพศทั้งชายและหญิง เกิดจากข้อสอบเป็นข้อคำถามที่ผู้สอบได้รับการ

Prince of Songkla University Pattani Campus

ฝึกฝนเฉพาะ จะมีโอกาสตอบถูกมากกว่า เป็นเรื่องราวที่กลุ่มนั้นๆสนใจและเป็นข้อสอบที่ถามความจำ

เพ็ญพนา สุขสม (2540 : บทคัดย่อ) เปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบที่แตกต่างกัน3 วิธี การวิจัยครั้งนี้มีความมุ่งหมายเพื่อเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบ ด้วยวิธีวิเคราะห์3 วิธี คือวิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบที่มี 3 พารามิเตอร์และวิธีแมนเทล – แฮนส์เซล ตามตัวแปรเพศและที่ตั้งของโรงเรียนและเพื่อศึกษาความสัมพันธ์ระหว่างดัชนีความลำเอียงทั้ง3 วิธี โดยใช้แบบทดสอบประเมินคุณภาพและวัดผลปลายปี วิชาภาษาไทย จำนวน50 ข้อ เป็นเครื่องมือในการเก็บรวบรวมข้อมูล กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่6 ปีการศึกษา2538 ของโรงเรียนในสังกัดสำนักงานการประถมศึกษาจังหวัดนครศรีธรรมราช จำนวน24,000 คน

ผลการวิจัยพบว่า1. วิธีวิเคราะห์ความลำเอียงแต่ละวิธีตรวจพบจำนวนข้อสอบที่ลำเอียงแตกต่างกัน วิธีแมนเทล – แฮนส์เซลเป็นวิธีที่ตรวจพบข้อสอบที่ลำเอียงมากที่สุด รองลงมาคือวิธีโค้งลักษณะข้อสอบที่มี3 พารามิเตอร์ และวิธีแปลงค่าความยากตรวจพบข้อสอบที่ลำเอียงน้อยที่สุด สำหรับความสอดคล้องนั้นดัชนีความลำเอียงทั้ง3 วิธีมีความสัมพันธ์กันทางบวกอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ซึ่งหมายความว่าวิธีวิเคราะห์ความลำเอียงทั้ง3 วิธีให้ผลสอดคล้องกัน เมื่อพิจารณาตามตัวแปรเพศพบว่าความสัมพันธ์มีค่าอยู่ระหว่าง4011 ถึง .6662 โดยที่วิธีโค้งลักษณะข้อสอบ3 พารามิเตอร์ มีความสัมพันธ์กับวิธีแมนเทล– แฮนส์เซล สูงกว่าวิธีแปลงค่าความยากของข้อสอบ และเมื่อพิจารณาตามตัวแปรที่ตั้งของโรงเรียนพบว่า ควมสัมพันธ์มีค่าอยู่ระหว่าง .5676 ถึง .7847 โดยที่วิธีโค้งลักษณะข้อสอบที่มี3 พารามิเตอร์มีความสัมพันธ์กับวิธีแมนเทล - แฮนส์เซลสูงกว่าวิธีแปลงค่าความยากของข้อสอบเช่นเดียวกัน

นพมาศ พิพัฒนสุข(2541 : บทคัดย่อ) เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีแมนเทลแฮนส์เซลกับวิธีถดถอยโลจิสติกในแบบสอบชนิดพหุมิติเมื่อใช้เกณฑ์จับคู่เปรียบเทียบแตกต่างกัน3 เกณฑ์ ได้แก่ คะแนนรวม คะแนนแบบสอบย่อย และคะแนนหลายแบบสอบย่อย โดยข้อมูลที่ใช้ในการศึกษาเก็บรวบรวมจากผลการตอบแบบทดสอบความสามารถทางคณิตศาสตร์ชั้นประถมศึกษาปีที่ 6 ที่ผู้วิจัยสร้างขึ้น กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่6 ในสังกัดสำนักงานการประถมศึกษากรุงเทพมหานคร จำนวน1,076 คน วิเคราะห์ค่าสถิติพื้นฐานโดยใช้โปรแกรม SPSS วิเคราะห์คุณภาพของข้อสอบโดยใช้โปรแกรม IRT(BAY) และ CTIA ตรวจสอบความตรงเชิงโครงสร้างด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันโดยใช้โปรแกรม LISREL 8.0 และตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยใช้โปรแกรม SIBTEST และ SPSS

