ผลของการจัดเรียงข้อสอบที่มีต่อค่าความคลาดเคลื่อนมาตรฐาน ในการวัดที่ระดับคะแนนเฉพาะ

ปริญญานิพนธ์

ของ

อำนวยพร คำพันธ์

ว. 2542

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา
มีนาคม 2542

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

คณะกรรมการควบคุมและคณะกรรมการสอบ ได้พิจารณาปริญญานิพนธ์ฉบับนี้แล้วเห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา ของมหาวิทยาลัยศรีนครินทรวิโรฒได้

คณะกรรมการควบคุม

.....ประธาน
(รองศาสตราจารย์ ดร.บุญเชิด ภิญโญอนันตพงษ์)

.....กรรมการ
(รองศาสตราจารย์นิภา ศรีไพโรจน์)

คณะกรรมการสอบ

.....ประธาน
(รองศาสตราจารย์ ดร.บุญเชิด ภิญโญอนันตพงษ์)

.....กรรมการ
(รองศาสตราจารย์นิภา ศรีไพโรจน์)

.....กรรมการที่แต่งตั้งเพิ่ม
(ผู้ช่วยศาสตราจารย์เชาวนา ชวลิตธำรง)

บัณฑิตวิทยาลัยอนุมัติให้รับปริญญานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษาของมหาวิทยาลัยศรีนครินทรวิโรฒ

.....คณบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ ดร.เสริมศักดิ์ วิศาลาภรณ์)
วันที่ 12 เดือน พฤศจิกายน พ.ศ. 2542

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้ได้รับความอนุเคราะห์ช่วยเหลืออย่างดียิ่งจาก รองศาสตราจารย์ ดร.บุญเชิด ภิญโญอนันตพงษ์ และรองศาสตราจารย์นิภา ศรีไพโรจน์ ที่กรุณาให้คำปรึกษา แนะนำ ชี้แนะแนวทางแก้ไขข้อบกพร่อง ตลอดจนข้อคิดเห็นต่าง ๆ ตั้งแต่ต้นจนสำเร็จ ลุล่วงด้วยดี ผู้วิจัยขอกราบขอบพระคุณ เป็นอย่างสูง มา ณ โอกาสนี้

ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์เชาวนา ชวลิตธำรง กรรมการสอบปริญญานิพนธ์ ที่ได้กรุณาให้ข้อเสนอแนะเพิ่มเติม ทำให้ปริญญานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น

ขอกราบขอบพระคุณ ผู้เชี่ยวชาญทุกท่าน ที่ช่วยประเมินความสอดคล้องของข้อคำถามกับนิยามของแบบทดสอบวัดความสามารถด้านตัวเลขแบบอนุกรมธรรมดาที่ใช้ในการวิจัย

ขอขอบพระคุณ ผู้อำนวยการ ผู้ช่วยผู้อำนวยการฝ่ายวิชาการ และคณาจารย์ทุกท่านของโรงเรียนที่เป็นกลุ่มตัวอย่าง ที่ให้ความร่วมมือในการเก็บรวบรวมข้อมูลเป็นอย่างดี

ขอขอบพระคุณ ท่านศึกษานิเทศก์อำเภอที่เป็นกลุ่มตัวอย่าง อาจารย์ทัตยา ยาชมภู ศึกษานิเทศก์จังหวัดสระบุรี และอาจารย์พนิดา ฟุ้งเกียรตินำสุข ตลอดจนพี่ ๆ เพื่อน ๆ และน้อง ๆ ชาววัดผลทุกท่าน ที่ให้ความช่วยเหลือด้วยดีมาโดยตลอด

ขอขอบคุณ คุณศิริพร ชื่นแช่ม ที่ให้ความช่วยเหลือในการจัดพิมพ์ จนสำเร็จเป็นรูปเล่มที่สวยงาม

สุดท้ายนี้ขอน้อมรำลึกถึงพระคุณของมารดา และญาติพี่น้องทุกท่านที่คอยให้กำลังใจช่วยเหลือตลอดเวลา

อำนวยพร คำพันธ์

สารบัญ

บทที่		หน้า
1	บทนำ	1
	ภูมิหลัง	1
	ความมุ่งหมายของการศึกษาค้นคว้า	3
	ความสำคัญของการศึกษาค้นคว้า	3
	ขอบเขตของการศึกษาค้นคว้า	4
	นิยามศัพท์เฉพาะ	5
2	เอกสาร และงานวิจัยที่เกี่ยวข้อง	7
	การจัดเรียงข้อสอบ	7
	ความคลาดเคลื่อนมาตรฐานในการวัด	10
	การประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ	12
	งานวิจัยที่เกี่ยวข้อง	28
3	วิธีดำเนินการศึกษาค้นคว้า	36
	ประชากร	36
	กลุ่มตัวอย่าง	36
	เครื่องมือที่ใช้ในการศึกษาค้นคว้า	37
	วิธีดำเนินการสร้างเครื่องมือที่ใช้ในการวิจัย	38
	วิธีดำเนินการเก็บรวบรวมข้อมูล	41
	สถิติที่ใช้ในการวิเคราะห์ข้อมูล	41
4	ผลการวิเคราะห์ข้อมูล	44
	สัญลักษณ์ และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล	44
	การเสนอผลการวิเคราะห์ข้อมูล	44
	ผลการวิเคราะห์ข้อมูล	45

บทที่ **หน้า**

5 สรุป อภิปรายผล และข้อเสนอแนะ..... 54

ความมุ่งหมายของการศึกษาค้นคว้า..... 54

กลุ่มตัวอย่าง..... 54

เครื่องมือที่ใช้ในการศึกษาค้นคว้า..... 54

วิธีดำเนินการเก็บรวบรวมข้อมูล..... 55

การเสนอผลการวิเคราะห์ข้อมูล..... 55

สรุปผลการวิจัย..... 55

อภิปรายผล..... 58

ข้อเสนอแนะ..... 59

บรรณานุกรม..... 61

ภาคผนวก..... 65

ประวัติย่อของผู้วิจัย..... 86

บัญชีตาราง

ตาราง		หน้า
1	จำนวนนักเรียนที่เป็นกลุ่มตัวอย่างจำแนกตามอำเภอและโรงเรียน	37
2	ค่าสถิติพื้นฐานของแบบทดสอบที่จัดเรียงจากง่ายไปหายากและเรียงแบบสุ่ม	45
3	ค่าความเชื่อมั่นของแบบทดสอบที่จัดเรียงจากง่ายไปหายากและเรียงแบบสุ่ม ที่คำนวณจากสูตร 2 สูตร	46
4	ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงจากง่ายไปหายาก	47
5	ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงแบบสุ่ม	48
6	ค่าดัชนีความสอดคล้อง (IOC) ของข้อคำถามกับนิยามศัพท์เฉพาะของแบบทดสอบ	67
7	ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบทดสอบเป็นรายข้อ	68
8	ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบทดสอบที่จัดเรียงข้อคำถามจากง่ายไปหายาก	69
9	ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบทดสอบที่จัดเรียงข้อคำถามแบบสุ่ม	70
10	ค่าไอเกน (มากกว่า 1) ของแบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายาก และเรียงแบบสุ่ม	72

บัญชีภาพประกอบ

ภาพประกอบ		หน้า
1	แสดงค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ โดยการคำนวณค่าจาก 4 วิธี.....	16
2	แผนผังแสดงขั้นตอนการดำเนินการสร้างแบบทดสอบ.....	38
3	การแสดงเส้นภาพ ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ ของแบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายาก และจัดเรียงข้อสอบแบบสุ่ม ที่คำนวณด้วยวิธีของธอร์นไดค์.....	49
4	การแสดงเส้นภาพ ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ ของแบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายาก และจัดเรียงข้อสอบแบบสุ่ม ที่คำนวณด้วยวิธีของลอร์ด.....	50
5	การแสดงเส้นภาพ ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ ของแบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายาก และจัดเรียงข้อสอบแบบสุ่ม ที่คำนวณด้วยวิธีของคีตส์.....	51
6	การแสดงเส้นภาพ ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงจากง่ายไปหายาก ที่คำนวณด้วยวิธีของธอร์นไดค์ ลอร์ด และคีตส์...	52
7	การแสดงเส้นภาพค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ ของแบบทดสอบที่จัดเรียงแบบสุ่มที่คำนวณด้วยวิธีของ ธอร์นไดค์ ลอร์ด และคีตส์.....	53

บทที่ 1
บทนำ

ภูมิหลัง

การวัดผลการศึกษา ไม่ว่าจะเป็นการวัดทางพฤติกรรมศาสตร์ หรือลักษณะทางจิตวิทยาของมนุษย์ สิ่งที่ต้องการวัดส่วนมากเป็นคุณลักษณะภายใน เช่น ความสามารถ ความถนัด ความสนใจและเจตคติ เป็นต้น สิ่งเหล่านี้ไม่สามารถวัดได้โดยตรงเหมือนกับการวัดทางกายภาพ ต้องอาศัยการวัดทางอ้อมโดยใช้เครื่องมือ ซึ่งได้แก่ ข้อสอบหรือแบบสอบถามต่าง ๆ เป็นสิ่งเร้าเพื่อกระตุ้นให้บุคคลแสดงพฤติกรรมออกมาด้วยการตอบคำถามเหล่านั้น แล้วจึงอนุมานผลการตอบจากคะแนนที่ได้ไปอธิบาย หรือทำนายคุณลักษณะภายในของบุคคล การวัดคุณลักษณะภายในนี้มักจะมีความคลาดเคลื่อนในการวัดรวมอยู่ด้วยซึ่งเรียกว่า ความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measurement) โดยสมมติค่ามาจากค่าความเบี่ยงเบนมาตรฐานของคะแนนผลการสอบจากการนำแบบทดสอบนั้น ๆ ไปทดสอบนักเรียนซ้ำหลาย ๆ ครั้ง (อนันต์ ศรีโสภา. 2525 : 62 - 64) และนิยมใช้ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่คำนวณมาจากผู้สอบทั้งหมดเพียงค่าเดียว มารายงานความคลาดเคลื่อนในการวัดของคะแนนแต่ละค่าคะแนน ซึ่งประมาณค่าความคลาดเคลื่อนได้จากค่าเบี่ยงเบนมาตรฐานของคะแนนสอบกับค่าความเชื่อมั่นของคะแนน จากสูตร $S_E = S_x(1 - r_{xx'})^{1/2}$ โดยอาศัยทฤษฎีแบบมาตรฐานเดิม (Classical Test Theory) การนำสูตรนี้มาใช้รายงานความไม่แม่นยำของคะแนนสอบมีข้อจำกัดบางประการ ประการแรกค่าที่คำนวณได้อาจจะขาดความแม่นยำเมื่อข้อมูลมีลักษณะการกระจายไม่เป็นโค้งปกติ (Non - Normal Distribution) ประการที่สอง ค่าที่คำนวณได้ไม่สามารถตีความหมายที่ถูกต้องแม่นยำในทุกระดับคะแนน จะใช้ได้ดีเฉพาะกับคะแนนเฉลี่ยทั้งกลุ่มประชากร แต่ใช้ได้ไม่ดีกับคะแนนของผู้สอบกลุ่มต่ำ กลุ่มกลาง หรือกลุ่มสูง จากผลดังกล่าวจึงทำให้ค่าความคลาดเคลื่อนมาตรฐานในการวัดจากค่าเฉลี่ยทั้งกลุ่มไม่เหมาะสมที่จะนำมาใช้ในการรายงานความแม่นยำของคะแนนแต่ละระดับคะแนน ด้วยเหตุผลสำคัญนี้คณะกรรมการกำหนดมาตรฐานในการทดสอบทางการศึกษาและจิตวิทยา (Standards for Education and Psychological Testing) ฉบับปี 1985 จึงสนับสนุนให้ประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดแบบเงื่อนไขเพื่อรายงาน

ความแม่นยำของคะแนนผู้สอบในแต่ละระดับคะแนนหรือช่วงคะแนน โดยมีการศึกษาวิจัยวิธีการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ ตามลำดับดังนี้

ในปี 1984 ลอร์ด (Lord. 1984) ได้ศึกษาค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนของความสามารถเฉพาะจากสูตรสี่สูตร โดยวิเคราะห์จากแบบทดสอบเพียงฉบับเดียว คือ SCAT (School and College Ability Test) จึงได้ข้อสังเกตว่าค่าความคลาดเคลื่อนมาตรฐานในการวัด สำหรับแต่ละคน หรือแต่ละระดับคะแนนขึ้นอยู่กับค่าสถิติ และหลักเหตุผลของคะแนนจริงของผู้สอบ ไม่ใช่คะแนนสอบ ดังนั้น ในการพูดถึงค่าความคลาดเคลื่อนมาตรฐานในการวัด"ที่ระดับคะแนนที่ได้" จึงหมายถึงคะแนนจริง ไม่ใช่คะแนนสอบ ต่อจากนั้นในปี 1985 เฟลดต์ สเตฟเฟน และกุปตา (Feldt, Steffen and Gupta. 1985) ได้ศึกษาเปรียบเทียบการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดด้วยวิธีห้าวิธี พบว่าค่าความคลาดเคลื่อนมาตรฐานในการวัดสูงสุดที่ได้ จะอยู่ช่วงกลางของระดับคะแนน และจะลดลงอย่างมากที่ช่วงปลายของคะแนน คือช่วงของค่าความคลาดเคลื่อนมาตรฐานในการวัดของคะแนนดิบหรือคะแนนสอบจากแบบทดสอบมาตรฐานด้านผลสัมฤทธิ์ทางการเรียน แต่ถ้าเป็นแบบทดสอบชนิดอื่น อาจจะมีแนวโน้มที่แตกต่างไปจากนี้ ต่อมาในปี 1986 บลิกต์ และชามา (Blixt and Chama. 1986) ได้ศึกษาเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับความสามารถที่แตกต่างกัน โดยศึกษาว่าวิธีประมาณค่าความคลาดเคลื่อนมาตรฐาน ทั้งสี่วิธีที่ลอร์ด (Lord. 1984) ใช้ศึกษานั้น เมื่อนำมาใช้ในการประมาณค่า กับคะแนนของแบบทดสอบชนิดอื่น ๆ จะให้ผลเช่นเดียวกันหรือไม่ และเพื่อศึกษาให้รู้แน่ชัดว่า ค่าความคลาดเคลื่อนมาตรฐานในการวัดจากสูตรเดิม หรือค่าความคลาดเคลื่อนมาตรฐานในการวัดทั้งฉบับมีเพียงหนึ่งค่า ซึ่งเป็นค่าคงที่มีความถูกต้องเหมาะสมกับทุก ๆ คะแนนเช่นเดียวกับค่าความคลาดเคลื่อนมาตรฐานในการวัดที่คำนวณจากแต่ละระดับคะแนนหรือไม่ ซึ่งพบว่าค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ได้จากสูตรเดิมหรือวิธีที่นิยมนั้นเป็นค่าที่เกินความจริง (Overestimate) ที่ช่วงปลายของคะแนนระดับต่ำ และสูง และมีค่าต่ำกว่าเล็กน้อยในช่วงกลางของระดับคะแนน เมื่อเปรียบเทียบกับค่าความคลาดเคลื่อนมาตรฐานในการวัด ด้วยวิธีของลอร์ด ทั้งสี่วิธี แต่เมื่อพิจารณาจากค่าประมาณที่ปัดเศษเป็นเลขจำนวนเต็มแล้วมีค่า

ไม่แตกต่างกัน จึงสรุปว่าสามารถใช้ค่าความคลาดเคลื่อนมาตรฐานในการวัดเฉลี่ยรวมแทนค่าความคลาดเคลื่อนมาตรฐานในการวัดของแต่ละระดับคะแนนได้

จากแนวคิดและผลการวิจัยข้างต้น ทำให้ผู้วิจัยมีความสนใจที่จะศึกษาให้แน่ชัดต่อไปว่า วิธีการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะซึ่งเป็นวิธีการคำนวณที่ง่าย ไม่ซับซ้อน แต่ให้ค่าที่น่าเชื่อถือ สามวิธี คือวิธีของธอร์นไดค์ (Thorndike) วิธีของลอร์ด (Lord) และวิธีของคีตส์ (Keats) จะให้ผลเป็นอย่างไร เมื่อมีการจัดเรียงข้อสอบแตกต่างกันสองแบบ คือจัดเรียงข้อสอบตามความยากรายข้อจากง่ายไปหายากและจัดเรียงข้อสอบแบบสุ่มโดยทำการศึกษาจากแบบทดสอบวัดความสามารถด้านตัวเลขแบบอนุกรมธรรมดา

ความมุ่งหมายของการศึกษาค้นคว้า

1. เพื่อศึกษาค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายากกับการจัดเรียงข้อสอบแบบสุ่ม
2. เพื่อเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบที่มีการจัดเรียงข้อสอบจากง่ายไปหายากกับการจัดเรียงข้อสอบแบบสุ่ม โดยใช้วิธีการประมาณค่า 3 วิธี คือ

วิธีของ ธอร์นไดค์

วิธีของ ลอร์ด

วิธีของ คีตส์

ความสำคัญของการศึกษาค้นคว้า

ผลจากการศึกษาในครั้งนี้ จะเป็นประโยชน์ต่อนักวัดผล และครูทั่วไปในเชิงทฤษฎี และการนำไปใช้ในเรื่องของการจัดเรียงข้อสอบ และการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดด้วยวิธีที่เหมาะสม

ขอบเขตของการศึกษาค้นคว้า

1. กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาครั้งนี้ เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2540 ของโรงเรียนในสังกัดสำนักงานการประถมศึกษาจังหวัดสระบุรี จำนวน 1,019 คน ซึ่งเลือกมาโดยวิธีการสุ่มแบบหลายขั้นตอน

2. ตัวแปรที่ศึกษาค้นคว้า

2.1 ตัวแปรอิสระ ได้แก่

2.1.1 การจัดเรียงข้อสอบ 2 แบบ คือ

จัดเรียงข้อสอบจากง่ายไปหายาก

จัดเรียงข้อสอบแบบสุ่ม

2.1.2 การประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัด ซึ่งแบ่งออกเป็น 3 วิธี คือ

วิธีของ ธอร์นไดค์

วิธีของ ลอร์ด

วิธีของ คีตส์

2.2 ตัวแปรตาม คือ ค่าความคลาดเคลื่อนมาตรฐานในการวัด

3. เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาครั้งนี้ เป็นแบบทดสอบวัดความสามารถด้านตัวเลขแบบอนุกรมธรรมดา ซึ่งผู้วิจัยสร้างขึ้นเองเป็นแบบปรนัยชนิดเลือกตอบ 4 ตัวเลือก จำนวน 40 ข้อ ซึ่งจัดเรียงข้อสอบ 2 แบบ คือจัดเรียงข้อสอบจากง่ายไปหายาก และจัดเรียงข้อสอบแบบสุ่ม

นิยามศัพท์เฉพาะ

1. การจัดเรียงข้อสอบ หมายถึง การเรียงลำดับข้อสอบตามเกณฑ์ความยาก และการสุ่ม คือ

1.1 การเรียงข้อสอบจากง่ายไปหายาก หมายถึง การจัดเรียงข้อสอบตามลำดับ โดยใช้ค่าความยากง่ายเป็นเกณฑ์ ในการศึกษาครั้งนี้ได้จัดเรียงตามลำดับจากง่ายไปหายาก

1.2. การจัดเรียงข้อสอบแบบสุ่ม หมายถึง การจัดเรียงข้อสอบ ตามลำดับโดยใช้การสุ่มเลือกเป็นเกณฑ์ ในการศึกษาครั้งนี้ใช้วิธีจับฉลากจากจำนวนข้อสอบทั้งหมด จะได้แบบทดสอบที่มีลักษณะคละกันระหว่างข้อที่ง่ายกับข้อที่ยาก

2. แบบทดสอบวัดความสามารถด้านตัวเลขแบบอนุกรมธรรมดา คือ แบบทดสอบวัดมโนภาพทางคณิตศาสตร์ ในแง่ความสัมพันธ์ของปริมาณตัวเลขโดยอาศัยวิธีการทางคณิตศาสตร์ ได้แก่ วิธีบวก ลบ คูณ และหาร

3. ค่าความยากง่ายของข้อสอบ (Item Difficulty) หมายถึง ค่าที่คำนวณจากอัตราส่วนของจำนวนผู้ที่ตอบคำถามข้อนั้นถูก กับจำนวนผู้ที่เข้าสอบทั้งหมด

4. ค่าความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Easurement) คือ ค่าคะแนนที่เบี่ยงเบนไปจากคะแนนจริง ในการศึกษาครั้งนี้ใช้ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ ซึ่งใช้วิธีประมาณค่าสามวิธี คือวิธีของ ธอร์นไดค์ วิธีของลอร์ด และวิธีของคีตส์

4.1 วิธีประมาณค่าของธอร์นไดค์ หมายถึง การคำนวณค่าความคลาดเคลื่อนมาตรฐานในการวัด จากความแตกต่างระหว่างคะแนนของแบบทดสอบครึ่งฉบับ คือ $X_1 - X_2$ โดยใช้สูตร

$$S_E = \left(S_{E_1}^2 + S_{E_2}^2\right)$$

4.2 วิธีประมาณค่าของลอร์ด หมายถึง การคำนวณค่าความคลาดเคลื่อนมาตรฐานในการวัดจากสูตรความน่าจะเป็นทวินามคูณด้วย (1-K) โดยหาค่า K จากสูตร

$$K = \frac{n(n-1)s_t^2}{\bar{X}\left(n-\bar{X}\right) - s_t^2 - ns_p^2}$$

แล้วนำ K ไปแทนค่าในสูตร

$$S_E = \left[\frac{X(k-X)}{k-1}(1-K)\right]^{1/2}$$

4.3 วิธีประมาณค่าของคีตส์ หมายถึง การคำนวณค่าความคลาดเคลื่อน-มาตรฐานในการวัดจากสูตร

$$S_E = \left[\left(\frac{X(k-X)}{k-1}\right)\left(\frac{1-r_{xx'}}{1-r_{21}}\right)\right]^{1/2}$$

5. ระดับคะแนนเฉพาะ (Specific Levels) หมายถึง การแบ่งระดับคะแนนออกเป็นช่วง ๆ เพื่อความเหมาะสมในการคำนวณค่าความคลาดเคลื่อนมาตรฐานในการวัด ระดับคะแนนเฉพาะที่ใช้ในการวิจัยครั้งนี้ คือ 5-7, 8-10, 11-13, 14-16, 17-19, 20-22, 23-25, 26-28, 29-31, 32-34, 35-37 และ 38-40

6. ผู้เชี่ยวชาญ หมายถึง ผู้ที่มีวุฒิทางการศึกษาอย่างต่ำปริญญาโททางการวัดผลการศึกษา จำนวน 5 คน

บทที่ 2
เอกสาร และงานวิจัยที่เกี่ยวข้อง

เอกสาร และงานวิจัยที่เกี่ยวข้องกับการศึกษาครั้งนี้ ผู้วิจัยขอเสนอเนื้อหาตามลำดับ ดังนี้

1. การจัดเรียงข้อสอบ
2. ความคลาดเคลื่อนมาตรฐานในการวัด
3. การประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ
4. งานวิจัยที่เกี่ยวข้อง

การจัดเรียงข้อสอบ

ทราเวอร์ส (เตือนใจ เสรษฐสักโก. 2511 : 10; อ้างอิงมาจาก Travers. 1950 : 127–128)ได้เสนอแนะวิธีเรียงข้อสอบไว้ 4 วิธี คือ

1. จัดเรียงตามลำดับความยากง่าย (Arrengement in Order of Difficulty) วิธีนี้เป็นวิธีการจัดเรียงข้อแบบธรรมดาทั่วไป ประโยชน์ส่วนใหญ่ของการเรียงข้อสอบวิธีนี้ก็คือนักเรียนจะได้พบข้อสอบข้อง่าย ๆ ก่อน และไม่เกิดความเหนื่อยมากนักเมื่อพบข้อยาก ๆ แต่การเรียงข้อสอบด้วยวิธีนี้มีข้อเสียอยู่ 2 ประการคือ ประการแรกการเรียงข้อสอบจากง่ายไปยากนั้น ข้อคำถามที่มีเนื้อหาเดียวกันก็จะต้องกระจายกันออกไปตามลำดับความยากง่าย ซึ่งจะมีผลทำให้ความคิดของนักเรียนเปลี่ยนแปลงอยู่เสมอ ข้อเสียประการที่สองก็คือ การเรียงข้อสอบแบบนี้นั้น ถ้านักเรียนพบข้อยากเสียแล้วในตอนต้น ๆ ก็จะทำให้เกิดความท้อใจที่จะทำข้อสอบต่อไป เนื่องจากนักเรียนรู้ว่าข้อสอบจะยิ่งยากขึ้นไปอีก

2. จัดเรียงความยากง่ายแบบหมุนเวียน (Arrangement in Cyclic Order of Difficulty) เพื่อจะได้หลีกเลี่ยงข้อเสียประการที่สองของการจัดเรียงข้อสอบตามลำดับความยากง่าย ทราเวอร์สจึงได้เสนอการเรียงข้อสอบแบบนี้ขึ้นเพื่อเป็นการกระตุ้นให้เด็กอ่านข้อสอบให้หมดทุกข้อ เพราะนักเรียนรู้ว่าถ้าทำข้อสอบไประยะหนึ่งก็จะพบข้อง่าย ๆ อีก วิธีนี้จึงมีข้อดี แต่ก็จะมีจุดอ่อนที่นักเรียนจะต้องเปลี่ยนความคิดในการแก้ปัญหาเร็วเกินไป

