

๖๗๑.๒๖๐๑๒

๙๖ ๖๕๔ ๐

๗.๖

ความคลาดเคลื่อน และความลำเอียงทางสถิติในการเทียบคะแนนเชิงเส้นตรง
แบบคะแนนจริงสัมพันธ์ในกลุ่มตัวอย่างขนาดเล็ก

๒๑ ก.พ. ๒๕๔๐

ปริญญาพนธ์
ของ
นิภาพร ยอดเมือง

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา
กันยายน ๒๕๓๙
ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

6๕๐๐๔

คณะกรรมการควบคุมและคณะกรรมการสอบได้พิจารณาปฏิญานพนธ์ฉบับนี้แล้ว
เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอก
การวัดผลการศึกษา ของมหาวิทยาลัยศรีนครินทรวิโรฒได้

คณะกรรมการควบคุม

..... ประธาน
(รศ.ดร.บุญเชิด ภิญโญนนท์พงษ์)

..... กรรมการ
(อ.ยุพา มานะจิตต์)

คณะกรรมการสอบ

..... ประธาน
(รศ.ดร.บุญเชิด ภิญโญนนท์พงษ์)

..... กรรมการ
(อ.ยุพา มานะจิตต์)

..... กรรมการที่แต่งตั้งเพิ่มเติม
(ผศ.เชาวนา ชวลิตธำรง)

บัณฑิตวิทยาลัยอนุมัติให้รับปฏิญานพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา ของมหาวิทยาลัยศรีนครินทรวิโรฒ

..... คณบดีบัณฑิตวิทยาลัย
(ดร.ศิริยุภา พูลสุวรรณ)

วันที่ 28. เดือนกันยายน พ.ศ.2539

ประกาศคุณประการ

ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงด้วยดี ด้วยความกรุณาจากรองศาสตราจารย์ ดร.บุญเชิด ภิญโญนันตพงษ์ อาจารย์ยุพา มานะจิตต์ ที่ได้ให้คำปรึกษา แนะนำ ตลอดจนแก้ไขข้อบกพร่องต่าง ๆ อีกทั้งได้ให้กำลังใจ และความห่วงใยแก่ผู้วิจัยตลอดมา ผู้วิจัยรู้สึกซาบซึ้งในความเมตตาของท่าน จึงขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอกราบขอบพระคุณผู้ช่วยศาสตราจารย์เชาวนา ชวลิตธำรง ที่กรุณาแก้ไขข้อบกพร่องเพิ่มเติม ให้ปริญญานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น

ขอกราบขอบพระคุณคณาจารย์ทุกท่าน ที่ประสิทธิ์ประสาทวิชาความรู้แก่ผู้วิจัย จนทำให้ปริญญานิพนธ์ฉบับนี้สำเร็จสมบูรณ์ด้วยดี

ขอขอบพระคุณผู้อำนวยการ ผู้ช่วยผู้อำนวยการ คณาจารย์ ทุกท่านที่ให้ความช่วยเหลือในการเก็บข้อมูลเป็นอย่างดี และขอขอบคุณนักเรียนของโรงเรียนที่เป็นกลุ่มตัวอย่างทุกคน ที่ให้ความร่วมมือเป็นอย่างดี

ขอขอบคุณยิ่งสำหรับ คุณอุทัยวรรณ สายพัฒนะ ที่ให้ความช่วยเหลือในด้านเครื่องคอมพิวเตอร์ เรื่องงานพิมพ์ เรื่องการวิเคราะห์ข้อมูล และเป็นกำลังใจด้วยดีตลอดมา

ขอขอบคุณ อาจารย์คำนิ้ง พุ่มจันทร์ อาจารย์ศศิธร นุลทอง ที่ให้ความช่วยเหลือในการเก็บข้อมูลเป็นอย่างดี

ขอขอบคุณ อาจารย์ยุรี ไทยวรรณ คุณจิตสุภา ไทยวรรณ ที่ให้ความอนุเคราะห์เครื่องพิมพ์สี ทำให้ปริญญานิพนธ์ฉบับนี้สำเร็จด้วยดี

ขอขอบคุณ คุณธีรวิทย์ นกทอง ที่ให้ความช่วยเหลือในการวิเคราะห์ข้อมูลการวิจัยเป็นอย่างดี

และขอขอบคุณน้อง ๆ วัดผล รุ่น 4 รุ่น 5 รุ่น 6 ทุกท่าน ที่ได้ช่วยเหลือ และเป็นกำลังใจแก่ผู้วิจัยด้วยดี

คุณค่า และประโยชน์ของปริญญานิพนธ์ฉบับนี้ มอบเป็นเครื่องบูชาพระคุณของบิดา-มารดา และผู้มีพระคุณทุกท่านของผู้วิจัย

นิภาพร ยอดเมือง

ปริญญานิพนธ์ฉบับนี้ได้รับทุนสนับสนุนการวิจัยงบประมาณแผ่นดิน

ประจำปีงบประมาณ 2539

ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูง

สารบัญ

บทที่		หน้า
1	บทนำ	1
	ภูมิหลัง	1
	ความมุ่งหมายของการศึกษาค้นคว้า	3
	ความสำคัญของการศึกษาค้นคว้า	4
	ขอบเขตของการศึกษาค้นคว้า	4
	นิยามศัพท์เฉพาะ	6
2	เอกสารและงานวิจัยที่เกี่ยวข้อง	8
	ความหมายของการเทียบคะแนน	8
	มโนทัศน์ของการเทียบคะแนน	9
	แบบจำลองของการวัด	12
	รูปแบบการเก็บรวบรวมข้อมูล	12
	วิธีการเทียบคะแนน	20
	การเทียบคะแนนเชิงเส้นตรง โดยใช้แบบทดสอบหลัก	22
	แบบทดสอบหลัก	31
	ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน	35
	ความลำเอียงในการเทียบคะแนน	<u>39</u>
	ทฤษฎีเซวานปัญหาและความถนัด	42
	ความหมายของการรับรู้	45
	ปัญหาเกี่ยวกับการรับรู้	46
	องค์ประกอบที่มีส่วนเกี่ยวกับการรับรู้	46
	แบบทดสอบวัดการรับรู้ตามแนวคิดของเทอร์สโตน	46
	งานวิจัยที่เกี่ยวข้อง	47

3	วิธีดำเนินการศึกษาค้นคว้า	58
	ประชากร	58
	กลุ่มตัวอย่าง	58
	เครื่องมือที่ใช้ในการศึกษาค้นคว้า	59
	วิธีดำเนินการสร้างเครื่องมือ	59
	วิธีดำเนินการเก็บรวบรวมข้อมูล	62
	ขั้นตอนการวิเคราะห์ข้อมูล	62
	สถิติที่ใช้ในการวิเคราะห์ข้อมูล	63
4	ผลการวิเคราะห์ข้อมูล	67
	สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล	67
	การเสนอผลการวิเคราะห์ข้อมูล	67
	ค่าสถิติพื้นฐานของแบบทดสอบ	68
	ค่าสถิติของกลุ่มตัวอย่างแต่ละขนาด	68
	เปรียบเทียบคะแนนเทียบเฉลี่ยของทุกขนาดกลุ่มตัวอย่าง	70
	ค่าความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน	70
	ค่าความลำเอียงทางสถิติในการเทียบคะแนน	73
	ภาพแสดงความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน	77
	ภาพแสดงความลำเอียงทางสถิติในการเทียบคะแนน	78
5	สรุป อภิปรายผล และข้อเสนอแนะ	79
	ความมุ่งหมายของการศึกษาค้นคว้า	79
	กลุ่มตัวอย่าง	79
	เครื่องมือที่ใช้ในการศึกษาค้นคว้า	80
	วิธีดำเนินการเก็บรวบรวมข้อมูล	80
	การเสนอผลการวิเคราะห์ข้อมูล	82
	สรุปผลการวิเคราะห์ข้อมูล	83
	อภิปรายผล	85
	ข้อเสนอแนะ	87

บรรณานุกรม	88
ภาคผนวก	93
ภาคผนวก ก คะแนนเทียบเฉลี่ยของทุกกลุ่มตัวอย่างและประชากรเทียบ	94
ภาคผนวก ข ค่าความยากและค่าอำนาจจำแนกของแบบทดสอบทั้งสามฉบับ	98
ภาคผนวก ค แบบทดสอบทั้งสามฉบับ	102
ประวัติย่อของผู้วิจัย	139
บทคัดย่อ	140

บัญชีตาราง

ตาราง	หน้า
1 แสดงจำนวนประชากร	5
2 สถิติพื้นฐานของแบบทดสอบ 3 ฉบับ	68
3 ค่าสถิติของกลุ่มตัวอย่างแต่ละขนาด	69
4 เปรียบเทียบคะแนนเทียบเฉลี่ยของทุกขนาดกลุ่มตัวอย่าง	70
5 ค่าความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน	71
6 ค่าความลำเอียงทางสถิติในการเทียบคะแนน	74

บัญชีภาพประกอบ

ภาพประกอบ

หน้า

1 ลำดับของการเทียบคะแนนของแบบทดสอบ และความสัมพันธ์ของแต่ละลำดับ	10
2 แสดงการเชื่อมต่อของแบบทดสอบ 2 ฉบับ	32
3 แสดงการเชื่อมแบบ 2 ลูกโซ่	33
4 แสดงจุดเชื่อมต่อของการเชื่อมแบบ 2 ลูกโซ่	33
5 แสดงตารางการเชื่อมของแบบทดสอบ 10 ฉบับ 7 ระดับความสามารถ	34
6 แสดงตารางการเชื่อมต่อโดยแบบทดสอบภายนอก	34
7 แสดงลำดับขั้นตอนการสร้างเครื่องมือ	60
8 แสดงความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนของคะแนนสิ่งเกิดทุกตัว	77
9 แสดงความลำเอียงทางสถิติในการเทียบคะแนนของคะแนนสิ่งเกิดทุกตัว	78

บทนำ

การทดสอบเพื่อการวัดผลมีความสำคัญอย่างยิ่ง เป็นกระบวนการที่ใช้มากที่สุด ในวงการศึกษานั้นนับได้ว่าเป็นหลักของการวัดผลในโรงเรียนก็ได้ การทดสอบที่ดีจะช่วยให้ครูทราบสภาพของนักเรียน และของครูเองได้ง่ายกว่าแบบอื่น ๆ มาก (ชวาล แพริตกุล.2518 : 88) ปัจจุบันการสร้างแบบทดสอบทำได้หลายรูปแบบ รูปแบบหนึ่งที่น่าสนใจในขณะนี้ ก็คือ รูปแบบที่ใช้แบบทดสอบต่างฉบับ แต่วัดในสิ่งเดียวกัน การเรียนการสอนในปัจจุบันจะพบว่าในรายวิชาหนึ่ง ๆ จะมีผู้สอนหลายท่าน ใช้แบบทดสอบต่างฉบับกัน จะทราบได้อย่างไรว่า การตัดสินผลการเรียนของแต่ละท่านยุติธรรมเพียงใด นักเรียนที่ได้คะแนนเท่ากันจากแบบทดสอบต่างฉบับกัน สมควรได้รับคะแนนเท่ากันจริงหรือ การเทียบคะแนน (test equating) จึงเป็นวิธีการอย่างหนึ่งที่จะช่วยให้การตัดสินเป็นไปอย่างยุติธรรม เพราะการเทียบคะแนนจะเป็นการแปลงคะแนนดิบที่นักเรียนแต่ละคนสอบได้ให้เป็นคะแนนมาตรฐานเดียวกัน นอกจากนี้การเทียบคะแนนยังป้องกันการรั่วของข้อสอบ ถ้าใช้ข้อสอบฉบับเดียว และเมื่อจำเป็นต้องสอบซ้ำ การสอบครั้งหลังก็จะเป็นข้อสอบที่ไม่สามารถนำความรู้ที่แท้จริงของผู้สอบออกมาได้เช่นครั้งแรก (Petersen, Kolen and Hoover.1989: 241-242) ในบางสถานการณ์ จำเป็นต้องใช้แบบทดสอบต่างฉบับกันวัดสิ่งเดียวกันสอบกับเด็กกลุ่มใหญ่ แม้เด็กนั่งใกล้กัน เด็กสอบคนละห้อง คนละเวลา หรือเด็กสอบไม่พร้อมกัน เพื่อป้องกันการลอกคำตอบ การบอกข้อสอบแก่ผู้ยังไม่สอบ (สงบ ลักษณะ.2525:23)

แบบทดสอบที่จะใช้ในการเทียบคะแนนนั้นจะต้องมีความคล้ายคลึงกัน ซึ่งจำแนกเป็น 3 แบบตามข้อตกลงของระดับความคล้ายคลึงของแบบทดสอบ คือ แบบจำลองคะแนนจริงสัมพันธ์ (Congeneric Model) แบบจำลองคะแนนจริงสมมูล (Tau-Equivalent Model) และแบบจำลองคู่ขนาน (Parallel Model) แบบจำลองคะแนนจริงสัมพันธ์เป็นแบบจำลองที่มีข้อตกลงที่ผ่อนปรนมากที่สุด คือ กำหนดว่าแบบทดสอบแต่ละฟอร์มต้องวัดฟังก์ชันทางจิตวิทยา ลักษณะเดียวกัน (The Same Psychological Function) หรือวัดมิติเดียวกัน (Unidimension) หรือวัดคุณลักษณะเดียวกัน (The Same Trait) ซึ่งพิจารณาได้จากคะแนนจริงของแต่ละฟอร์มมีความสัมพันธ์กันเป็นเส้นตรง ดังนั้นแบบทดสอบที่ตกลงให้มีข้อกำหนดดังกล่าวเรียกว่าแบบทดสอบคะแนน

จริงสัมพันธ์ (Congeneric test) สำหรับแบบจำลอง 2 แบบหลังเป็นกรณีเฉพาะของแบบจำลองคะแนนจริงสัมพันธ์ เพราะมีข้อตกลงที่เคร่งครัดกว่า และในทางปฏิบัติส่วนมากแล้ว แบบทดสอบคะแนนจริงสัมพันธ์และแบบทดสอบคู่ขนานไม่อาจกำหนดให้เป็นไปตามข้อตกลงของมันได้ ดังนั้น การนำข้อตกลงของแบบจำลองคะแนนจริงสัมพันธ์ หรือแบบทดสอบคะแนนจริงสัมพันธ์ไปใช้ในงานวัดผลต่าง ๆ จึงมีความเหมาะสมในทางปฏิบัติ (บุญเชิด ภิญโญนนตพงษ์. 2538:1)

การเปลี่ยนคะแนนที่ได้จากแบบทดสอบแต่ละฉบับให้อยู่ในมาตราเดียวกันและเทียบกันได้ที่นิยมใช้กันมาก ปีเตอร์เซน, มาร์โคและสจ๊วต (Petersen, Marco and Stewart. 1982:73) ได้นำเสนอไว้ 3 วิธี คือ

1. การเทียบจากเปอร์เซ็นต์ไทล์ที่เท่ากัน (Equipercetile equating) เป็นการแปลงคะแนนจากแบบทดสอบ 2 ฉบับ คะแนนแต่ละคะแนนจะสามารถเทียบเท่ากันได้ ก็ต่อเมื่อ อยู่ในตำแหน่งของเปอร์เซ็นต์ไทล์เดียวกัน

2. การเทียบเชิงเส้นตรง (Linear equating) เป็นการแปลงคะแนนจากแบบทดสอบ 2 ฉบับ คะแนนแต่ละคะแนนจะเทียบเท่ากันได้ ก็ต่อเมื่อ มีการกระจายของค่าเฉลี่ยของคะแนนแต่ละกลุ่มเหมือนกัน

3. การเทียบโดยทฤษฎีการตอบข้อคำถาม (Item Characteristic Curve equating) เป็นการแปลงคะแนนจริงจากแบบทดสอบ 2 ฉบับ คะแนนแต่ละคะแนนจะเทียบเท่ากันได้ ก็ต่อเมื่อค่าประมาณของระดับคุณลักษณะของแบบทดสอบทั้ง 2 เหมือนกัน

4. การเทียบคะแนนโดยใช้เทคนิคการวิเคราะห์องค์ประกอบ (Equating Using the Confirmatory Factor Analysis) ใช้ความเป็นไปได้ของการวิเคราะห์ตัวประกอบสูงสุดในการประมาณค่าพารามิเตอร์ ซึ่ง รอค (Rock. 1982 : 247) เป็นผู้เสนอไว้เป็นรูปแบบใหม่อีกวิธีหนึ่ง

การเทียบคะแนนเชิงเส้นตรงเป็นวิธีการที่ให้น่าเชื่อถือได้มากวิธีหนึ่งซึ่งโคลเลน และ ไวท์เนย์ (Kolen & Whitney. 1982) ได้ให้คำแนะนำว่าถ้าขนาดของกลุ่มตัวอย่างเล็ก ควรใช้วิธีเทียบคะแนนเชิงเส้นตรง และในการเทียบคะแนนได้มีการเสนอสูตรไว้มากมาย แต่สูตรของแมคแคน (MacCann. 1989)

$$Y' = \begin{bmatrix} r_{yz2} & S_{y2} & S_{z1} \\ r_{xz1} & S_{x1} & S_{z2} \end{bmatrix} X + \bar{Y}_2 + \begin{bmatrix} S_{y2} \\ S_{z2} \end{bmatrix} \left[\begin{bmatrix} \bar{Z}_1 - \bar{Z}_2 \end{bmatrix} - \frac{1}{r_{xz1}} \frac{S_{z1}}{S_{x1}} \bar{X}_1 \right]$$

เป็นสูตรที่ใช้ข้อตกลงของความคู่ขนานแบบคะแนนจริงสัมพันธ์

สำหรับขนาดของกลุ่มตัวอย่างก็ได้มีผู้ศึกษาไว้ต่าง ๆ กัน เช่น โคลเลน (Kolen.1985) ได้ใช้กลุ่มตัวอย่างขนาด 100 คนและ 250 คน ในการศึกษาความคลาดเคลื่อนของวิธีการเทียบคะแนนแบบทักเคอร์ (Tucker Equating) พบว่าขนาด 250 คนมีค่าความคลาดเคลื่อนน้อยกว่าขนาด 100 คน พาแชล, ฮอจตัน และครอมเรย์ (Parshall, Houghton and Kromrey.1995) ใช้กลุ่มตัวอย่างขนาด 15 คน 25 คน 50 คน 100 คนและ 500 คน ในการศึกษาความลำเอียงทางสถิติ และความคลาดเคลื่อนมาตรฐานในการเทียบคะแนน พบว่าความคลาดเคลื่อนน้อยมากทุกขนาดของกลุ่มตัวอย่างเมื่อคะแนนใกล้เคียงค่าเฉลี่ย สำหรับความลำเอียงทางสถิติจะมีผลกับกลุ่มตัวอย่างน้อยเมื่อคะแนนเบี่ยงเบนออกจากค่าเฉลี่ย 1.5 SD

ผู้วิจัยจึงสนใจศึกษาสูตรของแมคเคน ในแง่ของกลุ่มตัวอย่างที่มีขนาดเล็กกว่ากลุ่มตัวอย่างขนาดเท่าไรจึงจะมีความเหมาะสม หรือเพียงพอที่ให้ผลการเทียบคะแนนได้ใกล้เคียงกับการเทียบคะแนนจากกลุ่มตัวอย่างขนาดใหญ่ โดยพิจารณาความเหมาะสมจากคะแนนเทียบเฉลี่ย ค่าความลำเอียงทางสถิติ และค่าความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน

ความมุ่งหมายของการศึกษาค้นคว้า

การวิจัยครั้งนี้มีจุดมุ่งหมายที่จะตอบคำถามวิจัยที่ว่า การเทียบคะแนนเชิงเส้นตรงแบบคะแนนจริงสัมพันธ์ด้วยสูตรของแมคเคน กลุ่มตัวอย่างที่มีขนาดเล็กเท่าไร จึงจะสามารถให้ผลการเทียบคะแนนใกล้เคียงกับกลุ่มตัวอย่างขนาดใหญ่ ดังนั้นจึงมีจุดมุ่งหมายในการศึกษาเฉพาะต่อไปนี้ คือ

1. เพื่อศึกษาว่าคะแนนเทียบเฉลี่ยจากแต่ละกลุ่มแตกต่างกันหรือไม่
2. เพื่อศึกษาความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนในกลุ่มตัวอย่างขนาด 30, 60, 120 และ 250 ว่าแตกต่างกันหรือไม่
3. เพื่อศึกษาความลำเอียงทางสถิติ ในกลุ่มตัวอย่างขนาด 30, 60, 120 และ 250 ว่าแตกต่างกันหรือไม่

ความสำคัญของการศึกษาค้นคว้า

การศึกษาค้นคว้าครั้งนี้เน้นที่ผลการศึกษา กลุ่มตัวอย่างขนาดเล็กเพื่อใช้ในการเทียบคะแนน ว่ากลุ่มตัวอย่างขนาดใดใช้แทนกลุ่มตัวอย่างขนาดใหญ่ได้ โดยพิจารณาจากความลำเอียงทางสถิติและความคลาดเคลื่อนมาตรฐานในการเทียบคะแนนที่มีค่าน้อยที่สุด ผลการวิจัยครั้งนี้จึงเป็นประโยชน์อย่างยิ่งในการนำหลักการเทียบคะแนนไปใช้ เพราะเมื่อทราบขนาดกลุ่มตัวอย่างที่เหมาะสมเพียงพอก็จะเป็นการประหยัดค่าใช้จ่ายและเวลาได้เป็นอย่างดี

ขอบเขตของการศึกษาค้นคว้า

1. ประชากร

ประชากรที่ใช้ในการศึกษาค้นคว้าครั้งนี้ คือนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดกรุงเทพมหานคร เขตบางเขน เขตจตุจักร และเขตดอนเมือง

2. ประชากรเทียม (Pseudopopulations)

ประชากรเทียมที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2538 สังกัดกรุงเทพมหานคร เขตบางเขน จำนวน 42 โรงเรียน เขตจตุจักร จำนวน 1 โรงเรียน และเขตดอนเมือง จำนวน 1 โรงเรียน จำนวนนักเรียน 1,826 คน โดยแบ่งเป็น 2 กลุ่ม ตามขนาดของโรงเรียน เพื่อให้จำนวนนักเรียนมีขนาดใกล้เคียงกัน แล้วทำการสุ่มประชากรเทียมแต่ละกลุ่ม เป็นกลุ่มตัวอย่างขนาด 30 คน 60 คน 120 คนและ 250 คน ในกลุ่มตัวอย่างแต่ละขนาดสุ่มทั้งหมด 200 ครั้ง ได้กลุ่มตัวอย่างที่เป็นกลุ่มย่อยทั้งหมด 800 กลุ่มตัวอย่าง ดังตารางต่อไปนี้

ตาราง 1 แสดงจำนวนประชากรเทียม

กลุ่มที่ 1		กลุ่มที่ 2	
ชื่อโรงเรียน	จำนวนนักเรียน	ชื่อโรงเรียน	จำนวนนักเรียน
1. ไทยนิยมสงเคราะห์	336	1. ฤทธิยะวรรณาลัย	483
2. ประชานิเวศน์	265	2. วัดราชบูรณะนิมิตธรรม	110
3. ประชาภิบาล	110	3. บ้านคลองบัว	99
4. วัดเกาะสุวรรณาราม	96	4. ประชานุกูล	75
5. สายไหม	56	5. บางเขน (ไว้สา lions สรณ์)	71
6. บ้านบัวมล	25	6. หนองใหญ่	49
7. พระพรหมประสิทธิ์	25	7. ออเงิน	26
รวม	913	รวม	913

2. กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2538 สังกัดกรุงเทพมหานคร เขตบางเขน เขตจตุจักร เขตดอนเมือง ซึ่งได้มาโดยการสุ่มอย่างง่ายจากประชากรเทียมทั้งหมด โดยมีนักเรียนเป็นหน่วยการสุ่ม ซึ่งมีวิธีการสุ่มดังนี้

2.1 จากประชากรเทียมกลุ่มที่ 1 สุ่มแบบใส่คืน (With-replacement) โดยสุ่มกลุ่มตัวอย่างขนาด 30 คน 60 คน 120 คน และ 250 คน ขนาดละ 200 ครั้งรวมกลุ่มตัวอย่างทั้งหมด 800 กลุ่มตัวอย่าง

2.2 จากประชากรเทียมกลุ่มที่ 2 ก็กระทำการสุ่มเช่นเดียวกับในกลุ่มที่ 1 จะทำให้ได้กลุ่มตัวอย่างจากประชากรเทียมกลุ่มที่ 2 จำนวน 800 กลุ่มตัวอย่าง เช่นเดียวกัน

2.3 กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้จึงมีทั้งสิ้น 1,600 กลุ่มตัวอย่าง

3. ตัวแปรที่ใช้ในการศึกษาครั้งนี้

3.1 ตัวแปรอิสระ คือ ขนาดของกลุ่มตัวอย่างซึ่งแบ่งเป็น 4 ขนาด

3.1.1 ขนาด 30 คน

3.1.2 ขนาด 60 คน

3.1.3 ขนาด 120 คน

3.1.4 ขนาด 250 คน

3.2 ตัวแปรตาม คือ

3.2.1 คะแนนเทียบเฉลี่ย

3.2.2 ความลำเอียงทางสถิติ

3.2.3 ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน

นิยามศัพท์เฉพาะ

1. ความลำเอียงทางสถิติ หมายถึง ความแตกต่างระหว่างค่าเฉลี่ย ของคะแนนเทียบเฉลี่ยของกลุ่มตัวอย่าง 200 กลุ่ม กับคะแนนเทียบของประชากร

2. ความคลาดเคลื่อนมาตรฐานในการเทียบคะแนน หมายถึง รากที่สองของความแปรปรวนของคะแนนเทียบ ที่คำนวณจากกลุ่มตัวอย่าง 200 กลุ่ม

3. การเทียบคะแนน (Test Equating) หมายถึง การแปลงคะแนนดิบ ที่ทดสอบด้วยแบบทดสอบวัดการรับรู้ 2 ฉบับ ที่ต่างกันมาอยู่ในคะแนนมาตรฐานเดียวกันหรือหน่วยวัดเดียวกัน

4. การเทียบคะแนนเชิงเส้นตรงแบบคะแนนจริงสัมพันธ์ (Congeneric Linear Equating) เป็นวิธีการเปรียบเทียบคะแนนจากแบบทดสอบ 2 ฉบับ ที่มีความสัมพันธ์เป็นเส้นตรง และมีความเชื่อมั่นเท่ากัน การวิจัยครั้งนี้ผู้วิจัยใช้แบบทดสอบวัดการรับรู้ 2 ฉบับ

5. การรับรู้ (Perception) หมายถึง ความสามารถที่เกิดจากการได้สัมผัสทางตา หู จมูก ลิ้น ผิวกาย แล้วแสดงออกในรูปการกระทำใดการกระทำหนึ่งที่มีความหมายในงานวิจัยนี้ ใช้ความสามารถที่เกิดจากสัมผัสทางตา คือการมองเห็นความแตกต่างของรูปภาพต่าง ๆ อย่างรวดเร็วและแม่นยำ ซึ่งวัดได้โดยแบบทดสอบวัดการรับรู้ที่แบ่งเป็น 2 ตอน

ตอนที่ 1 ให้หาภาพที่เหมือนกับภาพที่กำหนดให้

ตอนที่ 2 ให้หาภาพที่ไม่เหมือนกับภาพที่กำหนดให้

6. แบบทดสอบเชื่อมโยง (External Anchor Test) หมายถึงแบบทดสอบที่นำไปสอบกับกลุ่มประชากรทั้ง 2 กลุ่ม ที่ใช้เป็นเกณฑ์เชื่อมโยงคะแนนในการแปลงคะแนนจากแบบทดสอบฉบับหนึ่ง ไปยังอีกฉบับหนึ่ง การวิจัยครั้งนี้ใช้แบบทดสอบวัดการรับรู้ จำนวน 30 ข้อ

7. กลุ่มตัวอย่างขนาดเล็ก (Small Sample) หมายถึง กลุ่มตัวอย่างที่มีขนาดเล็กกว่าประชากรที่ยอมรับเลือกมาศึกษา โดยในงานวิจัยนี้ใช้ฐานจากจำนวนห้องเรียนขนาด 1, 2, 4 และ 8 ห้องเรียน เป็นฐานกำหนดจำนวนนักเรียนขนาด 30 คน 60 คน 120 คน และ 250 คน

8. คะแนนเทียบเฉลี่ย หมายถึง คะแนนที่เกิดจากการแปลงคะแนนดิบแต่ละตัวของกลุ่มตัวอย่างด้วยสูตรของแมคเคน จากการสุ่มแต่ละครั้งรวมทั้งหมด 200 ครั้ง แล้วนำมาหาค่าเฉลี่ย

9. คะแนนสังเกต หมายถึง ผลการสอบของนักเรียนที่ทำแบบทดสอบเทียบคะแนนฉบับ X ที่มีค่าตั้งแต่ 0 ถึง 60 คะแนน

เอกสารและงานวิจัยที่เกี่ยวข้อง

การศึกษครั้งนี้ ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องดังต่อไปนี้

1. ความหมายของการเทียบคะแนน
2. มีนัยสำคัญของการเทียบคะแนน
3. แบบจำลองของการวัด
4. รูปแบบการเก็บรวบรวมข้อมูล
5. วิธีการเทียบคะแนน
6. การเทียบคะแนนเชิงเส้นโดยใช้แบบทดสอบหลัก
7. แบบทดสอบหลัก
8. ความคลาดเคลื่อนของการเทียบคะแนน
9. ความลำเอียงทางสถิติ
10. ทฤษฎีเซาวันปัญหาและความถนัด
11. ความหมายของการรับรู้
12. ปัญหาเกี่ยวกับการรับรู้
13. องค์ประกอบที่มีส่วนเกี่ยวข้องกับการรับรู้
14. แบบทดสอบวัดการรับรู้ตามแนวคิดของเทอร์สตัน
15. งานวิจัยที่เกี่ยวข้อง

ความหมายของการเทียบคะแนน

กัลลิกเซน (ภาวิณี ศรีสุขวัฒนา. 2529:174; อ้างอิงมาจาก Gulliksen 1950:298-304) ได้ให้ความหมายของการเทียบระดับคะแนนว่า เป็นวิธีการทำคะแนนที่ได้จากแบบทดสอบสองชุดของวิชาเดียวกันให้เป็นคะแนนที่สมมูลกันเปรียบเทียบกันโดยตรง

ฟลานาแกน (Flanagan. 1951:751-753) ได้กล่าวว่า การเทียบระดับคะแนนเป็นวิธีการทำคะแนนจากแบบทดสอบต่างชุดกันให้มีคุณลักษณะที่เทียบกันได้

แองกอฟ (Angoff. 1982:56) ได้ให้คำจำกัดความว่า การเทียบระดับคะแนนเป็นวิธี

การแปลงระบบของหน่วยหนึ่งของแบบทดสอบ ไปสู่ระบบของหน่วยวัดอีกหน่วยหนึ่งในรูปฟังก์ชันทางด้านจิตวิทยา ที่สามารถนำไปประยุกต์ใช้กับกลุ่มอื่น ๆ ได้

ลอร์ด (Lord, 1980:195) ได้กล่าวว่าการเทียบระดับคะแนน เป็นการนำคะแนนจากแบบทดสอบ X และแบบทดสอบ Y ที่มีความยากต่างกันจากความสามารถเดียวกันมาเทียบกัน