Prince of Songkla University Pattani Campus

ผลการวิจัยพบว่า วิธีแมนเทล-แฮนส์เซลเมื่อใช้คะแนนรวมเป็นเกณฑ์จับคู่เปรียบเทียบตรวจพบข้อสอบทำหน้าที่ต่างกัน จำนวน15 ข้อ (20%) และเมื่อใช้คะแนนแบบสอบย่อยเป็นเกณฑ์จับคู่ตรวจพบจำนวน14 ข้อ (18.67%) สำหรับวิธีถดถอยโลจิสติกเมื่อใช้คะแนนรวมเป็นเกณฑ์จับคู่ตรวจพบข้อสอบทำหน้าที่ต่างกันมากที่สุด จำนวน20 ข้อ (26.67%) รองลงมาคือ ใช้คะแนนแบบสอบย่อยเป็นเกณฑ์จับคู่ตรวจพบ จำนวน17 ข้อ (22.67%) และตรวจพบน้อยที่สุดคือ เมื่อใช้คะแนนหลายแบบสอบย่อยเป็นเกณฑ์จับคู่ตรวจพบ จำนวน13 ข้อ (17.33%) ส่วนวิธีแมนเทล-แฮนส์เซลมีประสิทธิภาพมากกว่าวิธีถดถอยโลจิสติก ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบชนิดพหุมิติเมื่อใช้เกณฑ์จับคู่คะแนนรวม และมีประสิทธิภาพไม่แตกต่างกันเมื่อใช้เกณฑ์จับคู่คะแนนแบบสอบย่อยและวิธีถดถอยโลจิสติกเมื่อใช้เกณฑ์จับคู่เปรียบเทียบคะแนนหลายแบบสอบย่อยมีความเหมาะสมในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบชนิดพหุมิติ

รัชฎาพร แก้วชาฎก (2544 : บทคัดย่อ) เปรียบเทียบความลำเอียงของข้อสอบโจทย์ปัญหาคณิตศาสตร์โดยใช้วิธีวิธีแปลงค่าความยาก วิธีแมนเทล-แฮนส์เซลและวิธีโค้งลักษณะข้อสอบ โดยกลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักเรียนชั้นประถมศึกษาปีที่6 ปีการศึกษา 2542 ในจังหวัดยะลา จำนวน 1,339 คน ซึ่งได้มาโดยการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) โดยมีกลุ่มนักเรียนที่ใช้ภาษาไทยในชีวิตประจำวันและกลุ่มที่ไม่ใช้ภาษาไทยในชีวิตประจำวันเป็นชั้น (Strata) และมีโรงเรียนเป็นหน่วยการสุ่ม (Sampling Unit) เครื่องมือที่ใช้ในการรวบรวมข้อมูลเป็นแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ชนิด 4 ตัวเลือก จำนวน40 ข้อ

ผลการวิจัยพบว่า

1. วิธีแปลงค่าความยากของข้อสอบตรวจพบจำนวนข้อสอบที่มีความลำเอียงแตกต่างจากวิธีโค้งลักษณะข้อสอบและวิธีของแมนเทลแฮนส์เซล อย่างมีนัยสำคัญทางสถิติ ที่ระดับ.001 ส่วนวิธีโค้งลักษณะข้อสอบกับวิธีของแมนเทล-แฮนส์เซล ตรวจพบจำนวนข้อสอบที่มีความลำเอียงไม่แตกต่างกัน

2. ค่าความเชื่อมั่นของแบบทดสอบหลังการคัดเลือกข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบและวิธีของแมนเทลแฮนส์เซล มีค่าไม่แตกต่างกัน