3. การจัดเรียงตามกลุ่มเนื้อหาวิชา (Arrangement According to Subject Matter Area) การเรียงข้อสอบโดยวิธีนี้เป็นการเรียงโดยรวมข้อสอบที่มีเนื้อหาเดียวกันเข้าไว้ด้วยกัน แล้วจัดเรียงข้อสอบจากง่ายไปยาก ซึ่งจะสามารถช่วยให้นักเรียนได้ใช้ความคิดอย่างแน่วแน่ในการแก้ปัญหาในหนึ่งเนื้อหาก่อนที่จะเปลี่ยนไปคิดในเนื้อหาอื่น ๆ

4. การจัดเรียงข้อสอบตามจุดมุ่งหมายของการวัด (Arrangement According to the Goals Measured) ในการจัดเรียงข้อสอบ ผู้จัดทำข้อสอบบางคนนิยมเรียงข้อ โดยรวมพฤติกรรมที่ต้องการที่วัดประเภทเดียวกัน เช่น ในข้อทดสอบ โค-ออพเพอเรทิฟ (Co-operative Tests) การเรียงข้อสอบที่วัดเกี่ยวกับคำศัพท์ และความคิดรวบยอด (Term and Concepts) เข้าไว้เป็นกลุ่มเดียวกัน ส่วนการวัดด้านความเข้าใจนั้นก็แยกไว้เป็นอีกพวกหนึ่ง การเรียงข้อสอบวิธีนี้มีประโยชน์สำหรับครูในการตรวจสอบการวัดว่าตรงจุดมุ่งหมายที่ต้องการหรือไม่ แต่มีข้อเสียคือ ไม่สามารถจะรวมเนื้อหาเดียวกัน หรือที่คล้ายกันเข้าไว้ด้วยกันได้

ศศิธร (ศศิธร สุวรรณสุข. 2519 : 8 ; อ้างอิงมาจาก Lindquist. 1955 : 179) ก็มีความเห็นในด้านการเรียงข้อสอบคล้ายคลึงกับข้อเสนอแนะของทราเวอร์ส โดยได้เสนอแนะวิธีจัดกลุ่มข้อคำถามเข้าไว้ด้วยกันว่า ถ้าข้อคำถามเหล่านั้นอยู่ในเนื้อหาเดียวกันและมีความยากง่ายพอ ๆ กันแล้ว ก็อาจสลับข้อกันได้ ไม่จำเป็นต้องเรียงลำดับกันตามความยากง่าย แต่ถ้าข้อคำถามเหล่านั้นอยู่ในเนื้อหาที่ต่างกัน และให้คะแนนแต่ละข้อเท่ากันก็จะต้องแยกกลุ่มข้อคำถามออกเป็นฉบับ ๆ และแยกเวลาในการทดสอบในแต่ละฉบับด้วย การเรียงลำดับข้อในแต่ละฉบับนั้น อาจจะขึ้นอยู่กับความสัมพันธ์ของเนื้อหาวิชา คือ อาจให้สอบเนื้อหาทั่ว ๆ ไปก่อน แล้วจึงค่อยสอบเนื้อหาที่จำเพาะเจาะจง

แมคนิคอล (เตือนใจ เศรษฐสักโก. 2511 : 13 ; อ้างอิงมาจาก Mac Nical. 1956 : 1 - 21) ได้เปรียบเทียบลักษณะการเรียงข้อสอบของแบบทดสอบ ซึ่งแต่เดิมเรียงจากง่ายไปหายาก ไปเป็นการเรียงแบบสุ่ม (Random Arrangement) กับเรียงจากข้อยากไปหาข้อง่าย ผลจากการศึกษาครั้งนี้ชี้ให้เห็นว่า การเรียงข้อสอบจากยากไปหาง่ายจะทำให้ข้อสอบยากขึ้นกว่าเดิมอย่างมีนัยสำคัญทางสถิติ

แฟลงเกอร์, เมลตอน และไมเออร์ส (เตือนใจ เศรษฐสักโก. 2511 : 14 ; อ้างอิงมาจาก Flangher, Melton and Myers. 1966 : 1-21) ได้นำแบบทดสอบความถนัดของ CEEB ด้านภาษา และคณิตศาสตร์ ซึ่งเดิมจัดแบ่งข้อสอบไว้เป็นตอน ๆ ตอนหนึ่งมี 5 ข้อ มาจัดเรียงเสียใหม่ในแบบต่าง ๆ ดังนี้ คือ

1. จัดเรียงตามแบบมาตรฐานเดิม
2. จัดเรียงข้อภายในแต่ละตอนเสียใหม่ ส่วนลำดับตอนยังคงเดิม
3. จัดเรียงข้อภายในแต่ละตอนเหมือนเดิม ส่วนลำดับตอนจัดใหม่
4. จัดเรียงทั้งข้อภายในแต่ละตอน และลำดับตอนเสียใหม่

หลังจากนั้นนำข้อสอบที่จัดเรียงทั้ง 4 แบบไปทดลองสอบกับนักเรียนระดับวิทยาลัย ผลปรากฏว่าการจัดเรียงข้อทดสอบดังกล่าว มีผลทำให้คะแนนจากการทดสอบด้านภาษาแตกต่างกัน สำหรับข้อทดสอบคณิตศาสตร์ จะไม่มีผลแตกต่างกันเลยไม่ว่าจะเรียงข้อทดสอบด้วยวิธีใด

ชวาล แพรัตกุล (2518 : 179-180) ได้เสนอแนะว่าการจัดระเบียบข้อสอบซึ่งได้แก่ การพิมพ์ การจัดลำดับข้อ การเขียนคำชี้แจง การตรวจให้คะแนน มีความสำคัญมากไม่ยิ่งหย่อนไปกว่าตัวข้อสอบเองเหมือนกัน บางเรื่องมีผลต่อการสอบโดยตรง อาจทำให้คะแนนนักเรียนแต่ละคนสูง หรือต่ำไป ซึ่งจะมีผลทำให้ความเที่ยงตรง และความเชื่อมั่นของข้อสอบเปลี่ยนแปลงไปด้วย และยังเสนอแนะอีกว่า การเรียงข้อสอบในวิชาหนึ่ง ๆ นั้น ควรเรียงข้อสอบโดยเริ่มจากข้อง่ายไปหายากเสมอ

อนันต์ ศรีโสภา (2525 : 40) ได้กล่าวถึง การจัดหมวดหมู่ของข้อสอบว่าจะจัดลำดับก่อนหลังตามเนื้อหาวิชา หรือตามลำดับก่อนหลัง หรือตามลำดับจากง่ายไปหายาก หรือตามลำดับจากคำถามทั่วไปไปหาคำถามที่เฉพาะ โดยควรระวังการจัดลำดับข้อสอบเหล่านี้ต้องไม่เป็นทางที่บอกคำตอบแก่ข้อหนึ่งข้อใดโดยเฉพาะ

<u>ความคลาดเคลื่อนมาตรฐานในการวัด</u> (Standard Error of Measurement)

ความคลาดเคลื่อนมาตรฐานในการวัด เป็นการวัดความเปลี่ยนแปลงภายในของนักเรียนแต่ละคน โดยทฤษฎีแล้วความคลาดเคลื่อนมาตรฐานในการวัด หาได้จากค่าเบี่ยงเบนมาตรฐานของคะแนนผลการสอบ จากการนำแบบทดสอบนั้นไปทดสอบนักเรียนกลุ่มเดียวกันหลาย ๆ ครั้ง เพื่อหาค่าเฉลี่ยของคะแนนจริง ตามทฤษฎีกล่าวว่าคะแนนจริงของนักเรียนแต่ละคนจะไม่เปลี่ยนแปลงถ้าเรานำข้อสอบนั้นไปทดสอบนักเรียนคนเดียวกันหลาย ๆ ครั้ง แต่อาจมีความคลาดเคลื่อนบางอย่างเกิดขึ้นได้ ซึ่งเป็นสาเหตุทำให้คะแนนผลการสอบของนักเรียนเปลี่ยนแปลงไป บางครั้งการเปลี่ยนแปลงนี้มีค่ามากกว่าคะแนนจริง บางครั้งมีค่าน้อยกว่าคะแนนจริง เราสมมุติให้ความคลาดเคลื่อน ของคะแนนมีลักษณะแบบสุ่ม (Random) ดังนั้น ความคลาดเคลื่อนทางบวก และความคลาด-เคลื่อนทางลบหักล้างกันหมดไป ค่าเฉลี่ยของคะแนนคลาดเคลื่อนจึงมีค่าเท่ากับศูนย์ และ ค่าเฉลี่ยของคะแนนสอบที่ได้จากการสอบซ้ำหลาย ๆ ครั้ง ก็จะมีค่าเท่ากับคะแนนจริง

สมมติให้คะแนนผลการสอบเหล่านี้มีการแจกแจงแบบปกติ (Normal Distribution) จะได้ว่า

$$\sigma_0^2 = \sigma_t^2 + \sigma_e^2$$

เมื่อ σ_o^2 แทน ความแปรปรวนของคะแนนสอบ

σ_t^2 แทน ความแปรปรวนของคะแนนจริง

σ_e^2 แทน ความแปรปรวนของคะแนนคลาดเคลื่อน

ถ้าคิดว่าคะแนนผลการสอบเหล่านี้ของนักเรียนแต่ละคนได้มาจากการทดสอบหลาย ๆ ครั้ง คะแนนจริงจะไม่เปลี่ยนแปลง ดังนั้น $\sigma_{ti}^2 = o$ ถ้า σ_{oi} แทนความเบี่ยงเบน-มาตรฐานของคะแนนผลการสอบที่ได้จากการทดสอบซ้ำหลาย ๆ ครั้ง จะได้

$$\sigma_{oi}^2 = o + \sigma_e^2$$
$$\sigma_{oi} = \sigma_e$$

ถ้าทราบความคลาดเคลื่อนมาตรฐานในการวัด จะช่วยให้ทราบว่าคะแนนจริงของแต่ละคะแนนผลการสอบมีค่าเท่าใด อนึ่ง การที่เราไม่สามารถสอบนักเรียนซ้ำหลายครั้งได้ เนื่องจากเป็นวิธีการที่ไม่ประหยัด และคะแนนที่ได้มักจะเปลี่ยนแปลงไป ซึ่งเกิดจากเหนื่อย หรือผลจากการเรียนรู้ในการทำข้อสอบครั้งแรกเป็นต้น การที่เราคำนวณค่าความเบี่ยงเบนมาตรฐานในการวัดจากกลุ่มคะแนนทั้งหมด แล้วใช้ค่านี้ในการตีความหมายของแต่ละคะแนนไม่สู้จะดีนักเพราะถ้าหากกลุ่มของคะแนนเป็นวิวิธพันธ์ (Heterogeneous) ก็จะทำให้ค่าที่ตีความหมายนั้น ผิดไปจากความเป็นจริง ดังนั้น ลอร์ด จึงได้สร้างสูตรเพื่อคำนวณค่าความเบี่ยงเบนมาตรฐานในการวัดของคะแนนแต่ละค่า ซึ่งมีสมการดังนี้ (อนันต์ ศรีโสภา, 2525 : 62 – 64)

$$\sigma_{xi} = \sqrt{\frac{X_i(n-X_i)}{n-1}}$$

เมื่อ σ_{xi} แทน ความคลาดเคลื่อนมาตรฐานของคะแนนที่ i

n แทน จำนวนข้อในแบบทดสอบ

X_i แทน คะแนนสอบของนักเรียนคนที่ i

จากสมการจะเห็นว่า ความคลาดเคลื่อนมาตรฐานนั้นขึ้นอยู่กับองค์ประกอบที่สำคัญ 2 ประการ คือ คะแนนของนักเรียน และจำนวนข้อในแบบทดสอบถ้าใช้สูตรนี้ คะแนนจะมีความคลาดเคลื่อนมาตรฐานที่แตกต่างกัน ซึ่งจะช่วยให้ตีความหมายของคะแนนได้ถูกต้องมากกว่า อย่างไรก็ดี บริษัทสร้างข้อสอบนิยมใช้สูตรนี้ แต่ครูที่สอนตามโรงเรียนไม่นิยมใช้ เพราะว่าจะต้องคำนวณค่าความคลาดเคลื่อนมาตรฐานของทุก ๆ คะแนน ซึ่งเสียเวลามาก

ค่าความคลาดเคลื่อนมาตรฐานในการวัด ช่วยให้เราทราบว่าความเปลี่ยนแปลงของคะแนนผลการสอบแต่ละคะแนนว่ามีช่วงของการเปลี่ยนแปลงกว้างหรือแคบเพียงใด เนื่องจากสมมติให้ความคลาดเคลื่อนมีลักษณะแบบสุ่ม และคะแนนผลการสอบของนักเรียนแต่ละคนมีการแจกแจงแบบปกติ อาจจะกล่าวได้ว่า โอกาสที่คะแนนจริงจะอยู่ระหว่าง $\pm 1\sigma_e$ ของคะแนนผลการสอบประมาณ 68 เปอร์เซ็นต์ และ $\pm 2\sigma_e$ ประมาณ 95 เปอร์เซ็นต์ (อนันต์ ศรีโสภา. 2525 : 62- 64)

<u>การประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ</u>

(Estimating of Standard Error of Measurement at Specific Score Levels)

ในช่วง 40 ปีมานี้ มีผู้ให้ความสนใจในการศึกษาเรื่อง วิธีการคำนวณหาค่าความคลาดเคลื่อนมาตรฐานในการวัดของแต่ละค่าคะแนน และยังมีการวางเงื่อนไขต่าง ๆ เข้าไปร่วมในการศึกษาด้วย ในปี 1949 โมเลนคอปฟ์ เป็นบุคคลแรกที่เสนอค่าความแปรปรวนคลาดเคลื่อนในการวัด ($\sigma^2_{E \cdot X}$) ซึ่งจะได้ค่าที่แตกต่างกันไปตามคะแนนสอบ ในปี 1951 ธอร์นไดค์ เสนอวิธีประมาณค่า $\sigma^2_{E \cdot X}$ ที่ทุกระดับคะแนน โดยยึดทฤษฎีมาตรฐานเดิม และแสดงให้เห็นว่าภายใต้เงื่อนไขที่แน่นอนผู้สอบแต่ละคนที่มีคะแนนอยู่ภายในอันตรภาคชั้นความแปรปรวนของความแตกต่างระหว่างแบบทดสอบแบ่งครึ่งก็คือค่าความแปรปรวนคลาดเคลื่อนของแบบทดสอบทั้งฉบับ (Feldt and Qualls. 1996)

ในปี 1984 ลอร์ด (Lord. 1984) ใช้สูตรในการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัด โดยใช้แบบทดสอบเพียงฉบับเดียว ที่คำนวณจากคำตอบที่ถูกต้อง ณ ระดับความสามารถเฉพาะ หรือคะแนนจุดตัด ทั้ง 4 วิธี ได้นำมาเปรียบเทียบ ซึ่งอาศัยข้อมูลจริงชุดเดียวกัน ดังนี้

วิธีที่1 แบ่งแบบทดสอบออกเป็น 2 ส่วน เป็นคะแนน X_1 และ X_2 สำหรับการสุ่มเลือกแต่ละบุคคล, คะแนน X_1 และ X_2 อาจจะพิจารณาที่ขนาดของกลุ่มตัวอย่าง $m = 2$ ซึ่งถูกสุ่มจากคะแนนของประชากร นักเรียนที่ได้จากแบบทดสอบครึ่งฉบับ ถ้าคิดจากคะแนนทั้ง 2 ตัว (X_1 กับ X_2 ทั้ง 2 ฉบับ) ขนาดของกลุ่มตัวอย่างจะเป็น $m/(m\text{-}1) = 2$ ซึ่งก็คือตัวประมาณค่าความแปรปรวนที่ไม่ลำเอียงของคะแนนแบบทดสอบครึ่งฉบับในนักเรียนแต่ละคน ลอร์ด และ โนวิค (Lord and Novick) ได้เสนอสูตรตัวประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ไม่มีความลำเอียงสำหรับแบบทดสอบครึ่งฉบับ คือ $\frac{1}{2}(X_1 - X_2)^2$

ค่าความคลาดเคลื่อนมาตรฐานในการวัด ของคะแนนรวมทั้งหมด X หรือ $X = X_1 + X_2$ ก็คือ $S.E.^2(X_1 + X_2) = S.E.^2(X_1) + S.E.^2(X_2)$ จากข้อตกลง

เบื้องต้นที่ว่าแบบทดสอบแบ่งครึ่งทั้ง 2 ส่วน จะมีค่าความคลาดเคลื่อนมาตรฐานในการวัดเท่ากัน เราได้ค่า $S.E.^2(X)=2S.E.^2(X_1)+2S.E.^2(X_2)$ ดังนั้น $(X_1-X_2)^2$ ก็คือตัวประมาณค่าที่ไม่มีความลำเอียง ที่ใช้สำหรับหาค่าความคลาดเคลื่อนมาตรฐานในการวัดคะแนน X ในแบบทดสอบเต็มฉบับ

โมเลนคอปฟ์ (MollenKopf) และ ลิวิงสตัน (Livingston) ใช้วิธีการจัดกลุ่มผู้สอบ ตามคะแนนที่ได้จากการวัด (X) และเฉลี่ยโดย $(X_1-X_2)^2$ ในผู้สอบทุกคน ณ คะแนนที่ได้ ถ้าแบบทดสอบมีความเชื่อมั่นที่เหมาะสม ผู้สอบทุกคนจะได้ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่เท่ากัน ดังนั้นการประมาณค่าโดยวิธีนี้จึงมีข้อคัดค้าน 2 ประการ คือ

1. ผู้สอบที่ได้คะแนนจากการสอบ (X) อาจจะไม่ได้เลือกมาโดยการสุ่มสมมติว่า ข้อสอบมี 50 ข้อ เมื่อ X คือ 49 มันมีทางเป็นไปได้เพียง 2 ประการ คือ X_1, X_2 จะได้ (24, 25) หรือไม่ก็ (25, 24) ค่าความคลาดเคลื่อนมาตรฐานในการวัดก็คือ 1.0 โดยไม่คำนึงถึง สหสัมพันธ์ระหว่างข้อสอบที่สูงเกินไปหรือไม่มีสหสัมพันธ์เลย

2. ถ้ากรณีทราบค่าความคลาดเคลื่อนมาตรฐานในการวัด ของกลุ่มผู้สอบที่ระดับคะแนนที่สอบได้ เราก็จะไม่ทราบคะแนนจริง (True Score) ของนักเรียน และผลที่ได้ไม่สามารถบ่งบอกค่าความคลาดเคลื่อนมาตรฐานในการวัด ณ ทุกระดับคะแนนตัวอย่าง เช่น เราอาจประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดสำหรับกลุ่มข้อสอบที่ได้ 49 คะแนน แต่ในความเป็นจริง เราจะไม่ทราบเลยว่า ไม่มีผู้สอบคนใดในกลุ่มนี้ที่มีคะแนนจริง (True Score) เหนือ 47 คะแนน

กิลฟอร์ด และเฟรชเชอร์ ได้เสนอแนะให้ใช้คะแนนจากแบบทดสอบคู่ขนาน (Parallel Test Froms) มากกว่าใช้คะแนนจากแบบทดสอบแบ่งครึ่ง และการประมาณค่าวิธีนี้จะมีความเที่ยงตรงในทางปฏิบัติ ก็ต่อเมื่อศึกษากับกลุ่มตัวอย่างขนาดใหญ่ โดยสรุป วิธีที่ 1 ให้ผลการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดสำหรับผู้สอบที่ถูกเลือกมาโดยการสุ่มค่าความคลาดเคลื่อนมาตรฐานในการวัด นี้เหมาะสมสำหรับแบบทดสอบคู่ขนาน

วิธีที่ 2 ภายใต้เงื่อนไข ของ IRT (Item – Response Theory) ความแปรปรวนของคะแนนที่สอบได้ ณ ระดับความสามารถที่กำหนด (θ) คือ

$$\sum^{n} P_i(\theta)Q_i(\theta)$$

เมื่อ $P_i(\theta)$ คือ ความน่าจะเป็นของการตอบคำถาม ข้อ I ถูกต้อง $(i = 1, 2\ldots, n)$ ณ ระดับความสามารถ (θ)

$$Q_i(\theta) = 1 - P_i(\theta)$$

ความแปรปรวนนี้ก็คือ ค่าความคลาดเคลื่อนมาตรฐานในการวัด ณ ระดับความสามารถ (θ) คะแนนจริง ξ คำนวณจาก

$$\xi = \sum_{i=1}^{n} P_i(\theta)$$

ดังนั้น $\sum^{n} P_i(\theta)Q_i(\theta)$ ก็คือ ค่าความคลาดเคลื่อนมาตรฐานในการวัด ณ คะแนนจริง $\xi = \Sigma^{n} P_i(\theta)$ ในทางปฏิบัติการกำหนดค่าพารามิเตอร์ $P_i(\theta)$ คือ การประมาณจากกระดาษคำตอบ แล้วนำมาแทนค่าในสูตรเพื่อหาค่าความคลาดเคลื่อนมาตรฐานในการวัด โดยใช้โลจิสติก โมเดล แบบสามพารามิเตอร์ ดังนี้

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + exp[-A_i(\theta - b_i)]}$$

เมื่อ A_i แทน ค่าอำนาจจำแนกของข้อสอบ

b_i แทน ค่าความยากของข้อสอบ

c_i แทน ค่าการเดา

วิธีที่ 2 นี้ แบบทดสอบ จะต้องเป็นมิติเดียว (Unidimension) ข้อตกลงนี้ง่ายในทางทฤษฎี แต่ค่อนข้างมีปัญหายุ่งยากในทางปฏิบัติ ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ได้จะเป็นตัวแทนคะแนน X ของผู้สอบในแบบทดสอบคู่ขนาน

วิธีที่ 3 วิธีนี้ถ้าแบบทดสอบคู่ขนานมีการสุ่มข้อสอบของแต่ละฉบับจำนวน n ข้อ ในกรณีนี้ค่าความคลาดเคลื่อนมาตรฐานในการวัด ก็คือ $\xi(n - \xi)/n$ สำหรับผู้สอบที่ถูกเลือกมาโดยการสุ่ม ตัวประมาณค่าที่ไม่ลำเอียง คือ $X(n - X)/(n - 1)$ วิธีนี้

ไม่คำนึงถึงข้อมูล และข้อตกลงเบื้องต้นต่าง ๆ การประมาณค่าที่ไม่ลำเอียงนี้จะเที่ยงตรงสำหรับกรณีที่ผู้สอบถูกเลือกก่อนที่จะทราบคะแนนของผู้สอบ

วิธีที่ 4 การประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดด้วยวิธีนี้ จะเหมาะสมเมื่อแบบทดสอบถูกสุ่มให้เป็นแบบทดสอบคู่ขนาน ซึ่งเป็นการยากที่จะให้แบบทดสอบคู่ขนานทั้ง 2 ฉบับ มีข้อสอบแต่ละข้อมีค่าความยากเท่ากัน หรือใกล้เคียงกัน สำหรับ วิธีที่ 4 นี้ มีข้อตกลงเบื้องต้นว่า ค่าความคลาดเคลื่อนมาตรฐานในการวัดสำหรับแบบทดสอบคู่ขนาน สามารถประมาณค่าโดยการเพิ่มในวิธีที่ 3 โดยคูณด้วย *(1-K)* เมื่อ K คือค่าคงที่ที่ถูกกำหนดขึ้นมา