ชูศักดิ์ ชัมภลิต (2527 : 2) ได้สรุปเกี่ยวกับการเทียบระดับคะแนน ว่าเป็นกระบวนการที่เกี่ยวข้องกับกิจกรรม 2 ประการ คือ

1. กระบวนการที่ทำให้แบบทดสอบ 2 ฉบับใด ๆ มีความเท่าเทียมกันหรือเท่ากันในเชิงของโครงสร้าง

2. การใช้วิธีการทางสถิติเพื่อปรับ (Adjust) คะแนนที่ได้จากแบบทดสอบแต่ละฉบับให้อยู่ในมาตราเดียวกัน และเทียบกันได้

สงขลิก ละครณะ (2525 : 22) ได้กล่าวว่า คะแนนจากแบบทดสอบ 2 ฉบับวัดสิ่งเดียวกัน แต่ไม่จำเป็นต้องเป็นข้อสอบคู่ขนานกัน จะถือว่าเทียบเท่ากันได้ถ้าคะแนนจากแบบทดสอบทั้ง 2 นั้นมาจากคะแนนจริง (True Score) หรือความสามารถแท้ (True ability) ที่เท่ากัน

วิรัช วรณรัตน์ (2530 : 69) กล่าวว่า การเทียบคะแนนเป็นการนำเอาคะแนนที่สังเกตได้จากแบบสอบต่างฉบับ ซึ่งแต่ละฉบับต่างก็วัดในสิ่งเดียวกันมาเทียบกัน

จากความหมายดังกล่าวสรุปได้ว่า การเทียบคะแนนเป็นวิธีการนำคะแนนจากแบบทดสอบต่างชุดกัน ซึ่งวัดคุณลักษณะเดียวกัน นำไปแปลงคะแนนให้มีหน่วยวัดเดียวกัน แล้วนำมาเปรียบเทียบกัน

มโนทัศน์ของการเทียบคะแนน

ในการเทียบคะแนนสอบจากแบบทดสอบฉบับหนึ่ง ไปยังอีกฉบับหนึ่งนั้น มีหลักการสำคัญดังมีรายละเอียดต่อไปนี้

1. สมมติให้ประชากร P แทนประชากรของผู้สอบ ซึ่งอาจสอบแบบสอบฉบับ X หรือแบบสอบฉบับ Y
2. กำหนดให้ $F(x)$ และ $G(y)$ แทน ฟังก์ชันการกระจาย (distribution function) ของคะแนนฉบับ X และฉบับ Y ตามลำดับ

3. สมมติให้ฟังก์ชันการเทียบมาตรา e_x แทนฟังก์ชันของการเทียบคะแนนจากแบบสอบฉบับ Y ไปสู่แบบสอบฉบับ X

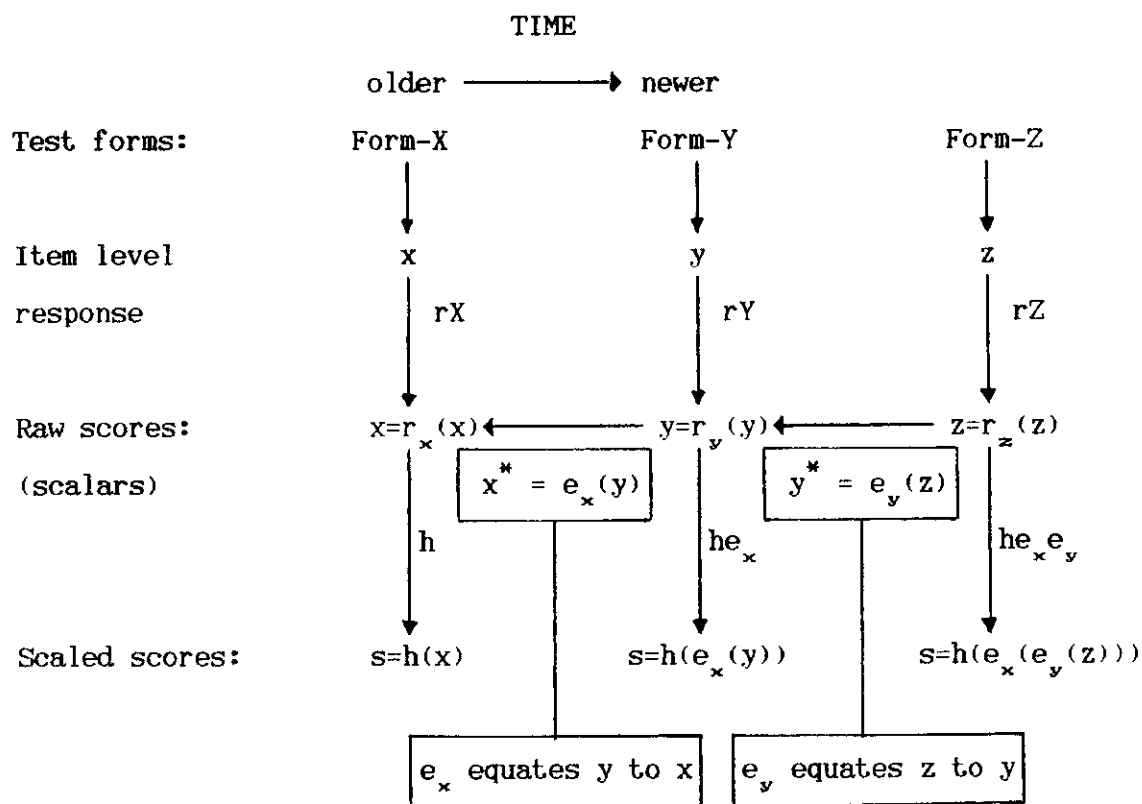
$$\text{ฉะนั้น } X^* = e_x(y)$$

4. ให้ $F^*(x)$ เป็นฟังก์ชันการกระจายของ X^*

$$\therefore F^*(x) = G(e_x^{-1}(x))$$

คำนิยาม

"ถ้า $F = F^*$ แล้ว แบบสอบฉบับ X และฉบับ Y สามารถเทียบโดยฟังก์ชัน e_x ในประชากร p "



ภาพประกอบ 1 ลำดับของการเทียบคะแนนของแบบทดสอบและความสัมพันธ์ของแต่ละลำดับ (Braun and Holland. 1982 : 11)

จากภาพที่ 1 อธิบายความหมายได้ดังนี้

"ระดับการตอบรายข้อ" (Item level responses) เป็นเวกเตอร์ของคะแนนรายข้อ โดยส่วนใหญ่แล้วจะใช้ 1 แทนการตอบถูก 0 แทนการตอบผิด และเลขอื่นๆแทนข้อที่เว้นว่าง (omitted) ฉะนั้นแบบทดสอบฉบับ x จะมีเวกเตอร์ของการตอบรายข้อ x แทนคะแนนของแต่ละคนที่สอบแบบทดสอบฉบับ x

"คะแนนดิบ" (Raw score) เป็นปริมาณตัวเลขที่ได้จากการตอบรายข้อของแต่ละคน ซึ่งคำนวณจากฟังก์ชัน r โดยทั่วไป ฟังก์ชัน r ใช้โดยวิธีใดวิธีหนึ่ง คือ

$$r = \text{จำนวนข้อถูก} \quad \text{หรือ}$$

$$r = \text{จำนวนข้อถูก} - \text{คะแนนการเดา} \quad (\text{แบบสอบชนิดเลือกตอบ})$$

ในที่นี้คะแนนดิบใช้สัญลักษณ์ x แทน คะแนนของแบบทดสอบฉบับ X , y แทน คะแนนของแบบทดสอบฉบับ Y และ z แทน คะแนนของแบบทดสอบฉบับ Z

"ฟังก์ชันการเทียบคะแนน" (Equating function) เป็นกฎเกณฑ์หรือฟังก์ชันที่กำหนดในการเทียบคะแนนจากแบบทดสอบฉบับใหม่ (Y) ให้อยู่ในมาตราเดียวกับแบบทดสอบฉบับเดิม (X) ใช้สัญลักษณ์ e เช่น

e_x เป็นฟังก์ชันที่เทียบคะแนนจากฉบับ Y ไปสู่ฉบับ X

e_y เป็นฟังก์ชันที่เทียบคะแนนจากฉบับ Z ไปสู่ฉบับ Y

คะแนน y ใด ๆ เมื่อใช้ฟังก์ชัน e_x เปลี่ยนคะแนนแล้วมักใช้สัญลักษณ์ x^* แทนเพื่อเป็นการง่ายต่อความเข้าใจว่า $x^* = e_x(y)$ คือ คะแนน x แต่มาจากแบบทดสอบฉบับอื่น

"ฟังก์ชันของมาตราวัด" (The Scaling Function) บางครั้งการรายงานคะแนนการสอบ นิยมรายงานคะแนนในรูปแบบของคะแนนมาตราวัด (Scaled score) ที่ง่ายต่อการเข้าใจและแปลความหมาย ฉะนั้นการเทียบคะแนนจำเป็นต้องเทียบในรูปแบบของคะแนนมาตราวัดแต่มีข้อจำกัดว่า ฟังก์ชันของมาตราวัดที่ใช้จะต้องเป็นฟังก์ชันของมาตราวัดเดียวกัน ในภาพประกอบ 1 ฟังก์ชันของมาตราวัดคือ s

"ฟังก์ชันการแปลงคะแนน" (The Conversion Function) คือ ฟังก์ชันที่แปลงคะแนนดิบเป็นคะแนนมาตรฐาน ฉะนั้นสำหรับการสอบฉบับ X ฟังก์ชันของมาตราวัด และฟังก์ชันการแปลงคะแนน ให้เป็นฟังก์ชันเดียวกัน ดังนี้

$$\text{Test X : } s = h(x)$$

$$\text{Test Y : } s = h(e_x(y))$$

$$\text{Test Z : } s = h(e_x(e_y(z)))$$

แบบจำลองของการวัด

แนวคิดของแบบทดสอบคู่ขนาน เป็นหัวใจสำคัญที่ชี้บ่งว่าแบบทดสอบ 2 ฉบับใดที่วัดในสิ่งเดียวกัน แต่ระดับความคู่ขนานของแบบทดสอบนั้น อาจจำแนกเป็น 3 ระดับ คือ

ความคู่ขนานระดับคะแนนจริงมาตรฐานเดิม (Classical Parallel Test) คือ แบบทดสอบสองฉบับ ที่มีข้อตกลงว่า 1. วัดคุณลักษณะเดียวกัน 2. คะแนนจริงแต่ละฉบับต้องเท่ากัน 3. มีค่าความยากเท่ากัน 4. มีการกระจายของคะแนนเท่ากัน 5. มีความแปรปรวนร่วมของคะแนนสอบแต่ละคู่เท่ากัน และ 6. มีความแปรปรวนร่วมของแบบทดสอบทั้งสองฉบับเท่ากัน

ความคู่ขนานระดับคะแนนจริงสมมูล (Tau-Equivalent Test) คือ แบบทดสอบสองฉบับ ต้องมีลักษณะต่อไปนี้ 1. ความแปรปรวนใกล้เคียงกัน 2. ค่าเฉลี่ยของคะแนนใกล้เคียงกัน 3. ความแปรปรวนระหว่างตอนเท่ากัน 4. ความแปรปรวนร่วมภายในเท่ากัน

และความคู่ขนานระดับคะแนนจริงสัมพันธ์ (Congeneric Test) ซึ่งกำหนดว่าแบบทดสอบแต่ละฉบับต้องวัดฟังก์ชันทางจิตวิทยาลักษณะเดียวกัน (The Same Psychological Function) หรือ วัดมิติเดียวกัน (Unidimensionality) หรือ วัดคุณลักษณะเดียวกัน (The Same Trait) โดยที่คะแนนจริงแต่ละฉบับมีความสัมพันธ์เป็นเส้นตรงเท่านั้น

ความคู่ขนานของ 2 ระดับแรก มีเงื่อนไขที่ค่อนข้างจะปฏิบัติตามได้ยาก ดังนั้นการเทียบคะแนนโดยใช้ข้อตกลงของความคู่ขนานแบบคะแนนจริงสัมพันธ์ จึงมีความเป็นไปได้ในทางปฏิบัติมากที่สุด (บุญเชิด บุญโณนตพงษ์. 2538 : 1)

รูปแบบการเก็บรวบรวมข้อมูล

ข้อมูลที่ได้จากแบบทดสอบต่างฉบับกัน บางครั้งไม่สามารถนำมาเทียบคะแนนกันได้จะต้อง

มีวิธีการเก็บรวบรวมที่เหมาะสม จึงจะสามารถนำคะแนนมาเทียบกันได้ ซึ่ง มาร์โค (Marco. 1981) ได้รวบรวมไว้ทั้งหมด 5 วิธี ดังนี้ คือ

1. กลุ่มนักเรียนกลุ่มเดียวตอบแบบทดสอบ 2 ชุด คือชุดใหม่กับชุดเก่าการทดสอบจะกระทำในวันเดียวหรือ 2 วันติดต่อกัน แต่หลักการที่สำคัญคือ ช่วงระยะห่างของการทำแบบทดสอบ 2 ชุด ต้องเป็นเวลาล้นๆ เพื่อมิให้เกิดประสบการณ์แทรกซ้อนที่มีผลกระทบต่อคะแนน ลำดับของการทดสอบชุดใดก่อนไม่ใช่ปัญหา เพราะถือว่าองค์ประกอบต่าง ๆ อัน ได้แก่ การเรียนรู้ การฝึกฝน ความล้า มีผลต่อคะแนนน้อยมาก

2. กลุ่มนักเรียน 2 กลุ่ม แต่ละกลุ่มได้รับการทดสอบทั้ง 2 ชุด ในลักษณะการจัดลำดับการสอบก่อน-หลัง สลับให้เกิดความสมดุลกัน แบบแผนที่ 2 นี้ ดัดแปลงมาจากแบบแผนที่ 1 เป็นแบบแผนที่จะรับประกันอิทธิพลขององค์ประกอบการเรียนรู้ การฝึกฝน และความล้า เป็นการขจัดความลำเอียงที่จะเกิดขึ้นกับผลการสอบชุดใดชุดหนึ่ง

3. กลุ่มนักเรียน 2 กลุ่มที่ได้มาโดยการสุ่มแต่ละกลุ่มแล้วสอบแบบทดสอบฉบับใดฉบับหนึ่งเพียงฉบับเดียว สิ่งสำคัญของแบบที่ 3 คือต้องมีกลุ่มสุ่ม 2 กลุ่ม ที่มีความคล้ายคลึงกันด้านความรู้ความสามารถ หรือ ทักษะที่ต้องการวัด เพื่อให้คะแนนสอบที่ได้มา ไม่ใช่เป็นผลของความแตกต่างของความสามารถของกลุ่ม The Graduate Management Admission Test, The Law School Admission Test และ The GRE Aptitude Test ได้ใช้รูปแบบการเก็บข้อมูลชนิดนี้ ในการเทียบคะแนน

4. กลุ่มนักเรียน 2 กลุ่มที่ได้มาจากการสุ่ม โดยแต่ละกลุ่มจะตอบแบบทดสอบฉบับใดฉบับหนึ่งเพียงฉบับเดียว และทั้งสองกลุ่มต้องตอบแบบทดสอบหลัก (Anchor Test) แบบทดสอบหลักนี้อาจแยกสอบต่างหากจากแบบทดสอบที่ใช้เทียบคะแนน เรียกว่าแบบทดสอบเชื่อมโยง (External Anchor Test) หรืออาจนำมารวมกับแบบทดสอบเทียบคะแนนแต่ละฉบับ เรียกว่าข้อสอบร่วม (Internal Anchor Test) ส่วนที่เป็นแบบสอบหลักควรประกอบด้วยคำถามที่คล้ายคลึงกับคำถามในแบบทดสอบที่ต้องการเทียบมาตราทั้งสองชุด ประโยชน์จากแบบสอบหลักที่มีต่อการเทียบมาตราของแบบสอบทั้งสองชุดจะมีมากหรือน้อย ขึ้นอยู่กับความสัมพันธ์ของแบบสอบหลักที่มีต่อแบบทดสอบสองชุดนั้น เหตุที่แบบที่ 4 ได้เปรียบ เพราะคะแนนชุดเก่าและใหม่สามารถปรับ เพื่อสะท้อนให้เห็นความแตกต่าง ที่ปรากฏอยู่ในกลุ่มสุ่มทั้งสองตามผลของแบบสอบหลักได้

5. กลุ่มนักเรียนไม่ได้มาจากการสุ่ม 2 กลุ่มแต่ละกลุ่มทำแบบทดสอบเพียงชุดเดียว และทำแบบสอบหลักเหมือนกันอีก เช่นเดียวกับ แบบแผนที่ 4 กลุ่มผู้สอบที่ไม่ได้มาจากการสุ่ม มักจะ

เกิดขึ้นกับการสอบตามธรรมชาติ ปัจจัยสำคัญของแบบแผนนี้อยู่ที่แบบสอบหลัก ซึ่งต้องมีความคล้ายคลึงมากที่สุด กับแบบทดสอบ 2 ชุด ที่ต้องการเทียบมาตราโดยปกติแล้วไม่มีวิธีการทางสถิติใดที่จะทำการปรับคะแนนในกรณีที่กลุ่มตัวอย่างไม่ได้มาจากการสุ่ม แต่จะทำได้ ก็ต่อเมื่อมีแบบสอบหลักที่มีความเป็นคู่ขนานกับแบบทดสอบชุดใหม่และเก่า หน้าที่ของแบบสอบหลัก คือ กำหนดจุดอ้างอิงให้เกิดขึ้นกับทั้งสองชุดในทำนองเดียวกับการกำหนดจุดตัด หรือ จุดเยือกแข็งของน้ำเพื่อใช้กำหนดมาตรานเทอร์โมมิเตอร์แบบฟาร์เรนไฮต์ให้ เท่ากับ องศาบนเทอร์โมมิเตอร์แบบเซนเซียสได้ แบบนี้ใช้กับการเทียบมาตราในแบบทดสอบ SAT มาแล้ว

วิธีการเก็บรวบรวมข้อมูลดังกล่าวคล้ายคลึงกับวิธีการที่ ปีเตอร์เซน และคนอื่น ๆ

(Petersen and others. 1989:244-246) ได้กล่าวไว้คือ

1. วิธีเก็บรวบรวมข้อมูลแบบผู้สอบกลุ่มเดียว (Single Group Design) รูปแบบนี้เป็นรูปแบบที่ง่ายที่สุด การเก็บรวบรวมข้อมูลแบบนี้ ใช้กลุ่มผู้สอบกลุ่มเดียวโดยให้ผู้สอบทำแบบทดสอบฉบับหนึ่ง แล้วตามด้วยแบบทดสอบต่างฉบับ แต่การสอบครั้งหลัง ผู้สอบอาจได้รับผลจากการสอบครั้งแรกเนื่องจากการเรียนรู้ ความเมื่อยล้า และการฝึกฝน

กลุ่มตัวอย่าง	แบบทดสอบ	
	ฉบับที่ 1	ฉบับที่ 2
P1	/	/

2. วิธีการเก็บรวบรวมข้อมูลแบบผู้สอบกลุ่มเดียว ที่ได้รับการจัดให้สมดุล (Counterbalanced Random Groups Design) ในทางปฏิบัติลำดับของการสอบก่อน-หลัง ย่อมส่งผลต่อผลการสอบ ดังนั้นรูปแบบนี้จึงปรับมาจากรูปแบบแรกโดยจัดผู้สอบเป็นกลุ่มสุ่ม 2 กลุ่ม แต่ละกลุ่มย่อยได้รับการสอบด้วยแบบทดสอบทั้ง 2 ฉบับ ในลักษณะการจัดลำดับก่อน-หลัง สลับให้เกิดความสมดุลการจัดรูปแบบนี้ เพื่อให้การสอบโดยใช้แบบทดสอบทั้ง 2 ฉบับ ได้รับอิทธิพลจากองค์ประกอบในเรื่องการเรียนรู้ความเมื่อยล้า และการฝึกฝนในลักษณะที่สมดุลกัน

กลุ่มตัวอย่าง	แบบทดสอบ			
	ฉบับที่ 1		ฉบับที่ 2	
	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 1	ครั้งที่ 2
P1	/			/
P2		/	/	

3. วิธีการเก็บรวบรวมข้อมูลแบบผู้สอบกลุ่มเท่าเทียมกัน (Equivalent Group Design) รูปแบบนี้จำเป็นต้องจัดกลุ่มผู้สอบให้มีความคล้ายคลึงกันมากเท่าที่จะทำได้ ผู้สอบในแต่ละกลุ่มจะทำแบบทดสอบฉบับเดียว ดังนั้นจึงหลีกเลี่ยงปัญหาเรื่องการเรียนรู้ ความเมื่อยล้า และการฝึกฝนจากการสอบครั้งแรกได้ แต่จะไม่มีข้อมูลที่จะนำมาใช้เปรียบเทียบความแตกต่างของกลุ่มตัวอย่างทั้งสอง วิธีการที่จะลดความแตกต่างระหว่างกลุ่ม ทำได้โดยใช้กลุ่มตัวอย่างขนาดใหญ่

กลุ่มตัวอย่าง	แบบทดสอบ	
	ฉบับที่ 1	ฉบับที่ 2
P1	/	
P2		/

4. วิธีการเก็บรวบรวมข้อมูลแบบผู้สอบกลุ่มละสองกลุ่ม โดยมีแบบทดสอบหลัก (Anchor Test Random Groups Design) รูปแบบกลุ่มละแต่ละกลุ่ม ทำแบบทดสอบเทียบคะแนนกลุ่มละ 1 ฉบับ และทั้ง 2 กลุ่ม จะได้รับแบบทดสอบหลักเพื่อปรับความแตกต่างระหว่างกลุ่ม

กลุ่มตัวอย่าง	แบบทดสอบ		
	ฉบับที่ 1	ฉบับที่ 2	แบบทดสอบหลัก
P1	/		/
P2		/	/

5. วิธีการเก็บรวบรวมข้อมูลแบบผู้สอบกลุ่มไม่เท่าเทียมกัน โดยมีแบบทดสอบหลัก (Anchor Test Non-equivalent Groups Design) เป็นรูปแบบที่ผู้สอบไม่ได้มาจากประชากรกลุ่มเดียวกัน สอบโดยใช้แบบทดสอบเทียบคะแนนต่างฉบับกัน เพื่อปรับความแตกต่างระหว่างกลุ่ม และลดความลำเอียงในการเทียบคะแนน จึงใช้แบบทดสอบหลัก

กลุ่มตัวอย่าง	แบบทดสอบ		
	ฉบับที่ 1	ฉบับที่ 2	แบบทดสอบหลัก
P1	/		/
Q1		/	/

6. วิธีการเก็บรวบรวมข้อมูลแบบผู้สอบกลุ่มเท่าเทียมกัน โดยแบ่งแบบทดสอบออกเป็น ส่วนๆ (Section Pre-equating Design with One Variable Section) วิธีการนี้จะแบ่งแบบทดสอบออกเป็น ส่วนๆ ใช้กับกลุ่มตัวอย่างที่มาจากประชากรเดียวกัน ลักษณะของแบบทดสอบแต่ละ ส่วนของแต่ละฉบับ จะเป็นแบบทดสอบคู่ขนานกันทั้งรูปแบบและเนื้อหา ผู้สอบจะทำแบบทดสอบ ฉบับที่ 1 ทุกส่วน แต่จะทำแบบทดสอบฉบับที่ 2 เพียงหนึ่งส่วน

			แบบทดสอบ					
			ฉบับที่ 1			ฉบับที่ 2		
กลุ่มตัวอย่าง			ส่วนที่ 1	ส่วนที่ 2	ส่วนที่ 3	ส่วนที่ 1	ส่วนที่ 2	ส่วนที่ 3
P1			/	/	/	/		
P2			/	/	/		/	
P3			/	/	/			/

7. วิธีการเก็บรวบรวมข้อมูลแบบผู้สอบกลุ่มเท่าเทียมกัน โดยสอบกับแบบทดสอบฉบับที่สอง 2 ส่วน (Section Pre-equating Design with Two Variable Section) ลักษณะของแบบทดสอบคล้ายกับแบบที่ 6 ต่างกันที่ผู้สอบทุกกลุ่มจะทำฉบับที่ 2 สองส่วน

แบบทดสอบ								
ฉบับที่ 1					ฉบับที่ 2			
กลุ่มตัวอย่าง	ส่วนที่ 1	ส่วนที่ 2	ส่วนที่ 3	ส่วนที่ 4	ส่วนที่ 1	ส่วนที่ 2	ส่วนที่ 3	ส่วนที่ 4
P1	/	/	/	/	/	/		
P2	/	/	/	/	/		/	
P3	/	/	/	/	/			/
P4	/	/	/	/		/	/	
P5	/	/	/	/		/		/
P6	/	/	/	/			/	/

8. วิธีการเก็บรวบรวมข้อมูลแบบผู้สอบกลุ่มไม่เท่าเทียมกัน โดยมีแบบทดสอบแบ่งเป็นส่วนๆ (Item Pre-Equating Design) วิธีการนี้ใช้กลุ่มประชากรที่แตกต่างกันกับแบบทดสอบหลายฉบับ โดยแต่ละฉบับแบ่งออกเป็น 2 ส่วนที่คู่ขนานกันทั้งเนื้อหาและรูปแบบ กลุ่มตัวอย่างแต่ละกลุ่มจะทำแบบทดสอบที่ใช้เทียบคะแนน 2 ส่วน และ อีก 1 ส่วนที่แปรเปลี่ยนไปจากฉบับอื่น วิธีนี้เหมาะที่จะใช้กับการเทียบคะแนนแบบ IRT

กลุ่มตัวอย่าง	แบบทดสอบ								
	V1	W1	W2	X1	X2	Y1	Y2	Z1	Z2
P1	/	/	/						
P2		/	/			/			
P3		/	/						/
Q1	/			/	/				
Q2				/	/		/		
Q3				/	/			/	

จากการศึกษาวิธีเก็บรวบรวมข้อมูลดังกล่าวมาแล้ว ผู้วิจัยใช้รูปแบบวิธีรวบรวมข้อมูลแบบใช้แบบทดสอบหลักที่มีประชากรแตกต่างกัน (รูปแบบที่ 5) ในงานวิจัยนี้

วิธีการเทียบคะแนน

การจะนำคะแนนของแบบทดสอบฉบับ X และฉบับ Y มาเทียบกันได้ั้น ลอร์ด (Petersen, Kolen and Hoover. 1989 : 242 ; citing Lord. 1980) ได้กล่าวว่า ต้องเป็นไปตามเงื่อนไข 4 ประการ คือ

1. วัดความสามารถเดียวกัน หมายถึง แบบทดสอบทั้ง 2 ฉบับต้องวัดในคุณลักษณะเดียวกัน (คุณลักษณะภายใน, ความสามารถหรือทักษะ)
2. มีความยุติธรรม หมายถึง ผู้สอบทุกกลุ่มมีความสามารถเหมือนกัน โดยความถี่ของการกระจายคะแนนในแบบทดสอบฉบับ Y หลังจากการแปลงคะแนนจะเท่ากับ การกระจายในแบบทดสอบฉบับ X
3. ไม่ผันแปรไปตามกลุ่มประชากร คือ การแปลงคะแนนจะให้ค่าเดียวกันไม่ว่าจะใช้ประชากรกลุ่มใด
4. มีความสมมาตร คือคะแนนแปลงต้องสามารถเปลี่ยนกลับไปได้ กล่าวคือการจับคู่ของคะแนนที่แปลงจาก X ไป Y ต้องเหมือนกับที่แปลงจาก Y ไป X

ฟลานาแกน (Flanagan. 1951:751-753) ได้กล่าวถึงการเทียบคะแนนทางสถิติไว้ 4 วิธีคือ 1. การเทียบค่าเฉลี่ย 2. การใช้เทคนิคและสมการถดถอย 3. ใช้คะแนนมาตรฐาน 4. ใช้เทียบคะแนนที่ใช้ตำแหน่งที่เท่ากัน (Equipercntile Method)

สำหรับ เรวดี อินทสระ (2530 : 10-21) ได้สรุป วิธีการเทียบคะแนนไว้มีดังต่อไปนี้

1. วิธีเทียบคะแนนเชิงเส้นตรง

วิธีนี้ได้รับความนิยมเป็นเวลานานมาแล้วโดยใช้สมการเส้นตรง เป็นรูปแบบอยู่ในนิยามที่ว่า คะแนนจากแบบทดสอบ 2 ชุด โดยที่แบบทดสอบทั้ง 2 ชุด อาจมีความคู่ขนานแบบคะแนนจริงมาตรฐานเดิม หรือมีความคู่ขนานแบบคะแนนจริงสมมูลแล้ว กำหนดให้คะแนนมาตรฐานทั้งสองชุดมีค่าเท่ากัน คือ

$$(Y - M_y) / S_y = (X - M_x) / S_x$$

เมื่อ X คือคะแนนจากแบบทดสอบชุดที่ 1

Y คือคะแนนจากแบบทดสอบชุดที่ 2

- M_x คือค่าเฉลี่ยของแบบทดสอบชุดที่ 1
 M_y คือค่าเฉลี่ยของแบบทดสอบชุดที่ 2
 S_x คือค่าความเบี่ยงเบนของแบบทดสอบชุดที่ 1
 S_y คือค่าความเบี่ยงเบนของแบบทดสอบชุดที่ 2

ในการนิรูปแบบการแจกแจงของคะแนนแบบทดสอบชุดที่ 1 เหมือนกับแบบทดสอบชุดที่ 2 ใช้การเทียบรูปแบบเชิงเส้นตรงได้สะดวก ง่ายและมีความเป็นปรนัยเพราะจะเป็นไปตามข้อตกลงเบื้องต้น

2. วิธีเทียบคะแนนแบบอิกวิเปอร์ เซนไทล์

มีขั้นตอนดังนี้ รวบรวมข้อมูลที่ได้จากแบบทดสอบ ด้วยแบบทดสอบ 2 ชุดแล้วนำมาเทียบคะแนนตามรูปแบบ คือ

- 2.1 ทำตารางประมาณการแจกแจงความถี่ของกลุ่มตัวอย่างทั้งสองกลุ่ม
- 2.2 นำค่าความถี่ประมาณของกลุ่มตัวอย่างทั้งสองกลุ่มหาตำแหน่ง เปอร์เซนต์ไทล์ของแต่ละชั้นคะแนน

2.3 เขียนกราฟแสดงความสัมพันธ์ของคะแนนชุดที่ 1 และชุดที่ 2 เพื่อแสดงคะแนนที่สมมูล โดยการลากเส้นกราฟให้ผ่านจุดต่าง ๆ ให้มากที่สุดแล้วเกลารเส้นโค้งให้เรียบ

- 2.4 ทำตารางคะแนนสมมูลของคะแนนชุดที่ 1 และชุดที่ 2 โดยอ่านจากกราฟ

3. วิธีเทียบคะแนนแบบอิงทฤษฎีการตอบข้อคำถาม

มีขั้นตอนของการเทียบคะแนนดังนี้

3.1 ขั้นตอนการประมาณคะแนนจริง โดยกะประมาณค่าพารามิเตอร์ของข้อสอบและค่าความสามารถของข้อสอบรายข้อ ใช้กะประมาณค่าความเป็นไปได้สูงสุด และหาค่าประมาณคะแนนจริงของผู้สอบ