3. ค่าฟังก์ชันสารสนเทศของแบบทดสอบหลังการคัดเลือกข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบและวิธีของแมนเทลแฮนส์เซล มีค่าแตกต่างกันทุกระดับความสามารถ โดยที่ช่วงความสามารถปานกลางถึงสูง ($0.59 \leq \theta \leq 1.46$) วิธีโค้งลักษณะข้อสอบมีค่าฟังก์ชันสารสนเทศสูงสุด รองลงมาคือวิธีแปลงค่าความยากของข้อสอบ

Prince of Songkla University Pattani Campus

และวิธีของแมนเทล – แฮนส์เซล แต่ในช่วงความสามารถต่ำถึงปานกลาง($-3.02 \leq \theta < 0.59$) วิธีแปลงค่าความยากของข้อสอบมีค่าฟังก์ชันสานสนเทศสูงสุด รองลงมาคือวิธีของแมนเทล–แฮนส์เซล และวิธีโค้งลักษณะข้อสอบ

วรรณา รอดตัว (2544 : บทคัดย่อ) เปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบที่แตกต่างกัน 3 วิธี โดยมีวัตถุประสงค์เพื่อค้นหาความลำเอียงของแบบทดสอบวิชาคณิตศาสตร์ I ซึ่งเป็นแบบทดสอบที่ใช้ในการสอบคัดเลือกเข้าศึกษาต่อระดับปริญญาตรี ในสถาบันอุดมศึกษาของรัฐและเอกชน ที่สังกัดทบวงมหาวิทยาลัย ประจำปีการศึกษา 2542 โดยวิธีวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีค่าอำนาจจำแนกของข้อสอบ วิธีแมนเทล แฮนส์เซลและวิธีซิปเทสท์ เปรียบเทียบผลการวิเคราะห์ความลำเอียงและศึกษาความสัมพันธ์ของวิธีวิเคราะห์ความลำเอียงทั้ง 3 วิธี กลุ่มตัวอย่างที่ใช้เป็นผู้เข้าสอบคัดเลือกเพื่อเข้าศึกษาต่อระดับปริญญาตรี ในสถาบันอุดมศึกษาของรัฐและเอกชน ประจำปีการศึกษา 2542 จำนวน 2540 คน โดยใช้วิธีสุ่มอย่างง่ายมาจากผู้เข้าสอบคัดเลือกทั้งสิ้นจำนวน 99,562 คน ผู้วิจัยนำผลการสอบที่ได้มาค้นหาความลำเอียงด้วยวิธีค่าอำนาจจำแนกของข้อสอบ วิธีแมนเทล– แฮนส์เซลและวิธีซิปเทสท์ ตามตัวแปรเพศและเขตที่ตั้งของสถานศึกษา หาค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงที่ได้จากการวิเคราะห์ โดยใช้สูตรเพียร์สัน โพรดักโมเมนต์ และเปรียบเทียบจำนวนข้อสอบที่ลำเอียงที่ได้จากการวิเคราะห์ด้วย 3 วิธีนี้

ผลการวิจัยพบว่า

1. ผลการวิเคราะห์ความลำเอียงของข้อสอบรายข้อด้วยวิธีวิเคราะห์ 3 วิธี ระหว่างตัวแปรเพศและเขตที่ตั้งของสถานศึกษา จากข้อสอบทั้งหมด 28 ข้อ พบว่า เขตที่ตั้งของสถานศึกษาพบจำนวนข้อสอบที่ลำเอียงสูงกว่าเพศ ดังนี้ 1.) วิธีค่าอำนาจจำแนกของข้อสอบพบข้อสอบที่ลำเอียงตามเพศ จำนวน 3 ข้อ ลำเอียงตามเขตที่ตั้งของสถานศึกษา จำนวน 12 ข้อ 2.) วิธีแมนเทล – แฮนส์เซล พบข้อสอบที่ลำเอียงตามเพศ จำนวน 3 ข้อ ลำเอียงตามเขตที่ตั้งของสถานศึกษา จำนวน 8 ข้อ 3.) วิธีซิปเทสท์ พบข้อสอบที่ลำเอียงตามเพศจำนวน 4 ข้อ ลำเอียงตามเขตที่ตั้งของสถานศึกษา จำนวน 6 ข้อ