การกระจายของคะแนน X ถูกจำกัดโดย θ หรือ ξ ก็คือ คอมพาวด์ ไบโนเมียล (Compound binomial) ค่าคงที่ K ถูกกำหนด โดยให้คะแนน X มีการกระจายแบบคอมพาวด์ ไบโนเมียล และหาค่า K ซึ่งจะมีความสัมพันธ์ระหว่าง Xและ ξ จะเท่ากับรากที่สองของค่าความเชื่อมั่นสูตร $KR-20$ หาค่า K โดยสูตรนี้

$$K=\frac{n(n-1)s_p^2}{\overline{X}(n-\overline{X})-s_x^2-ns_p^2}$$

เมื่อ $\overline{X}$ แทน ค่าเฉลี่ย

S_x^2 แทน ค่าความแปรปรวนของคะแนนผู้สอบทุกคน

S_p^2 แทน ความแปรปรวนของข้อสอบ ของ p_i

p_i แทน สัดส่วนผู้ตอบคำถามถูกต้อง

เมื่อได้ค่า K แล้วนำ K ไปแทนค่าในสูตร

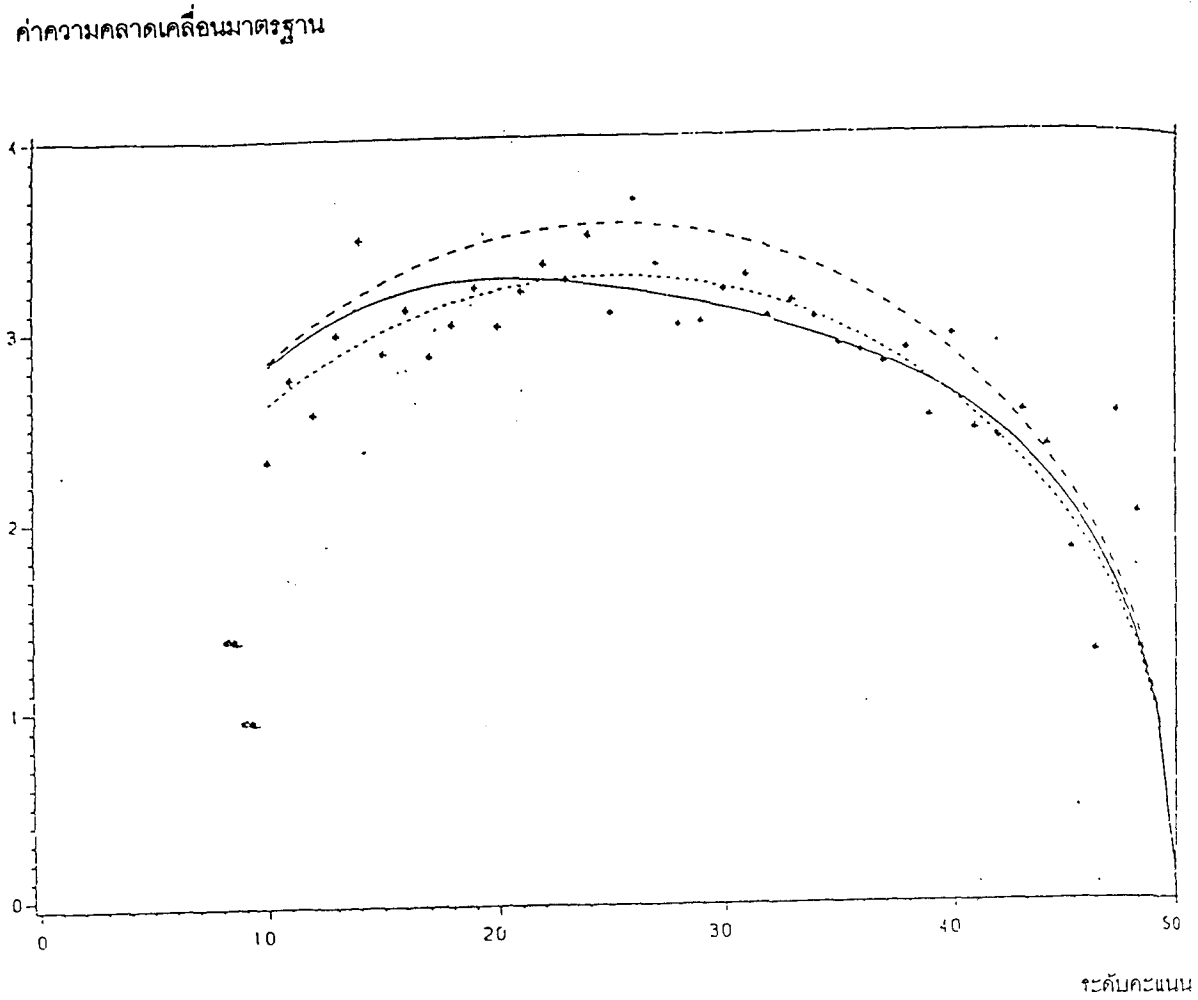
$$\sigma_E=\left[\frac{X(k-X)}{k-1}(1-K)\right]^{1/2}$$

เมื่อ σ_E แทน ค่าความคลาดเคลื่อนมาตรฐาน

X แทน ค่าระดับคะแนนเฉพาะ

k แทน จำนวนข้อคำถาม

วิธีนี้ผู้สอบจะถูกจัดกลุ่มตามคะแนนรวมที่ได้จากค่า X ตั้งแต่ 1 ถึง ... และจะหาค่าความคลาดเคลื่อนมาตรฐานในแต่ละกลุ่มคะแนน และวิธีนี้มีข้อได้เปรียบตรงที่กลุ่มตัวอย่างไม่จำเป็นต้องมีขนาดใหญ่ ผลการวิเคราะห์เปรียบเทียบ 4 วิธี ดังนี้



ภาพประกอบ 1 แสดงค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะโดยการคำนวณค่าจาก 4 วิธี

สรุปผล เนื่องจากว่า ข้อสอบ SCAT เป็นข้อสอบแบบหาคำสัมพันธ์กันผู้วิจัยเชื่อว่า ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ได้จากวิธีที่ 2 มีความถูกต้อง วิธีที่ 1 และ 4 สามารถใช้กับแบบทดสอบที่ไม่เป็นมิติเดียว (Unidimension) แต่ในทางปฏิบัติ วิธีที่ 4 จะเป็นวิธีที่ดีกว่า เพราะความง่าย และขนาดของกลุ่มตัวอย่างที่ไม่จำเป็นต้องใหญ่มาก

ในปี 1985 เฟลดต์ สเตเฟน และกุปตา (Feldt, Steffen and Gupta. 1985) ได้ศึกษาเปรียบเทียบการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ ด้วยวิธี 5 วิธี ดังนี้

วิธีที่ 1 ธอร์นไดค์ เป็นผู้เสนอเทคนิคการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนน แตกต่างกัน ธอร์นไดค์ คิดจากคะแนนรวมของแต่ละบุคคล โดยที่ X คือคะแนนรวมของแบบทดสอบคู่ขนาน ทั้ง 2 ฉบับคือ X_1 และ X_2 โดยยึดรูปแบบทฤษฎีแบบมาตรฐานเดิม คะแนนของแบบทดสอบครึ่งฉบับ และคะแนนรวมของแบบทดสอบ ถูกพิจารณาว่าเป็นผลรวมของ "คะแนนจริง" และค่าความคลาดเคลื่อน

$$\begin{aligned} X_1 &= T_1 + E_1 \\ X_2 &= T_2 + E_2 \\ X &= X_1 + X_2 \\ &= (T_1 + T_2) + (E_1 + E_2) \\ &= T + E \end{aligned}$$

จากทฤษฎีมาตรฐานเดิม (Classical test Theory) สหสัมพันธ์ระหว่าง E_1 และ E_2 มีค่าเท่ากับ 0 หมายความว่า ไม่มีความสัมพันธ์กัน ดังเช่นสูตร

$$\sigma_E = \left(\sigma_{E_1}^2 + \sigma_{E_2}^2\right)^{1/2}$$

ความแตกต่างระหว่างคะแนนของแบบทดสอบครึ่งฉบับ คือ $X_1 - X_2$ เท่ากับ $(T_1 - T_2) + (E_1 - E_2)$ สำหรับแบบทดสอบคู่ขนาน $T_1 = T_2$ หรือ $T_1 - T_2 = 0$ ดังนั้น

$$\begin{aligned} \sigma_{(x_1 - x_2)} &= \sigma_{(E_1 - E_2)} \\ &= \left(\sigma_{E_1}^2 + \sigma_{E_2}^2\right)^{1/2} \\ &= \sigma_E \end{aligned}$$

ค่าความคลาดเคลื่อนมาตรฐานของแบบทดสอบทั้งฉบับเท่ากับผลต่างของค่าเบี่ยงเบนมาตรฐานของแบบทดสอบครึ่งฉบับ สำหรับผู้สอบกลุ่มใหญ่ถ้าผู้สอบถูกจำกัดโดยเฉพาะค่า X ซึ่งก็คือค่าเบี่ยงเบนมาตรฐานของแบบทดสอบครึ่งฉบับ ซึ่งสอบกับกลุ่มย่อยเพื่อหาค่าความคลาดเคลื่อนมาตรฐานในการวัดในกรณีที่ระดับคะแนนแตกต่างกัน

วิธีที่ 2 โมเลนคอปฟ์ ได้เสนอวิธีประมาณค่าในการทำนายโดยใช้สมการการถดถอย (Regression) จากคะแนนทั้งหมดโดยใช้สูตร

$$\hat{Y} = a_0 + a_1(X) + a_2(X^2) + a_3(X^3) + a_4(X^4) + \ldots$$

ภายใต้ทฤษฎีการถดถอย (Regression) $\hat{Y}$ หมายถึง ค่าประมาณของ Y (μ_Y) สำหรับประชากร ของแต่ละบุคคลซึ่งหาได้จากค่า X อย่างไรก็ตาม ค่าเฉลี่ยของ Y ซึ่งกำหนดในกรณีนี้เป็นค่าเฉลี่ยของผลต่างของแบบทดสอบครึ่งฉบับยกกำลังสอง ของแบบทดสอบคู่ขนาน ค่าเฉลี่ยของผลต่างยกกำลังสอง ก็คือค่าความแปรปรวนของความคลาดเคลื่อนในการวัดสำหรับแบบทดสอบทั้งฉบับ $(\hat{Y})^{1/2}$ สำหรับค่าของ X คือ การประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนน X

การประยุกต์ใช้วิธีนี้สามารถทำได้ง่ายด้วยโปรแกรมวิเคราะห์ทางสถิติด้วยคอมพิวเตอร์ สัมประสิทธิ์ที่ถูกกำหนด จะนำเอามาแทนค่าของ X ในสมการเพื่อหาค่าความคลาดเคลื่อนมาตรฐาน

วิธีที่ 3 วิธีนี้มีพื้นฐานมาจากรูปแบบวิธีของลอร์ด (Lord. 1955) โดยลอร์ดสุ่มเลือกข้อสอบจำนวน k ข้อ จาก แบบทดสอบทั้งหมดที่ประกอบขึ้นเป็นแบบทดสอบ 1 ฉบับ และชุดข้อสอบจำนวน k ข้อ ประกอบขึ้นเป็นฉบับย่อย ๆ บุคคลแต่ละคน (i) สามารถตอบคำถามได้ถูกในสัดส่วนที่แน่นอน ϕ_i จากจำนวนข้อทั้งหมด ภายใต้แนวความคิดนี้คะแนนของผู้สอบคนที่ i ในแบบทดสอบ 1 ฉบับ จะได้จากการนับความถี่ Q ที่เกิดขึ้นในการสุ่มกลุ่มตัวอย่าง k ข้อ แนวความคิดเดิมของค่าความคลาดเคลื่อนมาตรฐานในการวัดก็คือค่าเบี่ยงเบนมาตรฐานของคะแนนที่ได้จากผู้สอบในแบบทดสอบฉบับคู่ขนาน ดังนั้นค่าความคลาดเคลื่อนมาตรฐานในการวัด คือ การเปรียบเทียบได้โดยตรงในด้านสถิติของความถี่ของค่าความคลาดเคลื่อนมาตรฐานที่กำหนดจากกลุ่มตัวอย่างขนาด k คือ $[k(\phi)(1-\phi)]^{1/2}$ สำหรับคนที่ i ค่าความคลาดเคลื่อนมาตรฐานในการวัดเท่ากับค่าพารามิเตอร์ของคนที่ i ลอร์ดได้เสนอการปรับแก้ค่าความลำเอียงในการหาค่าความแปรปรวนจากลุ่มตัวอย่างก็คือ การประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัด ที่ระดับคะแนน X สำหรับแบบทดสอบทั้งฉบับ โดยใช้สูตร

$$S_E = \left[\frac{X(k-X)}{k-1}\right]^{1/2}$$

X แทน ระดับคะแนน

k แทน จำนวนข้อของแบบทดสอบ

ต่อมาภายหลังได้มีการอภิปรายถึงสมการนี้ว่ามีความชัดเจนน้อย เนื่องจากเป็นการยากที่จะจับคู่ระหว่างแบบทดสอบให้คู่ขนานกันในด้านเนื้อหา ค่าความยาก และคุณสมบัติอื่น ๆ ได้ ดังนั้น สมการนี้จึงดูเหมือนว่าจะประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนน X เกินความจริง ซึ่งในความจริงผลเฉลี่ยกำลังสองของสมการจะกลายเป็นสูตร KR - 21 ซึ่งก็คือสัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบนั่นเอง

คีตส์ (Keats. 1957) ได้ปรับแก้สมการนี้ให้มีความเหมาะสมมากขึ้น โดยเพิ่มสูตรเข้าไปทางขวาของสมการ ซึ่งประกอบด้วยสัมประสิทธิ์ความเชื่อมั่น ดังสูตรนี้

$$S_E = \left[\left(\frac{X(k-X)}{k-1}\right)\left(\frac{1-r_{xx'}}{1-r_{21}}\right)\right]^{1/2}$$

เมื่อ $r_{xx'}$ แทน ค่าความเชื่อมั่นที่เหมาะสมของแบบทดสอบ

r_{21} แทน ค่าความเชื่อมั่นของแบบทดสอบที่คำนวณด้วย KR-21

คีตส์ ได้เสนอให้ใช้ค่าความเชื่อมั่นที่เหมาะสมจากสัมประสิทธิ์ความคู่ขนาน แต่ในทางปฏิบัติอาจจำเป็นต้องใช้แบบทดสอบแบ่งครึ่ง หรือสัมประสิทธิ์แอลฟา ของครอนบัค (Cronbach)

วิธีที่ 4 ลอร์ด ได้เสนอการใช้ค่าความคลาดเคลื่อนมาตรฐานในการวัดให้เป็นประโยชน์โดยการจับคู่แบบทดสอบ โดยการสุ่มแบบแบ่งชั้นของข้อคำถาม ซึ่งจะดีกว่าการสุ่มแบบสมบูรณ์แบบ และได้เสนอสูตรหาค่าความคลาดเคลื่อนมาตรฐานจากกลุ่มตัวอย่างแบบแบ่งชั้น ดังสมการนี้

$$S_{E(i)}=\left[\sum_{h=1}^{c}\frac{X_{ih}(k_h-X_{ih})}{k_h-1}\right]^{1/2}$$

เมื่อ $S_{E(i)}$ แทน ค่าความคลาดเคลื่อนมาตรฐานของคนที่ 1

X_{ih} แทน คะแนนของคนที่ i ในข้อสอบกลุ่มที่ตรงกันในประเภทชุด h ของแบบทดสอบ

c แทน จำนวนของชุดข้อคำถาม

k_h แทน จำนวนข้อคำถามในชุด h (k_h ต้องมากกว่า 1)

วิธีนี้ แบบทดสอบต้องได้คะแนนจากทุก ๆ ชุด ถ้าแต่ละชุดเป็นแบบทดสอบย่อยของเครื่องมือวัดทั้งหมด ค่าความคลาดเคลื่อนมาตรฐานของผู้สอบแต่ละคน และค่าเฉลี่ยของค่าความคลาดเคลื่อนมาตรฐานในการวัดจะได้มาจากแต่ละช่วงระดับคะแนน

วิธีที่ 5 วิธีนี้ใช้การวิเคราะห์ความแปรปรวน (ANOVA) คิดขึ้นโดย ฮอยท์ (Hoyt. 1941) โดยใช้เมตริกซ์คะแนน ในการวิเคราะห์เพื่อหาค่าเฉลี่ยของผลรวมยกกำลังสองของผู้สอบ (MS_s), ข้อสอบ (MS_1), และผลของความแตกต่างของผู้สอบ และ ข้อสอบ($MS_{s\times1}$) ซึ่งค่าความเชื่อมั่นของแบบทดสอบจะประมาณค่าจาก ค่าเฉลี่ยของผลรวมยกกำลังสอง ในทางปฏิบัติค่าความแปรปรวนคลาดเคลื่อนสำหรับแบบทดสอบที่มี k ข้อจะประมาณโดย $k(MS_{s\times1})$ ถ้าแต่ละบุคคลถูกจัดเข้ากลุ่มในระดับคะแนนต่าง ๆ ค่าความคลาดเคลื่อนมาตรฐานในการวัดสำหรับจุดกึ่งกลางคะแนนของระดับช่วงคะแนนอาจจะประมาณค่าโดยสูตร

$$S_E=[k(MS_{s\times1})]^{1/2}$$

วิธีนี้ให้ผลใกล้เคียงกับวิธีของ ธอร์นไดค์ และมีสิ่งที่เหมือนกัน คือ หลักของการประมาณค่า มีข้อเสียก็คือจะมีค่าที่ไม่แน่นอน เมื่อประมาณจากผู้สอบจำนวนน้อย ซึ่งมักจะพบที่ปลายสุดของระดับคะแนนด้านมาก และด้านน้อย

วิธีนี้ยึดหลักทฤษฎี Item Response Curve Theory (IRT) ซึ่งกล่าวถึงปัญหาความน่าจะเป็นในการประมาณค่า บุคคลแต่ละคน *i* กับ คะแนนความสามารถจริง θ_i จะตอบคำถามข้อ *j* ได้อย่างถูกต้อง ความน่าจะเป็นนี้ ถูกประมาณค่าโดย ฟังก์ชัน $P_j(\theta)$ ซึ่งเหมือนกับทั่วไปคือ ค่าพารามิเตอร์บางตัวประมาณค่าโดยการกำหนดฟังก์ชันของแต่ละข้อคำถาม ถ้าค่าพารามิเตอร์เหล่านี้เป็นค่าประมาณที่ดี ซึ่งความแปรปรวนคลาดเคลื่อน-ในการวัด ประมาณค่าจากผู้สอบคนที่ *i* โดยมีขั้นตอนมี 2 ขั้นตอนคือ

(1) ประมาณค่า θ_i ของผู้สอบ

(2) ประเมินฟังก์ชัน ดังนี้

$$S^2_{E(i)} = \left\{ \sum_j^k \left[P_j(\theta_1) \right] \left[1 - P_j(\theta_i) \right] \right\}^{1/2}$$

$P_j(\theta_i)$ คือ ค่าของฟังก์ชันสำหรับข้อที่ j เมื่อถูกประเมินที่ระดับความสามารถ θ_i ของผู้สอบคนที่ *i*

วิธีนี้ให้ผลการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดสำหรับผู้สอบแต่ละคน การใช้วิธีนี้ผู้สอบถูกจัดแบ่งเป็นกลุ่มตามระดับคะแนนค่าเฉลี่ยของค่าความคลาด-เคลื่อนมาตรฐานในการวัดคำนวณจากผู้สอบในแต่ละระดับคะแนน และการหาค่าระหว่างค่า X และค่าของ θ แต่ละคู่ที่แตกต่างกัน คือ (X_0, θ_0) ได้มาโดยใช้สูตร

$$X_0 = \sum_j^k P_j(\theta_0)$$

ในการศึกษาปัจจุบัน สามารถหาค่าพารามิเตอร์ 2 ตัวนี้ โดยใช้ Logistic Model ซึ่งจะต้องใช้โปรแกรมคอมพิวเตอร์ LOGIST ของ วูด, วินเจอร์สกี และลอร์ด (Wood, Wingersky and Lord)

ในการศึกษาครั้งนี้ เฟลดต์, สเตฟเฟน และกุปตา ได้ทำการวิเคราะห์เบื้องต้น 3 ข้อประการแรกเพื่อตรวจสอบความเป็นมิติเดียวของแบบทดสอบเพื่อให้ตรงกับข้อตกลง

เบื้องต้นของทฤษฎี IRT ประการที่สองเปรียบเทียบรูปแบบทางเลือกสำหรับโค้งคุณลักษณะของข้อสอบ ICCS (Item Characteristic Curves) ประการที่สาม วิเคราะห์การถดถอยภายใต้วิธีของ โพลิโนเมียล

ผลสรุปว่า ค่าความคลาดเคลื่อนมาตรฐานในการวัดทั้ง 5 วิธี มีลักษณะร่วมกัน ถึงแม้ว่าผลการประมาณค่าจะมีความแตกต่างกันทั้ง 5 วิธี แต่ก็มีแนวโน้มที่เหมือนกัน โดยมีแนวคิดเหมือนกันคือ ถ้าผู้สอบถูกแบ่งเป็นชั้น คือคะแนนจริง (Ture Score), ความแปรปรวนคลาดเคลื่อน ณ คะแนนจริง จะเท่ากับ $\sum P_j(1-P_j)$ เมื่อ P_j คือค่าความยากของข้อสอบข้อที่ j สำหรับแต่ละคน ณ คะแนนจริง ด้านวิธีการ วิธีคอมพาวด์ ไบโนเมียล (Compound binomial) และ ANOVA เป็นวิธีที่น่าสนใจ แต่วิธีคอมพาวด์ ไบโนเมียล อาจจะยากถ้าจำนวนข้อสอบในแต่ละช่วงมีจำนวนน้อย และขนาดของผู้เข้าสอบมีไม่มาก

ต่อมาในปี 1986 บลิกต์ และ ชามา (Blixt and Chama. 1986) ได้ศึกษาเปรียบเทียบวิธีในการประมาณค่าความคลาดเคลื่อนที่ระดับความสามารถแตกต่างกัน ผลปรากฏว่าเมื่อใช้สูตรต่างกันในการหาค่าความคลาดเคลื่อนมาตรฐาน ณ ระดับความสามารถหนึ่ง พบว่าแตกต่างอย่างไม่มีนัยสำคัญ

จุดมุ่งหมายของการศึกษาครั้งนี้ ก็คือ ศึกษาว่าผลจากการศึกษาของ ลอร์ด (Lord. 1984) สามารถจำลองมาใช้กับแบบทดสอบอื่น ๆ ที่แตกต่างจากเดิมได้หรือไม่ จุดมุ่งหมายประการที่ 2 ก็คือ เพื่อให้รู้แน่ว่ามีความแตกต่างระหว่าง ค่าความคลาดเคลื่อนมาตรฐานที่ถูกประมาณจากสูตรเดิม กับค่าความคลาดเคลื่อนมาตรฐานที่คำนวณหาจากแต่ละระดับความสามารถที่ต่างกันหรือไม่

กลุ่มตัวอย่างที่ใช้ในครั้งนี้ 1,833 คน ถูกตัดออก 2 คน เนื่องจากได้ " 0 " คะแนน และตัด 7 คน ที่ได้คะแนนเต็ม "35" คะแนน จึงเหลือข้อมูลที่ใช้จริงเพียงแค่ 1,824 คน แบบทดสอบที่ใช้มี 35 ข้อ ข้อละ 4 ตัวเลือก ค่าเฉลี่ยคือ 16.41 และค่าเบี่ยงเบนมาตรฐาน คือ 7.45 ค่าความเชื่อมันของแบบทดสอบที่หาโดยวิธี $KR-20$ ได้ค่าเท่ากับ .88 ค่าความคลาดเคลื่อนมาตรฐาน 2.54

เนื่องจากว่า วิธีหนึ่งในการศึกษาครั้งนี้ ซึ่งเป็นวิธีของลอร์ด มีข้อตกลงเบื้องต้นว่า แบบทดสอบต้องเป็นมิติเดียว (Unidimension) ดังนั้น จึงต้องทำการวิเคราะห์องค์ประกอบ โดยใช้วิธี PC (Principal Components) ได้ 6 องค์ประกอบ ที่มีค่าไอเกน 1 หรือมากกว่า จึงถูกนำมาสกัดองค์ประกอบ โดยทั้ง 6 องค์ประกอบที่ค่าไอเกน ดังนี้ 7.59, 1.62, 1.36, 1.08, 1.06 และ 1.01 ตามลำดับ (ค่าแรกมีค่ามากกว่าค่าที่ 2, 3, 4, 5, 6, หลายเท่า) จึงสรุปได้ว่า มีความเป็นมิติเดียว (Unidimension) คำนวณหาค่าความคลาดเคลื่อนมาตรฐานสำหรับแต่ละคะแนนจากผู้ที่ได้คะแนนจาก $X = 1$ ถึง $X = 34$ (ซึ่งแต่ละคะแนนอาจมีจำนวนคน n ต่างกันก็ได้) โดยใช้ทั้ง 4 วิธี ที่เสนอโดย ลอร์ด (Lord. 1984)

ผลจากการศึกษา พบว่าค่าความคลาดเคลื่อนมาตรฐานที่ได้จากการคำนวณ ทั้ง 4 วิธี จะมีค่าค่อย ๆ สูงขึ้นในช่วงกลางของคะแนน และต่ำลงเมื่ออยู่ในช่วงปลายของคะแนน ค่าสูงสุดสำหรับการศึกษาครั้งนี้ คือ ประมาณ 3 ส่วนการศึกษาของ ลอร์ด จะมีค่าลดน้อยลงในช่วงปลายคะแนน ขณะที่การประมาณค่าโดยใช้วิธีแบบเดิมจะให้ค่าที่คงที่ อย่างไรก็ตาม สามารถ สรุปได้ว่า การรักษาไว้ซึ่งค่าประมาณที่มาก โดยใช้วิธีเดิม คะแนนจริงที่ได้จะไม่ต่ำกว่าค่าประมาณจริงในคะแนนทั้งหมด ยกเว้นช่วงกลางของคะแนนที่มีคะแนนมาก ในทางปฏิบัติในการประยุกต์ใช้ ความสนใจไม่ได้ตั้งอยู่ที่ปลายสุดของคะแนน ค่าความคลาดเคลื่อนมาตรฐานตามวิธีดั้งเดิมจะได้ค่าที่ใกล้เคียงกันกับทั้ง 4 วิธีของ ลอร์ด