3.2 ขั้นตอนการเทียบมาตรา นำค่าคะแนนจริงของผู้สอบ มาจับคู่ที่ค่าความสามารถเดียวกันและปรับปรุงขยายคะแนนจริงที่ต่ำกว่าคะแนนการเดาให้มีความหมายขึ้น แล้วจึงนำคะแนนจริงของแบบทดสอบแต่ละชุดที่คะแนนความสามารถเดียวกันในแต่ละคู่สร้างเป็นตารางเทียบคะแนน

4. วิธีเทียบคะแนนแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ

มีขั้นตอนของการเทียบคะแนนดังนี้

- 4.1 รวบรวมข้อมูลที่ได้จากการสอบจากแบบทดสอบทั้งสองชุด

4.2 นำข้อมูลทั้งหมดไปวิเคราะห์องค์ประกอบสำคัญ เพื่อจะหาน้ำหนักขององค์ประกอบและค่าความร่วมกัน

4.3 กะประมาณคะแนนสมมูล (f) รายข้อ และรายบุคคลจากค่าสถิติในข้อ 2 เพื่อนำมาจัดตารางสมมูลระหว่างคะแนนดิบของแบบทดสอบทั้ง 2 ฉบับ

การเทียบคะแนนเชิงเส้นตรง โดยใช้แบบทดสอบหลัก

ผู้ที่คิดการเทียบคะแนนโดยใช้แบบทดสอบหลักในรูปแบบเชิงเส้นตรงคือเรดยาร์ด ทักเคอร์ (Ledyard Tucker) และได้ตั้งชื่อวิธีการเทียบคะแนนนี้ว่า วิธีเทียบคะแนนแบบทักเคอร์ (Tucker equating) เพื่อเป็นเกียรติแก่ผู้ริเริ่ม (Braun and Holland. 1982 : 23 ; citing Angoff. 1961) โดยเริ่มจากการตั้งข้อตกลงเบื้องต้นดังนี้

$$e_x(Y) = \mu_x + \frac{\delta_x}{\delta_y} (Y - \mu_y)$$

เมื่อ μ_x , δ_x , μ_y , และ δ_y คำนวณจากประชากรเทียบ T ในการคำนวณปริมาณมากนั้น ค่าเฉลี่ยและความแปรปรวนใช้สูตรคล้ายคลึงกัน

สูตรการหาค่าเฉลี่ยของ X ที่คำนวณจากประชากรเทียบคือ

$$(1) \quad \mu_x = \int E_p(X/v) dK_p(v)f + \int E_q(X/v) dK_q(v)(1-f)$$

ในขณะที่สูตรการหาความแปรปรวนของ X ที่คำนวณจากประชากรเทียบคือ

$$(2) \quad \sigma_x^2 = \int \text{Var}_p(X/v) dK_p(v)f + \int \text{Var}_q(X/v) dK_q(v)(1-f) + \int E_p^2(X/v) dK_p(v)f + \int E_q^2(X/v) dK_q(v)(1-f) - \mu_x^2$$

สูตรสำหรับ μ_y และ σ_y^2 ก็เช่นเดียวกันเพียงแต่เปลี่ยน X เป็น Y ในทุกที่ในสูตร

(1), (2)

$$E_p(X/v) = \int x dF_p(x/v)$$

$$\text{Var}_P(X/v) = \int [x - E_P(X/v)]^2 dF_P(x/v)$$

การหาค่าของ $E_Q(X/v)$ และ $\text{Var}_Q(X/v)$ ก็คล้ายกับสูตรที่คำนวณจากประชากร P การจะใช้สูตร (1)(2) ต้องมีข้อตกลงเบื้องต้น 2 ข้อ คือ

1. ค่าคาดหวัง และค่าความแปรปรวนของ X ที่มีผลต่อแบบทดสอบหลัก(V) ในประชากร P และ Q ต้องเหมือนกับ ค่าคาดหวังและค่าความแปรปรวนของ Y ที่มีผลต่อแบบทดสอบหลัก(V) ในประชากร P และ Q

2. การหาค่าของ $E_P(X/v)$, $E_Q(X/v)$, $\text{Var}_P(X/v)$ และ $\text{Var}_Q(X/v)$ ทำโดยการวิเคราะห์การถดถอยโดยมีค่าคาดหวังเป็นเส้นตรง และความแปรปรวนเป็นค่าคงที่ เช่น

$$E_P(X/v) = (a_{x/v,p})v + b_{x/v,p}$$

$$E_Q(Y/v) = (a_{x/v,q})v + b_{x/v,q}$$

$$\text{Var}_P(X/v) = \sigma_{x/v,p}^2$$

$$\text{Var}_Q(Y/v) = \sigma_{y/v,q}^2$$

สัญลักษณ์ที่ใช้ของการเทียบคะแนนแบบทักเคอร์ (Tucker Equating)

T	คือประชากรเทียบที่แบ่งเป็น 2 กลุ่ม คือกลุ่ม P และกลุ่ม Q
P	คือประชากรกลุ่มที่ 1 ที่ทำแบบทดสอบฉบับ X และแบบทดสอบหลัก V
Q	คือประชากรกลุ่มที่ 2 ที่ทำแบบทดสอบฉบับ Y และแบบทดสอบหลัก V
f	เป็นสัดส่วนน้ำหนักของประชากร P ใน T
1-f	เป็นสัดส่วนน้ำหนักของประชากร Q ใน T
E, Var	คือค่าคาดหวังและความแปรปรวนของประชากร
$K_P(v)$, $K_Q(v)$	เป็นฟังก์ชันการแจกแจงของแบบทดสอบหลักในประชากร P และ Q ตามลำดับ
$F_P(x/v)$, $F_Q(x/v)$	เป็นฟังก์ชันการกระจายอย่างมีเงื่อนไขของ X ที่มีผลต่อแบบทดสอบหลัก ในประชากร P และ Q ตามลำดับ

ขั้นตอนของการเทียบคะแนนเชิงเส้นตรง (Angoff. 1984 : 105)

1.1 หาค่าประมาณต่าง ๆ ตามสูตรดังนี้

$$\begin{aligned}\hat{\mu}_x &= M_{x1} + b_{xv1} (\hat{\mu}_v - M_{v1}) \\ \hat{\sigma}_x^2 &= S_{x1}^2 + b_{xv1}^2 (\hat{\sigma}_v^2 - S_{v1}^2) \\ \hat{\mu}_v &= M_{v2} + b_{yv2} (\hat{\mu}_x - M_{x2}) \\ \hat{\sigma}_v^2 &= S_{v2}^2 + b_{yv2}^2 (\hat{\sigma}_x^2 - S_{x2}^2)\end{aligned}$$

1.2 หาค่าประมาณของความชัน (A) และค่าคงที่ (B) จากสูตร

$$\begin{aligned}A &= b_{yv2} [S_{x1}^2 + b_{xv1}^2 (\hat{\sigma}_v^2 - S_{v1}^2)] \\ &\quad b_{xv1} [S_{v2}^2 + b_{yv2}^2 (\hat{\sigma}_x^2 - S_{x2}^2)] \\ B &= \hat{\mu}_x - A \hat{\mu}_v\end{aligned}$$

1.3 แทนค่าประมาณประชากรลงในสมการเส้นตรง

$$X^* = A_v + B$$

1.4 ใช้สมการสร้างตารางเทียบคะแนนจากแบบทดสอบชุดที่ 1 ไปสู่คะแนนของแบบทดสอบชุดที่ 2

แมคแคน (MacCann. 1989) ได้ศึกษาการใช้แบบทดสอบหลัก ที่มีความเท่าเทียมกันแบบคะแนนจริงสัมพันธ์ (Congeneric) และมีค่าความเชื่อมั่นเท่ากัน โดยใช้รูปแบบแบบสอบหลักที่มีประชากรแตกต่างกัน (The Common-Item Nonequivalent-Populations Design) ของวูดดรัฟ (Woodruff. 1986) โคลเลน และเบรนนอน (Kolen and Brennan. 1987) และ คูก และปีเตอร์เซน (Cook and Petersen. 1987) โดยใช้สูตรการเทียบคะแนนเชิงเส้นตรง 2 สูตร ซึ่งได้เสนอการได้มาของสูตร 2 สูตร ดังนี้ (บุญเชิด วิทยุณันตพงษ์. 2538 : 5-11)

การแสดงผลมาของการเทียบคะแนนสอบวิธีที่ 1 จะอาศัยข้อตกลงมาตรฐานของการแสดง

ที่มาของการเทียบคะแนนสอบทั้งหลายที่ใช้กัน คือ อาศัยข้อตกลงของเส้นถดถอยที่สมมูลกัน (Equivalent) ซึ่งส่วนมากมักเรียกว่า เส้นถดถอยที่ขนานกัน

สมมติให้ X และ Y เป็นแบบทดสอบสองฉบับ นำไปสอบนักเรียนกลุ่มที่ 1 และกลุ่มที่ 2 ตามลำดับ และให้ Z เป็นแบบทดสอบเชื่อมโยง ที่นำไปสอบนักเรียนกลุ่มรวมทั้งหมด (กลุ่ม 1 และกลุ่ม 2 รวมกัน)

การแสดงที่มาของสูตรจะใช้ข้อตกลงเกี่ยวกับเส้นถดถอย 2 เส้นขนานกันจากการกำหนดข้างต้นจะได้เส้นถดถอยของการพยากรณ์ X ด้วย Z สองเส้น ดังนั้น เส้นถดถอยเส้นหนึ่งได้จากคะแนนของนักเรียนกลุ่มรวมและเส้นถดถอยอีกเส้นหนึ่งได้จากคะแนนของนักเรียนกลุ่มที่ 1

จากข้อตกลงที่ว่าเส้นถดถอย 2 เส้นขนานกัน สามารถแตกออกเป็น 2 ข้อตกลงย่อย คือ

1. ข้อตกลงที่ว่า เส้นถดถอย 2 เส้นต้องมีความลาดชันเท่ากัน (Equal Slope)
2. ข้อตกลงที่ว่า เส้นถดถอย 2 เส้นต้องมีพิกัดแกนตั้งเท่ากัน (Equal Ordinate

Intercepts)

ข้อตกลงความลาดชันของเส้นถดถอย 2 เส้นจากนักเรียนกลุ่มรวมและนักเรียนกลุ่มที่ 1 เขียนเป็นสมการได้ดังนี้

$$(3) \quad b_{xz} = b_{xz1}$$

เมื่อ b_{xz} และ b_{xz1} เป็นสัมประสิทธิ์ของความลาดชันของคะแนนนักเรียนกลุ่มรวมและกลุ่มที่ 1 ตามลำดับ และมี

$$(4) \quad b_{xz} = r_{xz} \frac{S_x}{S_z}$$

$$(5) \quad b_{xz1} = r_{xz1} \frac{S_{x1}}{S_{z1}}$$

r_{xz} และ r_{xz1} เป็นสัมประสิทธิ์สหสัมพันธ์ระหว่างแบบทดสอบแต่ละฉบับของ

คะแนนนักเรียนกลุ่มรวม และกลุ่มที่ 1 ตามลำดับ

S_x และ S_z เป็นส่วนเบี่ยงเบนมาตรฐานของคะแนนนักเรียนกลุ่มรวมที่สอบแบบทดสอบฉบับ x และ z ตามลำดับ

S_{x1} และ S_{z1} เป็นส่วนเบี่ยงเบนมาตรฐานของคะแนนนักเรียนกลุ่มที่ 1 ที่สอบแบบทดสอบฉบับ x และ z ตามลำดับ

ข้อตกลงของพิกัดแกนตั้งของเส้นถดถอย 2 เส้นจากนักเรียนกลุ่มรวมและนักเรียนกลุ่ม 1 เขียนเป็นสมการได้ดังนี้

$$(6) \quad \bar{X} - b_{xz} \bar{Z} = \bar{X}_1 - b_{xz1} \bar{Z}_1$$

นำสมการ 3 แทนค่าในสมการ 6 และจัดเรียงค่าใหม่ จะประมาณค่าเฉลี่ยของ X จากนักเรียนกลุ่มรวมได้ดังนี้

$$(7) \quad \bar{X} = \bar{X}_1 + b_{xz1} (\bar{Z} - \bar{Z}_1)$$

ทำนองเดียวกัน นำสมการ 3 แทนค่าลงในสมการ 4 และจัดเรียงค่าใหม่จะประมาณค่าส่วนเบี่ยงเบนมาตรฐานของคะแนน X จากนักเรียนกลุ่มรวมได้ดังนี้

$$(7.1) \quad S_x = \frac{b_{xz1}}{r_{xz}} S_z$$

ทำนองเดียวกันถ้าใช้เส้นถดถอยของการพยากรณ์ Y ด้วย Z จากนักเรียนกลุ่มรวมเส้นหนึ่งและจากนักเรียนกลุ่มที่ 2 อีกเส้นหนึ่ง ตามข้อตกลงของเส้นถดถอยที่ขนานกัน จะได้สมการที่คล้ายกับสมการที่ 7 และ 7.1 ดังนี้

$$(8) \quad \bar{Y} = \bar{Y}_2 + b_{yz2} (\bar{Z} - \bar{Z}_2)$$

และ

$$(8.1) \quad S_y = \frac{b_{yz2}}{r_{yz}} S_z$$

คะแนนสอบสองคะแนน สามารถนำมาเทียบกันแบบเส้นตรงได้ โดยถือว่าคะแนนมาตรฐานที่ได้จากคะแนนทั้งสองเท่ากัน จากแนวคิดนี้จะสามารถแปลงคะแนนสอบได้จากสมการเส้นตรงดังนี้

$$(9) \quad Y = \bar{Y} + \frac{S_y}{S_x} (X - \bar{X})$$

เนื่องจากค่า X และ Y สามารถหาได้โดยตรงจากคะแนนสอบตั้งสมการ 7 และ 8 แต่ค่าอัตราส่วน S_y และ S_x ไม่สามารถหาค่าได้ เนื่องจากไม่ทราบค่า r_{xz} และ r_{yz} เพราะในทางปฏิบัติเราไม่ได้นำ X และ y ไปสอบกับนักเรียนกลุ่มรวม ดังนั้น การหาค่า S_y และ S_x จึงต้องใช้วิธีการประมาณค่า จากสมการ 7.1 และ 8.1 โดยนำสมการ 8.1 ไปหารสมการ 7.1 ดังนี้

$$(10) \quad \frac{S_y}{S_x} = \frac{b_{yz2} r_{xz}}{b_{xz1} r_{yz}}$$

และใช้สูตรในการปรับแก้สัมประสิทธิ์สหสัมพันธ์ที่ลดลงเนื่องจากความคลาดเคลื่อนแล้วจะได้สมการดังนี้

$$(11) \quad r_{xz} = r_{TxTz} \sqrt{r_{xx} r_{zz}}$$

ทำนองเดียวกันก็จะได้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง Y และ Z ดังนี้

$$(11.1) \quad r_{yz} = r_{TyTz} \sqrt{r_{yy} r_{zz}}$$

ทารสมการ 11 ด้วยสมการ 11.1 และใช้ข้อตกลงที่ว่า ความเชื่อมั่นของแต่ละฉบับ
เท่ากันตามสมการ 2 จะได้

$$(12) \quad \frac{r_{xz}}{r_{yz}} = \frac{r_{TxTz}}{r_{TyTz}}$$

ตามที่สมการ 1 ซึ่งเป็นข้อตกลงของคะแนนจริงสัมพันธ์มีความหมายว่าคะแนนจริงมีความ
สัมพันธ์เป็นเส้นตรง จากสมการ 1 นี้จะเห็นได้ว่า สหสัมพันธ์ของ T_z กับ T_x จะเท่ากับ
สหสัมพันธ์ของ T_z กับ T_y นั่นคือ

$$(13) \quad r_{TxTz} = r_{TyTz}$$

แทนค่าสมการ 12 และ 13 ลงในสมการ 10 จะได้

$$(14) \quad \frac{S_y}{S_x} = \frac{b_{yz2}}{b_{xz1}}$$

ขั้นสุดท้ายแทนสมการ 8 และสมการ 14 ลงในสมการ 9 และเขียนเป็นนิพจน์ของ
ความลาดชัน หรือสูตรสำหรับแปลงคะแนนสอบได้ดังนี้

$$(15) \quad Y = \begin{bmatrix} r_{yz2} & S_y & S_{z1} \\ r_{xz1} & S_x & S_{z2} \end{bmatrix} X + \bar{Y} + \begin{bmatrix} S_{y2} \\ S_{z2} \end{bmatrix} \left[(\bar{Z}_1 - \bar{Z}_2) - \frac{1}{r_{xz1} S_{x1}} \bar{X}_1 \right]$$

การแสดงให้เห็นของการเทียบคะแนนสอบวิธีที่ 2

การแสดงให้เห็นของวิธีที่ 1 ดังกล่าวแล้วนั้นได้ใช้ข้อตกลงของเส้นถดถอยของการพยากรณ์ X ด้วย Z สำหรับการแสดงให้เห็นของวิธีที่ 2 จะใช้เส้นถดถอยของการพยากรณ์ Z ด้วย X โดยกำหนดข้อตกลงของเส้นถดถอยว่าความลาดชันของเส้นถดถอยที่ได้จากคะแนนกลุ่ม 1 เท่ากับความลาดชันที่ได้จากนักเรียนกลุ่มรวม และพิกัดแกนตั้งของเส้นถดถอยที่ได้จากนักเรียนกลุ่มรวมเท่ากับพิกัดแกนตั้งของเส้นถดถอยที่ได้จากนักเรียนกลุ่มที่ 1

จากข้อตกลงของความลาดชันจะได้

$$(16) \quad b_{zx} = b_{zx1}$$

และมี

$$(17) \quad b_{zx} = r_{zx} \frac{S_z}{S_x}$$

$$(17.1) \quad b_{zx1} = r_{zx1} \frac{S_{z1}}{S_{x1}}$$

และใช้ข้อตกลงที่ว่า พิกัดแกนตั้งมีค่าเท่ากัน จะเขียนเป็นสมการได้ดังนี้

$$(18) \quad \bar{Z} - b_{zx} \bar{X} = \bar{Z}_1 - b_{zx1} \bar{X}_1$$

จากสมการที่ 16 และ 18 จะประมาณค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของคะแนน X ที่สอบจากนักเรียนกลุ่มรวม ซึ่งจะได้สมการที่คล้ายคลึงกับในวิธีที่ 1

$$(19) \quad \bar{X} = \bar{X}_1 + \frac{1}{b_{zx1}} (\bar{Z} - \bar{Z}_1)$$

$$(19.1) \quad S_x = \frac{1}{b_{zx1}} r_{xz} S_z$$

ทำนองเดียวกัน ถ้าใช้เส้นถดถอยของการพยากรณ์ Z ด้วย Y จากนักเรียนกลุ่มที่ 2 จะได้ค่าสมการดังนี้

$$(20) \quad \bar{Y} = \bar{Y}_2 + \frac{1}{b_{zy2}} (\bar{Z} - \bar{Z}_2)$$

และ

$$(20.1) \quad S_y = \frac{1}{b_{zy2}} r_{zy} S_z$$

ทำนองเดียวกันกับการได้มาของสมการ 14 ในวิธีที่ 1 จะได้

$$(21) \quad \frac{S_y}{S_x} = \frac{b_{zx1}}{b_{zy2}}$$

แทนค่าสมการ 19.1 และ 21 ลงในสมการ 9 และเขียนเป็นนิพจน์ของความลาดชันทั้งหมด หรือสมการเทียบคะแนนสอบดังนี้

$$(22) \quad Y = \begin{bmatrix} r_{xz1} & S_{y2} & S_{z1} \\ r_{yz2} & S_{x1} & S_{z2} \end{bmatrix} X + \bar{Y}_2 + \begin{bmatrix} 1 & S_{y2} \\ r_{yz2} & S_{z2} \end{bmatrix} \begin{bmatrix} (\bar{Z}_1 - \bar{Z}_2) - r_{xz1} \frac{S_{z1}}{S_{x1}} \bar{X}_1 \end{bmatrix}$$

ถ้าเปรียบเทียบสมการ 22 กับ 15 จะเห็นว่าวิธีที่ 2 แตกต่างจากวิธีที่ 1 เฉพาะที่ค่าสหสัมพันธ์เท่านั้น คือ ค่าสหสัมพันธ์ในสมการ 22 จะเขียนเป็นส่วนกลับของสหสัมพันธ์ในสมการที่ 15 และในงานวิจัยครั้งนี้จะนำการเทียบคะแนนด้วยสมการ 15 มาใช้ในการศึกษา

แบบทดสอบหลัก (Anchor Test)

แบบจำลองของการเทียบคะแนนของแบบทดสอบอาจจำแนกเป็น 2 กลุ่ม คือแบบจำลองกลุ่มที่ใช้ข้อสอบร่วม(Common Items) และแบบจำลองกลุ่มที่ใช้เฉพาะแบบทดสอบที่ต้องการเทียบคะแนนอย่างเดียว แบบจำลองกลุ่มแรกเป็นแบบที่มีการนำมาใช้อยู่เสมอ ซึ่ง วูดดรัฟ, โคลเลน และ เบรนนอน (Woodruff, 1986 ; Kolen and Brennan 1987) แห่ง The American College Testing Program (ACT) เรียกแบบจำลองนี้ว่า แผนแบบข้อสอบร่วมที่มีประชากรแตกต่างกัน(Common-Item Nonequivalent-Population Design)แต่ คูก และปีเตอร์เซน (Cook and Petersen 1987) แห่ง Educational Testing Service (ETS) เรียกแบบจำลองนี้ว่า แผนแบบที่ใช้แบบทดสอบหลัก(Anchor-Test Design) ซึ่งแบบหลังได้รับความนิยมมากกว่า (บุญเชิด วิทยุโณนันทพงษ์. 2538 : 2-3)

ลักษณะของแบบทดสอบหลักแบ่งเป็น 2 ลักษณะ

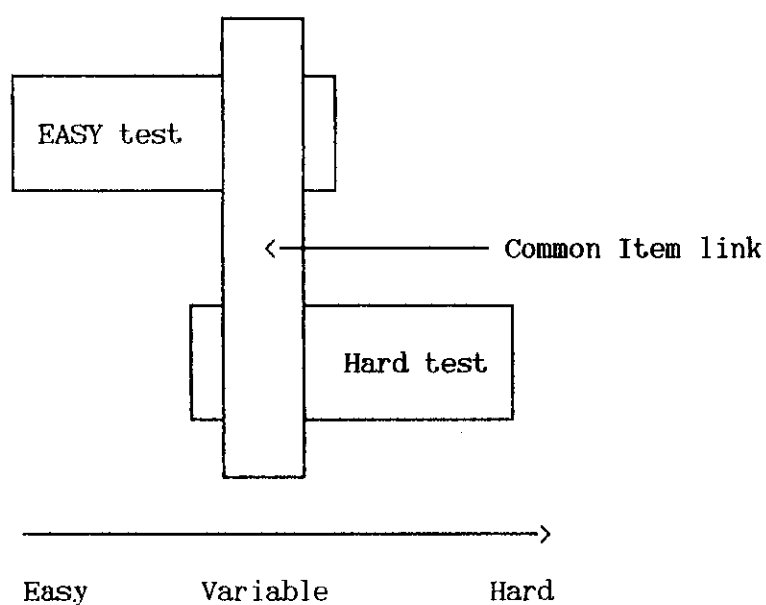
1.Internal Anchor Test เป็นแบบทดสอบหลักฉบับหนึ่งที่แฝงอยู่ในแบบทดสอบที่ใช้ในการเทียบคะแนนทั้งสองฉบับ

2.External Anchor Test เป็นแบบทดสอบหลักฉบับหนึ่งที่แยกจากแบบทดสอบที่ใช้ในการเทียบคะแนนทั้งสองฉบับ

ซึ่งแบบทดสอบหลักนี้ จะเป็นแบบทดสอบที่ช่วยในการปรับแก้คะแนนที่สอบได้ เนื่องจาก การนำคะแนนดิบ หรือคะแนนที่สอบได้จากแบบทดสอบสองฉบับมาเทียบกัน มีปัญหาอยู่ว่า ถ้าฉบับใดฉบับหนึ่งยาวกว่าจะทำให้เกิดการได้เปรียบเสียเปรียบได้ จึงใช้แบบทดสอบหลักมาช่วยแก้ปัญหาดังกล่าว

รูปแบบของการเชื่อมต่อของแบบทดสอบหลัก (อากรณ์ กาญจนกิจโสภณ. 2531:22-27) สามารถทำได้หลายวิธีดังนี้

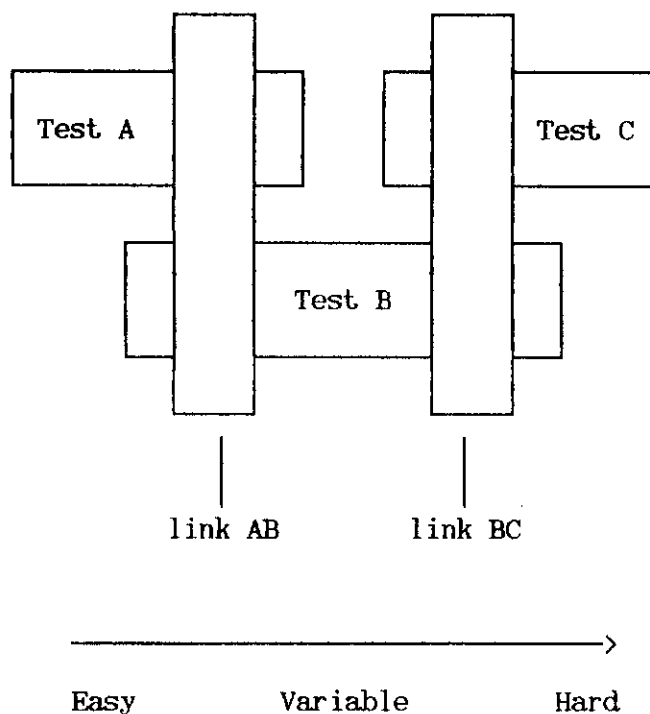
การเชื่อมต่อแบบธรรมดา หรือ 1 ลูกโซ่



ภาพประกอบ 2 แสดงการเชื่อมต่อของแบบทดสอบ 2 ฉบับ

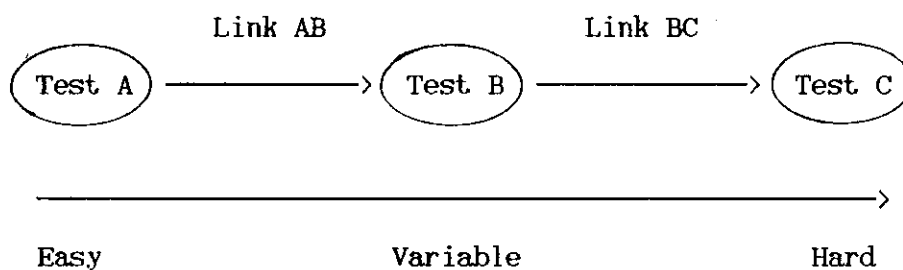
จากภาพประกอบ 2 เป็นการเชื่อมต่อของแบบทดสอบ 2 ฉบับ ที่มีฉบับหนึ่งง่าย อีกฉบับหนึ่งยาก โดยมีข้อสอบร่วมที่ประกอบด้วยข้อยากในฉบับง่ายร่วมกับข้อง่ายในฉบับยาก

การเชื่อมแบบ 2 ลูกโซ่



ภาพประกอบ 3 แสดงการเชื่อมแบบ 2 ลูกโซ่

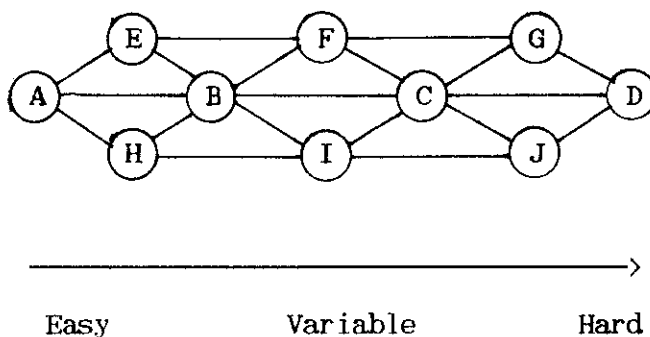
จากภาพประกอบ 3 แสดงให้เห็นการเชื่อมต่อของแบบทดสอบ 3 ฉบับ ที่มีจุดเชื่อมต่อ 2 จุด ซึ่งสามารถแสดงให้เห็นจุดเชื่อมต่อได้ชัดเจนดังภาพ



ภาพประกอบ 4 แสดงจุดเชื่อมต่อของการเชื่อมแบบ 2 ลูกโซ่

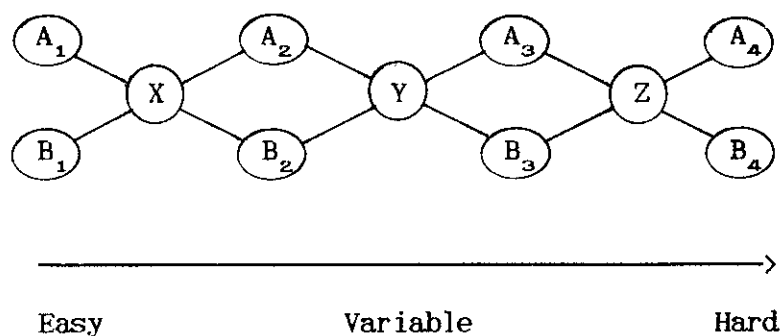
กรณีที่มีแบบทดสอบ 4 ฉบับหรือมากกว่า การเชื่อมแบบทดสอบเหล่านี้จะทำในรูปตาข่าย

รูปแบบการเชื่อมมากกว่า 2 ลูกโซ่ จะมีลักษณะเป็นตาข่าย



ภาพประกอบ 5 แสดงตาข่ายการเชื่อมของแบบทดสอบ 10 ฉบับ 7 ระดับความสามารถ

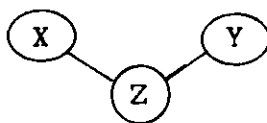
รูปแบบการเชื่อมโยงภายนอก



ภาพประกอบ 6 แสดงตาข่ายการเชื่อมต่อโดยแบบทดสอบภายนอก

จากภาพประกอบ 6 แสดงการเชื่อมต่อของแบบทดสอบ 2 ชุด คือชุดที่ 1 มีแบบทดสอบ 4 ฉบับคือ A_1, A_2, A_3, A_4 ชุดที่ 2 มีแบบทดสอบ 4 ฉบับ คือ B_1, B_2, B_3, B_4 แบบทดสอบที่เป็นตัวเชื่อม คือ แบบทดสอบฉบับ X, Y, Z

จากการศึกษาเรื่องรูปแบบการเชื่อมต่อของแบบทดสอบ ผู้วิจัยใช้รูปแบบการเชื่อมโยงภายนอกโดยมีแบบทดสอบ 2 ฉบับ (X,Y) และมีแบบทดสอบเชื่อมโยง (Z) 1 ฉบับ ดังภาพ



ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน (Standard Error of Equating)