2. ค่าสหสัมพันธ์ของดัชนีความลำเอียงของข้อสอบ จากการวิเคราะห์ด้วยวิธีค่าอำนาจจำแนกของข้อสอบ วิธีแมนเทล- แฮนส์เซลและวิธีซิปเทสท์ ระหว่างตัวแปรเพศและเขตที่ตั้งของสถานศึกษา พบว่า วิธีแมนเทล – แฮนส์เซลและวิธีซิปเทสท์ มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในตัวแปรเพศและมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในตัวแปรเขตที่ตั้งของสถานศึกษา

Prince of Songkla University Pattani Campus

3. เปรียบเทียบจำนวนข้อสอบที่ลำเอียงจากการใช้วิธีวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีค่าอำนาจจำแนกของข้อสอบ วิธีแมนเทล- แฮนส์เซลและวิธีซิปเทสท์ ตามตัวแปรเพศและเขตที่ตั้งของสถานศึกษา พบว่า วิธีการวิเคราะห์ความลำเอียงของข้อสอบทั้ง 3 วิธี พบจำนวนข้อสอบที่ลำเอียงไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ประเสริฐ จันต๊ะไพร (2546 : บทคัดย่อ) วิเคราะห์ความลำเอียงของข้อสอบตามทฤษฎีคลาสสิกคอล 5 วิธี ซึ่งมีวัตถุประสงค์เพื่อศึกษาการวิเคราะห์ความลำเอียงของข้อสอบคัดเลือกเข้าศึกษาต่อในระดับชั้นมัธยมศึกษาปีที่ 1 โรงเรียนสาธิตมหาวิทยาลัยรามคำแหง ประจำปีการศึกษา 2544 ในวิชาภาษาไทย จำนวน 50 ข้อ ตามทฤษฎีคลาสสิกคอล 5 วิธี ศึกษาสัมประสิทธิ์สหสัมพันธ์และเปรียบเทียบดัชนีความลำเอียงของข้อสอบระหว่างดัชนีความลำเอียง 5 วิธี กลุ่มประชากรที่ใช้เป็นนักเรียนที่เข้าสอบคัดเลือกศึกษาต่อในระดับชั้นมัธยมศึกษาปีที่ จำนวน 3,564 คน ผู้วิจัยนำผลการสอบมาค้นหาความลำเอียงด้วยวิธีแปลงค่าความยาก การวิเคราะห์ความแปรปรวน วิธีไคสแควร์ วิธีวิเคราะห์องค์ประกอบและวิธีการหาค่าอำนาจจำแนกของข้อสอบ ตามตัวแปรเพศ หาค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงที่ได้จากการวิเคราะห์ โดยใช้สูตรเพียร์สันโพรดักโมเมนต์ และเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียง จากการวิเคราะห์ความลำเอียงของข้อสอบทั้ง 5 วิธี โดยใช้วิธีไคสแควร์

ผลการวิจัยพบว่า

1. การวิเคราะห์ความลำเอียงของข้อสอบรายข้อด้วยวิธีวิเคราะห์ 5 วิธี ระหว่างตัวแปรเพศ จากข้อสอบทั้งหมด 50 ข้อ พบว่า วิธีวิเคราะห์ความแปรปรวน พบจำนวนข้อสอบที่ลำเอียงจำนวน 25 ข้อ รองลงมาคือ วิธีวิเคราะห์องค์ประกอบ พบจำนวนข้อสอบที่ลำเอียงจำนวน 24 ข้อ วิธีการหาค่าอำนาจจำแนกของข้อสอบ พบจำนวนข้อสอบที่ลำเอียงจำนวน 14 ข้อ วิธีแปลงค่าความยากและวิธีไคสแควร์ พบจำนวนข้อสอบที่ลำเอียงจำนวน 1 ข้อ