ต่อมาในปี 1992 ควอลส์ และ เพนี (Qualls and Payne. 1992) ได้ศึกษาเปรียบเทียบการประมาณค่าความคลาดเคลื่อนมาตรฐานที่ระดับคะแนนเฉพาะ โดยใช้วิธีคำนวณค่า 6 วิธี คือ วิธีของเฟลดต์, คีตส์, จารัวรา, วิธี 3 พารามิเตอร์ , วิธีของโมเลนคอปฟ์ และธอร์นไดค์

ในการศึกษาครั้งนี้ใช้แบบทดสอบ 2 ฉบับ คือ แบบทดสอบวัดความสามารถด้านภาษา และวัดความสามารถด้านคณิตศาสตร์ (ITBS) โดยการศึกษาเปรียบเทียบจากวิธีสอบซ้ำ (Test – Retest) กับวิธีแบบคู่ขนาน (Parallel From)

ผลสรุปค่าความคลาดเคลื่อนมาตรฐานในการวัด โดยเฉลี่ยแล้วแบบทดสอบแบบคู่ขนานมีค่าสูงกว่าแบบทดสอบชนิดสอบซ้ำ เมื่อเปรียบเทียบค่าที่ได้จากการประมาณค่าความคลาดเคลื่อนมาตรฐานทั้ง 6 วิธี พบว่า วิธีของ เฟลดต์ ได้ค่าสูงที่สุด และวิธี 3 พารามิเตอร์ ได้ค่าต่ำที่สุด อย่างไรก็ตามทั้ง 6 วิธี จะได้ค่าที่ใกล้เคียงกัน

ในปี 1996 เฟลดต์ และ ควอลส์ (Feldt and Qualls. 1996) ได้ศึกษาค่าความแปรปรวนของความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ

โดยแสดงความเห็นว่า วิธีประมาณค่าแบบเดิมวิธีเดียวที่ใช้กันอยู่ในการประมาณค่าความคลาดเคลื่อนมาตรฐานนั้น เป็นวิธีที่ค่อนข้างยากลำบากในการที่จะนำไปใช้กับผู้สอบทุกคน ดังนั้น ถ้าผู้สอบมีจำนวนมาก ความแปรปรวนคลาดเคลื่อนของแบบทดสอบทั้งฉบับ σ_E^2 จะ เท่ากับ ความแปรปรวนของความแตกต่างกันของแบบทดสอบครึ่งฉบับ $\sigma^2(X_1 - X_2)$ ถ้าผู้สอบถูกจัดกลุ่มตามคะแนนรวม, ความแปรปรวนของ $X_1 - X_2$ สำหรับแต่ละกลุ่มย่อยหาได้โดย การประมาณค่ากำลังสองของความคลาดเคลื่อนมาตรฐานในการวัดในกลุ่มย่อย รากที่สองจะเป็นค่าความคลาดเคลื่อนมาตรฐานในการวัด ณ ระดับคะแนนนั้น ๆ ซึ่งก็คือวิธีธอร์นไดค์ (Thornd. 1951) นั่นเอง

$$\sigma_E^2 = \sigma_{E_1}^2 + \sigma_{E_2}^2$$

ข้อจำกัดของวิธีนี้ คือ

1. จากการศึกษาของ วูดรัฟฟ์ (Woodruff. 1990) แสดงว่า เมื่อผู้สอบถูกจัดกลุ่มโดยยึดตามคะแนนที่สอบได้ (X) ความแปรปรวนของความแตกต่างกันของแบบทดสอบแบ่งครึ่ง สำหรับแต่ละกลุ่มอาจจะประมาณได้เกินจริง (Overestimate) หรือต่ำกว่าจริง (Under Estimate) เหตุที่เป็นเช่นนี้เพราะว่า E_1 และ E_2 ไม่อิสระจากกันภายใต้การกระจาย ซึ่งเป็นเงื่อนไขของค่า X ความลำเอียงจะขึ้นอยู่กับค่าความเชื่อมั่นของแบบทดสอบ ถ้าค่าความเชื่อมั่นมากขึ้น ความลำเอียง (Bias) จะลดน้อยลง

2. วิธีของธอร์นไดค์ นี้ จำนวนผู้สอบในบางคะแนนจะมีจำนวนน้อย โดยเฉพาะทางปลายคะแนน X ซึ่งอาจจะแก้ไขโดยจัดอันตรภาคชั้น ให้มีชั้นลดน้อยลง หรือช่วงปลายให้มีระยะกว้างมากขึ้นอย่างไรก็ตามค่า $\sigma^2_{E\cdot x}$ จะเปลี่ยนทันทีจากอันตรภาคชั้นหนึ่งไปอีกอันตรภาคชั้นหนึ่ง

ในปี 1949 โมเลนคอปฟ์ ได้เสนอวิธีที่สามารถเข้ากันได้กับวิธีของ ธอร์นไดค์ โดยการปรับแก้ให้ดีขึ้น, ใช้กำลัง 2 ของความแตกต่างกันระหว่างคะแนนของแบบทดสอบแบ่งครึ่งสำหรับผู้สอบแต่ละคน (i), การปรับแก้สำหรับความแตกต่างกันระหว่างคะแนนเฉลี่ยของแบบทดสอบแบ่งครึ่ง, ได้ถูกพิจารณาให้เป็นตัวแปรอิสระ Y_1 สำหรับผู้สอบแต่ละคน (i) คะแนนรวมทั้งฉบับของผู้สอบ i ก็คือ X_i กำลังสอง ของ X_i ,กำลังสามของ X_i และต่อไปด้วยการใช้ตัวพยากรณ์ในสมการถดถอย วิธีการถดถอย (Polynomial Regression)

การปรับแก้องค์ประกอบ, $\bar{X}_1 - \bar{X}_2$, ถูกทำให้เหมาะสมกับลักษณะของทฤษฎี Essential Tau Equivalent ของแบบทดสอบแบ่งครึ่ง ดังนี้

$$Y_i = \left(\left[X_{i1} - X_{i2}\right] - \left[\bar{X}_1 - \bar{X}_2\right]\right)^2$$

และ

$$\hat{Y} = a_0 + a_1(X) + a_2(X^2) + a_3(X^3) + \ldots$$

สำหรับกลุ่มย่อยที่ระดับคะแนน $X = X_0$ หรือกลุ่มย่อยภายในแต่ละระดับคะแนนคือ X และค่าเฉลี่ยก็คือ Y

$$\bar{Y} = \frac{\sum\left(\left[X_{i1} - X_{i2}\right] - \left[\bar{X}_1 - \bar{X}_2\right]\right)^2}{N}$$

จะเท่ากับความแปรปรวนของความแตกต่างกันระหว่างแบบทดสอบแบ่งครึ่ง ขณะที่ธอร์นไดค์ (Feldt and Qualls. 1996 ; citing Thorndike. 1951) ได้แสดงถึงการประมาณค่า σ_ε^2 สำหรับแบบทดสอบทั้งฉบับภายใต้ข้อตกลงเบื้องต้นของทฤษฎี Tau Equivalent แต่ภายใต้ทฤษฎีการถดถอย, ถ้ารูปแบบของการถดถอย ถูกเลือกมาอย่างดี, ค่า Y จะให้ค่าประมาณที่ดีกว่าค่าจากคะแนนสอบ $X = X_0$ ภายใต้เงื่อนไขค่าเฉลี่ยของกลุ่มประชากรย่อย ที่ ($\bar{Y}$) แต่ถึงกระนั้น รากกำลังสองของ $\hat{Y}$ จะประมาณค่าความคลาดเคลื่อนมาตรฐานดีกว่าที่ระดับคะแนน X_0

1. การศึกษาครั้งนี้จำลองข้อมูลขึ้นมา เป็นแบบทดสอบที่ประกอบด้วย 36 ข้อคำถาม แบ่งเป็น 6 ส่วนที่คู่ขนานกัน แต่ละส่วนมี 6 ข้อ ข้อสอบที่อยู่ภายในแต่ละส่วนมีค่าความยากตั้งแต่ .40 ถึง 0.60 เป็นแบบทดสอบที่ให้คะแนนแบบ 0, 1 ได้ค่าความเชื่อมั่น .785 คะแนนของข้อคำถาม คะแนนรวมทั้งฉบับได้มาจากกลุ่มตัวอย่างเทียม จำนวน 1,000 คน สุ่มข้อสอบออกเป็น 6 ส่วน ๆ ละ 6 ข้อ, 3 ส่วน ๆ ละ 12 ข้อ, และ 2 ส่วน ๆ ละ 18 ข้อ

2. แบบทดสอบที่ใช้เป็นการวัดความคิดรวบยอดทางด้านคำศัพท์ และคณิตศาสตร์ของ Iowa Test of Basic Skill สำหรับนักเรียนเกรด 4 และ 7

แบบทดสอบด้านคำศัพท์ (Vocabulary)	เกรด 4 มี 36 ข้อ (V-4)
	เกรด 7 มี 41 ข้อ (V-7)
แบบทดสอบด้านคณิตศาสตร์ (Mathematical)	เกรด 4 มี 32 ข้อ (M-4)
	เกรด 7 มี 41 ข้อ (M-4)

การประมาณค่าของแต่ละคนคำนวณจากสูตร

$$Y_i = L\tilde{\sigma}^2_{E_i} = \frac{K}{k}\left[\frac{\sum^{n}\left(X_{ij} - \bar{X}_i - \bar{X}_j + m\right)^2}{n-1}\right]$$

K แทน จำนวนข้อสอบทั้งฉบับ (36, 32, 41)

k แทน จำนวนข้อสอบในแต่ละส่วนที่ถูกแบ่ง(เช่น 6 ข้อ, 9 ข้อ, 18 ข้อ...)

X_{ij} แทน คะแนนสอบของคนที่ i ที่แบบทดสอบส่วนที่ j

$\bar{X}_j$ แทน คะแนนเฉลี่ยของแบบทดสอบส่วนที่ j ที่คำนวณจากนักเรียนทุกคน

$\bar{X}_i$ แทน คะแนนเฉลี่ยของผู้สอบ 1 คน จากคะแนนของแบบทดสอบทุกส่วน

m แทน คะแนนเฉลี่ยของแบบทดสอบทั้งฉบับ

จากการศึกษาในครั้งนี้ ค่าความแปรปรวนของความคลาดเคลื่อนมาตรฐานในการวัด ผลปรากฏว่าค่า $\tilde{\sigma}^2_{E\cdot x}$ ที่ระดับคะแนน X ตรงช่วงคะแนน จะมีค่าสูงในทุกฉบับ ซึ่งเป็น สิ่งที่เห็นได้อย่างชัดเจน

การประมาณค่าความคลาดเคลื่อนมาตรฐาน สามารถบรรลุผลได้ดี ก็โดยการแบ่งแบบทดสอบออกเป็นส่วน ๆ ที่มีลักษณะเป็นคู่ขนานกัน (Parallel) และการประยุกต์ใช้จะต้องเป็นไปตามข้อตกลงของ ทฤษฎี Essentially Tau Equivalent การประมาณค่าโดยใช้วิธีนี้จะให้ผลที่ดี ถ้าหากว่ามีจำนวนผู้สอบเพียง 10 เปอร์เซ็นต์ ที่มีคะแนนอยู่ทางช่วงปลาย ไม่ว่าจะเป็นคะแนนทางปลายด้านน้อยสุด หรือมากสุด ก็ควรจะตัดผู้สอบส่วนนี้ทิ้ง

งานวิจัยที่เกี่ยวข้อง

1. งานวิจัยภายในประเทศ

เตือนใจ เศรษฐสักโก (2511) ได้ศึกษาถึงการจัดระเบียบข้อทดสอบวิธีต่าง ๆ ที่จะส่งผลต่อความสามารถในการสอบของนักเรียน ชั้นประถมศึกษาปีที่ 7 ปรากฏผลว่า การเรียงข้อสอบจากฉบับที่ง่ายไปยาก และภายในแต่ละฉบับเรียงจากข้อง่ายไปยาก เป็นวิธีที่ทำให้ผลสัมฤทธิ์สูง ทั้งเมื่อใช้กับนักเรียนส่วนรวม และนักเรียนกลุ่มความสามารถสูงกับกลุ่มความสามารถต่ำส่วนนักเรียนกลุ่มความสามารถปานกลางนั้น จะใช้ข้อสอบที่เรียงแบบใดก็ได้

ศศิธร สุวรรณสุข (2519) ได้ศึกษาการเรียงลำดับคำถามแบบเรียงจากข้อง่ายไปหาข้อยาก เรียงจากข้อยากไปหาข้อง่าย เรียงตามหัวข้อเนื้อหาวิชา และเรียงจากข้อง่ายไปหาข้อยากในแต่ละหัวข้อเนื้อหาวิชา ในวิชาคณิตศาสตร์ ระดับชั้นประถมศึกษาปีที่ 7 พบว่าการเรียงข้อสอบต่างกัน ทำให้ค่าความเที่ยงตรงของแบบทดสอบวัดทักษะ การแก้ปัญหา และการหาเหตุผล แตกต่างกันอย่างไม่มีนัยสำคัญ

การเรียงข้อสอบต่างกัน ทำให้ความเชื่อมั่นของแบบทดสอบทักษะแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนความเชื่อมั่นแบบทดสอบการแก้ปัญหา และการหาเหตุผลให้ค่าความเชื่อมั่นแตกต่างกันอย่างไม่มีนัยสำคัญ

การเรียงข้อสอบต่างกัน ทำให้คะแนนสอบจากแบบทดสอบวัดทักษะ และการแก้ปัญหาแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบการหาเหตุผลให้ค่าความแตกต่างกันอย่างไม่มีนัยสำคัญ

การเรียงข้อสอบแบบจากข้อง่ายไปหายาก และการเรียงข้อสอบตามหัวข้อเนื้อหาวิชาจะเหมาะสมกับวิชาคณิตศาสตร์ ระดับชั้นประถมศึกษาปีที่ 7 มากที่สุด

นิโลบล นิ่มกิ่งรัตน์ (2519) ได้ศึกษาอิทธิพลของการจัดเรียงลำดับข้อคำถามที่จำแนกตามสมรรถภาพสมองที่มีต่อคุณภาพของแบบทดสอบคณิตศาสตร์ และภาษาไทย ในระดับชั้นมัธยมศึกษาปีที่ 3 ผลจากการศึกษาพบว่า

1. การจัดเรียงลำดับข้อคำถามที่จำแนกตามสมรรถภาพสมองที่จะวัดของแบบทดสอบภาษาไทยทั้ง 6 วิธี มีผลต่อคะแนนการสอบของนักเรียน และคุณภาพของแบบทดสอบด้านความเชื่อมั่น ความเที่ยงตรง และความยากมาตรฐาน แตกต่างกัน แต่ไม่มีผลทำให้ค่าอำนาจจำแนกแตกต่างกันอย่างมีนัยสำคัญ

2. การจัดเรียงลำดับข้อคำถามที่จำแนกตามสมรรถภาพสมองที่จะวัดของแบบทดสอบคณิตศาสตร์ทั้ง 6 วิธี ไม่มีผลทำให้คะแนนการสอบของนักเรียน และคุณภาพของแบบทดสอบทุกด้าน แตกต่างกันอย่างมีนัยสำคัญ

3. การศึกษาครั้งนี้พบว่า แบบทดสอบภาษาไทยชุดที่จัดเรียงข้อคำถามที่วัดสมรรถภาพสมองความรู้ความจำขึ้นเป็นอันดับแรกจะให้คะแนนการสอบของนักเรียนและคุณภาพของแบบทดสอบดีที่สุด ชุดที่จัดเรียงข้อสอบที่วัดสมรรถภาพสมองการนำไปใช้ขึ้นเป็นอันดับแรกจะให้คะแนนการสอบของนักเรียนและคุณภาพของแบบทดสอบต่ำสุด

4. แนวโน้มของการแปรเปลี่ยนของคุณภาพ และคะแนนการสอบของนักเรียนจากแบบทดสอบคณิตศาสตร์ และภาษาไทย มีลักษณะเกือบตรงกันข้าม ยกเว้นด้านความยากมาตรฐานที่มีลักษณะคล้ายคลึงกัน

5. แบบทดสอบภาษาไทยที่ควรใช้กับกลุ่มที่มีความถนัดทางการเรียนสูง และ ต่ำ เมื่อได้พิจารณาถึงความเหมาะสมของแบบทดสอบทั้งในด้านคุณภาพของแบบทดสอบ คะแนนการสอบของนักเรียน และลักษณะโค้งของการแจกแจงที่เป็นไปตามทฤษฎีแล้ว พบว่าควรใช้แบบทดสอบที่มีการจัดเรียงข้อคำถามดังนี้ คือ วัดความรู้ความจำ– ความเข้าใจ – การนำไปใช้ หรือสูงกว่านี้ มีแนวโน้มจะดี และเหมาะสมที่สุด

6. แบบทดสอบภาษาไทย และคณิตศาสตร์ ที่ควรใช้กับกลุ่มนักเรียนที่มีความถนัดทางการเรียนปานกลาง พบว่าใช้แบบทดสอบที่มีการเรียงแบบใดก็ได้

7. แบบทดสอบคณิตศาสตร์ที่ควรใช้กับกลุ่มที่มีความถนัดทางการเรียนสูง และต่ำ เมื่อได้พิจารณาถึงความเหมาะสมกับของแบบทดสอบทั้งในด้านคุณภาพของแบบทดสอบ คะแนนการสอบของนักเรียน และลักษณะโค้งของการแจกแจงที่เป็นไปตามทฤษฎี พบว่าควรใช้แบบทดสอบที่มีการจัดเรียงข้อคำถามดังนี้ คือวัดความเข้าใจ- การนำไปใช้ หรือสูงกว่านี้ – ความรู้ความจำ มีแนวโน้มจะดี และเหมาะสมที่สุด

แน่งน้อย เพชร์รัตน์ (2524) ได้ศึกษาว่า การเรียงลำดับข้อคำถามของแบบทดสอบด้วยวิธีต่างกันคือ เรียงจากง่ายไปหายาก เรียงตามลำดับความยากแบบหมุนเวียน และเรียงแบบสุ่ม โดยมิได้คำนึงถึงเนื้อหาของข้อคำถามนั้น จะมีผลต่อคุณภาพของแบบทดสอบวิชาวิทยาศาสตร์และแบบทดสอบทักษะขบวนการทางวิทยาศาสตร์อย่างไร กลุ่มตัวอย่างเป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2520 ในเขตจังหวัดนครปฐม จำนวน 300 คน

ผลการศึกษาค้นพบว่า การเรียงลำดับข้อคำถามส่งผลต่อคุณภาพของแบบทดสอบทั้ง 4 ด้าน คือ ความเที่ยงตรง ความเชื่อมั่น ความยากง่าย และอำนาจจำแนก ไม่แตกต่างกันในแบบทดสอบทั้งสองฉบับ เว้นแต่ในแบบทดสอบทักษะขบวนการทางวิทยาศาสตร์ พบว่าการเรียงแบบสุ่มส่งผลต่อความเชื่อมั่นสูงกว่าการเรียงคำถามแบบอื่น ๆ

วลีมาศ แซ่อึ้ง (2537) ได้ศึกษาความคลาดเคลื่อนมาตรฐาน และประสิทธิภาพสัมพันธ์ของการประมาณค่าเจตคติต่อวิชาคณิตศาสตร์ระหว่างมาตรวัดแบบลิเคอร์ทกับมาตรวัดแบบออสกูด โดยใช้โปรแกรม RSCALE version 2.0 ในการประมาณค่าเจตคติของผู้ตอบ และค่าคุณลักษณะประจำข้อ ตามแบบทฤษฎีการตอบข้อคำถามแบบเรทติ้งสเกลโมเดล พบว่าค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าเจตคติต่อวิชาคณิตศาสตร์ ระหว่างมาตรวัดแบบลิเคอร์ท กับมาตรวัดแบบออสกูด มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

รุ่งฤดี ทรัพย์นิธิ (2539) ได้ศึกษาเปรียบเทียบความคลาดเคลื่อนมาตรฐานในการประมาณค่าพารามิเตอร์ระหว่างการวิเคราะห์ด้วยโมเดลโลจิสติก แบบหนึ่ง สอง สาม พารามิเตอร์ และขนาดกลุ่มตัวอย่างต่างกัน โดยศึกษาจากกลุ่มตัวอย่างขนาด 250 คน, 500 คน, 750 คน และ 1,000 คน เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาคณิตศาสตร์ แบบปรนัย 5 ตัวเลือก ผลการวิจัยสรุปได้ดังนี้

1. ความคลาดเคลื่อนมาตรฐานในการประมาณค่าพารามิเตอร์ข้อสอบระหว่างกลุ่มตัวอย่างที่มีจำนวนไม่เท่ากัน เมื่อประมาณค่าด้วยโมเดลหนึ่งพารามิเตอร์ มีค่าแตกต่างกัน เมื่อประมาณค่าด้วยโมเดลสองพารามิเตอร์ ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความยากมีค่าไม่แตกต่างกัน ส่วนความคลาดเคลื่อนมาตรฐานในการ

ประมาณค่าอำนาจจำแนกมีค่าแตกต่างกันในบางกลุ่มตัวอย่าง เมื่อประมาณค่าด้วยโมเดลสามพารามิเตอร์ ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความยาก และการเดามีค่าไม่แตกต่างกัน ส่วนความคลาดเคลื่อนมาตรฐานในการประมาณค่าอำนาจจำแนกมีค่าแตกต่างกันในบางกลุ่มตัวอย่าง

2. ความคลาดเคลื่อนมาตรฐานในการประมาณค่าพารามิเตอร์ผู้สอน ระหว่างกลุ่มตัวอย่างที่มีจำนวนไม่เท่ากัน เมื่อประมาณค่าด้วยโมเดลหนึ่ง, สอง และสาม พารามิเตอร์ ปรากฏผลเช่นเดียวกันคือ มีค่าแตกต่างกันในบางกลุ่มตัวอย่าง

3. ความคลาดเคลื่อนมาตรฐานในการประมาณค่าพารามิเตอร์ข้อสอบระหว่างโมเดลเมื่อวิเคราะห์จากกลุ่มตัวอย่างที่มีจำนวนเท่ากัน ความคลาดเคลื่อนมาตรฐานในการประมาณค่าอำนาจมีค่าแตกต่างกันทุกขนาดกลุ่มตัวอย่าง ส่วนความคลาดเคลื่อนในการประมาณค่าความยากมีค่าแตกต่างกันในบางกลุ่มตัวอย่าง

4. ความคลาดเคลื่อนมาตรฐานในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดล เมื่อวิเคราะห์จากกลุ่มตัวอย่างที่มีจำนวนเท่ากัน พบว่ามีค่าแตกต่างกันจากทุกขนาด

จากการศึกษาพบว่า กลุ่มตัวอย่างที่นำมาใช้ในการวิเคราะห์และให้ค่าประมาณพารามิเตอร์ที่ดีที่สุด คือ กลุ่มตัวอย่างขนาน 750 คน

พิชญา เกตุสกุล (2539) ได้ศึกษาเปรียบเทียบความเที่ยงตรง และความคลาด-เคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง โดยศึกษากับกลุ่มตัวอย่างนักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษา จังหวัดลพบุรี จำนวน 419 คน ซึ่งได้มาจากการสุ่มแบบแบ่งชั้น เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ที่ผู้วิจัยสร้างขึ้นเอง ซึ่งมี 3 รูปแบบคือ แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง การประมาณค่าความยากของข้อสอบ และค่าความสามารถของผู้สอบใช้โปรแกรมไบคาล ตามทฤษฎีการตอบข้อสอบแบบราสช์ โมเดล ผลการวิจัยสรุปได้ดังนี้

1. ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันกับแบบทดสอบแยกกลุ่มสองครั้ง และแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันกับแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 ตามลำดับ ส่วนค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างกับแบบทดสอบแยกกลุ่มสองครั้ง มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

2. ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบกลุ่มที่มีความยากใกล้เคียงกันกับแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันกับแบบทดสอบแยกกลุ่มสองครั้ง และแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างกับแบบทดสอบแยกกลุ่มสองครั้ง มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกค่า

2. <u>งานวิจัยต่างประเทศ</u>

ลอร์ด (Lord. 1984) ได้ศึกษาค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะแบบทดสอบเพียงฉบับเดียว แล้วคำนวณจากจำนวนคำตอบที่ถูกต้อง ณ ระดับความสามารถเฉพาะ โดยใช้วิธีคำนวณ 4 วิธี คือ วิธีของ ลอร์ด, ราสย์ โมเดล , ลอร์ด และโนวิกส์ และวิธีของลอร์ด คอมพาวด์ ไบโนเมียล โดยใช้แบบทดสอบ SCAT เป็นข้อสอบแบบหาคำที่สัมพันธ์กัน จำนวน 50 ข้อ จุดมุ่งหมายในการศึกษาครั้งนี้ คือ เพื่อเรียกความสนใจในสูตรทั้ง 4 ให้คืนกลับมาในเรื่องการประมาณค่าความคลาดเคลื่อน-มาตรฐานในการวัด และเพื่อเปรียบเทียบสูตรทั้ง 4 โดยใช้ข้อมูลจริงในการวิเคราะห์จากกลุ่มตัวอย่างจำนวน 3,000 คน ซึ่งเป็นนักเรียนระดับมัธยมศึกษาจากการศึกษาพบว่า วิธีของลอร์ด (วิธีที่ 1) จะให้ค่าความคลาดเคลื่อนมาตรฐานในการวัดน้อยกว่าวิธีราสย์ โมเดล (วิธีที่ 2) และวิธีของลอร์ด คอมพาวด์ ไบโนเมียล (วิธีที่ 4)