ปีเตอร์เซน ,โคเลน และ โฮเวอร์ (Petersen,Kolen and Hoover.1984:255) กล่าวว่า ความคลาดเคลื่อนของการเทียบคะแนน(SEE)เกิดจากกลุ่มผู้สอบ ในการเทียบเชิงเส้นตรงเกิดเนื่องจากความไม่คงตัวของการสุ่มในค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานของคะแนนจาก X และ Y ความคลาดเคลื่อนของการเทียบคะแนน คือค่าเบี่ยงเบนมาตรฐานของการแปลงคะแนน $l_x(y_0)$ ในมาตราของ X สำหรับค่าคงที่ y_0 ของ Y เมื่อแต่ละ $l_x(y_0)$ คำนวณจากประชากรที่กำหนดจากการสุ่ม

วิธี bootstrap (Efron.1982) สามารถใช้ประมาณ ค่าความคลาดเคลื่อนของการเทียบเชิงเส้นตรง ในทางกลับกันค่าความคลาดเคลื่อนของกลุ่มตัวอย่างขนาดใหญ่ สามารถใช้เป็นประโยชน์ได้ เพราะความคลาดเคลื่อนของคะแนน ที่เทียบเชิงเส้นตรงของกลุ่มตัวอย่างขนาดใหญ่ และมีขนาดของกลุ่มตัวอย่างเท่ากันคาดว่าเป็น โค้งปกติ

ลอร์ด และแองกอฟฟ์ (Lord.1950; Angoff.1984)เสนอวิธีนี้สำหรับการรวบรวมข้อมูลแบบกลุ่มผู้สอบกลุ่มเดียว (Single-group design) กลุ่มผู้สอบกลุ่มเดียวที่ได้รับการจัดให้สมดุล(Counter-balanced random group design) กลุ่มผู้สอบเท่าเทียมกัน(Equivalent-group design) และกลุ่มผู้สอบเท่าเทียมกันโดยใช้แบบสอบหลัก (Anchor-Test random-groups design)

สำหรับกลุ่มผู้สอบเท่าเทียมกัน บรอน และฮอลแลนด์ (Braun and Holland. 1982) ได้เสนอสูตรการหาค่าความคลาดเคลื่อนของการเทียบคะแนนที่ครอบคลุมลักษณะการกระจายของคะแนนมากกว่าของ ลอร์ด หรือ แองกอฟฟ์ (Lord or Angoff) โดยเสนอสูตร ดังนี้

$$SE [l_x(y_o)] = \sqrt{\{(\hat{\sigma}_x^2/2)[2+(y_o - \mu_y)^2/\hat{\sigma}_y^2][1/(N_x) + (1/N_y)]\}}$$

ซึ่งบรอน และฮอลแลนด์ (Braun and Holland, 1982:32-36) ได้กล่าวถึงแหล่งของความลำเอียงที่มีต่อผลการเทียบคะแนน ดังนี้

1. ความแปรผันของการสุ่มและความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน ทุกทฤษฎีที่จะนำเสนอมีพื้นฐานในการสุ่มจากประชากร ด้วยเหตุนี้ การวัดทั้งหมดเป็นการแปรผันของการสุ่ม วิธีการทางสถิติมาตรฐานอาจจะใช้การประมาณความคลาดเคลื่อนในการสุ่ม สำหรับวิธีการเทียบคะแนน ความคลาดเคลื่อนในการเทียบคะแนน เป็นตัวแปรหนึ่งของการวัด

ความคลาดเคลื่อนของการเทียบคะแนนจากการสุ่ม

การแสดงให้เห็นของสูตรความคลาดเคลื่อน สำหรับการเทียบคะแนนเชิงเส้นตรงในการเทียบคะแนนของกลุ่มสุ่ม 2 กลุ่ม

ให้ n_x เป็นจำนวนผู้เข้าสอบจากการสุ่มจากประชากร P ด้วยแบบทดสอบฉบับ X และ n_y เป็นจำนวนผู้เข้าสอบจากการสุ่มจากประชากร P ด้วยแบบทดสอบฉบับ Y จำนวน n_x และ n_y อาจจะไม่เท่ากันก็ได้ X_i หมายถึง คะแนนของผู้สอบแต่ละคนในฉบับ X และ Y_j หมายถึงคะแนนของผู้สอบแต่ละคนในฉบับ Y การสุ่มโดยสมมุติว่า $\{X_i\}$ เป็นอิสระและการกระจายรูปทรงเดียวกับ $F(x)$ และ $\{Y_j\}$ ก็กระจายเหมือนกับ $G(y)$ X และ Y เป็นตัวอย่างที่เป็นอิสระ เราจะใช้สัญลักษณ์การกระจายของ X และ Y ดังนี้

$$\begin{aligned} \mu_x &= E(X) & , & & \mu_y &= E(Y) \\ \hat{\sigma}_x^2 &= \text{Var}(X) & , & & \hat{\sigma}_y^2 &= \text{Var}(Y) \end{aligned}$$

ฟังก์ชันการเทียบคะแนนเชิงเส้นตรง สำหรับการเทียบคะแนน Y ไป X ดังนี้

$$X^*(y) = \mu_x + \frac{\hat{\sigma}_x}{\hat{\sigma}_y} (y - \mu_y)$$

เมื่อ y แปรเปลี่ยนตามคะแนนดิบ Y และใช้การประมาณปริมาณประชากร $X^*(y)$ คือ

$$X^*(y) = \bar{X} + \frac{S_x}{S_y} (y - \bar{Y})$$

เมื่อ $\bar{X}$, S_x^2 เป็นค่าเฉลี่ยของกลุ่มตัวอย่าง และความแปรปรวนของ $\{X_1\}$
และ SEE คือ รากที่สองของความแปรปรวน

ทฤษฎี

$$\begin{aligned} \text{Var}(X^*(y)) = & \delta_x^2 \left\{ \frac{1}{n_x} + \frac{1}{n_y} + \left[\frac{\text{skew}(x)}{n_x} + \frac{\text{skew}(y)}{n_y} \right] Z(y) \right. \\ & \left. + \left[\frac{2+\text{kurt}(x)}{4n_x} + \frac{2+\text{kurt}(y)}{4n_y} \right] Z^2(y) \right\} + (\text{higher order terms}) \end{aligned}$$

$$\text{เมื่อ } \text{skew}(x) = \mu_{3x} / \delta_x^3, \quad \text{skew}(y) = \mu_{3y} / \delta_y^3$$

$$\text{kurt}(x) = (\mu_{4x} / \delta_x^4) - 3, \quad \text{kurt}(y) = (\mu_{4y} / \delta_y^4) - 3$$

$$\mu_{3x} = E(X - \mu_x)^3, \quad \mu_{3y} = E(Y - \mu_y)^3$$

$$\mu_{4x} = E(X - \mu_x)^4, \quad \mu_{4y} = E(Y - \mu_y)^4$$

$$\text{และ } X(y) = (y - \mu_y) / \delta_y$$

บทแทรกที่ 1 ถ้า X และ Y มีการกระจายปกติ เมื่อ $\text{Skew}(X)=0$ $\text{Skew}(Y)=0$
และ $\text{Kurt}(X) = \text{Kurt}(Y) = 0$ ดังนั้น

$$\text{Var}(X^*(y)) = (\sigma_x^2 / n_h)(2+Z^2(y)) + (\text{higher order terms})$$

เมื่อ $n_h = [1/2 (n_x^{-1} + n_y^{-1})]$ คือ ค่าเฉลี่ยฮาร์โมนิกของ n_x และ n_y

บทแทรกที่ 2 ถ้า X และ Y มีการกระจายปกติ และ $n_x = n_y = N/2$ แล้ว

$$\text{Var}(X^*(y)) = \frac{2\sigma_x^2}{N} (2+X^2(y)) + (\text{higher order terms})$$

2. ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน

สมมติให้ $\hat{X}^*(y)$ เป็นค่าประมาณของฟังก์ชันการเทียบคะแนนจาก Y ไป X

และ $\hat{Y}^*(z)$ เป็นค่าประมาณของฟังก์ชันการเทียบคะแนนจาก Z ไป Y

ดังนั้น

$$\hat{g}(z) = \hat{X}^*(\hat{Y}^*(z))$$

ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน สำหรับ $\hat{X}^*(y)$ และ $\hat{Y}^*(z)$ นิยามโดย

$$(\text{SEE}_x(y))^2 = \text{Var}(\hat{X}^*(y))$$

และ

$$(\text{SEE}_y(z))^2 = \text{Var}(\hat{Y}^*(z))$$

ค่าประมาณ $\hat{X}^*(y)$ และ $\hat{Y}^*(z)$ จะไม่ลำเอียง ถ้า $\hat{X}^*$ และ $\hat{Y}^*$ เป็นค่าทางสถิติที่อิสระ
ภายใต้เงื่อนไขตามทฤษฎี SEE สำหรับ $\hat{X}^*(\hat{Y}^*(z))$

ทฤษฎี

$$\text{Var} \left\{ \hat{X}^*(\hat{Y}(z)) \right\} = (\text{SEE}_x(Y^*(z)))^2 + \left[\frac{dx^*}{dy} (Y^*(z)) \right]^2 (\text{SEE}(z))^2 + (\text{higher order terms})$$

ความลำเอียงในการเทียบคะแนน (Bias in Equating)

บรอน และ ฮอลแลนด์ (Braun and Holland, 1982 : 36-39) กล่าวว่า ความลำเอียงในการเทียบคะแนน กำหนดในเทอมของความลำเอียงทางสถิติ (Statistical Bias) เป็นความแตกต่างระหว่างค่าเฉลี่ยของการประมาณการสุ่มซ้ำ จากประชากรเดียวกัน และค่าของการประมาณพารามิเตอร์ ความลำเอียงบางครั้งใช้อ้างถึงความแตกต่างระหว่างการประมาณค่าฟังก์ชันการเทียบคะแนน และค่าจริงของฟังก์ชันการเทียบคะแนน กล่าวคือ 1. จะคำนวณได้การกระจายของคะแนนจากประชากรถูกต้อง 2. การเขียนในเทอมของความลำเอียงเป็นสิ่งสำคัญโดยเหตุที่เกี่ยวข้องกับผลของการประมาณฟังก์ชันของการเทียบคะแนนมากกว่าการประมาณคุณสมบัติทางสถิติ

ความลำเอียงในการเทียบคะแนน เกิดจาก 2 แหล่งใหญ่ คือการแปรผันของประชากร และความลำเอียงของรูปแบบ

1. การแปรผันของประชากร (Population Variability)

เป็นไปได้ที่จะกำหนดให้การเทียบคะแนนเกิดจากประชากรเดียวกัน การเทียบคะแนนอาจจะเกิดจากประชากรที่แตกต่างกันก็ได้ ที่จะแนะนำต่อไปนี้เป็นความเป็นไปได้ สำหรับฟังก์ชันการเทียบคะแนนจากประชากรเดียวกัน เช่นถ้า $X_p^*(Y)$ เท่ากับคะแนนที่เทียบจาก Y ไป X ในประชากร P และ $Y_q^*(Z)$ เท่ากับคะแนนที่เทียบจาก Z ไป Y จากประชากร Q จะไม่เป็นฟังก์ชันของคะแนนรวมกัน $X_p^*(Y_q^*(Z))$ เท่ากับ Z ไป X บน P และ Q ซึ่งใช้ข้อสอบร่วมจะดีกว่า ประชากรที่ใช้ในข้อสอบร่วมเป็นผลรวมของ P และ Q ผลการเทียบจาก Y ไป X ใน P หรือ Q จะไม่รับรองว่าน่าพอใจ เพราะความสามารถในการเทียบขึ้นอยู่กับประชากรที่ใช้หรือ

อาจจะเรียกว่า การแปรผันของประชากร Y ในการเทียบคะแนน

ทฤษฎีทางสถิติของการแปรผันของประชากร ชัดต่อการแปรผันของการสลับแต่ก็ยัง ไม่มีการศึกษาในเรื่องนี้อย่างกว้างขวาง จะแสดงให้เห็น ดังนี้

สมมติว่า มี 2 ประชากร คือ P และ Q และมีแบบทดสอบ 3 ฉบับคือ X, Y, Z ฟังก์ชันการกระจายของคะแนน X, Y, Z คือ F, G และ H ฟังก์ชันการเทียบคะแนนแบบอิกวิเปอร์เซนไทล์ จาก Y ไป X ใน P คือ

$$X_P^*(y) = F_P^{-1}(G_P(y))$$

และ คล้ายกับการเทียบคะแนนจาก Z ไป Y ใน Q คือ

$$Y_Q^*(z) = G_Q^{-1}(H_Q(z))$$

ฟังก์ชันรวมเป็น

$$g(z) = X_P^*(Y_Q^*(z))$$

อาจเปลี่ยนเป็นเทียบคะแนนจาก Z ไป X ก็ได้ กับประชากร P หรือ Q ก็ได้

ทฤษฎี

แสดงว่า ถ้า $P=Q, g(z)$ เป็นการเทียบจาก Z ไป X ใน P ก็จะเป็นไปตามที่คาดหวังไว้

$$\text{ถ้า } P=Q \text{ แล้ว } X_P^*(Y_Q^*(z)) = F_P^{-1}(H_P(z))$$

ดังนั้น $g(z)$ จากฟังก์ชันรวมเป็นฟังก์ชันการเทียบคะแนนแบบอิกวิเปอร์เซนไทล์ สำหรับการเทียบจาก Z ไป X บน P

จากทฤษฎีที่กล่าวว ถ้า $P=Q$ จะไม่มีความลำเอียง แต่ถ้า $P \neq Q$

$$\text{bias}(z) = X_p^*(Y_p^*(z)) - (X_p^*(Y_q^*(z)))$$

เมื่อ ความลำเอียงน้อยจะประมาณโดยขยาย X_p^* จากชุดของ Taylor ดังนี้

$$\text{bias}(z) \simeq (dX_p^*/dy)(Y_p^*(z))\{Y_p^*(z)-Y_q^*(z)\}$$

เพื่อความเข้าใจที่ดีในความสัมพันธ์ของสูตรนี้ สมมติว่าฟังก์ชันของอิควิเปอร์เซนไทล์เป็นเส้นตรง
เมื่อ

$$X_p^*(y) = \mu_{xp} + (\partial_{xp} / \partial_{yp})(y - \mu_{yp}),$$

$$Y_q^*(z) = \mu_{yq} + (\partial_{yq} / \partial_{zq})(z - \mu_{zq}) = A_{qz} + B_q$$

$$Y_p^*(z) = \mu_{yp} + (\partial_{yp} / \partial_{zp})(z - \mu_{zp}) = A_{pz} + B_p$$

เมื่อ

$$dX_p^*/dy = \partial_{xp} / \partial_{yp}$$

และ

$$Y_p^*(z) - Y_q^*(z) = (A_p - A_q)z + (B_p - B_q)$$

สิ่งที่ต้องการคือหวังว่า ∂_{xp} , ∂_{yp} มีค่าใกล้เคียงกัน

ซึ่งจะเห็นว่า $\text{bias}(z)$ เป็นค่าประมาณความแตกต่างที่ $Z=z$ ระหว่างสองสมการ
เส้นตรงของการเทียบคะแนนจาก Z ไป Y ในประชากร P และ Q

2. ความคลาดเคลื่อนของรูปแบบ (Model Errors)

ความคลาดเคลื่อนนี้ เป็นความคลาดเคลื่อนที่เกิดจากการจัดกระทำโดยใช้ข้อตกลง
ง่าย ๆ และผิดรูปแบบ ซึ่งจำแนกได้ 2 ชนิด คือ

2.1 ความคลาดเคลื่อนที่ทดสอบได้ เช่น ความเป็นเส้นตรงของฟังก์ชันการถดถอย
สามารถทดสอบได้โดยใช้ข้อมูลจำนวนมาก จะสามารถแสดงความเพียงพอได้

2.2 ความคลาดเคลื่อนที่ทดสอบไม่ได้ เป็นกรณีที่ไม่สามารถหาข้อมูลมาพิสูจน์ข้อตกลงนั้น ๆ ได้ เช่น ความคลาดเคลื่อนในการประมาณค่าในการเทียบคะแนนของ Tucker โดยให้เงื่อนไขการกระจายของ X จาก V ใน P เหมือนกับการกระจายของ X จาก V ใน Q ซึ่งสามารถอธิบายความคลาดเคลื่อนของรูปแบบได้มากมาย

ผลของความคลาดเคลื่อนของรูปแบบนี้จะชี้ให้เห็นความลำเอียงในการประมาณฟังก์ชันของการเทียบคะแนน

ทฤษฎีเชาว์ปัญญาและความถนัด

แนวความคิดในเรื่องการจัดรูปแบบของเชาว์ปัญหามีมานานแล้ว โดยมีประวัติความเป็นมาดังนี้

1.ทฤษฎีองค์ประกอบเดียว (Uni-factor Theory) บางทีมีผู้เรียกว่า Global Theory ผู้คิดคือบีเน็ต และซิมอน (Binet and Simon, 1905) ทฤษฎีนี้เสนอโครงสร้างของเชาว์ปัญญา เป็นลักษณะอันหนึ่งอันเดียว ไม่แบ่งแยกออกเป็นส่วนย่อย คล้ายกับเป็นความสามารถทั่วไป (General ability) นั้นเอง

2.ทฤษฎีสององค์ประกอบ (Bi-Factor Theory) ผู้นำทฤษฎีนี้คือสเปียร์แมน (Charles Spearman) ในปี ค.ศ. 1927 เป็นทฤษฎีที่เกิดจากการวิเคราะห์คุณลักษณะ โดยกระบวนการทางสถิติ พบว่ากิจกรรมทางสมองทั้งหลาย เมื่อวิเคราะห์ดูแล้วมีองค์ประกอบร่วมอันหนึ่ง เรียกชื่อว่า องค์ประกอบทั่วไป (General Factor) หรือ G-factor แล้วยังมีองค์ประกอบย่อย เรียกว่าองค์ประกอบเฉพาะ (Specific factors) หรือ S-factors อีกหลายตัว ซึ่งเป็นแนวคิดใหม่

แต่อย่างไรก็ตามในบรรดาองค์ประกอบเฉพาะต่าง ๆ เหล่านี้ บางชนิดก็อาจมีคุณสมบัติร่วมกันเองในระหว่างพวกเดียวกัน เป็นกลุ่ม ๆ ได้เหมือนกัน แต่จะไม่สัมพันธ์กับความสามารถจำเพาะของชนิดอื่น ๆ สเปียร์แมนตั้งชื่อความสามารถที่เกาะกลุ่มกันเหล่านี้ว่า องค์ประกอบเฉพาะกลุ่ม (Group factor)

3.ทฤษฎีหลายองค์ประกอบ (Multiple-Factor Theory) ทฤษฎีนี้เป็นที่ยอมรับอย่างกว้างขวางของนักจิตวิทยาชาวอเมริกันผู้นำในการสร้างทฤษฎีนี้คือเทอร์สโตน (L.L. Thurstone) เสนอทฤษฎีเมื่อปี ค.ศ. 1933 เขาได้ทำการวิจัยโครงสร้างทางสมองอย่างกว้างขวางและได้ใช้

หลักการวิเคราะห์หลัมัยใหม่ ที่เรียกว่าการวิเคราะห์องค์ประกอบ (Factor Analysis) มาใช้ ทำให้สามารถแยกแยะความสามารถทางสมอง ออกเป็นส่วนย่อย ๆ ได้หลายอย่างทำให้เขามีความเชื่อว่า ความสามารถทางสมอง ไม่ได้ประกอบด้วยความสามารถร่วมเป็นแกนกลาง ดังเช่น G-factor ของสเปียร์แมน หากแต่ประกอบด้วยองค์ประกอบเป็นกลุ่ม ๆ หลาย ๆ กลุ่ม โดยแต่ละกลุ่มจะมีหน้าที่เป็นอย่างไร ๆ ไป โดยเฉพาะ หรืออาจจะทำงานร่วมกันบ้างก็ได้ ความสามารถทั่วไปของสเปียร์แมน เทอร์สโตนเห็นว่าเป็นเพียงองค์ประกอบทางภาษาเท่านั้น องค์ประกอบย่อย ๆ นี้ เทอร์สโตนให้ชื่อว่า ความสามารถปฐมภูมิของสมอง (Primary Mental Abilities) เขาแยกองค์ประกอบย่อย โดยยึดน้ำหนักองค์ประกอบเด่น ๆ (Loading factor) เป็นสำคัญ แต่จริง ๆ แล้วกลุ่มของความสามารถ หรือองค์ประกอบก็ยังทำหน้าที่เกี่ยวพันกันบ้างเหมือนกัน ดังเช่น องค์ประกอบด้านภาษา (Verbal factor) น้ำหนักองค์ประกอบที่มากที่สุดคือ คณิตศาสตร์ เหตุผล เทอร์สโตนพยายามวิเคราะห์องค์ประกอบ ความสามารถของมนุษย์ ออกมาได้หลายอย่าง แต่ที่เห็นได้ชัด และสำคัญ ๆ มีอยู่ 7 ประการ คือ

1. องค์ประกอบด้านภาษา (Verbal factor ใช้ย่อว่า V) องค์ประกอบส่วนนี้ของสมอง จะส่งผลให้รู้ถึงความสามารถด้านความเข้าใจในภาษา และการสื่อสารทั่วไป ผู้มีองค์ประกอบด้านนี้สูง จะมีความสามารถในการอ่านเอาเรื่อง อ่านแบบเข้าใจความหมาย รู้ความสำคัญของคำ รู้ความหมายของศัพท์ได้เป็นอย่างดี

2. องค์ประกอบด้านความคล่องแคล่วในการใช้ถ้อยคำ (Word Fluency factor ใช้ย่อว่า W) เป็นความสามารถที่จะใช้คำได้มากในเวลาจำกัด เช่นหาคำที่ขึ้นต้นด้วย "ด" มากที่สุดในเวลาจำกัด เป็นต้น ความสามารถด้านนี้จะส่งผลให้มีความสามารถด้านเจรจา และการประพันธ์ทั้งร้อยแก้วและร้อยกรอง ตอบโต้ทันทีทันใด เรียกว่ามีปฏิภาณไหวพริบดี ความสามารถนี้ไม่เหมือนข้อแรก ข้อนี้เน้นที่การพูดเป็นสำคัญ

3. องค์ประกอบด้านจำนวน (Number factor ใช้อักษรย่อ N) องค์ประกอบนี้ส่งผลให้มีความเข้าใจในวิชาคณิตศาสตร์ต่าง ๆ ได้ดี มีความสามารถมองเห็นความสัมพันธ์ และความหมายของจำนวนและมีความแม่นยำคล่องแคล่ว ในการบวก ลบ คูณ ทหารอย่างดี

4. องค์ประกอบด้านมิติสัมพันธ์ (Space factor ใช้อักษรย่อ S) ความสามารถด้านนี้จะส่งผลให้คนเข้าใจถึงขนาด และมิติต่าง ๆ อันได้แก่ ความสั้น ยาว ไกล ใกล้ และพื้นที่หรือทรวดทรงที่มีขนาด และปริมาตรแตกต่างกัน สามารถสร้างจินตนาการให้เห็นส่วนย่อยและส่วนผสมของวัตถุต่าง ๆ เมื่อนำมาซ้อนทับกัน สามารถรู้ความสัมพันธ์ของรูปทรงเรขาคณิต เมื่อเปลี่ยนแปลง

ที่อยู่

5.องค์ประกอบด้านความจำ (memory factor ใช้ตัวย่อ M) เป็นความสามารถด้านความทรงจำเรื่องราว และมีสติระลึกจำสามารถถ่ายทอดได้

6.องค์ประกอบด้านการรับรู้ (Perceptual factor ใช้ตัวย่อ P) องค์ประกอบของสมองด้านนี้ ได้แก่ความสามารถด้านเห็นรายละเอียด ความคล้อยคลึง หรือความแตกต่างระหว่างสิ่งของต่าง ๆ อย่างรวดเร็ว และถูกต้อง

7.องค์ประกอบด้านเหตุผล (Reasoning factor ใช้ตัวย่อ R) บางทีก็ใช้ induction หรือ General Reasoning องค์ประกอบนี้แสดงถึง ความสามารถด้านวิจารณ์เหตุผล หาเหตุ หาผล หาความสัมพันธ์ และหลักการมาสร้างกฎ หรือทฤษฎี

4.ทฤษฎีไฮราคิคัล (Hierarchical Theories) มีนักจิตวิทยากลุ่มหนึ่งได้จัดรูปแบบการประกอบกันขององค์ประกอบ อีกแบบหนึ่ง กลุ่มนี้คือ เบิร์ท (Burt) เวอร์นอน (Vernon) และฮัมเฟรย์ (Humphreys) โดยเฉพาะเวอร์นอน (Vernon) ได้เสนอโครงสร้างทางเชาว์ปัญญา ในปี ค.ศ.1960 โดยเริ่มต้นอธิบายตามแบบของสเปียร์แมน นั่นคือเวอร์นอน เริ่มที่ G-factor แบ่งออกเป็น 2 องค์ประกอบใหญ่ ๆ คือ Verbal-education (V:ed) และ Practical-mechanical (k-ed) องค์ประกอบใหญ่ 2 ชนิดนี้รวมเรียกว่า Major Group Factors องค์ประกอบใหญ่ 2 ชนิดนี้ ยังแยกย่อยลงไปอีกเป็น ด้านองค์ประกอบ Verbal-education แบ่งย่อยเป็นองค์ประกอบด้านภาษา ด้านตัวเลข และอื่น ๆ อีก ในทำนองเดียวกันองค์ประกอบ Practical-mechanical แบ่งย่อยเป็น Mechanical-information , Spatial, Manual และยังมีอื่น ๆ อีกแต่ยังไม่กำหนด กลุ่มองค์ประกอบนี้เรียกว่า Minor Group Factor ระดับที่ต่ำที่สุดเรียกว่าองค์ประกอบเฉพาะ (Specific factors)

5.ทฤษฎีโครงสร้างสามมิติของปัญญา (Three Faces of Intellect Model) ทฤษฎีนี้ สร้างขึ้นมาโดย กิลฟอร์ด (Guilford) เมื่อ ค.ศ.1967 มีชื่อเรียกหลายอย่าง เช่น Structure-of-Intellect Model หรือ Three-Dimensional Model of the structure of intellect กิลฟอร์ดได้วิเคราะห์ความสัมพันธ์ของคุณลักษณะ โดยจัดระบบของคุณลักษณะให้อยู่ในรูปใหม่เป็นลูกบาศก์รวมกัน 120 ก้อน และนิยามคุณลักษณะของเชาว์ปัญญาเป็น 3 มิติ ดังนี้

มิติที่ 1 ด้านกระบวนการหรือวิธีการของการคิด (Operation) มีส่วนประกอบย่อย 5 ส่วน

มิติที่ 2 ด้านเนื้อหา (Content) เป็นด้านที่ประกอบด้วยสิ่งเร้า และข้อมูลต่าง ๆ แบ่งได้ 4 อย่าง

มิติที่ 3 ผลของการคิด (Products) เป็นผลของกระบวนการจัดกระทำของความคิดกับข้อมูลจากเนื้อหา ผลผลิตของความคิดแยกได้เป็น 6 อย่าง

6. ทฤษฎีความสามารถทางสมองสองระดับ (Two-level theory of Mental Ability) ทฤษฎีนี้เสนอโดยเจนเซน (Jensen) เมื่อปี ค.ศ.1968 เจนเซน ได้เสนอทฤษฎีว่า ความสามารถด้านสมอง มี 2 ระดับ ระดับ I (Level I) เป็นความสามารถด้านการเรียนรู้และจำ โดยไม่ใช้วิธีการคิด ระดับ II (Level II) เป็นระดับการจัดกระทำทางสมอง เป็นขั้นสร้างมโนภาพ และแก้ปัญหา ซึ่งคล้ายกับองค์ประกอบทั่วไป (G-factor)

7. ทฤษฎีเชาว์ปัญญาของคัทเทิล ทฤษฎีนี้คิดโดยอาร์ บี คัทเทิล (R.B.Cattell) เสนอผลงานในวารสารปี ค.ศ.1963 และพิมพ์แนวคิดเป็นเล่มเมื่อปี ค.ศ.1971 เขาเสนอว่า โครงสร้างเชาว์ปัญญาประกอบด้วย 2 ส่วน คือ Fluid component คือความสามารถทั่วไป กับ Crystallized component คือความสามารถที่เชื่อมโยงกับ วัฒนธรรม และสิ่งแวดล้อม

ความหมายของการรับรู้

การรับรู้ หมายถึง ขบวนการที่อินทรีย์ต่าง ๆ พยายามที่จะแสดงความรู้สึกรู้จักจากสิ่งที่ตนได้รับรู้ออกมา ในรูปของการกระทำใดการกระทำหนึ่ง หรือในรูปของสิ่งใดสิ่งหนึ่ง ที่มีความหมาย ดังนั้น ความสามารถรับรู้ และแสดงการรับรู้ออกมาอย่างดี ต้องอาศัยจิตวิทยาทั่วไปและอวัยวะรับสัมผัสต่าง ๆ เช่น ตา หู จมูก ลิ้น ผิวกาย ถ้าขาดสิ่งหนึ่งสิ่งใดก็จะทำให้การรับรู้ขาดความสมบูรณ์

การรับรู้แบ่งออกเป็น 2 แบบ คือ

1. การรับรู้แบบอาศัยการประสานงานกันของอวัยวะต่าง ๆ เป็นการสร้างขบวนการโดยการรวมหน่วยต่าง ๆ ที่เกิดขึ้นขณะนั้น เป็นต้นว่า การเห็น การรู้สึก การสัมผัสเข้าด้วยกันเหมือนเป็นสิ่งเดียวกัน และเร้าให้อินทรีย์มีปฏิกิริยาตอบสนอง

2. การรับรู้แบบรวมส่วนต่าง ๆ ของสิ่งเร้าเข้าด้วยกัน แล้วรวมกับมโนภาพที่บุคคลนั้นวาดภาพขึ้น

ปัญหาเกี่ยวกับการรับรู้

การรับรู้เป็นเรื่องสลับซับซ้อนมาก จึงมักมีปัญหาเกิดขึ้นว่า

1. ทำไมสิ่งมีชีวิตจึงมองเห็นรูปร่างลักษณะ ขนาด ส่วนลึก การเคลื่อนไหว ระยะทางต่าง ๆ น้ำหนักของสิ่งของต่าง ๆ อะไรทำให้เกิดการรับรู้เช่นนั้น
2. จะมีวิธีการใดบ้างที่จะช่วยให้มนุษย์สามารถที่จะรับรู้สิ่งต่าง ๆ ที่ตนได้อ่าน หรือ ผ่านสายตาไป

องค์ประกอบที่มีส่วนเกี่ยวข้องกับการรับรู้

องค์ประกอบทางด้านประสบการณ์ดั้งเดิม คือบุคคลผู้นั้นจะมีประสบการณ์ดั้งเดิมเกี่ยวกับสิ่งนั้นมาก่อน

องค์ประกอบทางด้านสรีระ เนื่องจากความสามารถทางอวัยวะสัมผัสของมนุษย์ หรือของสิ่งมีชีวิตอยู่ในวงจำกัด ดังนั้นจึงไม่อาจตอบสนองสิ่งเร้าได้หมดทุกอย่าง คือสิ่งเร้าบางชนิดไม่อาจกระตุ้นความรู้สึกของเราได้ อาจทำให้เราไม่ได้ยิน ไม่เห็นได้