2. ค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงของข้อสอบจากการวิเคราะห์ความลำเอียงทั้ง 5 วิธี ระหว่างตัวแปรเพศ พบว่า วิธีแปลงค่าความยากกับวิธีการหาค่าอำนาจจำแนกของข้อสอบและวิธีวิเคราะห์ความแปรปรวนกับวิธีการหาค่าอำนาจจำแนกของข้อสอบ มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3. วิธีการวิเคราะห์ความลำเอียงทั้ง 5 วิธี ระหว่างตัวแปรเพศ พบว่า จำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ปิยะทิพย์ ตินวร (2549 : บทคัดย่อ) เปรียบเทียบประสิทธิภาพระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัดกับวิธีถดถอยโลจิสติกในแบบทดสอบพหุมิติภายใต้เงื่อนไข ขนาดของกลุ่มตัวอย่าง 3 ขนาด (2,000 คน 1,000 คน และ300 คน) ความยาวของแบบทดสอบ 3 ขนาด (40 ข้อ

Prince of Songkla University Pattani Campus

30 ข้อ และ 20 ข้อ) และเกณฑ์การจับคู่2 เกณฑ์ (คะแนนรวมทั้งฉบับ= TS และคะแนนแบบทดสอบย่อย = SS) เมื่อใช้ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยวิธีซิปเทสท์ เป็นเกณฑ์สำหรับการเปรียบเทียบประสิทธิภาพกลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่6 สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา2546 ที่เข้าสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ วิชาภาษาไทย จำนวน2,000 คนวิเคราะห์ค่าสถิติพื้นฐานโดยใช้โปรแกรม SPSS ตรวจสอบความตรงเชิงโครงสร้างด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองโดยใช้โปรแกรม LISREL 8.50 และตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยใช้โปรแกรม SIBTEST, LISREL 8.50 และ SPSS

ผลการวิจัยพบว่า

1. วิธีถดถอยโลจิสติก (LR) มีประสิทธิภาพไม่แตกต่างกันกับวิธีวิเคราะห์องค์ประกอบจำกัด(RFA) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสบในแบบทดสอบพหุมิติจาก2 เงื่อนไข (ขนาดกลุ่มตัวอย่างx ความยาวแบบทดสอบ x เกณฑ์การจับคู่ คือ 1,000 คน x 20 ข้อ x SS และ 300 คน x 20 ข้อ x SS

2. วิธี LR มีประสิทธิภาพมากกว่าวิธี RFA ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบพหุมิติ 16 เงื่อนไข คือ 1) 2,000 คน x 40 ข้อ x TS 2) 2,000 คน x 40 ข้อ x SS 3) 2,000 คน x 30 ข้อ x TS 4) 2,000 คน x 30 ข้อ x SS 5) 2,000 คน x 20 ข้อ x TS 6) 2,000 คน x 30 ข้อ x SS 7) 1,000 คน x 40 ข้อ x TS 8) 1,000 คน x 40 ข้อ x SS 9) 1,000 คน x 30 ข้อ x TS 10) 1,000 คน x 30 ข้อ x SS 11) 1,000 คน x 20 ข้อ x TS 12) 300 คน x 40 ข้อ x TS 13) 300 คน x 40 ข้อ x TS 14) 300 คน x 30 ข้อ x TS 15) 300 คน x 30 ข้อ x SS 16) 300 คน x 20 ข้อ x TS เนื่องจากวิธี LR มีอัตราความคลาดเคลื่อนประเภทที่1 น้อยกว่าวิธี RFA อย่างมีนัยสำคัญทางสถิติที่ระดับ05

7.2 งานวิจัยต่างประเทศ

จากการศึกษาเอกสารและงานวิจัยในต่างประเทศที่เกี่ยวข้องกับความลำเอียงนั้น มีผู้วิจัยหลายท่านที่ศึกษาเรื่องนี้ ดังเช่น

เมเซอร์และคณะ (Mazor and others, 1991 : 443-451 อ้างถึงใน วรรณา รอดตัว 2544 : 48) ศึกษาผลกระทบของขนาดกลุ่มตัวอย่างที่มีต่อการวิเคราะห์ความลำเอียงด้วยวิธี MH โดยใช้ข้อมูลที่จำลองโดยโปรแกรมDATAGEN แบบ 3 พารามิเตอร์ ขนาดกลุ่มตัวอย่างที่ศึกษามี5 ขนาด คือ 2,000 คน 1,000 คน 500 คน 200 คน และ 100 คน ความยาวของแบบทดสอบชุดละ 75 ข้อ