เนื่องจากว่า ข้อสอบ SCAT เป็นข้อสอบแบบหาคำที่สัมพันธ์กัน ผู้วิจัยเชื่อว่า ค่าความคลาดเคลื่อนมาตรฐานที่ได้จากวิธีราสย์ โมเดล วิธีที่ 2 มีความถูกต้อง วิธีที่ 1 และวิธีที่ 4 ของลอร์ด สามารถใช้กับแบบทดสอบที่ไม่เป็นมิติเดียว Unidimension ได้ แต่ถ้ามีการปฏิบัติ และหาเหตุผลสนับสนุน วิธีที่ 4 นี้จะเป็นวิธีที่ดีกว่า เพราะความง่าย และขนาดของกลุ่มตัวอย่างไม่จำเป็นต้องใหญ่มาก

บลิกต์ และชามา (Blixt and Shama. 1986) ได้ศึกษาเปรียบเทียบวิธีในการประมาณค่าความคลาดเคลื่อนมาตรฐานที่ระดับความสามารถแตกต่างกัน โดยใช้วิธีประมาณค่า 4 วิธี ที่เสนอโดย ลอร์ด ศึกษาจากกลุ่มตัวอย่าง 1,824 คน แบบทดสอบที่ใช้มี 35 ข้อ เป็นแบบปรนัย 4 ตัวเลือก ได้ค่าเฉลี่ยคือ 16.41 ค่าเบี่ยงเบนมาตรฐาน 7.45 ค่าความเชื่อมั่นของแบบทดสอบที่หาโดยวิธี KR. 20 ได้ค่าเท่ากับ .88 ค่าความคลาดเคลื่อนมาตรฐาน 2.54 ผลจากการศึกษาพบว่า เมื่อใช้สูตรต่างกันในการหาค่าความคลาดเคลื่อนมาตรฐานในการวัด ณ ระดับคะแนนความสามารถหนึ่ง มีความแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

เฟลดต์, สเตฟเฟน และ กุปตา (Feldt, Steffen and Gupta. 1985) สำนักทดสอบ The Standard for educational and Psychological ได้แนะนำว่า วิธีประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดมีมากมายหลายวิธี ในการศึกษาครั้งนี้ เป็นการเปรียบเทียบการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัด ด้วยวิธี 5 วิธี โดยใช้แบบทดสอบ 3 ฉบับ คือ ด้านคำศัพท์ (V)`, ด้านการอ่าน (L), และการใช้แหล่งข้อมูล (S_i) โดยทำการทดสอบกับเกรด 9 กับเกรด 11 ซึ่งชุดข้อสอบไม่เหมือนกัน ดังนั้นจึงมีแบบทดสอบทั้งหมด 9 ฉบับ ผลจากการศึกษา 5 วิธี คือ ค่าสูงสุดที่ได้จะอยู่ตรงกลางของระดับช่วงคะแนน และค่อย ๆ ลดน้อยลงเมื่ออยู่ปลายสุดของคะแนน แนวโน้มเหล่านี้อาจจะเป็นเครื่องหมายของค่าความคลาดเคลื่อนมาตรฐานของคะแนนดิบ ทั้งในแบบทดสอบวัดผลสัมฤทธิ์ และแบบทดสอบด้านความสามารถต่าง ๆ แต่แบบทดสอบลักษณะอื่นอาจจะมีแนวโน้มที่แตกต่างจากที่กล่าวมาข้างต้น อย่างไรก็ตาม การประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดจากการศึกษาครั้งนี้ พบว่ามี 2 วิธี ที่ได้รับการ

ยอมรับ และสนับสนุน คือวิธีของ ธอร์นไดค์ โดยยึดหลัก Polynomial และวิธีของคีตส์ ที่พัฒนาจากสูตรของลอร์ด โดยยึดหลักรูปแบบความคลาดเคลื่อนแบบ Binomial

ควอลส์และเพนี (Qualls and Payne. 1992) ได้ศึกษาเปรียบเทียบวิธีการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดของระดับคะแนน โดยใช้วิธีประมาณค่า 6 วิธี คือ วิธีของ เฟลดต์, คีตส์, จาร์รัวรา, สามพารามิเตอร์,โมเลนคอปฟ์ และ ธอร์นไดค์ โดยใช้แบบทดสอบวัดความถนัดด้านภาษา และคณิตศาสตร์ (ITBS) ซึ่งใช้วิธีสอบซ้ำ (Test - Retest) และแบบคู่ขนาน (Parallel From.) ผลจากการศึกษาปรากฏว่า ค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานในการวัดของแบบทดสอบแบบคู่ขนานมีค่าสูงกว่าแบบทดสอบชนิดสอบซ้ำ และค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ได้จากการประมาณค่าทั้ง 6 วิธี ได้ค่าใกล้เคียงกัน แต่วิธีของเฟลดต์ จะได้ค่าสูงกว่าวิธีอื่น ๆ

เฟลดต์และ ควอลส์ (Feldt and Qualls. 1996) ได้ศึกษา การประมาณค่าความ-แปรปรวนที่คลาดเคลื่อนในการวัดที่ระดับคะแนนเฉพาะ โดยใช้แบบทดสอบ 2 ประเภท คือแบบทดสอบที่วัดโดยตรง กับแบบทดสอบที่วัดโดยอ้อม แบบทดสอบวัดโดยตรงเป็นแบบทดสอบวัดคำศัพท์ และความสามารถทางคณิตศาสตร์ แต่ละฉบับแบ่งออกเป็นส่วน ๆ แต่ละส่วนเป็นแบบทดสอบคู่ขนานกัน ส่วนแบบทดสอบประเภทวัดโดยอ้อมเป็นแบบทดสอบคู่ขนานที่แบ่งเป็นส่วน ๆ เช่นกัน มีทั้งหมด 36 ข้อ โดยแบ่งเป็น 6 ส่วน, 3 ส่วน และ 2 ส่วน ค่าความยากอยู่ระหว่าง .40 - .60 ค่าความเชื่อมั่น .785 เป็นแบบทดสอบที่ให้คะแนน 0, 1 กลุ่มตัวอย่างเป็นนักเรียนระดับ 4 และ ระดับ 7 จำนวน 1,000 คน

ผลของการประมาณค่าความคลาดเคลื่อนมาตรฐาน สามารถบรรลุผลได้ดี โดยการแบ่งแบบทดสอบออกเป็นส่วน ๆ ที่มีลักษณะเป็นคู่ขนานกัน (Parallel) ค่าความแปรปรวนคลาดเคลื่อนของแต่ละระดับคะแนน โดยเฉลี่ยจะได้ 8.42, 8.16 และ 7.92 สำหรับแบบทดสอบที่แบ่งเป็น 2 ส่วน, 3 ส่วน และ 6 ส่วน ตามลำดับ ซึ่งค่าความแปรปรวนคลาดเคลื่อนที่ระดับคะแนน x จะมีค่าสูงตรงช่วงกลางของคะแนนในทุกฉบับ ซึ่งเป็นสิ่งที่เห็นได้อย่างชัดเจน แต่อย่างไรก็ตามภายในแต่ละส่วนจะต้องเป็นไปตามข้อตกลงของทฤษฎี Essentially Tau Equivalent การประมาณค่าโดยใช้การถดถอยแบบโพลีโนเมียล (Polynomial Regression) ก็จะได้ผลดี

จากการศึกษางานวิจัยที่เกี่ยวข้อง เกี่ยวกับความคลาดเคลื่อนมาตรฐานในการวัดนั้น จะเห็นว่าในประเทศไทยยังไม่มีผู้ใดศึกษาเกี่ยวกับค่าความคลาดเคลื่อนมาตรฐานในการวัด โดยใช้วิธีประมาณค่าหลาย ๆ วิธี ผู้วิจัยจึงเกิดแนวความคิดที่จะศึกษาเกี่ยวกับการจัดเรียงข้อสอบที่มีผลต่อการความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนน-เฉพาะ โดยใช้วิธี 3 วิธี ซึ่งได้แก่ วิธีของธอร์นไดค์, วิธีของลอร์ด, วิธีของคีตส์ เพื่อนำมาศึกษาว่าจะมีผลแตกต่างกันอย่างไร

บทที่ 3
วิธีดำเนินการศึกษาค้นคว้า

ประชากร

ประชากรที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ภาคเรียนที่ 2 ปีการศึกษา 2540 ของโรงเรียนในสังกัดสำนักงานการประถมศึกษาจังหวัดสระบุรี

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ภาคเรียนที่ 2 ปีการศึกษา 2540 ของโรงเรียนในสังกัดสำนักงานการประถมศึกษาจังหวัดสระบุรี ซึ่งเลือกมาโดยวิธีการสุ่มแบบหลายขั้นตอน (Multistage Random Sampling) โดยมีขั้นตอนการสุ่มดังนี้

ขั้นที่ 1 สุ่มอำเภอในจังหวัดสระบุรี ด้วยวิธีสุ่มอย่างง่าย (Simple Random Sampling) จาก 13 อำเภอ ได้กลุ่มตัวอย่าง 7 อำเภอ

ขั้นที่ 2 สุ่มโรงเรียนในแต่ละอำเภอที่เป็นกลุ่มตัวอย่าง ด้วยวิธีสุ่มอย่างง่าย ได้กลุ่มตัวอย่างที่ใช้ในการวิจัย 15 โรงเรียน จำนวน 1,019 คน

ตาราง 1 จำนวนนักเรียนที่เป็นกลุ่มตัวอย่าง จำแนกตามอำเภอ และโรงเรียน

อำเภอ	โรงเรียน	จำนวนนักเรียนที่ใช้ในการวิจัย
เมืองสระบุรี	วัดโนนสภาราม	49
	ศาลาแดง	123
	วัดหนองโนใต้	39
เสาไห้	วัดเจ้าฟ้า	34
	วัดสูง	59
พระพุทธบาท	วัดพระพุทธบาท	210
	ท้ายพิกุล(สำนักงานสลากกินแบ่งสงเคราะห์ 68)	96
วิหารแดง	ชุมชนวัดคลองไทร	52
	วัดหนองสรวง	29
หนองโดน	ชุมชนบ้านกลับ (สราญราษฎร์)	31
	หนองโดน (สนิทประชากร)	30
วังม่วง	บ้านวังม่วง	57
	บ้านมณีโชติสามัคคี	20
เฉลิมพระเกียรติ ฯ	หน้าพระลาน (พิบูลสงเคราะห์)	190
รวม		1,019

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นแบบทดสอบวัดความสามารถด้านตัวเลขแบบอนุกรมธรรมดา ซึ่งผู้วิจัยสร้างขึ้นเอง เป็นแบบปรนัยชนิดเลือกตอบ 4 ตัวเลือก จำนวน 40 ข้อ

ลักษณะของแบบทดสอบ

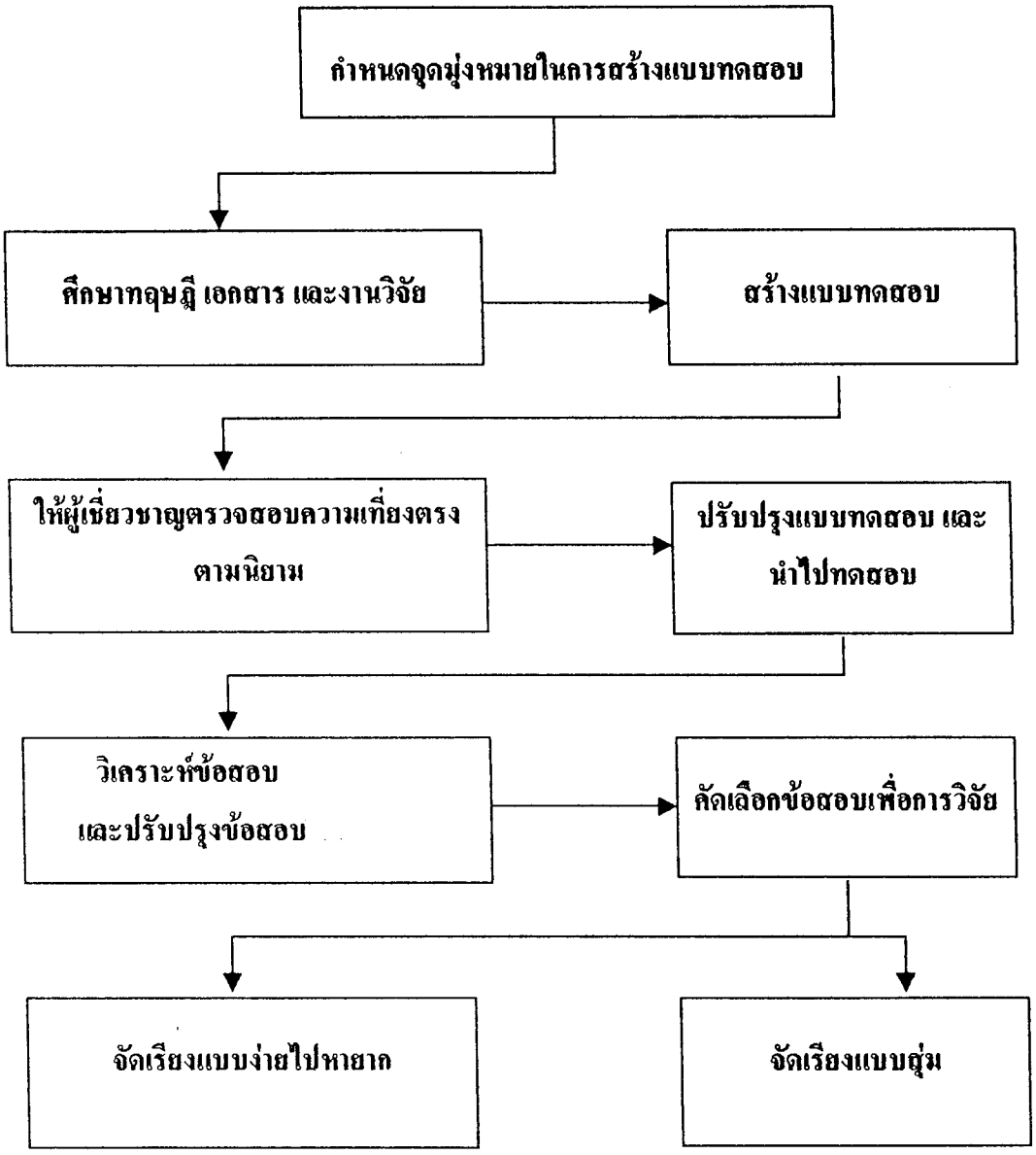
ลักษณะของข้อสอบ จะใช้หลักการเรียงตัวเลขตามกฎเกณฑ์ต่าง ๆ อย่างเป็นระบบ ซึ่งผู้ตอบต้องใช้ความคิดโดยอาศัยวิธีการทางคณิตศาสตร์ คือวิธีบวก ลบ คูณ และหาร โดยจัดเรียงข้อคำถาม 2 รูปแบบ คือ

ฉบับที่ 1 เรียงข้อสอบจากง่ายไปหายาก

ฉบับที่ 2 เรียงข้อสอบแบบสุ่ม

วิธีดำเนินการสร้างเครื่องมือที่ใช้ในการวิจัย

ในการสร้างเครื่องมือที่ใช้ในการวิจัยครั้งนี้ ได้ดำเนินการสร้างแบบทดสอบ เป็นลำดับขั้น ดังนี้



ภาพประกอบ 2 แผนผังแสดงขั้นตอนการดำเนินการสร้างแบบทดสอบ

รายละเอียดในการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้ามีดังนี้

1. กำหนดจุดมุ่งหมายในการสร้างแบบทดสอบ เพื่อวัดความสามารถด้านตัวเลขแบบอนุกรมธรรมดาของนักเรียนชั้นประถมศึกษาปีที่ 6 ว่ามีมโนภาพทางคณิตศาสตร์เพียงใด ในแง่ความสัมพันธ์ของปริมาณตัวเลข

2. ศึกษาทฤษฎี เอกสาร และงานวิจัยที่เกี่ยวข้อง จากหลักการสร้างแบบทดสอบความถนัดทางการเรียน แล้วนิยามเพื่อเป็นแนวทางในการสร้างแบบทดสอบ (ล้วน สายยศ และอังคณา สายยศ. 2527)

3. สร้างแบบทดสอบวัดความสามารถด้านตัวเลขแบบอนุกรมธรรมดา จำนวน 60 ข้อ เป็นแบบปรนัย ชนิดเลือกตอบ 4 ตัวเลือก

4. นำแบบทดสอบ (ฉบับร่าง) ไปให้ผู้เชี่ยวชาญจำนวน 5 คน พิจารณาตรวจสอบความเที่ยงตรงตามนิยามเป็นรายข้อ โดยใช้สัมประสิทธิ์ความสอดคล้องของผู้เชี่ยวชาญ (IOC) คือผู้เชี่ยวชาญ 3 ใน 5 คน เห็นว่าข้อสอบนั้นวัดได้ตรงตามที่นิยามถือว่าข้อสอบนั้นใช้ได้ผลการพิจารณาปรากฏว่า ข้อสอบทั้ง 60 ข้อ วัดได้ตรงตามนิยามทั้ง 60 ข้อ โดยมีค่าดัชนีความสอดคล้องอยู่ระหว่าง 0.6 – 1.00 ดังแสดงในภาคผนวก ก

5. นำแบบทดสอบที่ผู้เชี่ยวชาญพิจารณาแล้ว ไปทดสอบกับนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 150 คน แล้วนำผลการสอบมาตรวจให้คะแนน คือตอบถูกได้ 1 คะแนน ถ้าตอบผิด ไม่ตอบ หรือตอบมากกว่า 1 คำตอบ ให้ 0 คะแนน

6. นำผลในการตรวจให้คะแนนไปวิเคราะห์หาค่าความยาก และค่าอำนาจจำแนกของแบบทดสอบ โดยใช้เทคนิค 27 เปอร์เซ็นต์ และตารางสำเร็จรูป จุง เตห์ ฟาน (Chung-Teh-Fan) ได้ข้อสอบมีค่าความยากอยู่ระหว่าง .18 - .89 และ ค่าอำนาจจำแนกอยู่ระหว่าง .00 - .72 แล้วคัดเลือกข้อสอบที่มีค่าความยากอยู่ระหว่าง .20 - .80 และมีค่าอำนาจจำแนกตั้งแต่ .20 ขึ้นไป ให้ได้จำนวน 40 ข้อ ปรากฏว่าข้อสอบที่คัดเลือกไว้มีค่าความยากอยู่ระหว่าง .21 - .78 และค่าอำนาจจำแนกตั้งแต่ .20 - .64

7. นำข้อสอบที่คัดเลือกไว้จำนวน 40 ข้อ มาจัดเรียงเป็น 2 รูปแบบ คือ

ฉบับที่ 1 เรียงข้อสอบจากง่ายไปหายาก

ฉบับที่ 2 เรียงข้อสอบแบบสุ่ม

8. นำแบบทดสอบทั้ง 2 ฉบับ ไปทดสอบกับกลุ่มตัวอย่าง จำนวน 1,019 คน โดยแบ่งนักเรียนออกเป็น 2 กลุ่ม ด้วยวิธีสุ่มอย่างง่าย แล้วให้กลุ่มแรกจำนวน 514 คน ทำแบบทดสอบ ฉบับที่ 1 ที่จัดเรียงข้อสอบจากง่ายไปหายาก และกลุ่มที่สอง จำนวน 505 คน ทำแบบทดสอบ ฉบับที่ 2 ที่จัดเรียงข้อสอบแบบสุ่ม

ตัวอย่างข้อสอบ

คำชี้แจง แบบทดสอบชุดนี้เป็นแบบทดสอบวัดความสามารถด้านตัวเลขให้นักเรียนพิจารณาว่า ตัวเลขถัดไปคือข้อใด จาก ข้อ ก. ข. ค. ง. โดยใช้วิธีการทางคณิตศาสตร์ คือ บวก ลบ คูณ และหาร

1) 3 5 7 9 □ (วิธีบวก)

ก. 10

ข. 11

ค. 12

ง. 13

เฉลย ข้อ ข. 11

2) 25 22 19 16 □ (วิธีลบ)

ก. 11

ข. 12

ค. 13

ง. 14

เฉลย ข้อ ค. 13

3) 2 4 6 8 10 □ (วิธีคูณ)

ก. 12

ข. 14

ค. 16

ง. 18

เฉลย ข้อ ก. 12

4) 24 12 6 □ (วิธีหาร)

ก. 4

ข. 3

ค. 2

ง. 1

เฉลย ข้อ ข. 3

<u>วิธีดำเนินการเก็บรวบรวมข้อมูล</u>

ในการเก็บรวบรวมข้อมูล ผู้วิจัยดำเนินการตามขั้นตอน ดังนี้

1. นำแบบทดสอบไปดำเนินการสอบนักเรียนในโรงเรียนที่ใช้เป็นกลุ่มตัวอย่าง โดยขออนุญาตผู้บริหารโรงเรียนไว้แล้ว

2. ดำเนินการสอบ โดยอธิบายให้ผู้สอบทุกคนเข้าใจถึงวัตถุประสงค์ และผลประโยชน์ที่ได้รับจากการทำแบบทดสอบ

3. นำกระดาษคำตอบมาตรวจให้คะแนน ตอบถูกให้ 1 คะแนน ตอบผิดให้ 0 คะแนน

4. นำคะแนนที่ได้มาวิเคราะห์ทางสถิติ

<u>สถิติที่ใช้ในการวิเคราะห์ข้อมูล</u>

1. สถิติพื้นฐาน คือ

1.1 ค่าเฉลี่ย

1.2 ค่าเฉลี่ยร้อยละ

1.3 ส่วนเบี่ยงเบนมาตรฐาน

2. สถิติวิเคราะห์ข้อสอบเป็นรายข้อ โดยใช้เทคนิค 27 เปอร์เซ็นต์ และตารางสำเร็จรูปของ จุง เตห์ ฟาน (Chung-Teh-Fan. 1952) เพื่อหาค่าความยากง่าย (p) และค่าอำนาจจำแนก (r)

3. วิเคราะห์องค์ประกอบ เพื่อตรวจสอบความเป็นมิติเดี่ยว (Unidimension) โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป SPSS

4. หาค่าความเชื่อมั่นของแบบทดสอบ โดยใช้สูตร *KR* - 21 (บุญเชิด ภิญโญอนันตพงษ์ . 2523 : 256 ; อ้างอิงมาจาก Kuder – Richardson. 1937) โดยมีสูตรดังนี้

$$r_{tt} = \frac{n}{n-1}\left[1 - \frac{\bar{X}(n-\bar{X})}{ns_t^2}\right]$$

เมื่อ r_{tt} แทน ค่าความเชื่อมั่นของแบบทดสอบ

n แทน จำนวนข้อสอบของแบบทดสอบ

$\bar{X}$ แทน คะแนนเฉลี่ยของแบบทดสอบ

s_t^2 แทน คะแนนความแปรปรวนของแบบทดสอบ

5. หาค่าความเชื่อมั่น โดยใช้สูตร r_B แบบคะแนนจริงสัมพันธ์ (บุญเชิด ภิญโญอนันตพงษ์. 2538) มีสูตรดังนี้

$$r_B = \frac{1}{1-\sum \lambda_i^2}\left[1 - \frac{\sum p_i(1-p_i)}{S_x^2}\right]$$

เมื่อ p_i แทน ค่าความยากง่ายของข้อสอบแต่ละข้อ

S_x^2 แทน ความแปรปรวนของคะแนนทั้งฉบับ

$$\lambda_1 = \left[\sum B_{1i} - p_1(\bar{X}-1)\right]/S_x^2$$
$$\lambda_2 = \left[\sum B_{2i} - p_2(\bar{X}-1)\right]/S_x^2$$
$$\ldots\ldots$$
$$\lambda_k = \left[\sum B_{ki} - p_k(\bar{X}-1)\right]/S_x^2$$
$$\bar{X} = \sum p_i$$

B แทน สัดส่วนจำนวนผู้ตอบข้อสอบสองข้อใด ๆ ถูกทั้งคู่

6. วิเคราะห์ค่าความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measurement)