องค์ประกอบทั่วไปเกี่ยวกับการรับรู้ ได้แก่

1. การเคลื่อนไหว คือมนุษย์จะรับรู้ได้รวดเร็วต่อสิ่งที่มีการเคลื่อนไหว
2. การทำซ้ำซาก จะก่อให้เกิดการรับรู้ได้
3. การทำอะไรที่ใหม่ ๆ และแปลกย่อมทำให้คนสนใจมาก
4. การให้สีกลิ่นต่าง ๆ ย่อมได้รับความสนใจมากกว่าไม่มีสี หรือสีกลิ่นอย่างธรรมดา

แบบทดสอบวัดการรับรู้ตามแนวคิดของเทอร์สไตน์

สมรรถภาพด้านการรับรู้ เป็นความสามารถในการเก็บรายละเอียดต่าง ๆ ได้อย่างถูกต้อง รวดเร็ว นั่นคือเป็นคนช่างสังเกต ซึ่งอาจจะเป็นในรูปของการพิจารณาความคล้ายคลึงหรือความแตกต่างระหว่างสิ่งต่าง ๆ ก็ได้ ผู้ที่มีความถนัดด้านนี้มากจะสามารถทำงานที่ละเอียด ประณีต มีประสาทไวต่อสิ่งเร้า มีสายตาที่แหลมคม ซึ่งเหมาะกับอาชีพเกี่ยวกับงานตรวจสอบ งานเสมียน นักพิมพ์ดีด ช่างเขียน ช่างถ่ายรูป ช่างแกะนาฬิกา ช่างซ่อมเครื่องไฟฟ้า

แนวการเขียนข้อสอบวัดสมรรถภาพด้านนี้เป็นข้อสอบง่าย ๆ แต่ทำให้มาก ๆ ข้อ ภายใน

เวลาจำกัด ดังตัวอย่าง

1.แบบทดสอบที่กำหนดภาพให้ แล้วหาภาพที่เหมือนหรือที่นิยมเรียกย่อ ๆ ว่าภาพเหมือน ลักษณะของข้อสอบคือจะมีภาพให้ดูทางซ้ายมือ แล้วพิจารณาว่าภาพนี้เหมือนกับภาพใด ฉะนั้นผู้ตอบ ต้องมีความสามารถในการสังเกตได้รวดเร็วและถูกต้อง ในการสังเกตรายละเอียดลักษณะข้อสอบ เป็นได้ทั้งภาพจริง ภาพเสมือน สัญลักษณ์ และตัวอักษรต่าง ๆ

2.แบบทดสอบที่กำหนดภาพให้ แล้วให้หาภาพที่ต่างว่า ภาพ ก - จ ที่กำหนดไว้นั้น มี ภาพใดที่แตกต่างไปจากภาพทางซ้ายมือ นอกจากนั้นยังสามารถเขียนข้อสอบเป็นตัวเลข หรือ ตัวอักษรได้

3.แบบกำหนดให้เป็นคู่ ลักษณะของข้อสอบเป็นการให้พิจารณาว่า ภาพคู่ใดเหมือนกัน หรือแตกต่างกัน แล้วกาเครื่องหมายลงในช่องนั้น

งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการเทียบคะแนนเก่าที่ผ่านมานั้น ส่วนใหญ่เป็นการเปรียบเทียบ ผลการเทียบคะแนน โดยวิธีการเทียบคะแนนแบบต่าง ๆ โดยพิจารณาจากตัวแปรที่แตกต่างกัน เช่นลักษณะของกลุ่มตัวอย่าง สุ่ม-ไม่สุ่ม ขนาดเล็ก-ขนาดใหญ่ วิธีการรวบรวมข้อมูล ลักษณะของ แบบทดสอบ ว่าเป็นแบบทดสอบรวมหรือแบบทดสอบเชื่อมโยง เป็นต้น นอกจากนี้ยังได้มีการศึกษา ทางสถิติ ได้แก่การศึกษาความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนแบบต่าง ๆ การผ่านผื่น ข้อตกลง ตลอดจนความเพียงพอของการเทียบคะแนนหลาย ๆ วิธี ผู้วิจัยได้รวบรวมงานวิจัยที่ เกี่ยวข้องกับการใช้แบบทดสอบหลัก การศึกษากับประชากรที่แตกต่างกัน การศึกษาความคลาด เคลื่อนมาตรฐาน และการใช้แบบทดสอบแบบคะแนนจริงสัมพันธ์ ซึ่งมีดังต่อไปนี้

งานวิจัยต่างประเทศ

โคเลน (Kolen.1981 : 1-11) ได้เปรียบเทียบผลการเทียบคะแนนระหว่างรูปแบบ ตั้งเดิม 2 วิธี กับรูปแบบ IRT 7 วิธี พบว่าวิธีการเทียบคะแนนแบบต่าง ๆ ให้ผลแตกต่างกันอย่าง มีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 และพบว่าวิธี IRT ให้ผลคงเส้นคงวามากที่สุด และวิธี ที่ให้ผลคงเส้นคงวน้อยที่สุดคือวิธีการเชิงเส้นตรง

ปีเตอร์เซน, มาร์โค และสจวร์ต (Petersen, Marco and Stewart. 1982:71-185) ได้ศึกษาความเพียงพอของการเทียบคะแนนเชิงเส้นตรงหลาย ๆ วิธีและวิธีเปอร์เซนไทล์ที่เท่ากัน ในสถานการณ์ต่าง ๆ ผลปรากฏว่า

1. รูปแบบการเทียบคะแนนเชิงเส้นตรง โดยใช้เทคนิคของพอททอฟ (Potthoff) ให้ผลที่ไม่น่าพอใจในเกือบทุกสถานการณ์ที่ศึกษา
2. กรณีที่กลุ่มตัวอย่างขนาดเท่ากัน รูปแบบเปอร์เซนไทล์ที่เท่ากัน ให้ผลที่มีความคลาดเคลื่อนรวมมากกว่ารูปแบบเชิงเส้นตรง
3. รูปแบบการเทียบคะแนนเชิงเส้นตรง โดยใช้เทคนิคของทักเคอร์ 1 (Tucker1) ให้ผลการเทียบที่มีความคลาดเคลื่อนรวมมากกว่าวิธีเชิงเส้นตรงอื่น ๆ ในกรณีที่กลุ่มตัวอย่างแตกต่างกัน ยกเว้นสำหรับรูปแบบเชิงเส้นตรง โดยใช้เทคนิคของพอททอฟ
4. รูปแบบเปอร์เซนไทล์ที่เท่ากัน ให้ผลการเทียบดีกว่ารูปแบบเชิงเส้นตรงเมื่อแบบสอบ 2 ชุด มีความสัมพันธ์เชิงเส้นโค้ง อันเนื่องมาจากแบบสอบมีความยากแตกต่างกัน และจากกลุ่มตัวอย่างที่ไม่คล้ายคลึงกัน
5. ค่าความยากระหว่างแบบสอบหลักและแบบสอบที่เทียบคะแนนแตกต่างกันจะทำให้ความคลาดเคลื่อนในการเทียบมีมากกว่าความแตกต่างในเนื้อหา
6. ค่าความยากที่ต่างกันระหว่างแบบสอบทั้งสองชุดที่จะเทียบคะแนน จะทำให้ความคลาดเคลื่อนในการเทียบมีมากกว่าความแตกต่างในเนื้อหา
7. แบบสอบหลักที่สร้างให้มีลักษณะคล้ายคลึงกับแบบสอบที่จะเทียบคะแนนอย่างมาก หรืออาจเรียกว่าเป็นฉบับย่อส่วนแล้ว การเทียบคะแนนจะให้ผลดีที่สุด

เบรนนอน และโคเลน (Brennan and Kolen. 1987) ได้ศึกษาปัญหาในการเทียบคะแนน และพบว่า เงื่อนไขที่เหมาะสมในการเทียบคะแนนที่น่าพอใจ มีดังนี้

1. โครงสร้างของแบบสอบ
 - 1.1 อธิบายสถิติ และเนื้อหาชัดเจน
 - 1.2 สถิติและเนื้อหาต้องมั่นคงตลอด
 - 1.3 การสร้างแบบทดสอบ สถิติทุกตัวควรทดลองใช้ก่อน
 - 1.4 แบบทดสอบต้องยาวพอ คือ อย่างน้อย 35-40 ข้อ
 - 1.5 คะแนนเฉลี่ยต้องคงที่

1.6 มีข้อสอบรวม 2 ชุดชุดหนึ่งสอบมากกว่า 1 ปี อีกชุดหนึ่งสอบเดือนเดียวกับแบบสอบเทียบ

1.7 ข้อสอบรวมแต่ละชุดต้องมีอย่างน้อย $1/5$ ของแบบสอบรวม(อย่างน้อย 30 ข้อ)

1.8 ข้อสอบรวมต้องเป็นตัวแทนของแบบสอบรวมโดยเฉพาะเนื้อหาและลักษณะของสถิติ

1.9 ข้อสอบรวมมีตำแหน่งเดียวกัน ทั้งแบบทดสอบฉบับเก่าและฉบับใหม่

1.10 ข้อสอบรวมมีหน้าตาเหมือนกับแบบทดสอบฉบับเก่าและฉบับใหม่

2. กลุ่มผู้สอบ

2.1 กลุ่มผู้สอบ ต้องเป็นกลุ่มที่มั่นคง

2.2 กลุ่มผู้สอบมีขนาดใหญ่พอประมาณ ถ้าต่ำกว่า 400 คน จะมีปัญหา

3. เงื่อนไขการจัดการ

3.1 แบบทดสอบและข้อสอบมีความมั่นคง

3.2 การจัดการแบบทดสอบต้องได้รับการควบคุมอย่างดี

4. การศึกษาจริงและการทดลอง

4.1 หลักสูตรและเนื้อหาที่จะใช้ต้องคงที่ ทั้งตอนทดลองและกระทำจริง

บูเดสคู (Budesu.1985 : 13-20) ได้ศึกษาประสิทธิภาพของการเทียบคะแนนเชิงเส้นตรง ขึ้นอยู่กับความยาวของแบบทดสอบหลัก โดยใช้รูปแบบการรวบรวมข้อมูลแบบที่ 4 ของ Angoff และใช้วิธีการเทียบคะแนนของ Tucker ที่มีข้อตกลง ดังนี้

1. ประชากร 2 กลุ่ม คือ กลุ่ม α และ β มีระดับความสามารถใกล้เคียงกัน

2. สมการถดถอย(Slope และ Intercept) ทำนาย X จาก U เช่นเดียวกับในกลุ่ม T(กลุ่มรวม) และกลุ่ม α

3. การประมาณค่าความคลาดเคลื่อนของความแปรปรวนของ X จาก U เช่นเดียวกับกลุ่ม T และกลุ่ม α

4. สมการถดถอย ทำนาย Y จาก U ในกลุ่ม T และกลุ่ม β

5. การประมาณค่าความคลาดเคลื่อนของความแปรปรวนของ Y จาก U เช่นเดียวกับกลุ่ม T และกลุ่ม β

ใช้แบบทดสอบหลักที่มีความยาว .10, .20, .25, .30, .40 และ .50 ของแบบทดสอบรวม ผลปรากฏว่า แบบทดสอบหลักที่มีความยาว .50 ของแบบทดสอบรวมหรือครึ่งหนึ่งของแบบ

ทดสอบรวม มีขนาดของความสัมพันธ์ระหว่างแบบทดสอบหลักกับแต่ละส่วนของแบบทดสอบแต่ละฉบับมากที่สุดซึ่งหมายถึง มีประสิทธิภาพของการเทียบคะแนนมากที่สุดด้วย ซึ่งสอดคล้องกับ Angoff (1971) ที่กล่าวว่า เมื่อความสัมพันธ์ของแบบทดสอบหลักไม่มี แบบทดสอบหลักจะไม่มีประโยชน์สำหรับการเทียบคะแนนเลย ทำนองเดียวกัน ถ้าความสัมพันธ์เพิ่มขึ้นประสิทธิภาพและความถูกต้องของการประมาณค่าพารามิเตอร์ สำหรับกลุ่มรวม ก็เพิ่มขึ้น

คุก และ ปีเตอร์เซน (Cook and Petersen.1987 : 225-244) ได้ศึกษาปัญหาเกี่ยวกับการทำตามข้อตกลงและวิธีการเทียบคะแนนแบบ IRT โดยดูผลจาก 1.ความคลาดเคลื่อนของการสุ่ม 2.ลักษณะของกลุ่มตัวอย่าง 3. ลักษณะของข้อสอบรวม โดยใช้วิธีการเทียบคะแนนแบบอิกวิเพอร์เซนไทล์ทดสอบ กับกลุ่มตัวอย่าง 6 กลุ่ม คือ 2 กลุ่มจากการสุ่ม 1 กลุ่มเป็นผู้ชายทั้งหมด 1 กลุ่มเป็นผู้หญิงทั้งหมด อีก 1 กลุ่มเป็นผู้มีอายุ 21-23 ปี และกลุ่มสุดท้ายเป็นผู้มีอายุตั้งแต่ 29 ปีขึ้นไป กลุ่มตัวอย่างนี้จะคะแนนมาจากการสอบ GMAT ที่เทียบคะแนนโดยวิธี IRT การทดสอบใช้แบบทดสอบ 3 ฟอรม์ สอบต่างเวลากันโดยมีข้อสอบรวม 58 ข้อ ซึ่งมีความสัมพันธ์กับสูงผลของการศึกษา พบว่า ไม่ใช่ปัญหาที่เกิดจากแบบทดสอบที่มีโครงสร้างเหมือนกัน และชี้ให้เห็นว่าการสอบในเวลาที่แตกต่างกันมีผลต่อการเทียบคะแนน

โคป (Cope.1987 : 143-149) ได้เปรียบเทียบคะแนนเชิงเส้นตรง 5 วิธี ระหว่างวิธีของแองกอฟฟ์ (Angoff.1971)ตามรูปแบบที่ 5 ซึ่งมี 3 วิธี กับ วิธีของทักเคอร์ (Tucker) และ เลวีน (Levine) ในกรณีที่แบบสอบมีความเชื่อมั่นเท่ากัน โดยใช้แบบสอบหลัก แบบสอบที่ใช้เทียบคะแนนเป็นแบบสอบเพื่อให้ประกาศนียบัตรในวิชาชีพ จำนวน 5 ชุด สอบกับกลุ่มประชากรที่ไม่เท่าเทียมกัน 5 กลุ่ม กลุ่มละ 1 ชุด ปีละ 2 ครั้ง คือ แบบสอบชุด B, D และ E สอบในฤดูใบไม้ผลิ ส่วนแบบสอบชุด A กับชุด C สอบในฤดูฝน การเปรียบเทียบวิธีการเทียบคะแนนมี 2 แบบ คือ การเปรียบเทียบผลการเทียบคะแนนระหว่างแบบสอบแต่ละคู่ภายใต้วิธีการเชิงเส้นตรงที่แตกต่างกัน 5 วิธี ทั้งนี้โดยไม่เปรียบเทียบความถูกต้องของวิธีการต่าง ๆ เพราะว่ามีมาตรการคะแนนดิบของแบบสอบสองชุดถูกคาดว่าจะต่างกัน อีกแบบหนึ่งคือการเปรียบเทียบผลการเทียบคะแนนแบบสายโซ่ ซึ่งเป็นการเปรียบเทียบคะแนนแบบสอบแต่ละชุดกับตัวเอง โดยผ่านแบบสอบชุดอื่นที่เลือกมา ทั้งนี้เพื่อทำการเปรียบเทียบความถูกต้องของวิธีการเหล่านั้นด้วย

การวัดความคลาดเคลื่อนของการเทียบคะแนนใช้สถิติ 2 ตัว คือ ค่าประมาณความลำ

เอียง (estimated bias) และค่าประมาณที่ถ่วงน้ำหนักแล้วของรากที่สองของค่าเฉลี่ยยกกำลังสองของความคลาดเคลื่อนในการเทียบคะแนน (estimated weighted root squared error : RMSE) ซึ่งใช้เปรียบเทียบระหว่างวิธีที่แตกต่างกัน ซึ่งคำนวณจากสมการ ต่อไปนี้

$$\text{BIAS} = \bar{X}' - \bar{X}$$

เมื่อ $\bar{X}$ คือ ค่าเฉลี่ยของคะแนนดิบของแบบสอบชุดหนึ่ง

$\bar{X}'$ คือ ค่าเฉลี่ยของคะแนนดิบที่เท่าเทียมกัน โดยการเทียบแบบสอบชุดนั้นไปยังตัวเองแบบสายโซ่

$$\text{และ} \quad \text{RMSE} = \frac{n_1 (X'_1 - X_1)^{1/2}}{\Sigma n_1}$$

เมื่อ X_1 คือ คะแนนดิบของคนที่ i

X'_1 คือ คะแนนดิบที่เทียบได้

n_1 คือ จำนวนผู้สอบ

ในกรณีที่ผู้สอบจำนวนมาก การคำนวณ RMSE จะง่ายขึ้นโดยใช้สมการ

$$\text{RMSE} = [(m-1)^2 \text{Var}(X) + (\text{BIAS})^2]^{1/2}$$

เมื่อ m คือ ความชัน (slope) ของฟังก์ชัน $X' = mX + b$

และ $\text{Var}(X)$ คือ ความแปรปรวนของ X

ผลการวิจัยสรุปได้ว่า ในสถานการณ์ที่ศึกษาภายใต้รูปที่ 5 ตามวิธีของแองกอฟฟ์ ให้คะแนนดิบที่สัมพันธ์กันใกล้เคียงกับวิธีการของทศเคอร์และเลไวน์ โดยการเทียบคะแนนแบบสายโซ่ทั้ง 5 วิธี ให้ค่าประมาณความลำเอียงและรากที่สองของค่าเฉลี่ยยกกำลังสองของความคลาดเคลื่อนที่ใกล้เคียงกัน ส่วนการเทียบคะแนนแต่ละคู่ของแบบสอบ พบว่า คะแนนดิบที่เท่าเทียมกันจากรูปแบบ

ที่ 5 มีค่าอยู่ระหว่างคะแนนจากวิธีการของทักเคอร์และเลไวน์ ผู้วิจัยกล่าวว่าผลการวิจัยนี้สนับสนุนวิธีการเทียบคะแนนเชิงเส้นตรง ตามรูปแบบที่ 5 ซึ่งมีข้อตกลงที่มีข้อจำกัดน้อยกว่า วิธีของทักเคอร์และเลไวน์ ดังนั้น ในทางปฏิบัติควรใช้วิธีการตามรูปแบบที่ 5 วิธีการใดวิธีการหนึ่งใน 3 วิธี

คลิน และจาโจรา (Klein and Jarjoura, 1985 : 197-206) ได้ศึกษาถึงความสำคัญของเนื้อหาของข้อสอบรวม ที่ใช้เป็นตัวแทนของแบบทดสอบหลัก กับกลุ่มตัวอย่างที่แตกต่างกัน โดยใช้แบบทดสอบหลัก 2 แบบ แบบทดสอบหลักแบบที่ 1 เป็นข้อสอบรวมที่เป็นตัวแทนของ 60 ข้อและที่ไม่เป็นตัวแทนเนื้อหา 105 ข้อ แบบที่ 2 เป็นข้อสอบรวมที่เป็นตัวแทนของเนื้อหา 60 ข้อและที่ไม่เป็นตัวแทนของเนื้อหา 101 ข้อ โดยมีแบบทดสอบ ที่ใช้เทียบคะแนน เป็นแบบทดสอบเลือกตอบ 250 ข้อ แต่ละข้อกระจายอยู่ใน 6 เนื้อหา ผลการศึกษานพบว่า ความคลาดเคลื่อนจากการเทียบของข้อสอบรวมที่เป็นตัวแทนขนาดสั้น น้อยกว่าข้อสอบรวม ที่ไม่เป็นตัวแทนของเนื้อหาขนาดยาว และความลำเอียงก็เช่นเดียวกัน

แมคแคน (MacCann, 1989 : 263-276) ศึกษาการเปรียบเทียบวิธีเทียบคะแนนสอบ 2 วิธี ที่ใช้ข้อตกลงว่า มีความเชื่อมั่นเท่าเทียมกัน และใช้แบบทดสอบที่เป็นคะแนนจริงสัมพันธ์ทำการทดสอบกับกลุ่มตัวอย่างทั้งที่เป็นกลุ่มคล้ายคลึงกัน และกลุ่มที่มีความแตกต่างกัน ใช้แบบทดสอบ 5 วิชาแบ่งเป็นส่วนย่อยที่แตกต่างกัน ในแต่ละวิชา และใช้แบบทดสอบเชื่อมโยง ที่มีความยาวคิดเป็นสัดส่วนแตกต่างกัน ทั้งที่มากกว่าและน้อยกว่าแบบทดสอบที่นำมาใช้เทียบคะแนน แล้วเปรียบเทียบความคลาดเคลื่อนมาตรฐานของแต่ละวิธี (4 วิธี) คือวิธีของแมคแคน 2 วิธี และ วิธีของทักเคอร์ กับวิธีของเลไวน์ พบว่า วิธีที่ 1 ให้ผลใกล้เคียงกับวิธีของทักเคอร์ ส่วนวิธีที่ 2 ให้ผลใกล้เคียงกับวิธีของเลไวน์ และทั้ง 4 วิธีให้ผลความคลาดเคลื่อนมาตรฐาน ในระดับที่น่าพอใจเป็นส่วนมาก ผู้วิจัยสรุปว่าข้อดีของวิธีที่ 1 กับวิธีที่ 2 นั้นคือไม่จำเป็นต้องใช้ข้อตกลงที่ว่า การกระจายของคะแนนจากแบบทดสอบ 2 ฉบับที่นำมาเทียบคะแนนกันมีการแจกแจงเป็นโค้งปกติ

โคเลน (Kolen, 1985 : 209-223) ได้ศึกษาเรื่องความคลาดเคลื่อนของวิธีการเทียบคะแนนแบบทักเคอร์ กับกลุ่มตัวอย่างขนาด 100 และ 250 คน โดยใช้รูปแบบข้อสอบร่วมกับประชากรที่แตกต่างกัน แบบทดสอบที่ใช้เป็นแบบทดสอบเทียบคะแนน 2 ฉบับและแบบทดสอบรวมอีก

1 ฉบับ การคำนวณความคลาดเคลื่อนคำนวณโดยคอมพิวเตอร์ โดยใช้ข้อมูลจริงประมาณโดยใช้วิธี Bootstrap ของ Efron's (1982) ผลการศึกษาพบว่าความคลาดเคลื่อนของกลุ่มที่ไม่เป็นไปตามข้อตกลงปกติ และกลุ่มตัวอย่างขนาดใหญ่ (250 คน) มีความคลาดเคลื่อนน้อยกว่ากลุ่มตัวอย่างขนาดเล็ก (100 คน) และกลุ่มตัวอย่างที่เป็นไปตามข้อตกลง

โคเลน และเบรนนอน (Kolen and Brennan. 1987 : 263-277) ได้ศึกษาวิธีการเทียบคะแนนเชิงเส้นตรงกับแบบสอบหลักที่มีประชากรแตกต่างกัน โดยใช้วิธีการของทักเคอร์ และเลไวน์ ผลการเปรียบเทียบ พบว่าวิธีการของเลไวน์ ถ่วงน้ำหนักความแตกต่างของกลุ่มผู้สอบได้มากกว่า วิธีของทักเคอร์ ซึ่งจะช่วยอธิบายความแตกต่างระหว่างผลที่เกิด เมื่อใช้ข้อมูลจริง และจะเป็นประโยชน์ต่อการเทียบคะแนนเชิงเส้นตรงที่ใช้แบบสอบหลักที่มีประชากรแตกต่างกัน

วูดรUFF (Woodruff.1989 : 257-261) ได้ศึกษาเรื่องการเปรียบเทียบวิธีการเทียบคะแนนเชิงเส้นตรง 3 วิธี คือใช้รูปแบบคะแนนจริงสัมพันธ์ของเลไวน์ วิธีเทียบคะแนนของแองกอฟฟ์-เลไวน์ และวิธีเทียบคะแนนของทักเคอร์ โดยใช้แบบสอบหลักที่มีประชากรแตกต่างกัน โดยศึกษาจากความสัมพันธ์ระหว่างแบบสอบที่ใช้เทียบคะแนน กับแบบสอบหลัก ผลปรากฏว่าวิธีของแองกอฟฟ์-เลไวน์ มีความไวมากกว่าของทักเคอร์ ที่เนื้อหาไม่สมดุลระหว่างแบบสอบเทียบคะแนน กับแบบสอบหลัก วิธีการของทักเคอร์ไม่สามารถป้องกันความสามารถที่แตกต่างของผู้สอบได้ และวิธีของแองกอฟฟ์-เลไวน์ ก็ไม่สมควรใช้เมื่อความสัมพันธ์ของแบบทดสอบฉบับ Y และแบบทดสอบหลัก (V) มีค่าน้อยกว่า 1 แต่ถ้าใช้รูปแบบคะแนนจริงสัมพันธ์ จะมีข้อตกลงที่ยอมให้ความสัมพันธ์ของแบบทดสอบเทียบคะแนนกับแบบทดสอบหลักมีความสัมพันธ์น้อยกว่า 1 ได้ จากการพิสูจน์นี้ทำให้ทราบว่า วิธีที่ใช้รูปแบบคะแนนจริงสัมพันธ์ของเลไวน์ จะทำได้อย่างมีประสิทธิภาพในทางปฏิบัติ แต่ถ้าข้อมูลเป็นไปตามข้อตกลงอย่างเพียงพอวิธีทั้งสองจะทำได้ง่ายกว่าวิธีที่ใช้รูปแบบคะแนนจริงสัมพันธ์ของเลไวน์

พาแชล, ฮ็อจตัน และครอมเรย์ (Parshall, Houghton and Kromrey.1995 : 37-54) ศึกษาความคลาดเคลื่อนในการเทียบคะแนน และความลำเอียงทางสถิติในกลุ่มตัวอย่างขนาดเล็ก โดยวิธีเทียบคะแนนเชิงเส้นตรง โดยใช้แบบแผนข้อสอบร่วมที่มีประชากรแตกต่างกัน (The Common-Item Nonequivalent-groups design) ได้ศึกษากับกลุ่มตัวอย่างขนาด 15, 25,

50, 100 และ 500 คน จากประชากรเทียม(Pseudopopulations) 2 กลุ่ม กลุ่มละ 1,000 คน โดยวิธีสุ่มแบบไล่คั่น ขนาดกลุ่มตัวอย่างขนาดละ 1,000 ครั้ง เครื่องมือที่ใช้เป็นแบบทดสอบ 2 ฉบับที่มีลักษณะคู่ขนานกัน และมีข้อสอบรวมภายใน มี 5 เนื้อหาวิชา คือ

1. การศึกษาศิลปะ ฉบับ A มี 105 ข้อ ฉบับ B มี 109 ข้อ มีข้อสอบรวม 68 ข้อ
2. คณิตศาสตร์ ฉบับ A และฉบับ B มี 99 ข้อ มีข้อสอบรวม 68 ข้อ
3. การศึกษาดนตรี ฉบับ A มี 100 ข้อ ฉบับ B มี 99 ข้อ มีข้อสอบรวม 58 ข้อ
4. การศึกษาเด็กก่อนวัยเรียน ฉบับ A มี 146 ข้อ ฉบับ B มี 144 ข้อ ข้อสอบรวม

69 ข้อ และ

5. ภาษาสเปน ฉบับ A และ B มี 91 ข้อ มีข้อสอบรวม 47 ข้อ

ผลการศึกษาพบว่า

1. ความคลาดเคลื่อนจากการสุ่มน้อยมากเมื่อคะแนนใกล้เคียงค่าเฉลี่ย และจะเพิ่มขึ้นเรื่อยๆ เช่นเดียวกับการกระจายของคะแนนจากค่าเฉลี่ย

2. ความคลาดเคลื่อนมาตรฐานของคะแนนจะเพิ่มขึ้น เมื่อขนาดของกลุ่มตัวอย่างลดลง

3. ความลำเอียงทางสถิติ มีผลต่อกับขนาดของกลุ่มตัวอย่างน้อย เมื่อคะแนนเบี่ยงเบนออกจากค่าเฉลี่ย 1.5 SD

จากผลการวิจัยครั้งนี้ ทำให้ทราบข้อจำกัดที่สำคัญ คือ

1. การใช้ข้อทดสอบรวมประมาณ 50 % ทำให้ครอบคลุมเนื้อหาของแบบทดสอบ ที่ต้องการเทียบคะแนน

2. ตัวอย่างที่นำมาจากประชากรเทียม ถ้ามีความสามารถใกล้เคียงกัน จะทำให้ความคลาดเคลื่อนมาตรฐานน้อย

3. ถ้าใช้แบบทดสอบเชื่อมโยงเพียง 1 แบบทดสอบ (a single link) จะพบว่าการประมาณค่าความลำเอียงทางสถิติ และความคลาดเคลื่อนมาตรฐานจะมีขนาดเล็กกว่า การใช้แบบทดสอบเชื่อมโยงหลายฉบับ (multiple links)

งานวิจัยในประเทศ

เยาวดี ราชชัยกุล(ชูชีพ พงษ์สมบูรณ์. 2528 ; อ้างอิงมาจาก เยาวดี ราชชัยกุล. 2518) ได้ศึกษาการเทียบมาตรฐานโดยมีจุดมุ่งหมายสำรวจขอบเขตของแบบทดสอบแบบต่าง ๆ ใน

โปรแกรมการสอบมัธยมศึกษาปีที่ 5 ในประเทศไทย ว่ามีความเท่าเทียมกันเพียงใดใช้ข้อมูลจาก กระทรวงศึกษาธิการ ซึ่งเป็นผลการสอบของนักเรียนสายวิทยาศาสตร์ ที่สอบในเดือนมีนาคมและ สิงหาคม พ.ศ.2516 และ พ.ศ.2517 รวม 4 ครั้ง จำนวน 500 คน โดยใช้วิธีการเทียบเชิงเส้นตรง รูปแบบที่ 1 ของ Angoff ผลการศึกษาพบว่า เมื่อใช้สเกลร่วม (Common scale) ที่มีช่วงคะแนน 1-175 ค่าเฉลี่ย 100 ค่าเบี่ยงเบนมาตรฐาน 25 คะแนนที่ใช้เทียบทั้ง 4 ครั้ง มีคะแนนไม่เท่าเทียมกันทุก ๆ วิชา (ภาษาไทย, ภาษาอังกฤษ, สังคมศึกษา, คณิตศาสตร์ และ วิทยาศาสตร์) และยังพบว่าคะแนนจุดตัด (cut-off score) ของแต่ละครั้งไม่สามารถเปรียบเทียบกันได้

ซูชีน พงษ์สมบูรณ์ (2528) ได้เปรียบเทียบการเทียบมาตรฐานระหว่างรูปแบบที่ใช้ผู้สอบร่วมกับรูปแบบที่ใช้แบบสอบหลักกับกลุ่มตัวอย่างจากโครงการตรวจสอบคุณภาพการศึกษา ชั้นมัธยมศึกษาปีที่ 6 ที่สอบระหว่างวันที่ 13-14 ธันวาคม 2526 จำนวน 3,721 คนโดยใช้วิธีการเทียบมาตรฐาน 3 วิธี คือ การเทียบเชิงเส้นตรง การเทียบโดยใช้เปอร์เซ็นต์ไทล์ การเทียบโดยใช้โด่งลักษณะข้อสอบ ผลปรากฏว่า ประสิทธิภาพการเทียบมาตรฐานของ 3 วิธีทั้งผู้สอบร่วม และแบบสอบหลักไม่แตกต่างกัน