Prince of Songkla University Pattani Campus

ผลการวิจัยพบว่า เมื่อใช้กลุ่มตัวอย่าง2,000 คน วิธี MH จะตรวจค้นความลำเอียงได้ผิดพลาดร้อยละ 30 และเมื่อใช้จำนวนผู้สอบน้อยกว่า500 คน พบว่าจะตรวจค้นหาความลำเอียงผิดพลาดร้อยละ 50 และข้อสอบที่ไม่สามารถตรวจค้นความลำเอียงได้เป็นข้อสอบที่ยากมาก ข้อสอบที่มีค่าอำนาจจำแนกต่ำ และข้อสอบที่มีค่าความยากต่างกันเล็กน้อยสำหรับ2กลุ่ม

คลอเซอร์และคณะ (Clauser et. al, 1991 : 53-359) ศึกษาถึงอิทธิพลต่างๆที่มีผลต่อวิธีการตรวจสอบความลำเอียงของข้อสอบด้วยวิธีของแมนเทลแฮนส์เซล ข้อมูลที่ใช้เป็นข้อมูลที่จำลองขึ้นเพื่อใช้ในการศึกษาการตรวจสอบความลำเอียงแบบสม่ำเสมอข้อมูลที่จำลองขึ้นเป็นแบบทดสอบจำนวน 75 ข้อ ให้มีความลำเอียง16 ข้อ และมีระดับความแตกต่างของค่าความยากหลายระดับระหว่างกลุ่มอ้างอิงและกลุ่มเป้าหมาย กลุ่มตัวอย่างที่ใช้ศึกษามีจำนวน3,000 คน เป็นกลุ่มอ้างอิง 1,500 คน กลุ่มเป้าหมาย 1,500 คน มีการวิเคราะห์ระดับความยากของแบบทดสอบที่ต่างกันหลายระดับกับระดับความยากของแบบทดสอบ5 ระดับ อำนาจจำแนก4 ระดับ และการกระจายของความสามารถ 2 ลักษณะ

ผลการศึกษาพบว่า สิ่งที่มีอิทธิพลต่อการตรวจสอบความลำเอียงของข้อสอบด้วยวิธีของแมนเทล-แฮนส์เซล คือความแตกต่างในระดับความยากระหว่างผู้สอบทั้งสองกลุ่มรวมทั้งค่าอำนาจจำแนกและค่าความยากของข้อสอบด้วย จากการวิจัยนี้ทำให้ได้ข้อเสนอว่า วิธีแมนเทลแฮนส์เซลใช้ได้ดี สำหรับข้อสอบที่มีค่าอำนาจจำแนกสูงและมีแนวโน้มว่าไม่สามารถตรวจสอบความลำเอียงของข้อสอบที่มีค่าความยากสูง

โรเจอร์สและสวามินาทาน (Rogers and Swaminathan, 1993 : 105-116) ศึกษาเปรียบเทียบผลวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีถดถอยแบบโลจิสติก(วิธี LR) กับวิธี แมนเทล-แฮนส์เซล (วิธี MH) โดยใช้ข้อมูลที่จำลองขึ้น เงื่อนไขการศึกษาเป็น4 แบบ คือ ขนาดกลุ่มตัวอย่าง2 กลุ่ม ขนาด 250 คน และ 500 คนต่อกลุ่ม ระดับความเหมาะสมของโมเดลข้อมูล เป็น 2 ระดับ คือระดับที่เหมาะสมใช้รูปแบบ 2 พารามิเตอร์ ระดับที่ไม่เหมาะสมใช้รูปแบบ3 พารามิเตอร์ โดยค่าการเดาจะถูกกำหนดที่0.2 เครื่องมือที่ใช้เป็นข้อสอบจำนวน40 ข้อ ที่จำลองขึ้นมา7 แบบ ดังนี้ แบบที่ 1 ค่าความยากปานกลางและค่าอำนาจจำแนกต่ำ แบบที่3 ค่าความยากและค่าอำนาจจำแนกต่ำแบบที่ 4 ค่าความยากปานกลางและค่าอำนาจจำแนกสูง แบบที่5 ค่าความยากสูงค่าอำนาจจำแนกต่ำ แบบที่6 ค่าความยากและค่าอำนาจจำแนกสูง แบบที่7 เป็นแบบผสมคือ ค่าอำนาจจำแนกและค่าความยากต่างกันสำหรับ2 กลุ่ม โดยวิเคราะห์ร้อยละของความลำเอียงที่ตรวจค้นพบในความลำเอียงแบบยูนิฟอร์มและนอนยูนิฟอร์ม