6.1 วิธีของธอร์นไดค์ (Feldt, Steffen and Gupta. 1985; citing Thorndike. 1951)

$$S_E = \left(S_{E_1}^2 + S_{E_2}^2\right)$$

เมื่อ S_E แทน ค่าความคลาดเคลื่อนมาตรฐานของแบบทดสอบทั้งฉบับ

$S_{E_1}^2$ แทน ค่าความคลาดเคลื่อนมาตรฐานของครึ่งฉบับแรก

$S_{E_2}^2$ แทน ค่าความคลาดเคลื่อนมาตรฐานของครึ่งฉบับหลัง

6.2 วิธีของ ลอร์ด (Lord.1984) จากสูตรความน่าจะเป็นทวินามคูณด้วย (1-*K*) โดยหาค่า *K* ดังนี้

$$K = \frac{n(n-1)s_p^2}{\bar{X}(n-\bar{X}) - s_x^2 - ns_p^2}$$

เมื่อ n แทน จำนวนข้อคำถาม

$\bar{X}$ แทน คะแนนเฉลี่ยของแบบทดสอบทั้งฉบับ

s_x^2 แทน ค่าความแปรปรวนของแบบทดสอบทั้งฉบับ

s_p^2 แทน ค่าความแปรปรวนของชุดข้อคำถามที่ตอบถูกต้อง

แล้วนำค่า K ไปแทนค่าในสูตร

$$S_E = \left[\frac{X(k-X)}{k-1}(1-K)\right]^{1/2}$$

เมื่อ S_E แทนค่าความคลาดเคลื่อนมาตรฐาน

X แทนค่าระดับคะแนนเฉพาะ

k แทนจำนวนข้อคำถาม

6.3 วิธีของ คีตส์ (Keats.1957)

$$S_E = \left[\left(\frac{X(k-X)}{k-1}\right)\left(\frac{1-r_{xx'}}{1-r_{21}}\right)\right]^{1/2}$$

S_E แทน ค่าความคลาดเคลื่อนมาตรฐาน

$r_{xx'}$ แทน ค่าความเชื่อมั่นของแบบทดสอบที่คำนวณด้วยสัมประสิทธิ์ r_B

r_{21} แทน ความเชื่อมั่นของแบบทดสอบที่คำนวณด้วนสูตร *KR* - 21

7. ลงจุดเส้นภาพเพื่อแสดงแนวโน้มของค่าความคลาดเคลื่อนมาตรฐานในการวัด

บทที่4

ผลการวิเคราะห์ข้อมูล

<u>สัญลักษณ์ และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล</u>

ในการวิเคราะห์ข้อมูล และการแปลความหมายของผลการวิเคราะห์ข้อมูล ผู้วิจัยได้กำหนดสัญลักษณ์ และอักษรย่อ ดังนี้

k	แทน	จำนวนข้อในแบบทดสอบ
N	แทน	จำนวนนักเรียนในกลุ่มตัวอย่าง
$\bar{X}$	แทน	ค่าเฉลี่ยของคะแนนแบบทดสอบ
$\bar{X}(\%)$	แทน	ค่าเฉลี่ยคิดเป็นร้อยละของคะแนนแบบทดสอบ
SD	แทน	ค่าเบี่ยงเบนมาตรฐาน
KR-21	แทน	ค่าความเชื่อมั่นที่คำนวณจากสูตร KR-21
r_B	แทน	ค่าความเชื่อมั่นที่คำนวณจากสูตร สัมประสิทธิ์ r_B
SEM	แทน	ค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานในการวัด

<u>การเสนอผลการวิเคราะห์ข้อมูล</u>

การศึกษาค้นคว้าครั้งนี้ ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลตามลำดับขั้นดังนี้

1. ค่าสถิติพื้นฐานของแบบทดสอบ ทั้ง 2 ฉบับ

2. ค่าความเชื่อมั่นของแบบทดสอบทั้ง 2 ฉบับ

3. ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงจากง่ายไปหายากกับการจัดเรียงแบบสุ่ม โดยใช้วิธีคำนวณ 3 วิธี คือ วิธีของธอร์นไดค์ ลอร์ด และคีตส์

4. การเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ โดยใช้วิธีการวิเคราะห์เส้นภาพ

ผลการวิเคราะห์ข้อมูล

1. ค่าสถิติพื้นฐานของแบบทดสอบ

การวิเคราะห์ข้อมูลครั้งนี้ ผู้วิจัยได้นำแบบทดสอบทั้ง 2 ฉบับ ที่จัดเรียงข้อสอบจากง่ายไปหายาก กับการจัดเรียงข้อสอบแบบสุ่ม จำนวน 40 ข้อ ไปทดสอบกับกลุ่มตัวอย่างที่มีจำนวน 514 คน และ 505 คน ตามลำดับ และได้นำคะแนนผลการสอบมาคำนวณหาค่าสถิติพื้นฐาน ปรากฏผลดังแสดงในตาราง 2

ตาราง 2 ค่าสถิติพื้นฐานของแบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายาก และจัดเรียงข้อสอบแบบสุ่ม

ค่าสถิติพื้นฐาน	ฉบับที่จัดเรียงข้อง่ายไปยาก	ฉบับที่จัดเรียงข้อแบบสุ่ม
k	40	40
$\overline{X}$	20.480	19.028
$\overline{X}(\%)$	51.20	47.57
SD	8.890	8.535
N	514	505

ผลการวิเคราะห์ตามตาราง 2 คะแนนที่ได้จากแบบทดสอบ 2 ฉบับ จำนวน 40 ข้อ ที่จัดเรียงต่างกัน คือฉบับที่เรียงจากง่ายไปหายาก และฉบับที่เรียงแบบสุ่ม ได้ค่าเฉลี่ย 20.480 และ 19.028 ตามลำดับ คิดเป็นร้อยละ 51.20 และ 47.57 ตามลำดับ ส่วนค่าเบี่ยงเบนมาตรฐาน เท่ากับ 8.890 และ 8.535 ตามลำดับ แสดงว่าแบบทดสอบที่จัดเรียงข้อจากง่ายไปหายากโดยเฉลี่ยแล้วมีแนวโน้มง่ายกว่า แบบทดสอบที่จัดเรียงข้อแบบสุ่มเล็กน้อย แต่มีความกระจายของคะแนนใกล้เคียงกัน

2. ค่าความเชื่อมั่นของแบบทดสอบทั้ง 2 ฉบับ

ในการวิเคราะห์ข้อมูลครั้งนี้ ผู้วิจัยได้คำนวณค่าความเชื่อมั่นโดยใช้สูตร 2 สูตร คือ *KR – 21* และสูตรสัมประสิทธิ์ r_B เพื่อใช้ในการคำนวณหาค่าความคลาดเคลื่อนมาตรฐานในการวัดตามวิธีของ คีตส์ ปรากฏผลดังแสดงในตาราง 3

ตาราง 3 ค่าความเชื่อมั่นของแบบทดสอบที่จัดเรียงจากง่ายไปหายาก และการจัดเรียงแบบสุ่ม ที่คำนวณจากสูตร 2 สูตร

สูตร	ค่าความเชื่อมั่นของแบบทดสอบ	
	ฉบับที่จัดเรียงข้อง่ายไปยาก	ฉบับที่จัดเรียงข้อแบบสุ่ม
KR – 21	0.8960	0.8852
r_B	0.9070	0.8935

ผลการวิเคราะห์ตามตาราง 3 พบว่าค่าความเชื่อมั่นของแบทดสอบ 2 ฉบับ คือ ฉบับที่จัดเรียงข้อสอบง่ายไปหายาก และฉบับที่จัดเรียงข้อสอบแบบสุ่ม ที่คำนวณจากสูตร *KR* – 21 ได้ค่า 0.8960 และ 0.8852 ตามลำดับ ส่วนที่คำนวณจากสูตรสัมประสิทธิ์ r_B ได้ค่า 0.9070 และ 0.8935 ตามลำดับ แสดงว่าค่าความเชื่อมั่นของแบบทดสอบทั้ง 2 ฉบับ มีค่าสูงโดย ฉบับที่จัดเรียงข้อสอบจากง่ายไปหายากมีค่าสูงกว่าฉบับที่จัดเรียงข้อสอบแบบสุ่มเล็กน้อย

3. ค่าความคลาดเคลื่อนมาตรฐานในการวัด

ในการวิเคราะห์ข้อมูลครั้งนี้ เป็นการวิเคราะห์ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียง 2 รูปแบบ คือ จัดเรียงข้อสอบจากง่ายไปหายาก กับจัดเรียงข้อสอบแบบสุ่ม โดยใช้วิธีคำนวณ 3 วิธี คือวิธีของธอร์นไดค์ ลอร์ด และคีตส์ ปรากฏผลดังนี้

3.1 ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายากโดยใช้วิธีคำนวณ 3 วิธี ปรากฏผลดังแสดงในตาราง 4

ตาราง 4 ค่าความคลาดเคลื่อนมาตรฐานในการวัด ที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงข้อจากง่ายไปหายาก

ระดับคะแนนเฉพาะ	N	ค่าความคลาดเคลื่อนมาตรฐานในการวัด		
		วิธีของธอร์นไดค์	วิธีของลอร์ด	วิธีของคีตส์
5-7	12	2.2900	2.1844	2.1622
8-10	51	2.6260	2.5546	2.5286
11-13	70	2.6400	2.8035	2.7749
14-16	79	2.5390	2.9617	2.9316
17-19	56	2.7620	3.0435	3.0125
20-22	49	3.1600	3.0550	3.0239
23-25	50	2.5430	2.9970	2.9665
26-28	30	2.1090	2.8653	2.8362
29-31	34	1.9960	2.6490	2.6221
32-34	41	2.0010	2.3245	2.3008
35-37	24	1.5300	1.8353	1.8166
38-40	18	1.0980	0.9551	0.9454
SEM	**514**	**2.2745**	**2.5191**	**2.4934**

ผลการวิเคราะห์ตามตาราง 4 พบว่า ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่คำนวณโดยวิธีของ ธอร์นไดค์ ลอร์ด และคีตส์ ได้ค่าอยู่ระหว่าง 1.0980 – 3.1600, 0.9551 – 3.0550 และ 0.9454 – 3.0239 ตามลำดับ โดยวิธีของธอร์นไดค์ ได้ค่าสูงที่สุด และวิธีของคีตส์ ได้ค่าต่ำที่สุด

3.2 ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงข้อสอบแบบสุ่ม โดยใช้วิธีคำนวณ 3 วิธี ปรากฏผลดังแสดงในตาราง 5

ตาราง 5 ค่าความคลาดเคลื่อนมาตรฐานในการวัด ที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงแบบสุ่ม

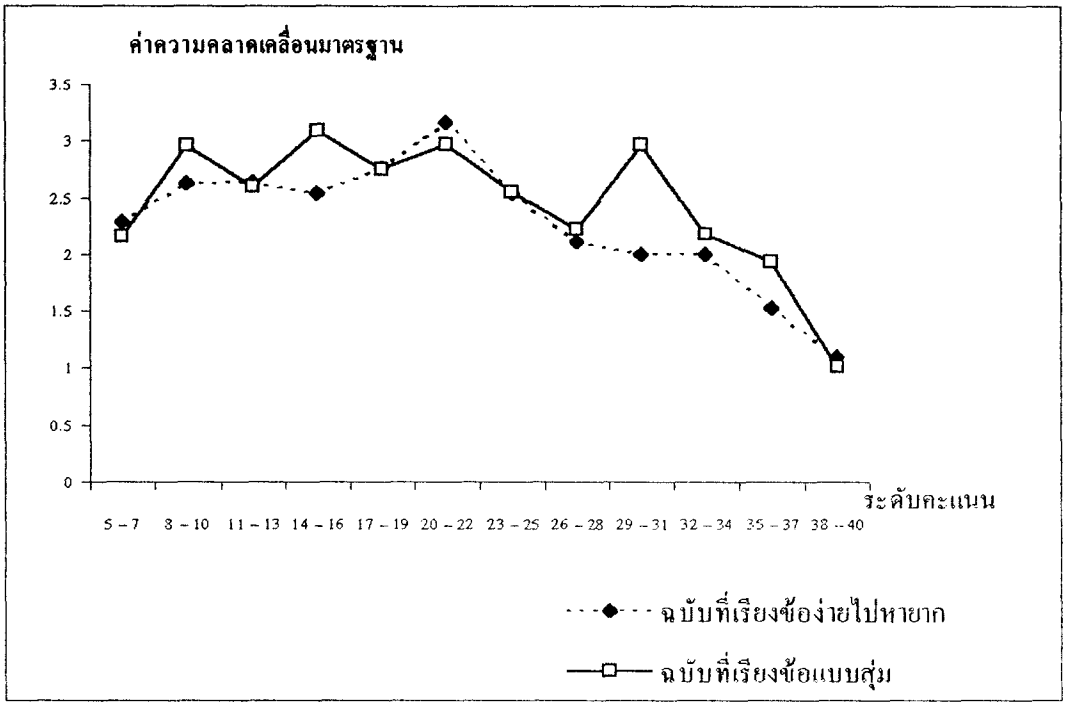
ระดับคะแนนเฉพาะ	N	ค่าความคลาดเคลื่อนมาตรฐานในการวัด		
		วิธีของธอร์นไดค์	วิธีของลอร์ด	วิธีของคีตส์
5 – 7	15	2.1540	2.2147	2.2025
8 – 10	59	2.9560	2.5900	2.5758
11 – 13	92	2.6010	2.8432	2.8267
14 – 16	81	3.0840	3.0027	2.9862
17 – 19	53	2.7500	3.0856	3.0687
20 – 22	51	2.9620	3.0973	3.0803
23 – 25	39	2.5560	3.0385	3.0218
26 – 28	23	2.2200	2.9050	2.8891
29 – 31	27	2.9650	2.6857	2.6709
32 – 34	30	2.1770	2.3567	2.3437
35 – 37	26	1.9370	1.8607	1.8505
38 – 40	9	1.0140	0.9683	0.9630
SEM	**505**	**2.4506**	**2.5540**	**2.5399**

ผลการวิเคราะห์ตามตาราง 5 พบว่า ค่าความคลาดเคลื่อนมาตรฐานในการวัดของแบบทดสอบที่จัดเรียงข้อแบบสุ่ม ซึ่งคำนวณโดยวิธีของธอร์นไดค์ ลอร์ด และคีตส์ ได้ค่าอยู่ระหว่าง 1.0140 – 3.0840, 0.9683 – 3.0973 และ 0.9630 – 3.0803 ตามลำดับ โดยวิธีของ ลอร์ด ได้ค่าสูงที่สุด และวิธีของคีตส์ ได้ค่าต่ำที่สุด

4. การเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบ 2 ฉบับ ที่คำนวณโดยวิธี 3 วิธี

ในการเปรียบเทียบครั้งนี้ ผู้วิจัยได้นำค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะมาลงจุดเป็นเส้นภาพ โดยแยกเป็นการเปรียบเทียบระหว่างการจัดเรียงข้อสอบ 2 แบบ ดังแสดงในภาพ 3 - 5 และการเปรียบเทียบระหว่างวิธีที่คำนวณ 3 วิธี ดังแสดงในภาพ 6 - 7

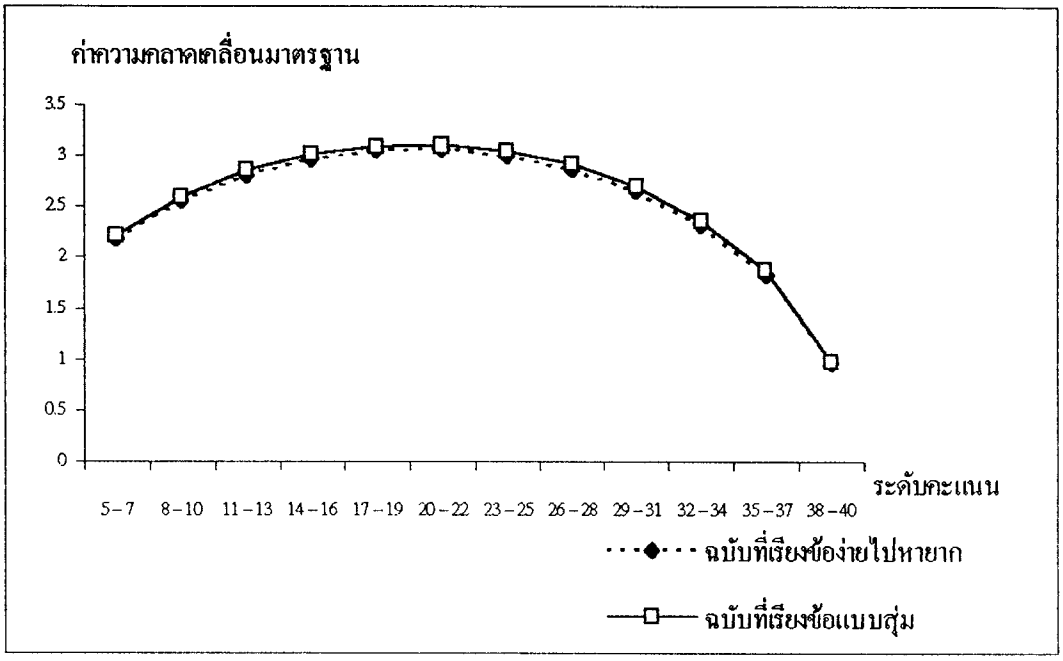
วิธีของธอร์นไดค์



ภาพประกอบ 3 การแสดงเส้นภาพค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายาก และจัดเรียงข้อสอบแบบสุ่ม ที่คำนวณด้วยวิธีของธอร์นไดค์

ผลการวิเคราะห์จากภาพประกอบ 3 พบว่า ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายาก มีค่าสูงสุด เท่ากับ 3.1600 และฉบับที่จัดเรียงข้อสอบแบบสุ่ม เท่ากับ 2.9620 ซึ่งทั้ง 2 ฉบับค่าสูงสุดจะอยู่ช่วงกลางของระดับคะแนน 20 – 22 ส่วนค่าที่ต่ำสุดจะอยู่ในช่วงปลายของระดับคะแนน 38 – 40 ซึ่งฉบับที่จัดเรียงข้อสอบแบบสุ่มจะได้ค่าต่ำที่สุด เท่ากับ 1.0140 และฉบับที่เรียงจากง่ายไปหายากได้ค่าต่ำสุดเท่ากับ 1.0980 และค่าความคลาดเคลื่อนมาตรฐานในการวัด ของฉบับที่เรียงข้อสอบจากง่ายไปหายากจะได้ค่าต่ำกว่าฉบับที่เรียงข้อสอบแบบสุ่มในทุกระดับคะแนน ยกเว้นช่วงกลางของระดับคะแนน

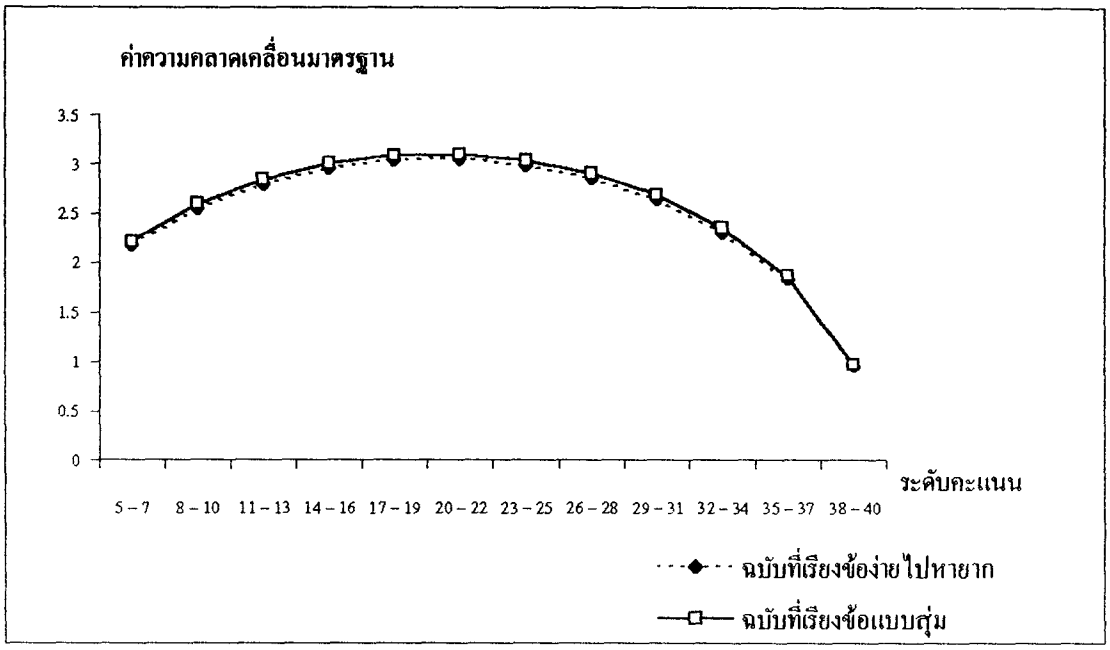
วิธีของลอร์ด



ภาพประกอบ 4 การแสดงเส้นภาพค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ ของแบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายาก และจัดเรียงข้อสอบแบบสุ่ม ที่คำนวณด้วยวิธีของลอร์ด

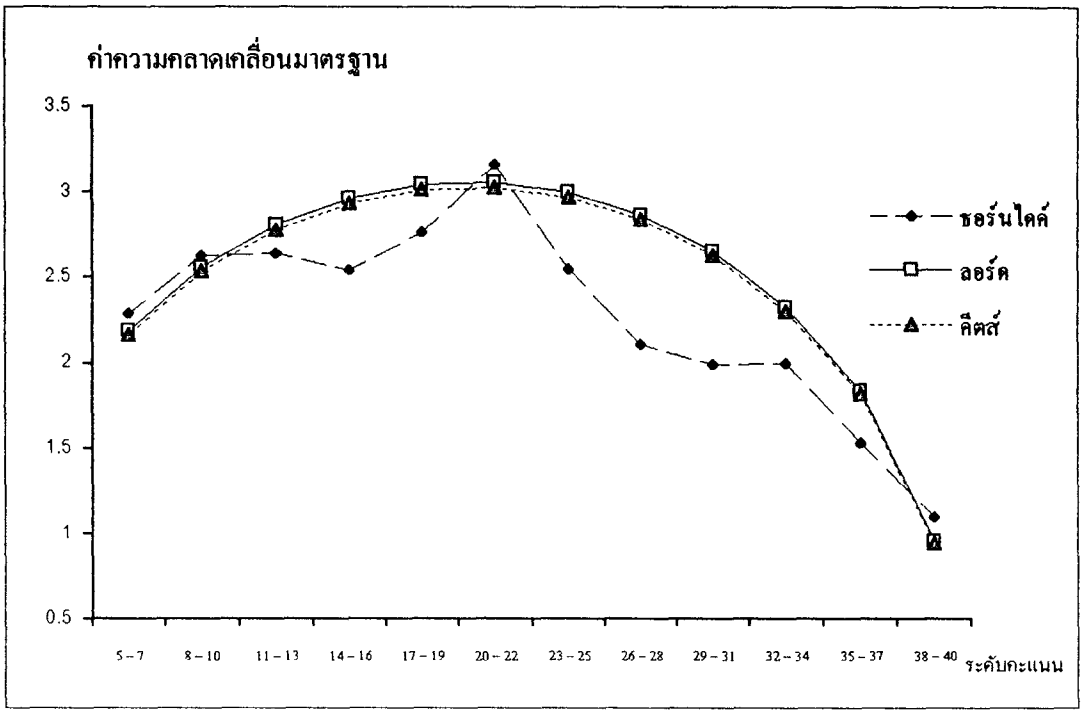
ผลการวิเคราะห์จากภาพประกอบ 4 พบว่า ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ ของแบบทดสอบฉบับที่จัดเรียงข้อสอบแบบสุ่มมีค่าสูงสุด เท่ากับ 3.0973 และฉบับที่จัดเรียงข้อสอบจากง่ายไปหายาก เท่ากับ 3.05550 ซึ่งทั้ง 2 ฉบับค่าสูงสุดจะอยู่ในช่วงกลางของระดับคะแนน 20-22 ส่วนค่าที่ต่ำที่สุดจะอยู่ในช่วงปลายของระดับคะแนน 38 – 40 ซึ่งฉบับที่จัดเรียงข้อสอบจากง่ายไปหายากจะได้ค่าต่ำสุด เท่ากับ 0.9551 และฉบับที่เรียงแบบสุ่มเท่ากับ 0.9683 และค่าความคลาดเคลื่อนมาตรฐานในการวัด ของแบบทดสอบที่เรียงจากง่ายไปหายากมีค่าต่ำกว่าแบบทดสอบที่เรียงแบบสุ่มเล็กน้อยในทุกระดับคะแนน