ภาวิณี ศรีสุขวัฒนานนท์(2529) ได้เปรียบเทียบผลของการใช้รูปแบบเทียบมาตรฐานที่ต่างกัน 3 รูปแบบ คือ รูปแบบอิกวิเปอร์เซ็นต์ไทล์ รูปแบบเชิงเส้นตรง และรูปแบบอิงทฤษฎีการตอบข้อสอบ แบบโลจิสต์ 3 พารามิเตอร์ โดยใช้ข้อสอบร่วมภายในร้อยละ 20, 40 และ 60 ตามลำดับของแบบสอบที่ใช้เทียบคะแนน 35 ข้อ ศึกษากับผู้สอบคัดเลือกเข้าทำงาน และนักศึกษาปริญญาตรีของสถาบันอุดมศึกษาแห่งหนึ่ง ผลปรากฏว่า การเทียบมาตรฐานกับกลุ่มผู้สอบคัดเลือก รูปแบบที่ได้ผลดี คือ รูปแบบอิกวิเปอร์เซ็นต์ไทล์ ส่วนกลุ่มนักศึกษาสอบวัดผลสัมฤทธิ์ รูปแบบที่ให้ผลดี คือ รูปแบบเชิงเส้นตรงและพบว่า ความยาวของแบบทดสอบร่วม มีผลต่อความแม่นยำ และความเพียงพอของการเทียบมาตรฐาน

เรวดี อินทสระ (2530) ได้เปรียบเทียบความคลาดเคลื่อนมาตรฐานของ การเทียบคะแนนและความเพียงพอของวิธีการเทียบคะแนน 2 วิธี คือรูปแบบอิงทฤษฎีการตอบข้อสอบ ที่ใช้พารามิเตอร์ 3 ตัว และรูปแบบที่ใช้เทคนิคการวิเคราะห์องค์ประกอบกับกลุ่มตัวอย่างคือนักเรียน

ชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2529 จังหวัดพัทลุง จำนวน 1,557 คนโดยใช้แบบทดสอบ 2 ชุด ชุดละ 30 ข้อ แต่ละชุดมีข้อสอบรวมอีก 15 ข้อ ผลปรากฏว่า การเทียบคะแนนรูปแบบอิงทฤษฎี การตอบข้อสอบให้ความเพียงพอในระดับที่น่าพอใจ ส่วนรูปแบบที่ใช้เทคนิคการวิเคราะห์องค์ประกอบ ให้ความเพียงพอในระดับปานกลาง สำหรับการเปรียบเทียบความคลาดเคลื่อนมาตรฐานพบว่า การใช้เทคนิคการวิเคราะห์องค์ประกอบ มีความคลาดเคลื่อนสูงกว่า รูปแบบอิงทฤษฎีการตอบข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อาภรณ์ กาญจนกิจโสภณ (2531) ได้ศึกษาการสร้างแบบทดสอบและตารางเทียบระดับคะแนนตามแนวคิดในวิชาคณิตศาสตร์ โดยใช้วิธีเทียบคะแนนแบบราสส์ กับนักเรียนมัธยมศึกษาตอนต้น (มัธยมศึกษาปีที่ 1, ปีที่ 2 และปีที่ 3) จำนวนระดับชั้นละ 490, 482 และ 415 คน ตามลำดับ ใช้แบบทดสอบที่สร้างขึ้น ระดับชั้นละ 40 ข้อ 30 ข้อ และ 30 ข้อ โดยมีแบบสอบร่วมฉบับละ 10 ข้อ ได้นำแบบทดสอบไปทดสอบกับกลุ่มตัวอย่างแล้วสร้างตารางเทียบระดับคะแนน เพื่อศึกษาพัฒนาการของนักเรียน

สุรัตน์ ขวัญญูจันทร์ (2531) ได้ศึกษาการสร้างแบบทดสอบและตารางการเทียบคะแนนตามแนวนอนในวิชาวิทยาศาสตร์ ด้วยวิธีราสส์โมเดล กับนักเรียนมัธยมศึกษาปีที่ 2 ปีการศึกษา 2531 จำนวน 1,391 คน โดยสร้างแบบทดสอบ 2 ฉบับฉบับละ 60 ข้อ และมีข้อสอบรวมภายใน 10 ข้อ เมื่อนำไปทดสอบกับกลุ่มตัวอย่างแล้วนำมาสร้างตารางการเทียบคะแนน พบว่า ภายใต้อาณาเขตความสามารถที่เท่ากัน ได้คะแนนดิบแต่ละค่าเท่ากันทุกคะแนน และได้ค่าดัชนีความเพียงพอของความคลาดเคลื่อนในการเทียบคะแนนอยู่ในระดับน่าพอใจมาก

สุจินดา ผ่องอักษร (2533) ได้ศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียน ในกลุ่มทักษะของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษาพุทธศักราช 2521 ในช่วงระยะเวลา 3 ปีการศึกษา โดยใช้การเทียบคะแนนรูปแบบราสส์ ข้อมูลได้จากโครงการประเมินคุณภาพนักเรียนชั้นประถมศึกษาปีที่ 6 ของการประถมศึกษาแห่งชาติ ปีการศึกษา 2529-2531 โดยใช้แบบสอบ 2 ฉบับ (คณิตศาสตร์กับภาษาไทย) ผลปรากฏว่า นักเรียนมีความก้าวหน้าในการเรียนกลุ่มทักษะอยู่ในระดับน่าพอใจ

จากการศึกษางานวิจัยที่ผ่านมา ทำให้ผู้วิจัยต้องการศึกษาถึงความคลาดเคลื่อนมาตรฐาน และความลำเอียงทางสถิติ ในการเทียบคะแนนเชิงเส้นตรงแบบคะแนนจริงสัมพันธ์ในกลุ่มตัวอย่างขนาดเล็ก โดยต้องการพิสูจน์ว่า กลุ่มตัวอย่างขนาดเล็ก ขนาดเท่าใดที่มีความคลาดเคลื่อนมาตรฐานและความลำเอียงทางสถิติ ใกล้เคียงกับกลุ่มตัวอย่างขนาดใหญ่ จนสามารถใช้แทนกลุ่มตัวอย่างขนาดใหญ่ได้

วิธีดำเนินการศึกษาค้นคว้า

ประชากร

ประชากร

ประชากรที่ใช้ในการศึกษาค้นคว้าครั้งนี้ คือนักเรียนชั้นประถมศึกษาปีที่ 6 ของโรงเรียนสังกัด กรุงเทพมหานคร เขตบางเขน เขตดอนเมือง เขตจตุจักร

ประชากรเทียม (Pseudopopulations)

ประชากรเทียมที่ใช้ในการศึกษาค้นคว้าครั้งนี้ คือนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรุงเทพมหานคร เขตบางเขน จำนวน 12 โรงเรียน เขตดอนเมือง จำนวน 1 โรงเรียน เขตจตุจักร จำนวน 1 โรงเรียน จำนวนนักเรียน 1,826 คน โดยแบ่งเป็น 2 กลุ่มตามขนาดของโรงเรียน เพื่อให้มีจำนวนนักเรียนใกล้เคียงกัน แล้วทำการสุ่มประชากรเทียมแต่ละกลุ่ม เป็นกลุ่มตัวอย่างขนาด 30 คน 60 คน 120 คน และ 250 คน ในกลุ่มตัวอย่างแต่ละขนาด สุ่มทั้งหมด 200 ครั้ง ได้กลุ่มตัวอย่างที่เป็นกลุ่มย่อยทั้งหมด 800 กลุ่มตัวอย่าง ดังแสดงไว้ในตาราง 1

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ คือนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2538 เขตบางเขน เขตดอนเมือง และเขตจตุจักร สังกัดกรุงเทพมหานคร ที่ผู้วิจัยทำการทดสอบไว้ แล้วนำคะแนนมาเก็บไว้ในแผ่นจานแม่เหล็ก แล้วทำการสุ่มอย่างง่าย (Simple Random Sampling) จากประชากรเทียมทั้งหมด โดยมีนักเรียนเป็นหน่วยการสุ่ม ซึ่งมีวิธีสุ่มดังนี้

1. จากประชากรเทียมกลุ่มที่ 1 ทำการสุ่มแบบใส่คืน (With-replacement) โดยสุ่มกลุ่มตัวอย่างขนาด 30 คน 60 คน 120 คน และ 250 คน ขนาดละ 200 ครั้ง รวมเป็นกลุ่มตัวอย่างทั้งหมด 800 กลุ่มตัวอย่าง

2. ทำการสุ่มจากประชากรเทียมกลุ่มที่ 2 เช่นเดียวกับข้อ 1

3. กลุ่มตัวอย่างที่ใช้ในการวิจัยนี้ จะมีทั้งสิ้น 1,600 กลุ่ม จาก 2 กลุ่มประชากรเทียบ

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นแบบทดสอบวัดความสามารถทางสมองตามทฤษฎีของเทอร์สโตน (Thurstone) ด้านการรับรู้ (Perception Factor) จำนวน 3 ฉบับ เป็นแบบทดสอบที่ทำให้คู่ขนานแบบคะแนนจริงสัมพันธ์ 2 ฉบับ คือฉบับ X และฉบับ Y ใช้เป็นแบบทดสอบเทียบคะแนน สำหรับฉบับที่ 3 หรือฉบับ Z เป็นแบบทดสอบเชื่อมโยง (External Anchor Test) ที่มีจำนวนข้อหรือความยาวเป็น 50 % ของแบบทดสอบที่ใช้เทียบคะแนน นอกจากนั้นลักษณะคำถามยังเหมือนกับแบบทดสอบที่ใช้ในการเทียบคะแนน เพราะจากการศึกษาพบว่าความยาว และความครอบคลุมเนื้อหาของแบบทดสอบเชื่อมโยง มีผลต่อความแม่นยำ ความเพียงพอ ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน (ภาวิณี ศรีสุขวัฒนานนท์, 2529 ; คลิน และจาโจรา (Klein and Jarjoura, 1985 : 197 - 206) ; พาร์แชล, ฮอจตัน และครอมเรย์ (Parshall, Houghton and Kromrey, 1995 : 37 - 54) ; สุมนา โสติดิผลอนันต์, 2539) ดังนั้น รายละเอียดดังนี้

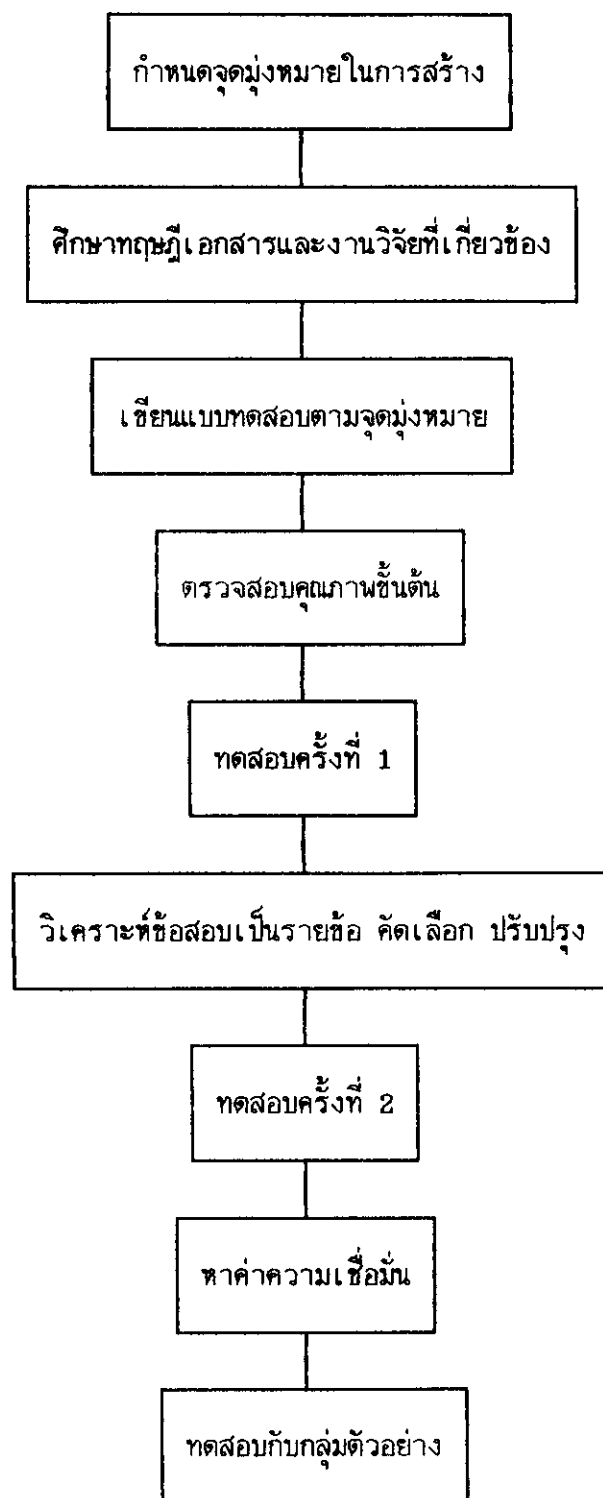
ฉบับ X มีสองตอน ตอนที่หนึ่งประกอบด้วยลักษณะคำถามประเภทถามเหมือน คือให้หารูปที่เหมือนกับรูปที่กำหนดให้ จำนวน 30 ข้อ ตอนที่สองเป็นคำถามประเภทถามไม่เหมือน คือให้หารูปที่ไม่เหมือนกับรูปที่กำหนดให้จำนวน 30 ข้อ รวมทั้งฉบับมี 60 ข้อ

ฉบับ Y ก็มีลักษณะคำถามเช่นเดียวกับฉบับ X คือตอนที่หนึ่งเป็นคำถามประเภทถามเหมือน คือให้หารูปที่เหมือนกับรูปที่กำหนดให้ จำนวน 30 ข้อ ตอนที่สองเป็นคำถามประเภทถามไม่เหมือน คือให้หารูปที่ไม่เหมือนกับรูปที่กำหนดให้จำนวน 30 ข้อ รวมทั้งฉบับมี 60 ข้อ

ฉบับ Z เป็นแบบทดสอบเชื่อมโยงโดยมีลักษณะคำถามเช่นเดียวกับฉบับ X และฉบับ Y แต่มีจำนวนข้อเพียงครึ่งหนึ่งของฉบับ X และฉบับ Y ดังนั้น ตอนที่หนึ่งเป็นคำถามประเภทถามเหมือน คือให้หารูปที่เหมือนกับที่กำหนดให้จึงมีจำนวน 15 ข้อ ตอนที่สองเป็นคำถามประเภทถามไม่เหมือน คือให้หารูปที่ไม่เหมือนกับที่กำหนดให้จำนวน 15 ข้อ รวมทั้งฉบับมีจำนวน 30 ข้อ

วิธีดำเนินการสร้างเครื่องมือ

การดำเนินการสร้างเครื่องมือสำหรับใช้ในการศึกษาค้นคว้าครั้งนี้ผู้วิจัยได้ดำเนินการสร้างตามขั้นตอนดังนี้



ภาพประกอบ 7 แสดงลำดับขั้นตอนการสร้างเครื่องมือ

จากภาพประกอบ 7 แสดงลำดับขั้นตอนการสร้างเครื่องมือ ซึ่งมีรายละเอียดในการปฏิบัติ ดังนี้

1. กำหนดจุดมุ่งหมายในการสร้างแบบทดสอบ เพื่อวัดความสามารถทางสมองหรือสติปัญญา ด้านการรับรู้ (Perception Factor) ตามทฤษฎีของเทอร์สโตน (Thurstone)

2. ศึกษาทฤษฎี เอกสารและงานวิจัยที่เกี่ยวข้อง โดยศึกษาทฤษฎีเข้าวนปัญญาและความถนัด ทฤษฎีโครงสร้างทางสมองของเทอร์สโตน ด้านการรับรู้ ความหมายของการรับรู้ องค์ประกอบ ต่างๆ ที่มีส่วนเกี่ยวข้องกับการรับรู้ การสร้างแบบทดสอบวัดการรับรู้

3. เขียนข้อสอบด้านการรับรู้ 3 ฉบับ แต่ละฉบับมี 2 ตอน ฉบับ X และฉบับ Y ตอนที่หนึ่ง ประกอบด้วยลักษณะคำถามประเภทถามเหมือน คือให้หารูปที่เหมือนกับรูปที่กำหนดให้ จำนวน 60 ข้อ ตอนที่สองลักษณะคำถามเป็นประเภทถามไม่เหมือน คือให้หารูปที่ไม่เหมือนกับรูปที่กำหนดให้ จำนวน 60 ข้อ รวมทั้งสองตอน ฉบับละ 120 ข้อ แต่ฉบับ Z ซึ่งเป็นแบบทดสอบเชื่อมโยง ตอนที่หนึ่งจะประกอบด้วยลักษณะคำถามประเภทถามเหมือน คือให้หารูปที่เหมือนกับรูปที่กำหนดให้ จำนวน 30 ข้อ ตอนที่สองเป็นประเภทถามไม่เหมือน คือให้หารูปที่ไม่เหมือนกับรูปที่กำหนดให้ จำนวน 30 ข้อ รวม 60 ข้อ ในการเขียนข้อสอบแต่ละข้อแต่ละฉบับใช้วิธีการเดียวกัน เช่น ต้องการให้ดูจำนวนวงที่ซ้อนกัน ในฉบับ X ใช้รูปวงกลมซ้อนกัน 5 วง ในฉบับ Y ใช้รูปสี่เหลี่ยมซ้อนกัน 5 รูป และในฉบับ Z ใช้รูปหกเหลี่ยมซ้อนกัน 5 รูป เป็นต้น

4. ตรวจสอบคุณภาพขั้นต้น โดยนำไปทดสอบกับกลุ่มทดลองที่ไม่ใช่กลุ่มตัวอย่างจำนวน 30 คน เพื่อตรวจสอบเรื่องภาษาที่ใช้ในคำถาม คำสั่งของแบบทดสอบ ความเหมาะสมของเวลาที่ใช้ในการสอบ พบว่าเวลาที่เหมาะสม ฉบับ X และฉบับ Y จำนวน 60 ข้อ ใช้เวลา 6 นาที ฉบับ Z จำนวน 30 ข้อ ใช้เวลา 3 นาที

5. ปรับปรุงและทดสอบครั้งที่ 1 กับกลุ่มทดลองที่ไม่ใช่กลุ่มตัวอย่างจำนวน 100 คน

6. วิเคราะห์ข้อสอบเป็นรายข้อ เพื่อหาค่าความยาก (P), ค่าอำนาจจำแนก (r) โดย

ฉบับ X คัดเลือกค่า P อยู่ระหว่าง .31 - .80 ค่า r อยู่ระหว่าง .24 - .72

ฉบับ Y คัดเลือกค่า P อยู่ระหว่าง .31 - .80 ค่า r อยู่ระหว่าง .22 - .66

ฉบับ Z คัดเลือกค่า P อยู่ระหว่าง .32 - .80 ค่า r อยู่ระหว่าง .20 - .70

7. นำแบบทดสอบทั้ง 3 ฉบับไปทดสอบกับกลุ่มทดลองจำนวน 100 คน แล้วหาค่าความเชื่อมั่นของแบบทดสอบ โดยใช้สูตรสัมประสิทธิ์แอลฟา (KR-20) ได้ค่าความเชื่อมั่นดังนี้ แบบทดสอบฉบับ X มีค่าความเชื่อมั่น .9162 แบบทดสอบฉบับ Y มีค่าความเชื่อมั่น .9148

และแบบทดสอบฉบับ Z มีค่าความเชื่อมั่น .8923 ซึ่งค่าความเชื่อมั่นของแบบทดสอบทั้ง 3 ฉบับมีค่าใกล้เคียงกัน

8. นำแบบทดสอบทั้ง 3 ฉบับ ซึ่งอยู่ในภาคผนวก ไปทดสอบกับกลุ่มตัวอย่าง

วิธีดำเนินการเก็บรวบรวมข้อมูล

วิธีดำเนินการเก็บรวบรวมข้อมูลได้ดำเนินการตามขั้นตอนดังต่อไปนี้

1. ติดต่อขอความร่วมมือกับโรงเรียนที่ใช้เป็นกลุ่มตัวอย่าง โดยกำหนดวันเก็บรวบรวมข้อมูลในระหว่างวันที่ 4 - 14 มีนาคม 2539
2. จัดเตรียมแบบทดสอบให้เพียงพอกับจำนวนประชากรเทียมที่ศึกษา
3. นำแบบทดสอบฉบับ X และฉบับ Z ไปทดสอบกับประชากรเทียมกลุ่มที่ 1 และนำแบบทดสอบฉบับ Y และฉบับ Z ไปทดสอบกับประชากรเทียมกลุ่มที่ 2
4. คัดเลือกกระดาษคำตอบที่สมบูรณ์ไว้ กลุ่มประชากรเทียมละ 913 ชุด
5. นำผลการสอบมาตรวจให้คะแนน โดยตอบถูกให้ 1 คะแนน ตอบผิดให้ 0 คะแนน
6. นำข้อมูลที่ได้จากการตรวจมาวิเคราะห์

ขั้นตอนการวิเคราะห์ข้อมูล

แบ่งเป็น 2 ตอน

1. การวิเคราะห์ข้อมูลในการทดลองเครื่องมือ (Try out) ได้ดำเนินการดังนี้

1.1 หาค่าความยาก(P) โดยใช้สูตรอย่างง่าย และค่าอำนาจจำแนก(r) โดยใช้ค่าสหสัมพันธ์แบบพอยท์-ไบซีเรียล ได้ค่าความยากและค่าอำนาจจำแนกดังนี้

ฉบับ X ได้ค่า P อยู่ระหว่าง .31 - .80 ค่า r อยู่ระหว่าง .24 - .72

ฉบับ Y ได้ค่า P อยู่ระหว่าง .31 - .80 ค่า r อยู่ระหว่าง .22 - .66

ฉบับ Z ได้ค่า P อยู่ระหว่าง .32 - .80 ค่า r อยู่ระหว่าง .20 - .70

1.2 หาค่าความเชื่อมั่น โดยใช้สูตรสัมประสิทธิ์แอลฟา (KR-20) ของคูเดอร์ และริชาร์ดสัน (Kuder and Richardson, 1937) (บุญเชิด ภิญโญนนตพงษ์ .2523:256) ได้ค่าความเชื่อมั่นดังนี้

ฉบับ X มีค่าความเชื่อมั่น .9162 ฉบับ Y มีค่าความเชื่อมั่น .9148

ฉบับ Z มีค่าความเชื่อมั่น .8923

2. การวิเคราะห์ข้อมูลการวิจัยได้ดำเนินการดังนี้

2.1 ทหาค่าสถิติพื้นฐานของทุกกลุ่ม คือ ค่าเฉลี่ย($\bar{X}$) และค่าเบี่ยงเบนมาตรฐาน(SD) ค่าสหสัมพันธ์ (r) ระหว่างแบบทดสอบที่ใช้เทียบคะแนน กับแบบทดสอบเชื่อมโยง เสนอผลไว้ในหน้า 69

2.2 ทหาคะแนนแปลงจากการเทียบคะแนนโดยใช้สูตรของ แมคแคน(MacCann. 1989) เสนอไว้ในภาคผนวก

2.3 เปรียบเทียบคะแนนเทียบเฉลี่ยของแต่ละขนาดของกลุ่มตัวอย่าง เสนอผลไว้ในหน้า 70

2.4 ทหาค่าความคลาดเคลื่อนมาตรฐาน ของคะแนนจากการเทียบ ของคะแนนสังเกต ทุกตัว โดยใช้สูตรของพาร์แชล, ฮอจตัน และครอมเรย์(Parshall, Houghton and Kromrey. 1995) เสนอผลไว้ในหน้า 71

2.5 ทหาค่าความลำเอียงทางสถิติของคะแนนสังเกตทุกตัว โดยใช้สูตรของพาร์แชล, ฮอจตัน และ ครอมเรย์ (Parshall, Houghton and Kromrey. 1995) เสนอผลไว้ในหน้า 74

2.6 สร้างตารางเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐาน และความลำเอียงทางสถิติ ของทุกขนาดกลุ่มตัวอย่าง เสนอผลไว้ในหน้า 77 และ 78

2.7 พิจารณากลุ่มตัวอย่างที่เหมาะสมจาก คะแนนเทียบเฉลี่ย ตารางเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐาน และความลำเอียงทางสถิติ

สถิติที่ใช้ในการวิเคราะห์ข้อมูล

1. ทหาค่าสถิติพื้นฐานของทุกกลุ่ม คือ ค่าเฉลี่ย($\bar{X}$) และค่าเบี่ยงเบนมาตรฐาน(SD) ค่าสหสัมพันธ์ (r) ระหว่างแบบทดสอบที่ใช้เทียบคะแนนกับแบบทดสอบเชื่อมโยง

2. ทหาค่าความยาก (P) โดยใช้สูตรอย่างง่าย (พวงรัตน์ ทวีรัตน์. 2535 : 136)

$$P = \frac{R}{N}$$

เมื่อ	P แทน	ค่าความยากของคำถามแต่ละข้อ
	R แทน	จำนวนผู้ตอบถูกในแต่ละข้อ
	N แทน	จำนวนผู้เข้าสอบทั้งหมด

3. หาค่าอำนาจจำแนก (r) โดยใช้ค่าสหสัมพันธ์แบบพอยท์-ไบซีเรียล (Allen and Yen. 1979 : 34)

$$r_{pbis} = \left[\frac{\bar{Y}_1 - \bar{Y}}{s_y} \right] \sqrt{\frac{p_x}{(1 - p_x)}}$$

เมื่อ	r_{pbis}	แทน ค่าอำนาจจำแนก
	$\bar{Y}_1$	แทน คะแนนเฉลี่ยของนักเรียนที่ทำข้อนั้นถูก
	$\bar{Y}$	แทน คะแนนเฉลี่ยของนักเรียนที่ทำข้อนั้นผิด
	s_y	แทน คะแนนเบี่ยงเบนมาตรฐานของแบบทดสอบฉบับนั้น
	p_x	แทน สัดส่วนของคนที่ทำข้อนั้น ได้

4. หาค่าความเชื่อมั่น โดยใช้สูตรสัมประสิทธิ์แอลฟา (KR-20) ของคูเดอร์ และ ริชาร์ดสัน (Kuder and Richardson.1937) (บุญเชิด ภิญโญนนตพงษ์ .2523:256)

$$r_{tt} = \frac{K}{K - 1} \left[\frac{S^2_t - \sum pq}{S^2_t} \right]$$

เมื่อ	r_{tt}	แทน ค่าความเชื่อมั่นของข้อสอบแต่ละฉบับ
	K	แทน จำนวนข้อสอบ
	P	แทน อัตราส่วนของคนที่ทำถูกในแต่ละข้อ
	q	แทน $1 - p$
	S^2_t	แทน ความแปรปรวนของคะแนนที่สอบได้

4. หาคะแนนแปลงจากการเทียบคะแนนโดยใช้สูตรของ แมคแคน (MacCann, 1989)

$$Y' = \frac{\begin{bmatrix} r_{yzz} & S_{y2} & S_{z1} \\ r_{xz1} & S_{x1} & S_{z2} \end{bmatrix}}{\begin{bmatrix} r_{yzz} & S_{y2} \\ r_{xz1} & S_{z2} \end{bmatrix}} \begin{bmatrix} 1 & S_{z1} \\ r_{xz1} & S_{x1} \end{bmatrix} X + \bar{Y}_2 + \begin{bmatrix} 1 & S_{z1} \\ r_{xz1} & S_{x1} \end{bmatrix} (\bar{Z}_1 - \bar{Z}_2) \bar{X}_1$$

เมื่อ Y'	คือ	คะแนนเทียบที่แปลงจากแบบทดสอบทั้งสามฉบับ
r_{xz1}, r_{xz2}	คือ	ค่าความเชื่อมั่นของแบบทดสอบระหว่างแบบทดสอบฉบับ X กับฉบับ Z และความเชื่อมั่นของแบบทดสอบฉบับ Y กับฉบับ Z ตามลำดับ
S_x, S_y	คือ	ค่าความเบี่ยงเบนมาตรฐานของแบบทดสอบฉบับ X และแบบทดสอบฉบับ Y ตามลำดับ
X, Y	คือ	คะแนนดิบจากแบบทดสอบเพื่อนำมาเทียบคะแนน 2 ฉบับ
Z	คือ	คะแนนดิบจากแบบทดสอบเชื่อมโยง

5. หาค่าความคลาดเคลื่อนมาตรฐานของคะแนนจากการเทียบโดยใช้สูตรของพาร์แชล, ฮัจตัน และ ครอมเวย์ (Parshall, Houghton and Kromrey, 1995)

$$SE(\theta_i) = \sqrt{\sum (\theta_{ij} - \theta'_i)^2 / n-1}$$

เมื่อ $SE(\theta_i)$	คือ	ความคลาดเคลื่อนมาตรฐานของคะแนนเทียบตัวที่ i
θ_{ij}	คือ	คะแนนเทียบคะแนนตัวที่ i ของกลุ่มตัวอย่าง j
θ'_i	คือ	คะแนนเทียบคะแนนตัวที่ i เฉลี่ยจาก 200 กลุ่มตัวอย่าง

6. ทาค่าความลำเอียงทางสถิติ โดยใช้สูตรของพาร์แชล, ฮ็อจตัน และ ครอมเรย์
(Parshall, Houghton and Kromrey. 1995)

$$\text{Bias}(\theta_1) = \theta'_1 - \theta_1$$

เมื่อ	$\text{Bias}(\theta_1)$	คือ ความลำเอียงทางสถิติของคะแนนเทียบคะแนนตัวที่ i
	θ'_1	คือ คะแนนเทียบคะแนนตัวที่ i เฉลี่ยจาก 200 กลุ่มตัวอย่าง
	θ_1	คือ คะแนนเทียบที่มาจากคะแนนตัวที่ i ซึ่งคำนวณจากกลุ่มตัวอย่างขนาดใหญ่ (913 คน)

ผลการวิเคราะห์ข้อมูล

สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล

เพื่อความเข้าใจที่ตรงกันในการแปลความหมายผลการวิเคราะห์ข้อมูล ผู้วิจัยจึงกำหนดสัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล ดังต่อไปนี้

$\bar{X}$	แทน ค่าเฉลี่ยของคะแนน
SD	แทน ค่าเบี่ยงเบนมาตรฐาน
X_1	แทน คะแนนจากแบบทดสอบฉบับ X ที่ทดสอบกับประชากรกลุ่มที่ 1
Y_2	แทน คะแนนจากแบบทดสอบฉบับ Y ที่ทดสอบกับประชากรกลุ่มที่ 2
Z_1	แทน คะแนนจากแบบทดสอบฉบับ Z ที่ทดสอบกับประชากรกลุ่มที่ 1
Z_2	แทน คะแนนจากแบบทดสอบฉบับ Z ที่ทดสอบกับประชากรกลุ่มที่ 2
r_{xz}	แทน ค่าสหสัมพันธ์ระหว่างแบบทดสอบฉบับ X และแบบทดสอบฉบับ Z
r_{yz}	แทน ค่าสหสัมพันธ์ระหว่างแบบทดสอบฉบับ Y และแบบทดสอบฉบับ Z
SE	แทน ค่าความคลาดเคลื่อนมาตรฐานของคะแนนจากการเทียบ
Bias	แทน ค่าความลำเอียงทางสถิติของคะแนนจากการเทียบ
SS	แทน ผลบวกของความแปรปรวน
MS	แทน ค่าเฉลี่ยของความแปรปรวน
df	แทน ชั้นแห่งความเป็นอิสระ