ผลการวิจัยพบว่า ร้อยละของความลำเอียงที่ตรวจค้นพบในความลำเอียงแบบยูนิฟอร์มสูงกว่าแบบนอนยูนิฟอร์มทั้ง2 วิธีและในทุกเงื่อนไข โดยวิธี LR จะตรวจพบร้อยละของความลำเอียง

Prince of Songkla University Pattani Campus

สูงที่ระดับความยากปานกลางค่าอำนาจจำแนกสูง สำหรับความลำเอียงแบบนอนยูนิฟอร์มพบว่า วิธี LR จะตรวจค้นความลำเอียงลดลงร้อยละ14 ในโมเดลข้อมูลที่ไม่เหมาะสม ส่วนวิธี MH จะตรวจค้นพบลดลงร้อยละ 7 และในขนาดกลุ่มตัวอย่างที่เพิ่มขึ้น วิธีLR ตรวจค้นพบความลำเอียงได้เพิ่มขึ้นร้อยละ 19 ส่วนวิธี MH จะตรวจพบเพิ่มขึ้นร้อยละ 11 และที่ความยากต่ำ วิธีMH ตรวจค้นพบได้น้อยกว่าวิธีLR แต่ที่ความยากสูงทั้ง 2 วิธีให้ผลใกล้เคียงกัน แต่วิธีMH ไม่สามารถตรวจสอบได้ในระดับความยากปานกลาง และวิธีLR จะตรวจค้นความลำเอียงได้ดีกว่าวิธีMH ในความลำเอียงทั้ง2 แบบ แต่ค่าใช้จ่ายที่ใช้ในการวิเคราะห์จะสูงกว่าวิธีMH

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องรวมไปถึงจุดเด่นจุดด้อยของวิธีการวิเคราะห์ความลำเอียงประกอบกับความพร้อมในเรื่องของเอกสาร เครื่องมือ อุปกรณ์ และโปรแกรมสำหรับการศึกษา การวิจัยครั้งนี้ผู้วิจัยเลือกการวิเคราะห์ความลำเอียงของข้อสอบโดยใช้วิธีแมนเทล แฮนส์เซล ซึ่งเป็นวิธีที่อยู่ในกลุ่มการใช้ตารางการจรณ์ในการวิเคราะห์สามารถนำไปวิเคราะห์ด้วยโปรแกรม Microsoft Excel และวิธีโค้งลักษณะข้อสอบ3 พารามิเตอร์โดยวัดค่าความแตกต่างพารามิเตอร์ความยาก เป็นวิธีที่อยู่ในกลุ่มการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ ใช้โปรแกรมBILOG MG 3.0 ในการวิเคราะห์ และทำการตรวจสอบความลำเอียงของแบบทดสอบคณิตศาสตร์ ระดับชั้นประถมศึกษาปีที่5 ที่ใช้ในการประเมินคุณภาพการศึกษาระดับท้องถิ่น ปีการศึกษา 2550 สังกัดสำนักงานเขตพื้นที่การศึกษาปัตตานี เขต2 เมื่อผู้สอบมีความแตกต่างในด้านของภาษาที่ใช้ในชีวิตประจำวันและวิชาคณิตศาสตร์เป็นวิชาที่ผู้วิจัยเป็นครูปฏิบัติการสอน

Prince of Songkla University Pattani Campus