วิธีของคีตส์



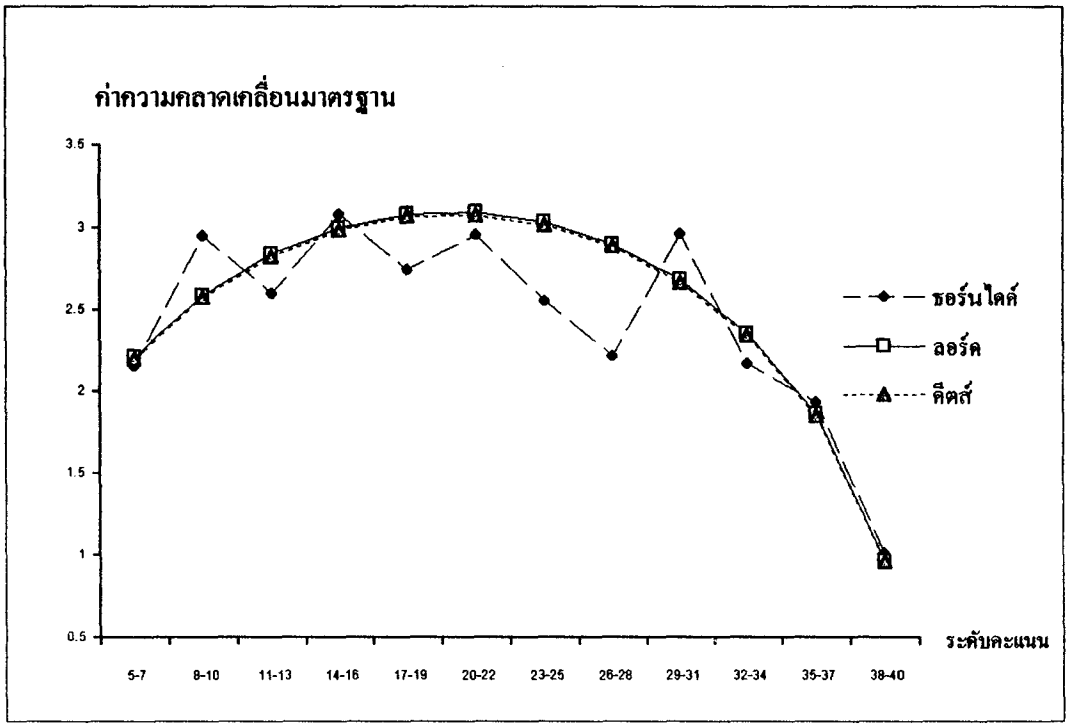
ภาพประกอบ 5 การแสดงเส้นภาพค่าความคลาดเคลื่อนมาตรฐานในการวัด ที่ระดับคะแนนเฉพาะ ของแบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายาก และจัดเรียงข้อสอบแบบสุ่ม ที่คำนวณด้วยวิธีของคีตส์

ผลการวิเคราะห์จากภาพประกอบ 5 พบว่า ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ ของแบบทดสอบฉบับที่จัดเรียงข้อสอบแบบสุ่มมีค่าสูงสุด เท่ากับ 3.0803 ส่วนฉบับที่จัดเรียงข้อสอบจากง่ายไปหายาก เท่ากับ 3.0239 ซึ่งทั้ง 2 ฉบับ ค่าสูงสุดจะอยู่ในช่วงกลางของระดับคะแนน 20 – 22 ส่วนค่าที่ต่ำที่สุดจะอยู่ช่วงปลายของระดับคะแนน 38 – 40 ซึ่งฉบับที่จัดเรียงข้อสอบจากง่ายไปหายากจะมีค่าต่ำสุดเท่ากับ 0.9454 และฉบับที่เรียงข้อสอบแบบสุ่มเท่ากับ 0.9630 ค่าความคลาดเคลื่อนมาตรฐานในการวัด ของแบบทดสอบที่เรียงจากง่ายไปหายากมีค่าต่ำกว่าฉบับที่เรียงแบบสุ่มเล็กน้อย ในทุกระดับคะแนน



ภาพประกอบ 6 การแสดงเส้นภาพค่าความคลาดเคลื่อนมาตรฐานในการวัด ที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงจากง่ายไปหายาก ที่คำนวณด้วยวิธีของธอร์นไดค์ ลอร์ด และคีตส์

ผลการวิเคราะห์จากภาพประกอบ 6 พบว่า ค่าความคลาดเคลื่อนมาตรฐาน-ในการวัดที่ระดับคะแนนเฉพาะ เมื่อเปรียบเทียบกัน 3 วิธี วิธีของลอร์ด และคีตส์ ให้ค่าใกล้เคียงกันมากที่สุด คือเป็นรูปพาราโบล่าคว่ำ ส่วนวิธีของ ธอร์นไดค์ ได้ค่าต่างไปจากสองวิธีแรก โดยมีค่าสูงต่ำสลับกันไป และมีแนวโน้มเป็นรูปพาราโบล่าคว่ำเช่นเดียวกัน ซึ่งวิธีของธอร์นไดค์ ให้ค่าสูงที่สุด เท่ากับ 3.1600 วิธีของลอร์ด ได้เท่ากับ 3.0550 และวิธีของคีตส์ เท่ากับ 3.0239 ตามลำดับ ทั้ง 3 วิธี จะให้ค่าความคลาดเคลื่อนมาตรฐาน-ในการวัดสูงที่สุดที่ช่วงคะแนนตรงกลาง คือ ที่ระดับคะแนน 20 – 22 และให้ค่าความคลาดเคลื่อนมาตรฐานในการวัดต่ำที่ช่วงปลายของคะแนนด้านมาก คือ ที่ระดับคะแนน 38 – 40 ซึ่งวิธีของคีตส์ให้ค่าต่ำที่สุดเท่ากับ .9454 วิธีของลอร์ด เท่ากับ .9551 และวิธีของธอร์นไดค์ เท่ากับ 1.0980 ตามลำดับ



ภาพประกอบ 7 การแสดงเส้นภาพค่าความคลาดเคลื่อนมาตรฐานในการวัด ที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงแบบสุ่ม ที่คำนวณด้วยวิธีของธอร์นไดค์ ลอร์ด และคีตส์

ผลการวิเคราะห์จากภาพประกอบ 7 พบว่า ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ เมื่อเปรียบเทียบกัน 3 วิธี วิธีของ ลอร์ดและคีตส์ ให้ค่าใกล้เคียงกันมากที่สุด เป็นรูปพาราโบล่าคว่ำ ส่วนวิธีของธอร์นไดค์ ได้ค่าต่างไปจากสองวิธีแรก คือ มีค่าสูงต่ำสลับกันไป แต่มีแนวโน้มเป็นรูปพาราโบล่าคว่ำเช่นเดียวกัน ซึ่งวิธีของลอร์ดให้ค่าสูงที่สุด เท่ากับ 3.0973 วิธีของธอร์นไดค์ เท่ากับ 3.0840 และวิธีของคีตส์ เท่ากับ 3.0803 ตามลำดับ ทั้ง 3 วิธี จะให้ค่าความคลาดเคลื่อนมาตรฐานในการวัดสูงในช่วงกลางของระดับคะแนน คือ ที่ระดับคะแนน 2.0–22 ในวิธีของลอร์ด และคีตส์ ส่วนวิธีของธอร์นไดค์ จะให้ค่าสูงที่ระดับคะแนน 14–16 และทั้ง 3 วิธี ให้ค่าต่ำที่ช่วงปลายของคะแนนด้านมาก คือ ที่ระดับคะแนน 38 – 40 ซึ่งวิธีของคีตส์ให้ค่าต่ำที่สุด เท่ากับ .9630 วิธีของลอร์ด เท่ากับ .9683 และวิธีของธอร์นไดค์ เท่ากับ 1.0140 ตามลำดับ

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

ความมุ่งหมายของการศึกษาค้นคว้า

การศึกษาค้นคว้าครั้งนี้ เป็นการศึกษาผลของการจัดเรียงข้อสอบที่มีต่อค่าความคลาดเคลื่อนมาตรฐานของการวัดที่ระดับคะแนนเฉพาะ โดยมีจุดมุ่งหมายดังนี้

1. เพื่อศึกษาค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงจากง่ายไปหายากกับการจัดเรียงแบบสุ่ม

2. เพื่อเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ ของวิธีคำนวณ 3 วิธี คือ

วิธีของ ธอร์นไดค์ (Thorndike. 1951)

วิธีของ ลอร์ด (Lord. 1984)

วิธีของ คีตส์ (Keats. 1957)

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2540 ของโรงเรียนในสังกัดสำนักงานการประถมศึกษาจังหวัดสระบุรี จำนวน 1,019 คน

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นแบบทดสอบวัดความสามารถด้านตัวเลขแบบอนุกรมธรรมดา ซึ่งผู้วิจัยสร้างขึ้นเองเป็นแบบปรนัย ชนิดเลือกตอบ 4 ตัวเลือก จำนวน 40 ข้อ โดยจัดเรียงข้อคำถาม 2 รูปแบบ คือ ฉบับที่ 1 จัดเรียงข้อคำถามจากง่ายไปหายากฉบับที่ 2 จัดเรียงข้อคำถามแบบสุ่ม

วิธีดำเนินการเก็บรวบรวมข้อมูล

ในการเก็บรวบรวมข้อมูล ผู้วิจัยดำเนินการตามขั้นตอนดังนี้

1. นำแบบทดสอบไปดำเนินการสอบนักเรียนในโรงเรียนที่เป็นกลุ่มตัวอย่าง โดยขออนุญาตผู้บริหารโรงเรียนไว้ล่วงหน้า

2. ดำเนินการสอบ โดยอธิบายให้ผู้เข้าสอบทุกคนเข้าใจถึงวัตถุประสงค์ และผลประโยชน์ที่ได้รับจากการทำแบบทดสอบ

3. นำกระดาษคำตอบมาตรวจให้คะแนน คือ ตอบถูกให้ 1 คะแนนตอบผิดให้ 0 คะแนน

4. นำคะแนนที่ได้มาวิเคราะห์ทางสถิติ

การเสนอผลการวิเคราะห์ข้อมูล

ในการศึกษาค้นคว้าครั้งนี้ ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลตามลำดับขั้นดังนี้

1. ค่าสถิติพื้นฐานของแบบทดสอบทั้ง 2 ฉบับ

2. ค่าความเชื่อมั่นของแบบทดสอบทั้ง 2 ฉบับ

3. ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบทั้ง 2 ฉบับ

4. เปรียบเทียบความแตกต่างโดยเสนอเป็นเส้นภาพ (Profile)

สรุปผลการวิจัย

ผลการวิจัยสรุปได้ดังนี้

1. แบบทดสอบ

ค่าสถิติพื้นฐานของแบบทดสอบทั้ง 2 ฉบับ คือแบบจัดเรียงข้อคำถามจากง่ายไปหายาก และการเรียงแบบสุ่ม ฉบับละ 40 ข้อ ได้ค่าเฉลี่ย 20.480 และ 19.028 หรือคิดเป็นร้อยละ 51.20 และ 47.57 ตามลำดับ ส่วนค่าเบี่ยงเบนมาตรฐานมีค่าเท่ากับ 8.890 และ 8.535 ตามลำดับ

2. <u>ค่าความเชื่อมั่นของแบบทดสอบ</u>

เนื่องจากการคำนวณค่าความคลาดเคลื่อนมาตรฐานในการวัดโดยวิธีของคีตส์ ต้องใช้ค่าความเชื่อมั่น 2 สูตร ผู้วิจัยจึงนำคะแนนผลการสอบของกลุ่มตัวอย่างจากแบบทดสอบ 2 ฉบับ ไปคำนวณหาค่าความเชื่อมั่นโดยใช้สูตร $KR - 21$ และ สูตรสัมประสิทธิ์ r_B ได้ค่าความเชื่อมั่น ฉบับที่ 1 เท่ากับ 0.8960 และ 0.9070 ฉบับที่ 2 เท่ากับ 0.8852 และ 08935 ตามลำดับ

3. <u>ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ</u>

จากการนำคะแนนผลการสอบจากกลุ่มตัวอย่าง ของแบบทดสอบ 2 ฉบับ โดยแบ่งเป็นช่วงคะแนน (ระดับคะแนนเฉพาะ) แล้วคำนวณหาค่าความคลาดเคลื่อน-มาตรฐานในการวัด โดยใช้วิธีคำนวณ 3 วิธี คือ ของ ธอร์นไดค์ ลอร์ดและคีตส์ ได้ผลดังนี้

3.1 ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ ของแบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายาก จำนวน 40 ข้อ โดยใช้ วิธีคำนวณ 3 วิธี พบว่า วิธีของธอร์นไดค์ ได้ค่าอยู่ระหว่าง 1.0980 – 3.1600 วิธีของลอร์ด ได้ค่า 0.9351 – 3.0550 และวิธีของ คีตส์ ได้ค่า 0.9454 – 3.0239 โดยวิธีของ ธอร์นไดค์ จะได้ค่าสูงที่สุด และวิธีของคีตส์ จะให้ค่าที่ต่ำที่สุด

3.2 ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ ของแบบทดสอบที่จัดเรียงข้อสอบแบบสุ่ม จำนวน 40 ข้อ โดยใช้วิธีคำนวณ 3 วิธี พบว่า วิธีของธอร์นไดค์ ได้ค่าอยู่ระหว่าง 1.0140 – 3.0840 วิธีของลอร์ด ได้ค่า 0.9683 – 3.0973 และวิธีของคีตส์ ได้ค่า 0.9630 – 3.0803 ซึ่งวิธีของลอร์ด จะได้ค่าสูงที่สุด และวิธีของคีตส์ จะได้ค่าต่ำที่สุด

3.3 ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ ของแบบทดสอบทั้ง 2 ฉบับ ที่ใช้วิธีคำนวณ 3 วิธี พบว่า การคำนวณโดยวิธีของลอร์ด และคีตส์ ได้ค่าความคลาดเคลื่อนมาตรฐานในการวัดของแบบทดสอบที่จัดเรียงจากง่ายไปหายาก มีค่าต่ำกว่าแบบทดสอบที่จัดเรียงแบบสุ่ม ในทุกระดับคะแนน แต่เมื่อใช้วิธีของธอร์นไดค์ พบว่าค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ได้จากแบบทดสอบที่เรียง

จากง่ายไปหายากมีค่าต่ำกว่า และสูงกว่าแบบทดสอบที่เรียงแบบสุ่มในบางระดับคะแนน แต่เมื่อวิเคราะห์จากค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานในการวัดพบว่า จากการคำนวณทั้ง 3 วิธี ฉบับที่เรียงจากง่ายไปหายาก จะมีค่าต่ำกว่า ฉบับที่เรียงแบบสุ่ม

4. <u>การเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ โดยวิธีคำนวณ 3 วิธี ได้ผลดังนี้</u>

4.1 ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายาก และจัดเรียงข้อสอบแบบสุ่มที่คำนวณด้วยวิธีของธอร์นไดค์ได้ค่าที่แตกต่างกันในทุกระดับคะแนน ซึ่งฉบับที่จะเรียงข้อสอบจากง่ายไปหายาก จะมีค่าความคลาดเคลื่อนมาตรฐานในการวัดต่ำกว่าฉบับที่เรียงแบบสุ่มในทุกระดับคะแนน ยกเว้นช่วงกลางของระดับคะแนน

4.2 ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายาก และจัดเรียงข้อสอบแบบสุ่มที่คำนวณด้วยวิธีของลอร์ด ปรากฏว่าฉบับที่เรียงจากง่ายไปหายากมีค่าต่ำกว่าฉบับที่เรียงแบบสุ่มในทุกระดับคะแนน

4.3 ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงจากง่ายไปหายาก และจัดเรียงข้อสอบแบบสุ่มที่คำนวณด้วยวิธีของคีตส์ ปรากฏว่าฉบับที่เรียงจากง่ายไปหายากมีค่าต่ำกว่าฉบับที่เรียงแบบสุ่มในทุกระดับคะแนน

4.4 ค่าความคลาดเคลื่อนมาตรฐานในการวัดมีระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายาก โดยใช้วิธีคำนวณ 3 วิธี พบว่า วิธีของลอร์ด และคีตส์ได้ค่าที่มีความใกล้เคียงกันมาก ส่วนวิธีของธอร์นไดค์ได้ค่าที่มีลักษณะสูงต่ำสลับกันในบางระดับคะแนน แต่วิธีของคีตส์ได้ค่าต่ำที่สุด ซึ่งอยู่ในช่วงปลายของระดับคะแนน ส่วนวิธีของธอร์นไดค์ได้ค่าสูงที่สุด ซึ่งอยู่ในช่วงกลางของระดับคะแนน

4.5 ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงข้อคำถามแบบสุ่มโดยใช้วิธีคำนวณ 3 วิธี พบว่า วิธีของลอร์ดและคีตส์ ได้ค่าที่ใกล้เคียงกันมาก ส่วนวิธีของธอร์นไดค์ได้ค่าที่แตกต่างจาก2 วิธีแรก

อย่างเห็นได้ชัด โดยวิธีของคีตส์ได้ค่าต่ำที่สุดในช่วงปลายของคะแนน ส่วนวิธีของลอร์ดได้ค่าสูงที่สุดอยู่ในช่วงกลางของระดับคะแนน

อภิปรายผล

การวิจัยครั้งนี้ มีจุดมุ่งหมายสำคัญเพื่อค้นหาว่า แบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายาก กับการจัดเรียงข้อสอบแบบสุ่มนั้นมีผลต่อค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะหรือไม่ เมื่อคำนวณด้วยวิธี 3 วิธี คือวิธีของธอร์นไดค์ ลอร์ด และคีตส์ ซึ่งสามารถอภิปรายผลได้ดังนี้

ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงจากง่ายไปหายาก กับการจัดเรียงแบบสุ่มโดยใช้วิธีคำนวณ 3 วิธี มีความแตกต่างกัน

การจัดเรียงข้อสอบจากง่ายไปหายาก และการจัดเรียงแบบสุ่มของแบบทดสอบวัดความสามารถด้านตัวเลข แบบอนุกรมธรรมดา ชนิดเลือกตอบ 4 ตัวเลือก มีผลทำให้ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะมีความแตกต่างกัน โดยฉบับที่เรียงจากง่ายไปหายากจะได้ค่าที่ต่ำกว่าฉบับที่เรียงแบบสุ่ม จากการคำนวณทั้ง 3 วิธี ทั้งนี้อาจเนื่องมาจากการเรียงข้อสอบจากง่ายไปหายากมีส่วนทำให้นักเรียนเกิดความมั่นใจในการตอบ จึงเป็นการวัดความสามารถที่แท้จริงของนักเรียน ทำให้ค่าความคลาดเคลื่อนต่ำกว่าฉบับที่เรียงแบบสุ่ม ซึ่งตรงกับความเห็นของ ชวาล แพรัตกุล (2518 : 193) ที่ว่าแบบทดสอบแต่ละฉบับควรเรียงข้อสอบจากข้อง่ายไปยากเสมอ และการพบข้อง่าย ๆ ในตอนแรกก็จะไม่ทำให้นักเรียนเกิดความท้อใจอีกด้วย

ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่คำนวณ โดยวิธี 3 วิธี มีค่าที่แตกต่างกัน โดยวิธีของลอร์ด และคีตส์มีค่าความคลาดเคลื่อนมาตรฐานในการวัดที่สอดคล้องกันเป็นรูปโค้งกลมแบบพาราโบลาคว่ำ และวิธีของคีตส์ จะมีค่าต่ำที่สุดในช่วงปลายสุดของระดับคะแนน ซึ่งสอดคล้องกับผลการวิจัยของ เฟลดต์ สเตฟเฟน และกุปตา (Feldt, Steffen and Gupta, 1985) และผลการวิจัยของ บลิกต์ และ ชามา (Blixt and Chama, 1986) คือทั้ง 3 วิธี จะมีค่าคลาดเคลื่อนมาตรฐานในการวัดค่อย ๆ สูงขึ้นจะสูงที่สุดในช่วงกลาง

ของระดับคะแนน และจะต่ำลงเมื่ออยู่ช่วงปลายของคะแนน ซึ่งวิธีของธอร์นไดค์ จะมีค่าสูงที่สุด

ความแตกต่างที่เห็นชัดเจนมากที่สุดคือ วิธีของธอร์นไดค์ซึ่งค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ได้จะมีลักษณะสูง ๆ ต่ำ ๆ สลับกันในบางระดับคะแนน ซึ่งอาจจะเป็นเพราะวิธีการคำนวณซึ่งผันแปรไปตามจำนวนของผู้สอบที่ได้ระดับคะแนนเฉพาะและค่าเบี่ยงเบนมาตรฐานของผลต่างของคะแนนครึ่งฉบับ แต่อย่างไรก็ตามในช่วงของระดับคะแนนส่วนใหญ่ ทั้ง 3 วิธี จะมีค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ใกล้เคียงกัน และโดยผลสรุปค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานในการวัด จากทั้ง 3 วิธี มีค่าที่ใกล้เคียงกันมาก

ข้อสังเกตก็คือ ค่าความคลาดเคลื่อนมาตรฐานในการวัดสำหรับแต่ละคนขึ้นอยู่กับค่าสถิติ และหลักเหตุผลของคะแนนจริง (True Score) ของผู้สอบ ไม่ใช่คะแนนที่ได้จากการสอบหรือวัด (Observed Score) ดังนั้นในการพูดถึงค่าความคลาดเคลื่อนมาตรฐานในการวัด "ที่ระดับคะแนนเฉพาะ" จะหมายถึง คะแนนจริง ณ ระดับคะแนนที่ได้ ไม่ใช่คะแนนจากการสอบ

ข้อเสนอแนะ

ผลการวิจัยครั้งนี้ สามารถสรุปเป็นข้อเสนอแนะในการนำไปใช้ และในการทำวิจัยครั้งต่อไป ดังนี้

1. ข้อเสนอแนะในการนำไปใช้

1.1 จากการวิเคราะห์ผลของการจัดเรียงข้อสอบจากง่ายไปหายาก กับการจัดเรียงแบบสุ่ม แสดงให้เห็นว่าแบบทดสอบที่จัดเรียงจากง่ายไปหายาก ทำให้ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะต่ำกว่าแบบทดสอบที่จัดเรียงแบบสุ่ม ดังนั้นครูควรใช้วิธีการเรียงข้อสอบจากง่ายไปหายากในการสร้างแบบทดสอบชนิดเลือกตอบ แบบปรนัย

1.2 การนำวิธีการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดไปใช้ ควรเลือกวิธีของลอร์ด และคีตส์ เนื่องจากค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ได้มีความสอดคล้องกันมีแนวโน้ม และทิศทางที่ชัดเจน นอกจากนั้นวิธีของลอร์ด และคีตส์ยังเป็นวิธีที่ไม่ยุ่งยากซับซ้อนและสะดวกในการคำนวณ กลุ่มตัวอย่างก็ไม่จำเป็นต้องมีขนาดใหญ่มากด้วย

2. ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

2.1 ควรมีการศึกษาถึงค่าความยากของแบบทดสอบที่ส่งผลต่อค่าความคลาดเคลื่อนมาตรฐานในการวัด

2.2 ควรมีการศึกษาค่าความคลาดเคลื่อนมาตรฐานในการวัด โดยการประมาณค่าด้วยวิธีอื่น ๆ เช่น วิธีโพลิโนเมียล ของโมเลนคอปฟ์, วิธีวิเคราะห์ความแปรปรวน (ANOVA), วิธีที่ยึดหลักทฤษฎี IRT และวิธีการถดถอย (Regression)

2.3 ควรมีการศึกษาเพื่อเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานในการวัด โดยใช้วิธีการประมาณค่าเช่นเดียวกันนี้ แต่ใช้กับกลุ่มตัวอย่างที่มีขนาดใหญ่กว่านี้ คือตั้งแต่ 1,000 คนขึ้นไป

บรรณานุกรม

บรรณานุกรม

ชวาล แพรัตกุล. เทคนิคการวัดผล. พิมพ์ครั้งที่ 6. กรุงเทพฯ : วัฒนาพานิช,2518.

เตือนใจ เศรษฐสักโก. การศึกษาถึงการจัดระเบียบข้อสอบด้วยวิธีต่าง ๆ ที่จะส่งผลต่อความสามารถของนักเรียน ชั้น ป.7. ปริญญานิพนธ์ กศ.ม. กรุงเทพ ฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2511. อัดสำเนา.

แน่งน้อย เพชรรัตน์. ผลของการเรียงลำดับข้อคำถามด้วยวิธีต่างกัน ที่มีผลต่อคุณภาพของแบบทดสอบวิชาทักษะกระบวนการทางวิทยาศาสตร์. ปริญญานิพนธ์ กศ.ม. กรุงเทพ ฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2524. อัดสำเนา.

นิโลบล นิ่มกิ่งรัตน์. อิทธิพลของการจัดเรียงลำดับข้อคำถามที่จำแนกตามสมรรถภาพทางสมองที่มีต่อคุณภาพของแบบทดสอบคณิตศาสตร์ และภาษาไทย ม.3. ปริญญานิพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2519. อัดสำเนา.

บุญเชิด ภิญโญอนันตพงษ์. การวัด และประเมินผลการศึกษา : ทฤษฎีการวัด และประเมินผลการประยุกต์ใช้ . กรุงเทพ : อักษรเจริญทัศน์, 2523.

บุญเชิด ภิญโญอนันตพงษ์. "แบบทดสอบคะแนนจริงสัมพันธ์ : สัมประสิทธิ์ r_B ," วัดผลการศึกษา. 16(48) : 43-56 ; มกราคม – เมษายน 2538.

พิชญา เกตุสกุล. การเปรียบเทียบความเที่ยงตรง และความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบระหว่างแบบทดสอบแยกกลุ่มที่มีรูปแบบต่างกัน. ปริญญานิพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2539. อัดสำเนา.

รุ่งฤดี ทรัพย์นิธิ. การศึกษาเปรียบเทียบความคลาดเคลื่อนมาตรฐานในการประมาณค่าพารามิเตอร์ระหว่างการวิเคราะห์ด้วยโมเดลโลจิสติก แบบหนึ่ง สอง สาม พารามิเตอร์ และขนาดกลุ่มตัวอย่างต่างกัน. ปริญญานิพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2539. อัดสำเนา.

ล้วน สายยศ และอังคณา สายยศ. หลักการสร้างแบบทดสอบความถนัดทางการเรียน. พิมพ์ครั้งที่ 1. สำนักพิมพ์ วัฒนาพานิช, 2527.