การเสนอผลการวิเคราะห์ข้อมูล

ในการศึกษาค้นคว้าครั้งนี้ ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลตามลำดับดังนี้

1. ค่าสถิติพื้นฐานของแบบทดสอบ
2. ค่าสถิติของกลุ่มตัวอย่างแต่ละขนาด
3. เปรียบเทียบคะแนนเทียบเฉลี่ยของทุกขนาดกลุ่มตัวอย่าง
4. ค่าความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน

5. ค่าความลำเอียงทางสถิติในการเทียบคะแนน
6. ภาพแสดงความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนของคะแนนสิ่ง เกิดทุกตัว
7. ภาพแสดงความลำเอียงทางสถิติในการเทียบคะแนนของคะแนนสิ่ง เกิดทุกตัว

1. ค่าสถิติพื้นฐานของแบบทดสอบ

ผู้วิจัยได้นำแบบทดสอบที่ใช้ในการเทียบคะแนน 2 ฉบับคือแบบทดสอบฉบับ X และฉบับ Y และนำแบบทดสอบเชื่อมโยง 1 ฉบับ คือฉบับ Z ไปทดสอบกับนักเรียนที่กำหนดไว้ 2 กลุ่ม ผลการทดสอบนำมาคำนวณค่าสถิติพื้นฐานได้ดังนี้

ตาราง 2 สถิติพื้นฐานของแบบทดสอบ 3 ฉบับ

แบบทดสอบ	กลุ่มประชากร	จำนวนคน	จำนวนข้อสอบ	$\bar{X}$	S.D.
ฉบับ X	1	913	60	38.8204	11.0845
ฉบับ Y	2	913	60	30.1139	11.2925
ฉบับ Z	1	913	30	16.3844	5.6318
ฉบับ Z	2	913	30	16.9792	6.9652

จากตาราง 2 แสดงว่าแบบทดสอบที่ใช้เทียบคะแนนฉบับ X แนวโน้มมีค่าเฉลี่ยสูงกว่าฉบับ Y แต่มีค่าเบี่ยงเบนมาตรฐานต่ำกว่า ฉบับ Y เล็กน้อย สำหรับแบบทดสอบเชื่อมโยงฉบับ Z ที่ใช้ทดสอบกับประชากรกลุ่มที่ 2 มีค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานสูงกว่า ที่ใช้ทดสอบกับประชากรกลุ่มที่ 1 เล็กน้อย

2. ค่าสถิติของกลุ่มตัวอย่างแต่ละขนาด

ค่าสถิติของกลุ่มตัวอย่างแต่ละขนาด ผู้วิจัยเสนอในรูปของพิสัยของค่าสถิติที่ได้จากการคำนวณในแต่ละครั้งของการสุ่ม เพราะในแต่ละขนาดของกลุ่มตัวอย่าง สุ่มมาจำนวน 200 ครั้ง

ดังนั้นค่าสถิติแต่ละตัวจะมีจำนวนตัวละ 200 ค่า รวมทั้งสิ้นจะได้ค่าสถิติ 2,000 ค่าในแต่ละขนาดกลุ่มตัวอย่าง ซึ่งเป็นจำนวนมาก ผู้วิจัยจึงนำเสนอในรูปของพิสัย ดังแสดงไว้ในตารางต่อไปนี้

ตาราง 3 ค่าสถิติของกลุ่มตัวอย่างแต่ละขนาด

ขนาด สถิติ	30คน	60 คน	120 คน	250 คน
$\bar{X}_{x1}$	32.8667- 45.0667	35.7500- 41.7833	36.0750- 41.2667	37.3960- 40.1560
S.D. _{x1}	3.3945- 14.7034	8.3716- 13.1292	9.1391- 12.6796	9.8979- 11.9576
$\bar{X}_{y2}$	24.5333- 35.5667	25.8500- 33.9667	27.7250- 33.7833	28.5240- 31.6760
S.D. _{y2}	7.6894- 14.0553	9.0953- 13.8468	9.9594- 12.9673	10.0593- 12.0892
$\bar{X}_{z1}$	14.3333- 18.9333	14.0833- 18.5167	15.1583- 17.5333	15.7440- 17.0400
S.D. _{z1}	4.0612- 7.0288	4.6032- 6.4951	4.8036- 6.3164	5.1368- 6.2489
$\bar{X}_{z2}$	12.8667- 20.7333	15.1833- 20.0500	15.6333- 19.1750	16.0360- 17.9040
S.D. _{z2}	5.1121- 8.4040	5.7581- 8.0200	6.1924- 7.6761	6.4631- 7.4862
r_{xz}	.1515- .8335	.3164- .7904	.3488- .7398	.4675- .6863
r_{yz}	.2569- .8841	.3494- .8018	.4380- .7774	.5174- .7213

จากตาราง 3 กลุ่มตัวอย่างขนาด 30 คน ค่าสถิติทุกค่า มีผลของค่าสถิติมากที่สุด รองลงมาคือ กลุ่มตัวอย่างขนาด 60 คน 120 คน และกลุ่มตัวอย่างขนาด 250 คน มีผลของค่าสถิติน้อยที่สุด

3. เปรียบเทียบคะแนนเทียบเฉลี่ยของทุกขนาดกลุ่มตัวอย่าง

จากจุดมุ่งหมายข้อที่ 1 ต้องการเปรียบเทียบคะแนนเทียบเฉลี่ยของทุกขนาดกลุ่มตัวอย่าง ผู้วิจัยได้นำคะแนนเทียบแต่ละค่าที่ได้จากการแปลงคะแนนดิบแต่ละตัวด้วยสูตรของแมคเคน ในการลุ่มแต่ละครั้ง จำนวน 200 ครั้งทำให้ได้คะแนนเทียบทั้งหมด 200 ค่าในแต่ละกลุ่มตัวอย่างแล้วนำมาหาค่าเฉลี่ย นำค่าเฉลี่ยของคะแนนเทียบ มาเปรียบเทียบกัน ในแต่ละขนาดกลุ่มตัวอย่าง มาวิเคราะห์ความแปรปรวน ดังแสดงในตาราง 4

ตาราง 4 เปรียบเทียบคะแนนเทียบเฉลี่ยของทุกขนาดกลุ่มตัวอย่าง

Source of Variation	SS	df	MS	F
Explained	135072	3	45023.905	1.294
Residual	27695665	796	34793.549	
Total	27830736	799	34831.960	

$$F_{.05}(3,796) = 8.53$$

จากตาราง 4 แสดงว่าคะแนนเทียบเฉลี่ยของกลุ่มตัวอย่างทั้ง 4 ขนาด แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

4. ค่าความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน

ค่าความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน เป็นค่าของความแตกต่างของคะแนนเทียบแต่ละตัวกับคะแนนเทียบเฉลี่ยของแต่ละขนาดของกลุ่มตัวอย่าง ซึ่งแสดงในตารางต่อไปนี้

ตาราง 5 ค่าความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน

ขนาด คะแนนสังเกต	30 คน	60 คน	120 คน	250 คน
0	18.60	8.59	6.43	3.61
1	18.14	8.37	6.28	3.52
2	17.68	8.16	6.12	3.43
3	17.22	7.95	5.96	3.34
4	16.76	7.73	5.80	3.25
5	16.30	7.52	5.64	3.16
6	15.84	7.31	5.48	3.07
7	15.39	7.09	5.33	2.98
8	14.93	6.88	5.17	2.89
9	14.47	6.67	5.01	2.80
10	14.02	6.45	4.85	2.72
11	13.56	6.24	4.70	2.63
12	13.10	6.03	4.54	2.54
13	12.65	5.82	4.39	2.45
14	12.19	5.61	4.23	2.36
15	11.74	5.40	4.07	2.27
16	11.29	5.19	3.92	2.19
17	10.84	4.98	3.76	2.10
18	10.39	4.78	3.61	2.01
19	9.94	4.57	3.46	1.93
20	9.49	4.37	3.31	1.84
21	9.05	4.16	3.15	1.75
22	8.60	3.96	3.00	1.67
23	8.16	3.76	2.85	1.59

ตาราง 5 (ต่อ)

ขนาด คะแนนสังเกต	30 คน	60 คน	120 คน	250 คน
24	7.72	3.56	2.71	1.50
25	7.29	3.37	2.56	1.42
26	6.86	3.17	2.41	1.34
27	6.43	2.98	2.27	1.26
28	6.01	2.80	2.13	1.18
29	5.59	2.62	2.00	1.11
30	5.19	2.44	1.86	1.04
31	4.79	2.27	1.73	.97
32	4.41	2.12	1.61	.90
33	4.04	1.97	1.50	.84
34	3.70	1.84	1.39	.79
35	3.39	1.72	1.30	.74
36	3.11	1.63	1.22	.70
37	2.88	1.56	1.16	.68
38	2.72	1.52	1.11	.66
39	2.62	1.51	1.09	.66
40	2.61	1.54	1.10	.67
41	2.68	1.59	1.12	.69
42	2.83	1.67	1.17	.72
43	3.03	1.77	1.24	.77
44	3.30	1.89	1.32	.82
45	3.60	2.03	1.42	.88
46	3.93	2.18	1.53	.94
47	4.29	2.34	1.64	1.01

ตาราง 5 (ต่อ)

ขนาด คะแนนสังเกต	30 คน	60 คน	120 คน	250 คน
48	4.67	2.51	1.77	1.08
49	5.06	2.69	1.90	1.15
50	5.47	2.87	2.03	1.23
51	5.88	3.06	2.17	1.31
52	6.30	3.25	2.31	1.39
53	6.72	3.45	2.45	1.47
54	7.15	3.64	2.60	1.55
55	7.59	3.84	2.74	1.63
56	8.03	4.04	2.89	1.72
57	8.47	4.25	3.04	1.80
58	8.91	4.45	3.19	1.89
59	9.35	4.66	3.35	1.97
60	9.80	4.86	3.50	2.06

จากตาราง 5 ค่าความคลาดเคลื่อนในการเทียบคะแนนของกลุ่มตัวอย่างขนาด 30 คน สูงสุดคือ 18.60 ต่ำสุดคือ 2.61 กลุ่มตัวอย่างขนาด 60 คน สูงสุดคือ 8.59 ต่ำสุดคือ 1.51 กลุ่มตัวอย่างขนาด 120 คน สูงสุดคือ 6.43 ต่ำสุดคือ 1.09 สำหรับกลุ่มตัวอย่างขนาด 250 คน ซึ่งมีค่าความคลาดเคลื่อนน้อยกว่ากลุ่มตัวอย่างขนาดต่าง ๆ คือสูงสุด 3.61 และต่ำสุด เพียง .66 เท่านั้น

5. ค่าความลำเอียงทางสถิติในการเทียบคะแนน

ค่าความลำเอียงทางสถิติในการเทียบคะแนน เป็นความแตกต่างระหว่างคะแนนเทียบเฉลี่ยจาก 200 ครั้ง กับคะแนนเทียบของกลุ่มประชากร ในตารางต่อไปนี้ได้นำเสนอค่าความลำเอียงทางสถิติของทุกขนาดของกลุ่มตัวอย่าง ของคะแนนสังเกตทุกตัว

ตาราง 6 ค่าความลำเอียงทางสถิติในการเทียบคะแนน

ขนาด คะแนนสังเกต	30 คน	60 คน	120 คน	250 คน
0	-2.33	-1.02	-1.18	.03
1	-2.27	-.99	-1.14	.03
2	-2.21	-.96	-1.11	.04
3	-2.16	-.93	-1.09	.03
4	-2.10	-.90	-1.05	.03
5	-2.05	-.88	-1.03	.02
6	-1.98	-.85	-.99	.03
7	-1.93	-.83	-.97	.02
8	-1.87	-.80	-.93	.02
9	-1.81	-.77	-.90	.03
10	-1.76	-.74	-.87	.02
11	-1.69	-.71	-.84	.02
12	-1.64	-.69	-.81	.01
13	-1.58	-.66	-.78	.02
14	-1.52	-.63	-.74	.02
15	-1.47	-.61	-.72	.01
16	-1.40	-.58	-.68	.01
17	-1.35	-.55	-.66	.01

ตาราง 6 (ต่อ)

ขนาด คะแนนสังเกต	30 คน	60 คน	120 คน	250 คน
18	-1.29	-.52	-.63	.01
19	-1.23	-.49	-.59	.01
20	-1.18	-.47	-.57	.00
21	-1.12	-.44	-.53	.01
22	-1.06	-.42	-.51	.00
23	-1.00	-.38	-.47	.00
24	-.94	-.35	-.44	.00
25	-.89	-.33	-.41	.00
26	-.83	-.30	-.38	.00
27	-.77	-.28	-.35	-.01
28	-.71	-.25	-.32	.00
29	-.65	-.22	-.28	.00
30	-.60	-.19	-.26	-.01
31	-.54	-.16	-.22	-.01
32	-.48	-.14	-.20	-.01
33	-.42	-.11	-.17	-.01
34	-.36	-.08	-.13	-.01
35	-.31	-.06	-.11	-.02
36	-.25	-.03	-.07	-.01
37	-.20	.00	-.05	-.02
38	-.13	.03	-.01	-.02
39	-.08	.05	.01	-.03
40	-.02	.08	.05	-.02
41	.04	.11	.08	-.02

ตาราง 6 (ต่อ)

ขนาด คะแนนสังเกต	30 คน	60 คน	120 คน	250 คน
42	.09	.13	.11	-.03
43	.16	.17	.14	-.03
44	.21	.19	.17	-.03
45	.27	.22	.20	-.03
46	.33	.25	.24	-.03
47	.38	.27	.26	-.03
48	.45	.30	.29	-.03
49	.50	.32	.32	-.04
50	.56	.36	.35	-.04
51	.62	.39	.39	-.03
52	.67	.41	.41	-.04
53	.73	.44	.45	-.04
54	.79	.46	.47	-.05
55	.85	.49	.51	-.04
56	.91	.52	.54	-.04
57	.96	.55	.57	-.05
58	1.02	.58	.60	-.05
59	1.08	.60	.63	-.05
60	1.14	.63	.66	-.05

จากตาราง 6 ค่าความลำเอียงทางสถิติในการเทียบคะแนน ของกลุ่มตัวอย่างขนาด 30 คน สูงสุดคือ -2.33 และต่ำสุดคือ -.02 กลุ่มตัวอย่างขนาด 60 คน สูงสุดคือ -1.02 และต่ำสุดคือ 0 กลุ่มตัวอย่างขนาด 120 คน สูงสุดคือ -1.18 ต่ำสุดคือ .01 สำหรับกลุ่ม

สรุป อภิปรายผล และข้อเสนอแนะ

ความมุ่งหมายของการศึกษาค้นคว้า

1. เพื่อศึกษาว่าคะแนนเทียบเฉลี่ยของแต่ละกลุ่มตัวอย่างแตกต่างกันหรือไม่
2. เพื่อศึกษาความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนในกลุ่มตัวอย่างขนาด 30 คน 60 คน 120 คน และ 250 คน ว่าแตกต่างกันหรือไม่
3. เพื่อศึกษาความลำเอียงทางสถิติในการเทียบคะแนนของกลุ่มตัวอย่างขนาด 30 คน 60 คน 120 คน และ 250 คน ว่าแตกต่างกันหรือไม่

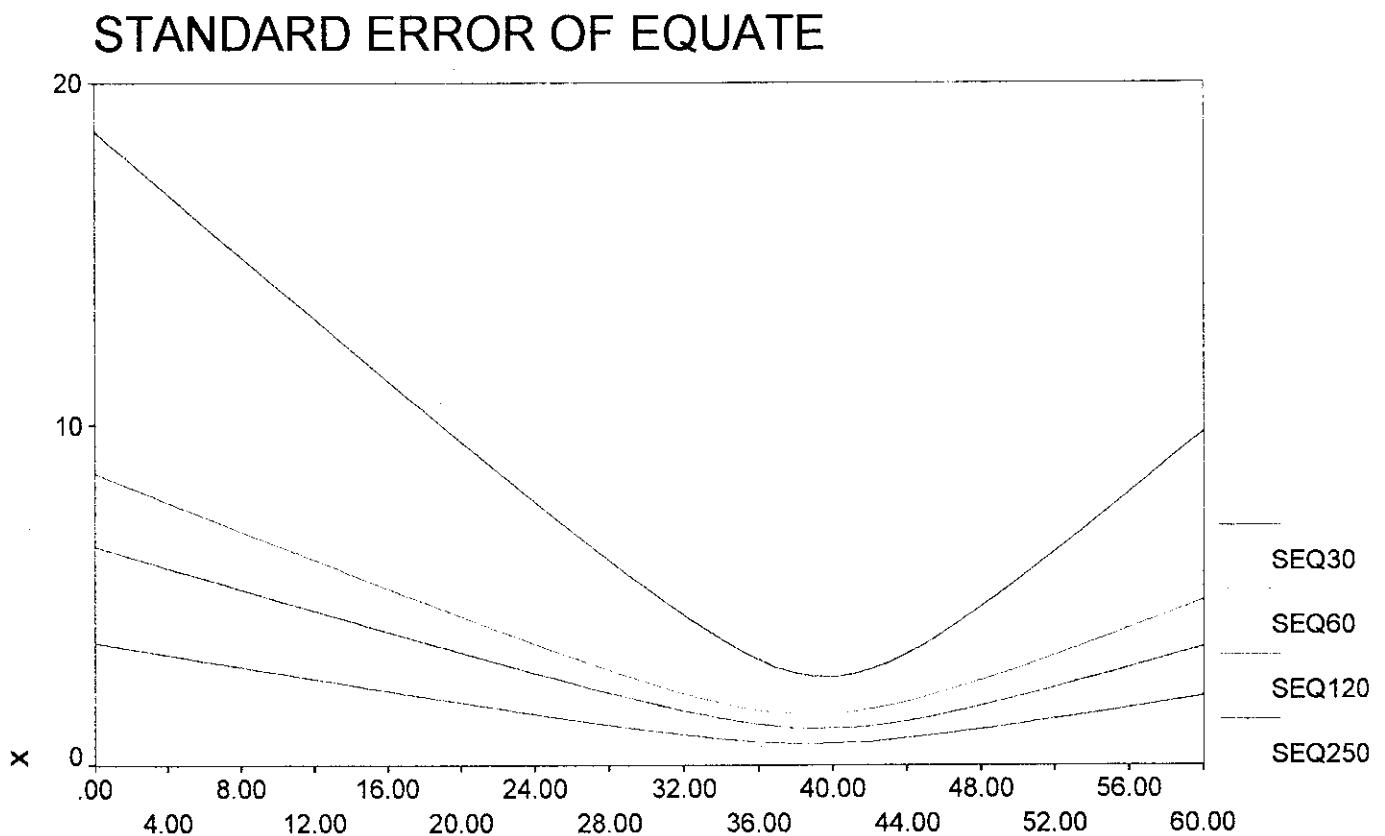
กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ คือนักเรียนชั้นประถมศึกษาปีที่ 6 เขตบางเขน เขตดอนเมือง และเขตจตุจักร สังกัดกรุงเทพมหานคร ที่ผู้วิจัยได้ทำการทดสอบไว้แล้วนำคะแนนมาเก็บไว้ในแผ่นงานแม่เหล็ก แล้วทำการสุ่มอย่างง่าย (Simple Random Sampling) จากประชากรเทียบทั้งหมด โดยมีนักเรียนเป็นหน่วยการสุ่ม ซึ่งมีวิธีการดังนี้

1. จากประชากรเทียบกลุ่มที่ 1 ทำการสุ่มกลุ่มตัวอย่างแบบใส่คืน (With-replacement) โดยสุ่มกลุ่มตัวอย่างขนาด 30 คน 60 คน 120 คน และ 250 คน ขนาดละ 200 ครั้ง รวมเป็นกลุ่มตัวอย่างทั้งหมด 800 กลุ่มตัวอย่าง
2. ทำการสุ่มจากประชากรเทียบกลุ่มที่ 2 เช่นเดียวกับข้อ 1
3. กลุ่มตัวอย่างที่ใช้ในการวิจัยนี้ จึงมีทั้งสิ้น 1,600 กลุ่ม จาก 2 กลุ่มประชากรเทียบ

ตัวอย่างขนาด 250 คน มีค่าความลำเอียงทางสถิติที่น้อยที่สุด คือค่าสูงสุดเพียง .03 และต่ำสุดคือ 0

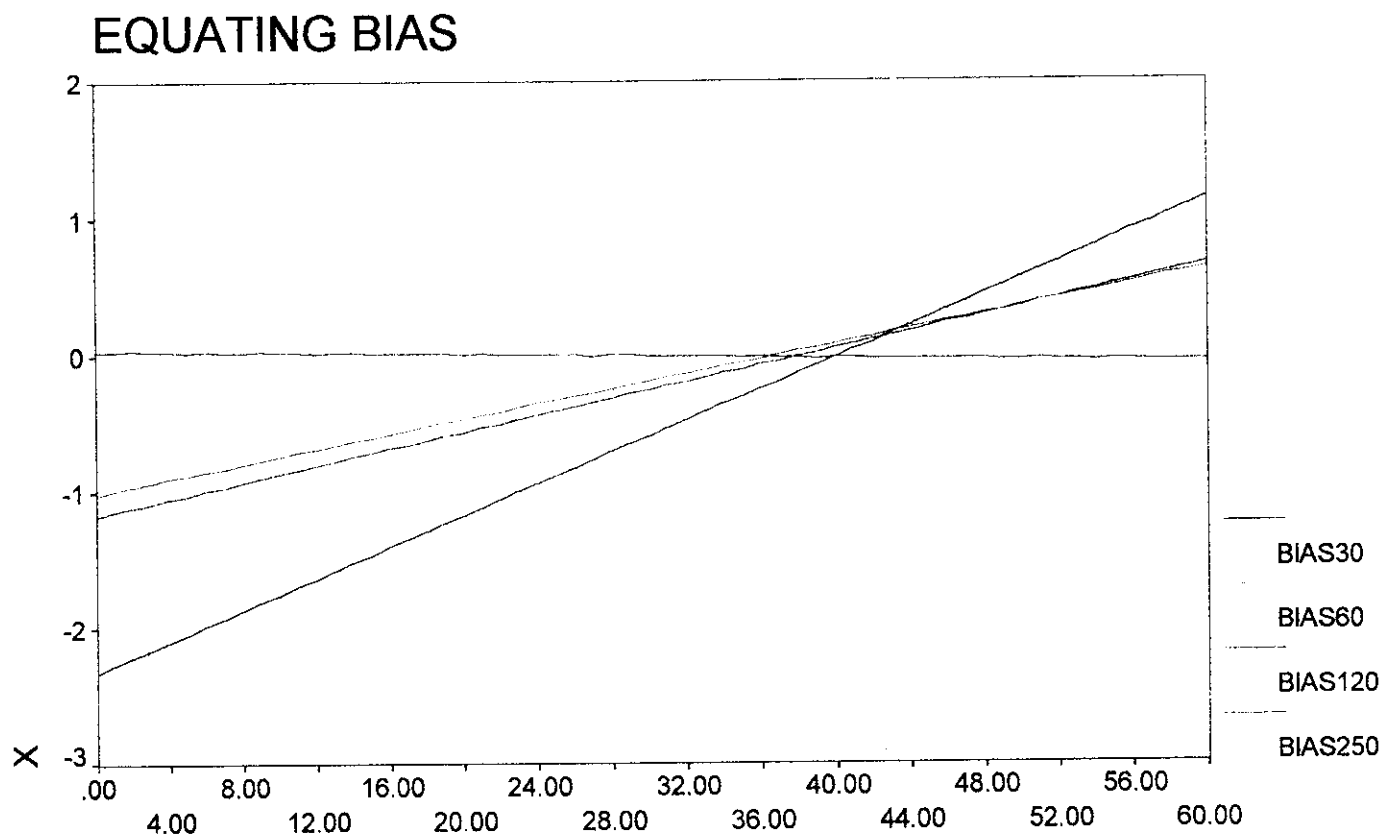
6. ภาพแสดงความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนของคะแนนสังเกตทุกตัว



ภาพประกอบ 8 แสดงความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนของคะแนนสังเกตทุกตัว

จากภาพประกอบ 8 พบว่าค่าความคลาดเคลื่อนมาตรฐานในการเทียบคะแนน ของกลุ่มตัวอย่างขนาดต่างๆ น้อยมากเมื่อคะแนนสังเกตเข้าใกล้ค่าเฉลี่ย (38.8204) โดยกลุ่มตัวอย่างขนาด 250 คน มีค่าความคลาดเคลื่อนต่ำสุด และกลุ่มตัวอย่างขนาด 30 คน มีค่าความคลาดเคลื่อนสูงสุด

7. ภาพแสดงความลำเอียงทางสถิติในการเทียบคะแนนของคะแนนสังเกตทุกตัว



ภาพประกอบ 9 แสดงความลำเอียงทางสถิติในการเทียบคะแนนของคะแนนสังเกตทุกตัว

จากภาพประกอบ 9 พบว่าค่าความลำเอียงทางสถิติในการเทียบคะแนน ของคะแนนสังเกตทุกตัว ของกลุ่มตัวอย่างขนาด 250 คน มีค่าต่ำที่สุด และค่าความลำเอียงทางสถิติในการเทียบคะแนนของคะแนนสังเกตทุกตัว ของกลุ่มตัวอย่างขนาด 30 คน มีค่าสูงที่สุด และพบว่าค่าความลำเอียงทางสถิติของทุกขนาดกลุ่มตัวอย่างจะมีค่าต่ำสุด เมื่อคะแนนสังเกตเข้าใกล้ค่าเฉลี่ย (38.8204)

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นแบบทดสอบวัดความสามารถทางสมองตามทฤษฎีของเทอร์สโตน (Thurstone) ด้านการรับรู้ (Perception Factor) จำนวน 3 ฉบับ เป็นแบบทดสอบที่ทำให้คู่ขนานแบบคะแนนจริงสัมพันธ์ 2 ฉบับ คือฉบับ X และฉบับ Y ใช้เป็นแบบทดสอบเทียบคะแนน สำหรับฉบับที่ 3 หรือฉบับ Z เป็นแบบทดสอบเชื่อมโยง (External Anchor Test) ที่มีจำนวนข้อ หรือความยาว 50 % ของแบบทดสอบเทียบคะแนน และมีลักษณะคำถามเช่นเดียวกับแบบทดสอบเทียบคะแนน ดังมีรายละเอียดดังนี้

ฉบับ X มีสองตอน ตอนที่หนึ่งประกอบด้วยลักษณะคำถามประเภทถามเหมือน คือให้หารูปที่เหมือนกับ รูปที่กำหนดให้ จำนวน 30 ข้อ ตอนที่ 2 เป็นคำถามประเภทถามไม่เหมือน คือให้หารูปที่ไม่เหมือนกับรูปที่กำหนดให้จำนวน 30 ข้อ รวมทั้งฉบับมีจำนวน 60 ข้อ

ฉบับ Y เป็นแบบทดสอบที่วัดสิ่งเดียวกันกับฉบับ X คือมีลักษณะคำถามเช่นเดียวกับฉบับ X ซึ่งมีสองตอน ตอนที่หนึ่งเป็นคำถามประเภทถามเหมือน คือให้หารูปที่เหมือนกับรูปที่กำหนดให้จำนวน 30 ข้อ ตอนที่สองเป็นคำถามประเภทถามไม่เหมือนคือให้หารูปที่ไม่เหมือนกับรูปที่กำหนดให้ จำนวน 30 ข้อ รวมทั้งฉบับมีจำนวน 60 ข้อ

ฉบับ Z เป็นแบบทดสอบเชื่อมโยงที่วัดสิ่งเดียวกันกับแบบทดสอบฉบับ X และฉบับ Y โดยมีลักษณะคำถามเช่นเดียวกับฉบับ X และฉบับ Y แต่มีจำนวนข้อเพียงครึ่งหนึ่งของฉบับ X และฉบับ Y ทั้งสองตอน ดังนั้นตอนที่หนึ่ง จึงเป็นคำถามประเภทถามเหมือน คือให้หารูปที่เหมือนกับรูปที่กำหนดให้ จำนวน 15 ข้อ ตอนที่สองเป็นคำถามประเภทถามไม่เหมือน คือให้หารูปที่ไม่เหมือนกับรูปที่กำหนดให้ จำนวน 15 ข้อ รวมทั้งฉบับมีจำนวน 30 ข้อ

วิธีดำเนินการเก็บรวบรวมข้อมูล

วิธีดำเนินการเก็บรวบรวมข้อมูลได้ดำเนินการตามขั้นตอนดังต่อไปนี้

1. ติดต่อขอความร่วมมือกับโรงเรียนที่ใช้เป็นกลุ่มตัวอย่าง โดยกำหนดวันทำการเก็บข้อมูลตั้งแต่วันที่ 4 - 14 มีนาคม 2539
2. จัดเตรียมแบบทดสอบให้เพียงพอกับจำนวนประชากรที่ศึกษา
3. นำแบบทดสอบฉบับ X และฉบับ Z ไปทดสอบกับประชากรเทียมกลุ่มที่ 1 และนำแบบทดสอบฉบับ Y และฉบับ Z ไปทดสอบกับประชากรเทียมกลุ่มที่ 2

4. คัดเลือกกระดาษคำตอบที่สมบูรณ์ไว้กลุ่มประชากรเทียมละ 913 ชุด
5. นำผลการสอบมาตรวจให้คะแนน โดยตอบถูกให้ 1 คะแนน ตอบผิดให้ 0 คะแนน
6. นำข้อมูลที่ได้จากการตรวจมาวิเคราะห์

ขั้นตอนการวิเคราะห์ข้อมูล

แบ่งเป็น 2 ตอน

1. การวิเคราะห์ข้อมูลในการทดลองเครื่องมือ (Try out) ได้ดำเนินการดังนี้

1.1 หาค่าความยาก(P) โดยใช้สูตรอย่างง่าย และค่าอำนาจจำแนก(r) โดยใช้ค่าสหสัมพันธ์แบบพอยท์-ไบซีเรียล

1.2 หาค่าความเชื่อมั่น โดยใช้สูตรสัมประสิทธิ์แอลฟา (KR-20) ของคูเดอร์ และ ริชาร์ดสัน (Kuder and Richardson. 1937) (บุญเชิด ภิญโญนนตพงษ์ .2523:256)

2. การวิเคราะห์ข้อมูลการวิจัยได้ดำเนินการดังนี้

2.1 หาค่าสถิติพื้นฐานของทุกกลุ่ม คือ ค่าเฉลี่ย($\bar{X}$) และค่าเบี่ยงเบนมาตรฐาน(SD) ค่าสหสัมพันธ์ (r) ระหว่างแบบทดสอบที่ใช้เทียบคะแนนกับแบบทดสอบเชื่อมโยง

2.2 หาคะแนนแปลงจากการเทียบคะแนน โดยใช้สูตรของ แมคแคน (MacCann. 1989)

2.3 เปรียบเทียบคะแนนเทียบเฉลี่ยของแต่ละขนาดของกลุ่มตัวอย่าง

2.4 หาค่าความคลาดเคลื่อนมาตรฐานของคะแนนจากการเทียบทุกตัว โดยใช้สูตรของ พาร์แชล, ฮัจตัน และ ครอมเรย์ (Parshall, Houghton and Kromrey. 1995)