วลีมาศ แซ่อึ้ง. การศึกษาความคลาดเคลื่อนมาตรฐาน และประสิทธิภาพสัมพันธ์ของการประมาณค่าเจตคติต่อวิชาคณิตศาสตร์ ระหว่างมาตรวัดแบบลิเคอร์ทกับมาตรวัดแบบออสกูด. ปริญญานิพนธ์ กศ.ม. กรุงเทพ ฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2537. อัดสำเนา.

ศศิธร สุวรรณสุข. ค่าความเที่ยงตรง ค่าความเชื่อมั่นของแบบทดสอบที่เรียงลำดับคำถามต่างกันในวิชาคณิตศาสตร์ ระดับชั้นประถมศึกษาปีที่ 7. ปริญญานิพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2519. อัดสำเนา.

อนันต์ ศรีโสภา. การวัดผลการศึกษา. พิมพ์ครั้งที่ 3. สำนักงานพิมพ์ไทยวัฒนาพานิช, 2525.

Blixt, S.L. and D.D. Shama. "An Empirical Investigation of the Standard Error of Measurement at Different Ability Levels," Educational and Psychological Measurement. 46 : 545-550; 1986.

Fan, Chung – Teh "Item Analysis Table," Educational Testing Service. Princeton, New Jersey : 1952.

Feldt, L.S., M. Steffen and N.C. Gupta. "A Comparison of Five Models for Estimating the Standard Error of Measurement at Specific Score Levels," Applied Psychological Measurement. 9 : 351-361; 1985.

Feldt, L.S. and A.L. Qualls. "Estimation of Measurement Error Variance at Specific Score Levels," Journal of Educational measurement. 33 : 141-156; 1996.

Hoyt, C. "Test Reliability Obtained by Analysis of Variance," Psychometrika. 6 : 153 – 160; 1941.

Keats, J.A. "Estimation of Error Variance of Test Scores," Psychometrika. 22 : 29 – 41; 1957.

Lord, F.M. "Estimating Test Reliability," Educational and Psychological Measurement. 15 : 325-336; 1955.

Lord, F.M. "Standard Error of Measurement at Different Ability Levels," Journal of Educational Measurement. 21 : 239-243; 1984.

Qualls and A. Payne. "A Comparison of Score Level Estimates of the Standard Error of Measurement," The University of Iowa. 213 – 225; 1992.

Woodruff, D.J. "Conditional Standard Error of Measurement in Prediction," Journal of Educational Measurement. 27 : 191-208; 1990.

ภาคผนวก

ภาคผนวก ก.

1. ค่าดัชนีความสอดคล้อง (IOC) ของข้อคำถามกับนิยามศัพท์เฉพาะของแบบทดสอบ
2. ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบทดสอบเป็นรายข้อเพื่อคัดเลือกข้อที่มีคุณภาพ

ตาราง 6 ค่าดัชนีความสอดคล้อง (IOC) ของข้อคำถามกับนิยามศัพท์เฉพาะ ของแบบทดสอบที่ใช้คุณภาพเป็นรายข้อ

ข้อที่	IOC	ข้อที่	IOC	ข้อที่	IOC
1	0.80	21	1.00	41	1.00
2	1.00	22	1.00	42	1.00
3	1.00	23	1.00	43	1.00
4	1.00	24	0.80	44	1.00
5	1.00	25	1.00	45	1.00
6	1.00	26	0.80	46	1.00
7	1.00	27	1.00	47	1.00
8	1.00	28	1.00	48	1.00
9	1.00	29	1.00	49	1.00
10	1.00	30	1.00	50	1.00
11	1.00	31	0.80	51	1.00
12	1.00	32	0.80	52	1.00
13	0.80	33	1.00	53	1.00
14	0.80	34	1.00	54	1.00
15	1.00	35	0.80	55	1.00
16	1.00	36	1.00	56	0.80
17	1.00	37	1.00	57	1.00
18	1.00	38	1.00	58	1.00
19	1.00	39	1.00	59	0.60
20	1.00	40	1.00	60	0.80

ตาราง 7 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบทดสอบเป็นรายข้อ เพื่อคัดเลือกข้อที่มีคุณภาพ

ข้อที่	p	r	ข้อที่	p	r	ข้อที่	p	r
1*	.89	.32	21	.70	.44	41	.37	.28
2	.66	.44	22*	.38	.18	42	.61	.44
3	.57	.40	23	.28	.22	43	.52	.36
4	.45	.40	24*	.22	.16	44	.53	.48
5	.53	.24	25*	.12	.12	45	.62	.60
6	.68	.22	26*	.13	.04	46	.28	.25
7*	.76	.16	27	.65	.32	47	.54	.36
8	.28	.24	28*	.85	.40	48	.32	.28
9	.57	.28	29*	.88	.44	49	.52	.40
10	.22	.40	30	.78	.40	50	.33	.24
11	.54	.64	31	.62	.40	51	.36	.72
12	.54	.48	32*	.82	.32	52	.54	.52
13*	.88	.28	33*	.14	.24	53	.26	.22
14*	.87	.40	34	.25	.22	54*	.25	.12
15	.41	.56	35	.28	.20	55*	.09	.04
16	.68	.60	36*	.14	.00	56*	.20	.38
17	.62	.56	37	.21	.22	57*	.18	.10
18	.73	.44	38	.58	.48	58*	.24	.16
19	.57	.40	39	.32	.44	59*	.29	.18
20	.49	.20	40	.52	.48	60*	.21	.08

* ข้อสอบที่ถูกคัดออก

ตาราง 8 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบทดสอบที่จัดเรียงข้อคำถามจากง่ายไปหายาก

ข้อที่	p	r	ข้อที่	p	r
1	.78	.40	**21**	.53	.24
2	.73	.44	**22**	.52	.48
3	.70	.44	**23**	.52	.40
4	.68	.60	**24**	.52	.36
5	.68	.22	**25**	.49	.20
6	.66	.44	**26**	.45	.40
7	.65	.32	**27**	.41	.56
8	.62	.60	**28**	.37	.28
9	.62	.56	**29**	.36	.72
10	.62	.40	**30**	.33	.24
11	.61	.44	**31**	.32	.44
12	.58	.48	**32**	.32	.28
13	.57	.40	**33**	.28	.25
14	.57	.40	**34**	.28	.24
15	.57	.28	**35**	.28	.22
16	.54	.64	**36**	.28	.20
17	.54	.52	**37**	.26	.22
18	.54	.48	**38**	.25	.22
19	.54	.36	**39**	.22	.40
20	.53	.48	**40**	.21	.22

ตาราง 9 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบทดสอบที่จัดเรียงข้อคำถามแบบสุ่ม

ข้อที่	p	r	ข้อที่	p	r
1	.66	.44	**21**	.62	.40
2	.57	.40	**22**	.25	.22
3	.45	.40	**23**	.28	.20
4	.53	.24	**24**	.21	.22
5	.68	.22	**25**	.58	.48
6	.28	.24	**26**	.32	.44
7	.57	.28	**27**	.52	.48
8	.22	.40	**28**	.37	.28
9	.54	.64	**29**	.61	.44
10	.54	.48	**30**	.52	.36
11	.41	.56	**31**	.53	.48
12	.68	.60	**32**	.62	.60
13	.62	.56	**33**	.28	.25
14	.73	.44	**34**	.54	.36
15	.57	.40	**35**	.32	.28
16	.49	.20	**36**	.52	.40
17	.70	.44	**37**	.33	.24
18	.28	.22	**38**	.36	.72
19	.65	.32	**39**	.54	.52
20	.78	.40	**40**	.26	.22

ภาคผนวก ข.

ผลการวิเคราะห์องค์ประกอบของแบบทดสอบที่จัดเรียงข้อคำถามจากง่ายไปหายาก และจัดเรียงแบบสุ่มเพื่อตรวจสอบความเป็นมิติเดียว (Unidimensional)

ตาราง 10 ค่าไอเกน (มากกว่า 1) ของแบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายากและเรียงข้อสอบแบบสุ่ม

องค์ประกอบที่	ค่าไอเกน ฉบับที่เรียงจากง่ายไปยาก	ค่าไอเกน ฉบับที่เรียงแบบสุ่ม
1	8.99436	8.09778
2	1.75579	1.99903
3	1.46763	1.65036
4	1.41732	1.41231
5	1.29174	1.25532
6	1.19702	1.23184
7	1.14709	1.16345
8	1.07156	1.13062
9	1.03073	1.10705
10	1.01379	1.05176

ภาคผนวก ค.

แบบทดสอบที่ใช้ในการวิจัย

ฉบับที่ 1 แบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายาก

ฉบับที่ 2 แบบทดสอบที่จัดเรียงข้อสอบแบบสุ่ม

แบบทดสอบวัดความสามารถด้านตัวเลข

ฉบับที่ 1

<u>คำชี้แจง</u>

1. แบบทดสอบฉบับนี้เป็นแบบทดสอบวัดความสามารถด้านตัวเลข โดยให้นักเรียนใช้ความคิดอย่างเป็นระบบ ผ่านกลวิธีทางคณิตศาสตร์ คือการบวก ลบ คูณ และหาร แล้วพิจารณาหาแนวโน้มของตัวเลขที่ต้องการ (.....)
2. แบบทดสอบฉบับนี้มี 40 ข้อ เป็นแบบเลือกตอบชนิด 4 ตัวเลือก
3. เวลาที่ใช้ทำข้อสอบ 40 นาที
4. วิธีตอบ ให้นักเรียนเลือกคำตอบที่ถูกที่สุดหรือเหมาะสมที่สุดเพียงคำตอบเดียว โดยทำเครื่องหมาย X ลงในช่องหมายเลขที่ต้องการในกระดาษคำตอบ

ตัวอย่างข้อสอบ

0.) 2 4 6 8 .....

ก. 9

ข. 10

ค. 11

ง. 12

00) 2 5 3 6 .....

ก. 4

ข. 6

ค. 8

ง. 10

กระดาษคำตอบ

	ก	ข	ค	ง
0)		X		
00)	X			

1. 29 27 24 22 19 .....
   ก. 18
   ข. 17
   ค. 16
   ง. 15

2. 16 12 8 4 .....
   ก. 3
   ข. 2
   ค. 1
   ง. 0

3. 19 15 12 10 .....
   ก. 9
   ข. 7
   ค. 5
   ง. 3

4. 4 7 6 7 8 7 .....
   ก. 7
   ข. 8
   ค. 9
   ง. 10

5. 1 3 1 6 . .....
   ก. 1
   ข. 2
   ค. 3
   ง. 4

6. 3 3 6 6 18 .....
   ก. 3
   ข. 6
   ค. 18
   ง. 72

7. 9 12 15 18 .....
   ก. 21
   ข. 24
   ค. 28
   ง. 30

8. 20 19 17 16 14 .....
   ก. 13
   ข. 12
   ค. 11
   ง. 10

9. 14 21 .....35 42
   ก. 23
   ข. 25
   ค. 27
   ง. 28

10. 16 14 12 8
   ก. 4
   ข. 6
   ค. 8
   ง. 10

11. 144 72 36 18 .....
ก. 15
ข. 13
ค. 11
ง. 9

12. 1 3 7 9 13 15 .....
ก. 17
ข. 19
ค. 21
ง. 23

13. 35 30 26 23 .....
ก. 22
ข. 21
ค. 20
ง. 19

14. 5 6 8 11 .....
ก. 13
ข. 14
ค. 15
ง. 16

15. 21 16 18 13 15 .....
ก. 5
ข. 10
ค. 15
ง. 20

16. 2 5 ..... 11 14
ก. 6
ข. 7
ค. 8
ง. 9

17. 1 5 8 10 .....
ก. 11
ข. 12
ค. 13
ง. 14

18. 4 9 16 25 .....
ก. 36
ข. 46
ค. 64
ง. 81

19. 1 3 5 8 11 15 .....
ก. 17
ข. 19
ค. 21
ง. 23

20. 24 20 16 12 .....
ก. 8
ข. 6
ค. 4
ง. 2

21. 25 20 16 13 .....
ก. 12
ข. 11
ค. 9
ง. 7

22. 4 6 9 13 .....
ก. 16
ข. 18
ค. 20
ง. 22

23. 49 36 25 16 .....
ก. 9
ข. 6
ค. 4
ง. 2

24. 25 16 9 4 .....
ก. 3
ข. 2
ค. 1
ง. 0

25. 7 8 10 11 14 .....
ก. 15
ข. 16
ค. 17
ง. 18

26. 24 23 21 18 .....
ก. 16
ข. 15
ค. 14
ง. 13

27. 25 16 9 4 .....
ก. 3
ข. 2
ค. 1
ง. 0

28. 4 9 15 22 .....
ก. 24
ข. 26
ค. 28
ง. 30

29. 11 8 9 6 7 .....
ก. 5
ข. 4
ค. 3
ง. 2

30. 7 12 4 9 .....
ก. 1
ข. 3
ค. 5
ง. 7

31. 162 54 18 6 .....

ก. 2

ข. 3

ค. 4

ง. 5

32. 2 5 4 8 6 11 .....

ก. 6

ข. 7

ค. 8

ง. 9

33. 15 10 13 7 11 .....

ก. 9

ข. 7

ค. 4

ง. 1

34. 2 4 12 24 72 .....

ก. 114

ข. 124

ค. 134

ง. 144

35. 4 12 36 324

ก. 72

ข. 108

ค. 180

ง. 252

36. 48 40 45 37 .....

ก. 35

ข. 39

ค. 42

ง. 46

37. 3 6 12 24 .....

ก. 48

ข. 36

ค. 30

ง. 28

38. 1 4 9 16 .....

ก. 19

ข. 21

ค. 23

ง. 25

39. 32 28 26 21 .....

ก. 20

ข. 19

ค. 18

ง. 17

40. 80 70 75 65 .....

ก. 80

ข. 70

ค. 60

ง. 50

แบบทดสอบวัดความสามารถด้านตัวเลข

ฉบับที่ 2

<u>คำชี้แจง</u>

1. แบบทดสอบฉบับนี้เป็นแบบทดสอบวัดความสามารถด้านตัวเลข โดยให้นักเรียนใช้ความคิดอย่างเป็นระบบ ผ่านกลวิธีทางคณิตศาสตร์ คือการบวก ลบ คูณ และหาร แล้วพิจารณาหาแนวโน้มของตัวเลขที่ต้องการ (.....)
2. แบบทดสอบฉบับนี้มี 40 ข้อ เป็นแบบเลือกตอบชนิด 4 ตัวเลือก
3. เวลาที่ใช้ทำข้อสอบ 40 นาที
4. วิธีตอบ ให้นักเรียนเลือกคำตอบที่ถูกที่สุดหรือเหมาะสมที่สุดเพียงคำตอบเดียว โดยทำเครื่องหมาย X ลงในช่องหมายเลขที่ต้องการในกระดาษคำตอบ

ตัวอย่างข้อสอบ

0.) 2 4 6 8 .....

ก. 9

ข. 10

ค. 11

ง. 12

00) 2 5 3 6 .....

ก. 4

ข. 6

ค. 8

ง. 10

กระดาษคำตอบ

	ก	ข	ค	ง
0)		X		
00)	X			

1. 1 3 1 6 1 9 .....

ก. 1

ข. 2

ค. 3

ง. 4

2. 7 8 10 11 14 .....

ก. 15

ข. 16

ค. 17

ง. 18

3. 4 7 6 7 8 7 .....

ก. 7

ข. 8

ค. 9

ง. 10

4. 1 3 7 9 13 15 .....

ก. 17

ข. 19

ค. 21

ง. 23

5. 1 5 8 10 .....

ก. 11

ข. 12

ค. 13

ง. 14

6. 5 6 8 11 .....

ก. 13

ข. 14

ค. 15

ง. 16

7. 1 3 5 8 11 15 .....

ก. 17

ข. 19

ค. 21

ง. 23

8. 2 5 4 8 6 11 .....

ก. 6

ข. 7

ก. 8

ง. 9

9. 1 4 9 16 .....

ก. 19

ข. 21

ค. 23

ง. 25

10. 2 5 ..... 11 14

ก. 6

ข. 7

ค. 8

ง. 9

11. 24 23 21 18 .....

ก. 16

ข. 15

ค. 14

ง. 13

12. 20 19 17 16 14 .....

ก. 13

ข. 12

ค. 11

ง. 10

13. 35 30 26 23 .....

ก. 22

ข. 21

ค. 20

ง. 19

14. 16 12 8 4 .....

ก. 3

ข. 2

ค. 1

ง. 0

15. 3 3 6 6 18 .....

ก. 3

ข. 6

ค. 18

ง. 36

16. 19 15 12 10 .....

ก. 9

ข. 7

ค. 5

ง. 3

17. 29 27 24 22 19 .....

ก. 18

ข. 17

ค. 16

ง. 15

18. 25 16 9 4 .....

ก. 3

ข. 2

ค. 1

ง. 0

19. 15 10 13 7 11 .....

ก. 9

ข. 7

ค. 4

ง. 1

20. 14 21 ..... 35 42

ก. 23

ข. 25

ค. 27

ง. 28

21. 3 6 12 24 .....

ก. 48

ข. 36

ค. 30

ง. 28

22. 32 28 26 21 .....

ก. 20

ข. 19

ค. 18

ง. 17

23. 162 54 18 6 .....

ก. 2

ข. 3

ค. 4

ง. 5

24. 25 16 9 4 .....

ก. 3

ข. 2

ค. 1

ง. 0

25. 16 14 12 8

ก. 4

ข. 6

ค. 8

ง. 10

26. 4 12 36 324

ก. 72

ข. 108

ค. 180

ง. 252

27. 144 72 36 18 .....

ก. 15

ข. 13

ค. 11

ง. 9

28. 4 6 9 13 .....

ก. 16

ข. 18

ค. 20

ง. 22

29. 9 12 15 18 .....

ก. 21

ข. 24

ค. 27

ง. 30

30. 2 4 12 24 72 .....

ก. 114

ข. 124

ค. 134

ง. 144

31. 49 36 25 16 .....
ก. 9
ข. 6
ค. 4
ง. 2

32. 24 20 16 12 .....
ก. 8
ข. 6
ค. 4
ง. 2

33. 25 20 16 13 .....
ก. 12
ข. 11
ค. 9
ง. 7

34. 11 8 9 6 7 .....
ก. 5
ข. 4
ค. 3
ง. 2

35. 4 9 15 22 .....
ก. 24
ข. 26
ค. 28
ง. 30

36. 21 16 18 13 15 .....
ก. 5
ข. 10
ค. 15
ง. 20

37. 7 12 4 9 .....
ก. 1
ข. 3
ค. 5
ง. 7

38. 48 40 45 37 .....
ก. 35
ข. 39
ค. 42
ง. 46

39. 80 70 75 65 .....
ก. 80
ข. 70
ค. 60
ง. 50

40. 4 9 16 25 .....
ก. 36
ข. 46
ค. 64
ง. 81

ภาคผนวก ง

รายชื่อผู้เชี่ยวชาญที่ประเมินความเที่ยงตรงตามนิยาม

รายชื่อผู้เชี่ยวชาญ

อาจารย์วิศิษฏ์ พหลยุทธ	โรงเรียนวังม่วงวิทยาคม จ.สระบุรี
อาจารย์อุทัยวรรณ สายพัฒนะ	โรงเรียนธรรมศาสตร์คลองหลวงวิทยาคม
อาจารย์นิภาพร ยอดเมือง	โรงเรียนบางเขน (ไว้สาลีอนุสรณ์) กทม.
อาจารย์พนิดา ฟุ้งเกียรตินำสุข	โรงเรียนชุมชนคลองเตย กทม.
อาจารย์อุไร จันทร์สว่าง	โรงเรียนสมบุญศุภบุญมี จ.พระนครศรีอยุธยา

ประวัติย่อของผู้วิจัย

ชื่อ นางสาวอำนวยพร คำพันธ์

เกิดวันที่ 24 เมษายน 2498

สถานที่เกิด อำเภอวิเศษชัยชาญ จังหวัดอ่างทอง

สถานที่อยู่ปัจจุบัน 28/1 หมู่ที่ 10 ตำบลพัฒนานิคม อำเภอพัฒนานิคม จังหวัดลพบุรี 15140

สถานที่ทำงานปัจจุบัน สำนักงานการประถมศึกษาอำเภอวังม่วง จังหวัดสระบุรี

ตำแหน่ง ศึกษานิเทศก์ ระดับ 7

ประวัติการศึกษา

พ.ศ. 2515 มัธยมศึกษาปีที่ 3
โรงเรียนพัฒนานิคม อำเภอพัฒนานิคม
จังหวัดลพบุรี

พ.ศ. 2519 ป.ก.ศ. สูง (วิชาเอกการประถมศึกษา)
วิทยาลัยครูเทพสตรี ลพบุรี

พ.ศ. 2524 กศ.บ. (วิชาเอกการประถมศึกษา)
มหาวิทยาลัยศรีนครินทรวิโรฒ บางแสน

พ.ศ. 2542 กศ.ม. (การวัดผลการศึกษา)
มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

ผลของการจัดเรียงข้อสอบที่มีต่อค่าความคลาดเคลื่อนมาตรฐาน ในการวัดที่ระดับคะแนนเฉพาะ

บทคัดย่อ

ของ

อำนวยพร คำพันธ์

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา
มีนาคม 2542

การวิจัยครั้งนี้ เป็นการศึกษาผลของการจัดเรียงข้อสอบที่มีต่อค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ โดยมีจุดมุ่งหมายในการศึกษาว่า แบบทดสอบที่จัดเรียงข้อสอบจากง่ายไปหายากกับการจัดเรียงข้อสอบแบบสุ่มมีผลต่อการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดหรือไม่ และค่าความคลาดเคลื่อนมาตรฐานในการวัดโดยใช้วิธีประมาณค่า 3 วิธี มีผลแตกต่างกันอย่างไร โดยศึกษากับกลุ่มตัวอย่างคือ นักเรียนชั้นประถมศึกษาปีที่ 6 ภาคเรียนที่ 2 ปีการศึกษา 2540 จำนวน 1,019 คน ซึ่งใช้วิธีสุ่มแบบหลายขั้นตอน เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวัดความสามารถด้านตัวเลขแบบอนุกรมธรรมดาซึ่งผู้วิจัยสร้างขึ้นเอง เป็นแบบปรนัยชนิดเลือกตอบ 4 ตัวเลือก จำนวน 40 ข้อ จัดเรียงข้อสอบเป็น 2 รูปแบบ คือ เรียงจากข้อง่ายไปหายาก และเรียงข้อแบบสุ่ม แล้วคำนวณค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ โดยใช้วิธีของ ธอร์นไดค์ (Thorndike. 1951), ลอร์ด (Lord. 1984) และคีตส์ (Keats. 1957)

ผลการวิจัยพบว่า ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะของแบบทดสอบที่จัดเรียงจากง่ายไปหายาก มีค่าต่ำกว่าแบบทดสอบที่จัดเรียงแบบสุ่ม และค่าความคลาดเคลื่อนมาตรฐานในการวัดที่คำนวณโดยวิธี 3 วิธี มีค่าที่แตกต่างกัน โดยวิธีของ ลอร์ด และคีตส์จะได้ค่าที่ใกล้เคียงสอดคล้องกัน เป็นรูปพาราโบลาคว่ำ ส่วนวิธีของธอร์นไดค์ ค่าที่ได้จะมีลักษณะสูง ๆ ต่ำ ๆ สลับกันในบางระดับคะแนน แต่ยังมีแนวโน้มเป็นรูปพาราโบลาคว่ำเช่นกัน จากการเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนเฉพาะ พบว่าวิธีของคีตส์จะได้ค่าที่ต่ำที่สุดอยู่ในช่วงปลายของระดับคะแนนทั้ง 2 ฉบับ และค่าที่สูงที่สุดคือวิธีของ ธอร์นไดค์จากฉบับที่เรียงจากง่ายไปหายาก และวิธีของลอร์ดจากฉบับที่เรียงแบบสุ่ม ซึ่งอยู่ในช่วงกลางของระดับคะแนน

THE EFFECT OF ITEM ARRANGEMENT ON THE STANDARD ERROR OF MEASUREMENT AT SPECIFIC SCORE LEVELS

AN ABSTRACT

BY

AMNUAYPORN KAMPAN

Presented in partial fulfillment of the requirements for the

Master of Education degree in Educational Measurement

At Srinakharinwirot University

March 1999

The Purposes of this research were to study whether the item arrangement would effect estimations standard error of measurement or not and to study the differences of standard error measured by three methods at specific score levels.

The samples, randomized by multistage random sampling, were 1,019 Pratom VI students studying in second semester of academic year 1997. The instruments were Numerical Ability Test on Number Series, 40 items of multiple choices with 1 alternatives, arranged the item into two types : the first one was arrangement in order of difficulty, the second one was random arrangement. The result were calculated standard error of measurement by three methods of Thorndike. (1951), Lord. (1984) and Keats (1957).

The main findings were as follows : The standard error of measurement at specific score levels of the first type item (Arrangement in Order of Difficulty) had lower value than the random arrangement item. The standard error measurement by three methods of specific score levels had different values. Lord and Keats method had Harmonious value as Parabolic. For Thorndike method, the value of some score levels alternated between high and low, but was still in Parabolic. The comparison of standard error of measurement at specific score levels was found that Keats method had the lowest value at the top of score level in both types, the highest value was at medium of score level both by Thorndike method from the first type item arrangement order of difficult and by Lord method from the random arrangement item.