2.5 หาค่าความลำเอียงทางสถิติทุกค่าของคะแนนสังเกต โดยใช้สูตรของพาร์แชล, ฮัจตัน และ ครอมเรย์ (Parshall, Houghton and Kromrey. 1995)

2.6 สร้างตารางเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐาน และความลำเอียงทางสถิติ ของทุกขนาดกลุ่มตัวอย่าง

2.7 พิจารณากลุ่มตัวอย่างที่เหมาะสมจากคะแนนเทียบเฉลี่ย ตารางเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานและความลำเอียงทางสถิติ

สถิติที่ใช้ในการวิเคราะห์ข้อมูล

1. หาค่าสถิติพื้นฐานของทุกกลุ่ม คือ ค่าเฉลี่ย($\bar{X}$) และค่าเบี่ยงเบนมาตรฐาน(SD) ค่าสหสัมพันธ์ (r) ระหว่างแบบทดสอบที่ใช้เทียบคะแนนกับแบบทดสอบเชื่อมโยง
2. หาค่าความยาก(P) โดยใช้สูตรอย่างง่าย และค่าอำนาจจำแนก(r) โดยใช้ค่าสหสัมพันธ์แบบพอยท์-ไบซีเรียล
3. หาค่าความเชื่อมั่น โดยใช้สูตรสัมประสิทธิ์แอลฟา(KR-20) ของคูเดอร์ และ ริชาร์ดสัน (Kuder and Richardson. 1937) (บุญเชิด ภิญโญนันตพงษ์ .2523:256)
4. หาคะแนนแปลงจากการเทียบคะแนนโดยใช้สูตรของ แมคแคน (MacCann.1989)
5. หาค่าความคลาดเคลื่อนมาตรฐานของคะแนนจากการเทียบโดยใช้สูตรของพาร์แชล, ฮ็อดจตัน และ ครอมเวรี่ (Parshall, Houghton and Kromrey.1995)
6. หาค่าความลำเอียงทางสถิติ โดยใช้สูตรของ พาร์แชล, ฮ็อดจตัน และ ครอมเวรี่ (Parshall, Houghton and Kromrey.1995)

การเสนอผลการวิเคราะห์ข้อมูล

ในการศึกษาค้นคว้าครั้งนี้ ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลตามลำดับขั้นดังนี้

1. ค่าสถิติพื้นฐานของแบบทดสอบ
2. ค่าสถิติของกลุ่มตัวอย่างแต่ละขนาด
3. เปรียบเทียบคะแนนเทียบเฉลี่ยของทุกขนาดกลุ่มตัวอย่าง
4. ค่าความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน
5. ค่าความลำเอียงทางสถิติในการเทียบคะแนน
6. ภาพแสดงความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนของคะแนนสังเกตทุกตัว
7. ภาพแสดงความลำเอียงทางสถิติในการเทียบคะแนนของคะแนนสังเกตทุกตัว

สรุปผลการวิเคราะห์ข้อมูล

1. ค่าสถิติพื้นฐานของแบบทดสอบ

คะแนนจากแบบทดสอบฉบับ X มีค่าเฉลี่ย 38.8204 ซึ่งสูงกว่าค่าเฉลี่ยของคะแนนจากแบบทดสอบฉบับ Y คือ 30.1139 แต่มีค่าความเบี่ยงเบนมาตรฐาน 11.0845 น้อยกว่าจากแบบทดสอบฉบับ Y คือ 11.2925 สำหรับคะแนนจากแบบทดสอบเชื่อมโยง ฉบับ Z มีค่าเฉลี่ยใกล้เคียงกันทั้งที่ทดสอบกับประชากรกลุ่มที่ 1 และที่ทดสอบกับประชากรกลุ่มที่ 2 คือค่าเฉลี่ยของแบบทดสอบฉบับ Z ที่สอบกับประชากรกลุ่มที่ 1 (Z_1) เท่ากับ 16.3844 และ ค่าเฉลี่ยของแบบทดสอบฉบับ Z ที่สอบกับประชากรกลุ่มที่ 2 (Z_2) เท่ากับ 16.9792 แต่สำหรับค่าเบี่ยงเบนมาตรฐานของคะแนนจากแบบทดสอบฉบับ Z ที่สอบกับประชากรกลุ่มที่ 1 (SD_1) มีค่า 5.6318 ซึ่งน้อยกว่า ค่าเบี่ยงเบนมาตรฐานของคะแนนจากแบบทดสอบฉบับ Z ที่สอบกับประชากรกลุ่มที่ 2 (SD_2) ซึ่งเท่ากับ 6.9652

2. ค่าสถิติของกลุ่มตัวอย่างแต่ละขนาด

ค่าสถิติที่ได้จากการสุ่ม 200 ครั้ง (Pseudopopulation) ประกอบด้วย ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐานของคะแนนจากแบบทดสอบฉบับ X ฉบับ Y และแบบทดสอบฉบับ Z ค่าสหสัมพันธ์ระหว่างแบบทดสอบฉบับ X และฉบับ Z และค่าสหสัมพันธ์ระหว่างแบบทดสอบฉบับ Y และฉบับ Z ผลปรากฏว่าในทุกค่าสถิติของกลุ่มตัวอย่างขนาด 30 คน มีพิสัยกว้างที่สุด รองลงมาคือกลุ่มตัวอย่างขนาด 60 คน และ 120 คนตามลำดับ สำหรับกลุ่มตัวอย่างขนาด 250 คน มีพิสัยแคบที่สุด นอกจากนั้นยังพบว่าค่าสหสัมพันธ์ระหว่างแบบทดสอบฉบับ X และฉบับ Z ของกลุ่มตัวอย่างขนาด 30, 60, 120 และ 250 คน มีค่าสูงสุดของแต่ละขนาดตามลำดับคือ .8335 .7904 .7398 .6863 และค่าสหสัมพันธ์ระหว่างแบบทดสอบฉบับ Y และฉบับ Z ของกลุ่มตัวอย่างขนาด 30, 60, 120 และ 250 คน มีค่าสูงสุดของแต่ละขนาดตามลำดับคือ .8841 .8018 .7774 .7213

3. เปรียบเทียบคะแนนเทียบเฉลี่ยของทุกขนาดกลุ่มตัวอย่าง

ผลปรากฏว่าคะแนนเทียบเฉลี่ยของกลุ่มตัวอย่างทั้ง 4 ขนาด แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ นั่นคือคะแนนเทียบเฉลี่ยของกลุ่มตัวอย่างขนาด 30 คน 60 คน 120 คน และ 250 คน

ไม่แตกต่างกัน

4. ค่าความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน

ค่าความคลาดเคลื่อนมาตรฐานของกลุ่มตัวอย่างขนาด 30 คน ค่าสูงสุดคือ 18.60 ค่าต่ำสุดคือ 2.61 กลุ่มตัวอย่างขนาด 60 คน ค่าสูงสุดคือ 8.59 ค่าต่ำสุดคือ 1.51 กลุ่มตัวอย่างขนาด 120 คน ค่าสูงสุดคือ 6.43 ค่าต่ำสุดคือ 1.09 กลุ่มตัวอย่างขนาด 250 คน ค่าสูงสุดคือ 3.61 ค่าต่ำสุดคือ 0.66

5. ค่าความลำเอียงทางสถิติในการเทียบคะแนน

ค่าความลำเอียงทางสถิติในการเทียบคะแนน ของกลุ่มตัวอย่างขนาด 30 คน สูงสุดคือ -2.33 ต่ำสุดคือ -.02 กลุ่มตัวอย่างขนาด 60 คน สูงสุดคือ -1.02 ต่ำสุดคือ 0 กลุ่มตัวอย่างขนาด 120 คน สูงสุดคือ -1.18 ต่ำสุดคือ .01 กลุ่มตัวอย่างขนาด 250 คน สูงสุดคือ .03 ต่ำสุดคือ 0

6. ภาพแสดงความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนของคะแนนสังเกตทุกตัว

จากภาพประกอบที่ 8 พบว่าค่าความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน ของคะแนนสังเกตทุกตัว มีค่าน้อยมากเมื่อคะแนนเข้าใกล้ค่าเฉลี่ย (38.8204) และมีค่าเพิ่มขึ้นเมื่อคะแนนกระจายออกจากค่าเฉลี่ย

7. ภาพแสดงความลำเอียงทางสถิติในการเทียบคะแนนของคะแนนสังเกตทุกตัว

จากภาพประกอบที่ 9 พบว่าความลำเอียงทางสถิติในการเทียบคะแนน ของคะแนนสังเกตทุกตัว ของกลุ่มตัวอย่างขนาด 250 คน มีค่าต่ำที่สุด ความลำเอียงทางสถิติในการเทียบคะแนน ของคะแนนสังเกตทุกตัว ของกลุ่มตัวอย่างขนาด 250 คน มีค่าสูงที่สุด และพบว่า ความลำเอียงทางสถิติในการเทียบคะแนนของทุกขนาดกลุ่มตัวอย่างมีค่าต่ำสุด เมื่อคะแนนเข้าใกล้ค่าเฉลี่ย

อภิปรายผล

1. ค่าสถิติพื้นฐานของแบบทดสอบ

จากค่าเบี่ยงเบนมาตรฐานและค่าเฉลี่ยของแบบทดสอบฉบับ X และฉบับ Y พบว่าค่าเฉลี่ยของแบบทดสอบฉบับ X มีแนวโน้มสูงกว่าฉบับ Y แต่มีค่าเบี่ยงเบนมาตรฐานต่ำกว่าฉบับ Y ที่เป็นเช่นนั้น แสดงว่าแบบทดสอบฉบับ X ค่อนข้างง่ายกว่าฉบับ Y และประชากรเทียมกลุ่มที่ 1 ที่ทำแบบทดสอบฉบับ X ค่อนข้างมีความสามารถใกล้เคียงกัน มากกว่าประชากรเทียมกลุ่มที่ 2 ที่ทำแบบทดสอบฉบับ Y

2. ค่าสถิติของกลุ่มตัวอย่างแต่ละขนาด

จากทุกค่าสถิติ สถิติของกลุ่มตัวอย่างขนาด 30 คน มีพิสัยกว้างที่สุด รองลงมาคือขนาด 60 คน 120 คน ตามลำดับ สำหรับกลุ่มตัวอย่างขนาด 250 คน มีพิสัยของทุกค่าสถิติแคบที่สุด สาเหตุที่เป็นเช่นนั้นเพราะ เมื่อลุ่มมาจากจำนวนประชากรที่เท่ากัน แต่ลุ่มมาจำนวนน้อย โอกาสที่ลุ่มได้นักเรียนแตกต่างกัน จึงมีมากกว่า ทำให้ค่าสถิติที่ได้มาจึงมีการกระจายมากทำให้พิสัยกว้าง แต่ถ้าลุ่มมาจำนวนมาก โอกาสที่ลุ่มได้นักเรียนแตกต่างกันมีน้อย ค่าสถิติที่ได้มาจึงมีการกระจายน้อย ทำให้พิสัยแคบ นอกจากนั้นยังพบว่า ค่าสหสัมพันธ์ระหว่างแบบทดสอบฉบับ X กับฉบับ Z และแบบทดสอบฉบับ Y กับฉบับ Z มีค่าสหสัมพันธ์กันสูง แสดงให้เห็นว่าคะแนนที่ได้จากการเทียบครั้งนี้ มีประสิทธิภาพ เพราะจากการศึกษาของ บูดেসคู (Budesu. 1985) พบว่าแบบทดสอบหลักที่มีความยาวเป็นครึ่งหนึ่งของแบบทดสอบที่ใช้เทียบคะแนน และมีค่าสหสัมพันธ์สูง แสดงถึงประสิทธิภาพของการเทียบคะแนนว่ามีสูงด้วย ซึ่งสอดคล้องกับ แองกอฟฟ์ (Angoff. 1971) ที่กล่าวว่า เมื่อความสัมพันธ์ของแบบทดสอบหลักไม่มี แบบทดสอบหลักจะไม่มีประโยชน์ สำหรับการเทียบคะแนนเลย ทำนองเดียวกันถ้าความสัมพันธ์เพิ่มขึ้น ประสิทธิภาพของความถูกต้องของการประมาณค่าพารามิเตอร์สำหรับกลุ่มที่ใช้ในการเทียบคะแนนก็เพิ่มขึ้นด้วย

3. เปรียบเทียบคะแนนเทียบเฉลี่ยของทุกขนาดกลุ่มตัวอย่าง

จากการเปรียบเทียบคะแนนเทียบเฉลี่ย พบว่ากลุ่มตัวอย่างทั้ง 4 ขนาดแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ หมายความว่า คะแนนเทียบเฉลี่ยของแต่ละกลุ่มมีค่าใกล้เคียงกัน ดังนั้นในการเทียบคะแนนอาจจะใช้กลุ่มตัวอย่างขนาด 30 คน 60 คน 120 คน หรือ 250 คน ก็ได้ เพราะ

ผลการเทียบคะแนนจะมีค่าใกล้เคียงกัน แต่กลุ่มตัวอย่างที่มีคะแนนเทียบใกล้เคียงกับกลุ่มประชากร เทียม คือกลุ่มตัวอย่างขนาด 250 คน

4. ค่าความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน

จากการศึกษาค่าความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนพบว่า กลุ่มตัวอย่างขนาด 30 คน มีค่าความคลาดเคลื่อนสูงที่สุด และกลุ่มตัวอย่างขนาด 250 คน มีค่าความคลาดเคลื่อนต่ำที่สุด ซึ่งสอดคล้องกับงานวิจัยของโคเลน (Kolen.1985) ที่พบว่ากลุ่มตัวอย่างขนาด 250 คน มีค่าความคลาดเคลื่อนต่ำกว่า กลุ่มตัวอย่างขนาด 100 คน และสอดคล้องกับงานวิจัยของ พาร์แชล ฮอจตัน และครอมเรย์ (Parshall, Houghton and Kromrey.1995) ที่พบว่า กลุ่มตัวอย่างขนาดใหญ่ 500 คน ค่าความคลาดเคลื่อนก็จะมีค่าน้อยกว่ากลุ่มตัวอย่างขนาด 15 คน 25 คน 50 คนและ100 คน

5. ค่าความลำเอียงทางสถิติในการเทียบคะแนน

จากการศึกษา ค่าความลำเอียงทางสถิติในการเทียบคะแนน ของทุกขนาดกลุ่มตัวอย่าง มีค่าต่ำมาก โดยในกลุ่มตัวอย่างขนาด 250 คน และ 60 คน มีค่าต่ำสุดที่ 0 ซึ่งหมายถึงไม่มีค่าความลำเอียงทางสถิติในการเทียบคะแนนเลย ซึ่งสอดคล้องกับงานวิจัยของ พาร์แชล ฮอจตันและ ครอมเรย์ (Parshall, Houghton and Kromrey.1995) ที่กล่าวว่า ขนาดของกลุ่มตัวอย่าง เกี่ยวพันกับค่าความลำเอียงทางสถิติน้อย

6. ภาพแสดงความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนของคะแนนสังเกตทุกตัว

แสดงให้เห็นว่า ค่าความคลาดเคลื่อนมาตรฐาน ของการเทียบคะแนนของคะแนนสังเกตทุกตัว มีค่าน้อยมากเมื่อคะแนนเข้าใกล้ค่าเฉลี่ย (38.8204) และมีค่าเพิ่มขึ้นเมื่อคะแนนกระจายออกจากค่าเฉลี่ย ซึ่งสอดคล้องกับงานวิจัยของ พาร์แชล ฮอจตัน และครอมเรย์ (Parshall, Houghton and Kromrey.1995) ที่พบว่าค่าความคลาดเคลื่อนมาตรฐานมีค่าน้อยมาก เมื่อคะแนนเข้าใกล้ค่าเฉลี่ย และจะเพิ่มมากขึ้นเรื่อย ๆ เช่นเดียวกับการกระจายของคะแนน จากค่าเฉลี่ย และพบว่าค่าความคลาดเคลื่อนมาตรฐานจะเพิ่มขึ้น เมื่อขนาดของกลุ่มตัวอย่างมีขนาดลดลง

7. ภาพแสดงความลำเอียงทางสถิติในการเทียบคะแนนของคะแนนสิ่งเกิดทุกตัว

แสดงให้เห็นว่า ความลำเอียงทางสถิติในการเทียบคะแนนของคะแนนสิ่งเกิดทุกตัว ของกลุ่มตัวอย่างขนาด 250 คน มีค่าต่ำที่สุด ความลำเอียงทางสถิติในการเทียบคะแนนของคะแนนสิ่งเกิดทุกตัว ของกลุ่มตัวอย่างขนาด 30 คน มีค่าสูงที่สุด และพบว่า ความลำเอียงทางสถิติในการเทียบคะแนนของทุกขนาดกลุ่มตัวอย่างมีค่าต่ำสุด เมื่อคะแนนเข้าใกล้ค่าเฉลี่ย ซึ่งสอดคล้องกับงานวิจัยของ พาร์แชล ฮัจตัน และครอมเรย์ (Parshall, Houghton and Kromrey, 1995) ที่พบว่าค่าความลำเอียงทางสถิติ มีผลกับขนาดของกลุ่มตัวอย่างน้อย เมื่อคะแนนเบี่ยงเบนออกจากค่าเฉลี่ย 1.5 SD

สรุปผลการวิจัย

จากการพิจารณาคะแนนเทียบเฉลี่ย ค่าความคลาดเคลื่อนมาตรฐาน ค่าความลำเอียงทางสถิติในการเทียบคะแนน พบว่ากลุ่มตัวอย่างที่สามารถนำมาใช้ในการเทียบคะแนนได้นั้น สามารถใช้ได้ทุกขนาดที่ใช้ในการศึกษา คือขนาด 30 คน 60 คน 120 คน และ 250 คน แต่ถ้าต้องการใช้กลุ่มตัวอย่างที่ให้ผลการเทียบคะแนนใกล้เคียงกับกลุ่มตัวอย่างขนาดใหญ่ และมีค่าความคลาดเคลื่อนมาตรฐาน และความลำเอียงทางสถิติน้อยที่สุดแล้ว กลุ่มตัวอย่างที่เหมาะสม คือกลุ่มตัวอย่างขนาด 250 คน

ข้อเสนอแนะ

1. ควรมีการศึกษาความคลาดเคลื่อน และความลำเอียงทางสถิติในการเทียบคะแนน เมื่อใช้กลุ่มตัวอย่างขนาดเล็ก โดยใช้สูตรการหาความคลาดเคลื่อน และความลำเอียงทางสถิติสูตรอื่นที่ไม่ใช่สูตรของพาร์แชล ฮัจตัน และครอมเรย์ เพื่อเป็นการเปรียบเทียบ และยืนยันว่าสอดคล้องกับงานวิจัยนี้หรือไม่
2. ควรมีการศึกษากับกลุ่มตัวอย่างที่มีความสามารถใกล้เคียงกัน เนื่องจากงานวิจัยของพาร์แชล ฮัจตัน และครอมเรย์ พบว่ากลุ่มตัวอย่างที่นำมาจากประชากรเดียวกัน จะทำให้ค่าความคลาดเคลื่อนมาตรฐานน้อย
3. ควรศึกษาโดยใช้แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน เพื่อเป็นการเปรียบเทียบว่าให้ผลต่างจากแบบทดสอบวัดความถนัดหรือไม่

บรรณานุกรม

บรรณานุกรม

จิตรา วสุวานิช และคณะ. จิตวิทยาเด็ก. กรุงเทพฯ : มหาวิทยาลัยรามคำแหง, 2530.

ชวาล แพรัตกุล. การทดสอบเพื่อค้นและพัฒนาสมรรถภาพ. สำนักทดสอบทางการศึกษาและจิตวิทยา วิทยาลัยวิชาการศึกษา ประสานมิตร, 2517.

✓ ชูชีพ พงษ์สมบูรณ์. การเปรียบเทียบการเทียบมาตราระหว่างรูปแบบที่ใช้ผู้สอบร่วมกับรูปแบบที่ใช้แบบสอบร่วม. วิทยานิพนธ์ ค.ม. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2528. อุดลำนเา.

ทองหล่อ วิภาวีน. การวัดความถนัด. กรุงเทพฯ : โอเดียนสโตร์, 2524.

บุญชม ศรีสะอาด. การวัดเชาวน์ปัญญาและความถนัด. มหาวิทยาลัยศรีนครินทรวิโรฒ มหาสารคาม, 2521.

บุญเชิด ภิญโญนนตพงษ์. การวัดและประเมินผลการศึกษา : ทฤษฎีและการประยุกต์. กรุงเทพฯ : อักษรเจริญทัศน์, 2523.

———. แบบทดสอบคะแนนจริงสัมพันธ์ : วิธีเทียบคะแนนสอบ. ภาควิชาวัดผลและวิจัยการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2538. อุดลำนเา.

———. "แบบทดสอบคะแนนจริงสัมพันธ์ : การวิเคราะห์ทางสถิติ." วารสารการวัดผลการศึกษา. 16(47) : 50-70 ; กันยายน-ธันวาคม 2537.

———. "แบบทดสอบคะแนนจริงสัมพันธ์ : สัมประสิทธิ์ r_B ." วารสารการวัดผลการศึกษา. 16(48) : 43-56 ; มกราคม-เมษายน 2538.

ประมวญ ดิคนันท์. จิตวิทยาเบื้องต้น. กรุงเทพฯ : ไทยวัฒนาพานิช, 2524.

———. จิตวิทยาการศึกษานิเทศ. กรุงเทพฯ : การศาสนา , 2532.

พวงรัตน์ ทวีรัตน์. วิธีการวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์. ครั้งที่ 4. กรุงเทพฯ : ฟิงเกอร์ปรีน แอนด์ มีเดียจำกัด, 2535.

ภาวิณี ศรีสุขวัฒนานนท์. การเปรียบเทียบผลจากการใช้รูปแบบการเทียบมาตราที่ต่างกันเมื่อแบบสอบร่วมมีความยาวต่างกัน. วิทยานิพนธ์ ค.ม. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2529. อุดลำนเา.

เรวดี อินทสระระ. การเปรียบเทียบความคลาดเคลื่อนของการเทียบมาตราระหว่างรูปแบบอิงทฤษฎีการตอบข้อสอบกับรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ. ปรินญาพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2530. อุดลำนเา.

ล้วน สายยศ. หลักการสร้างแบบทดสอบวัดความถนัด. สำนักทดสอบทางการศึกษาและจิตวิทยา, มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2522.

ล้วน สายยศ และ อังคณา สายยศ. หลักการสร้างแบบทดสอบวัดความถนัดทางการเรียน. กรุงเทพฯ : วัฒนาพานิช, 2527.

วัชร ทรัพย์มี. โครงการตำราและเอกสารทางวิชาการ คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. กรุงเทพฯ : คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2531.

✓ วัชร วรรณรัตน์. "การเทียบมาตรฐาน(Equating)," วารสารการวัดผลการศึกษา. 11(29): 69-73 ; กันยายน-ธันวาคม 2530.

✓ สงบ ลักษณะ. "การเทียบระดับคะแนนระหว่างแบบทดสอบ(Equating)," วารสารการวัดผลการศึกษา. 4(2) : 21-32 ; กันยายน-ธันวาคม 2525.

สมบูรณ์ ชิตพงษ์ และสำเริง บุญเรืองรัตน์. การวัดความถนัด. กรุงเทพฯ : ไทยวัฒนาพานิช, 2518.

สุจินดา ผ่องอักษร. การศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในกลุ่มทักษะ ของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษา พุทธศักราช 2521 ในช่วงระยะเวลา 3 ปีการศึกษา (ปีการศึกษา 2529 - 2531) โดยใช้การเทียบคะแนนรูปแบบราชส์. ปริญญาพันธ์ กศ.ด. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2533. อัดสำเนา. ๓๗ ๖, ๒๗,

✓ สุมนา โสติดิฉลอนันต์. การศึกษาวิธีการเทียบคะแนนแบบเส้นตรงตามแบบจำลองคะแนนจริง สัมพันธ. ปริญญาพันธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2539. อัดสำเนา.

โสภา ชูนิกุลชัย. จิตวิทยาทั่วไป. กรุงเทพฯ : ไทยวัฒนาพานิช, 2521.

สำเริง บุญเรืองรัตน์. "การพัฒนาทฤษฎีเลแทนเทรทเพื่อวิเคราะห์ข้อสอบ," วารสารการวัดผลการศึกษา. 7(21) : 41-68 ; มกราคม-เมษายน 2529.

✓ อภรณ์ กาญจนกิจโสภณ. การสร้างแบบทดสอบและตารางเทียบระดับคะแนนตามแนวทางในวิชาคณิตศาสตร์. กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2531. ๓๗, ๒๕๐, ๓ ๕๖๓, ๗
อัดสำเนา.

Angoff, William H. "Technical and Pratical Issues in Equating : A Discussion of Four Papers," Applied Psychology Measurement. 11 (3) : 291-300 ; September, 1987.

- Angoff, William H. "Summary and Derivation of Equating Methods Used at ETS," In Test Equating. edited by P.W.Holland and D.B.Rubin. New York : Acedemic Press, 1982.
- Braun, Henry I. and Holland, Paul W. "Observed-score Test Equating : A Mathematical Analysis of Some ETS Equating Procedures," In Test Equating. edited by P.W.Holland and D.B.Rubin. New York : Acedemic Press, 1982.
- Brennan, Robert L. and Kolen, Michael G. "Some practical issues in equating," Applied Psychological Measurement. 11(3): 279-290 ; September, 1987.
- Budescu, David. "Efficiency of Linear Equating as a Function of the Lenght of the Anchor Test," Journal of Education Measurement. 22(1) : 13-20 ; Spring, 1985.
- Cook, Linda L. and Petersen, Nancy S. "Problems Related to the Use of Conventional and Item Response Theory Equating Methods in Less Than Optimal Circumstances," Applied Psychology Measurement. 11(3) : 225-244 ; September, 1987.
- Cope, Ronald T. "How Well Do The Angoff Design V Linear Equating Methods Compare With the Tucker and Levine Methods," Applied Psycho-logical Measurement. 11(2) : 143-149 ; June, 1987.
- Flanagan, T.C. "Unit, Scores and Norm," In Education Mesurement. E.F. Lindquist(ED). Washington, DC : American Council on Education. 1951.
- Klein, Lawrence W. and Jarjoura, David. "The importance of content representation for common-item equating with nonrandom groups," Journal of Educational Measurement. 22(3) : 197-206 ; June, 1985.
- Kolen, Michael J. "Standard Errors of Tucker Equating," Applied Psycho-logical Measurement. 9(2) : 209-223 ; June, 1985.
- Kolen, Michael J. and Brennan Robert L. "Linear Equating Models for the Common Item Nonequivalent-Population Design," Applied Psychology Measurement . 11(3) : 263-277 ; September, 1987.

MacCann, Robert G. "A Comparison of two Observed-score Equating Methods that Assume Equally Reliable, Congeneric Tests," Applied Psychological Measurement. 13(3): 263-276 ; September, 1989.

Mary J. Allen and Wendy M. Yen. In Introduction to Measurement Theory. Brooks/Cole Publishing Company Monterey, California. 1979.

Parshall, Cynthia G., Houghton Pansy Du Bose. and Kromrey, Jeffrey D. "Equating Error and Statistic Bias in Small Sample Linear Equating," Journal of Educational Measurement. 32(1) : 37-54 ; Spring, 1995.

Petersen, Nancy S., Kolen, Michael J. and Hoover, H.D. "Scaling, Norming, and Equating," In Educational Measurement. edited by R.L. Linn, p.241-260. New York : American Council on Education and Macmillan Publishing Company, 1989.

Petersen, Nancy S., Marco, Gary L. and Stewart, E. Elizabeth. "A test of the Adequacy of Linear Score Equating Models," In Test Equating. edited by P.W. Holland and D.B. Rubin, New York: Academic Press, 1982.

Woodruff, David J. "A Comparison of three linear equating methods for common-item nonequivalent-populations design," Applied Psychological Measurement. 13(3):257-261; September, 1989.

អាជ្ញាធរ

คะแนนเทียบเฉลี่ยของกลุ่มตัวอย่างขนาดต่าง ๆ และประชากรเทียบทุกคะแนนสังเกต

ขนาด คะแนนสังเกต	30 คน	60 คน	120 คน	250 คน	913 คน
0	- 7.53	- 6.22	- 6.38	- 5.17	- 5.20
1	- 6.58	- 5.30	- 5.45	- 4.28	- 4.31
2	- 5.63	- 4.38	- 4.53	- 3.38	- 3.42
3	- 4.68	- 3.45	- 3.61	- 2.49	- 2.52
4	- 3.73	- 2.53	- 2.68	- 1.60	- 1.63
5	- 2.78	- 1.61	- 1.76	- .71	- .73
6	- 1.82	- .69	- .83	.19	.16
7	- .87	.23	.09	1.08	1.06
8	.08	1.15	1.02	1.97	1.95
9	1.03	2.07	1.94	2.87	2.84
10	1.98	3.00	2.87	3.76	3.74
11	2.94	3.92	3.79	4.65	4.63
12	3.89	4.84	4.72	5.54	5.53
13	4.84	5.76	5.64	6.44	6.42
14	5.79	6.68	6.57	7.33	7.31
15	6.74	7.60	7.49	8.22	8.21
16	7.70	8.52	8.42	9.11	9.10
17	8.65	9.45	9.34	10.01	10.00
18	9.60	10.37	10.26	10.90	10.89
19	10.55	11.29	11.19	11.79	11.78
20	11.50	12.21	12.11	12.68	12.68
21	12.45	13.13	13.04	13.58	13.57
22	13.41	14.05	13.96	14.47	14.47
23	14.36	14.98	14.89	15.36	15.36

ขนาด คะแนนสังเกต	30 คน	60 คน	120 คน	250 คน	913 คน
24	15.31	15.90	15.81	16.25	16.25
25	16.26	16.82	16.74	17.15	17.15
26	17.21	17.74	17.66	18.04	18.04
27	18.17	18.66	18.59	18.93	18.94
28	19.12	19.58	19.51	19.83	19.83
29	20.07	20.50	20.44	20.72	20.72
30	21.02	21.43	21.36	21.61	21.62
31	21.97	22.35	22.29	22.50	22.51
32	22.93	23.27	23.21	23.40	23.41
33	23.88	24.19	24.13	24.29	24.30
34	24.83	25.11	25.06	25.18	25.19
35	25.78	26.03	25.98	26.07	26.09
36	26.73	26.95	26.91	26.97	26.98
37	27.68	27.88	27.83	27.86	27.88
38	28.64	28.80	28.76	28.75	28.77
39	29.59	29.72	29.68	29.64	29.67
40	30.54	30.64	30.61	30.54	30.56
41	31.49	31.56	31.53	31.43	31.45
42	32.44	32.48	32.46	32.32	32.35
43	33.40	33.41	33.38	33.21	33.24
44	34.35	34.33	34.31	34.10	34.14
45	35.30	35.25	35.23	35.00	35.03
46	36.25	36.17	36.16	35.89	35.92
47	37.20	37.09	37.08	36.79	36.82
48	38.16	38.01	38.00	37.68	37.71

ขนาด คะแนนสังเกต	30 คน	60 คน	120 คน	250 คน	913 คน
49	39.11	38.93	38.93	38.57	38.61
50	40.06	39.86	39.85	39.46	39.50
51	41.01	40.78	40.78	40.36	40.39
52	41.96	41.70	41.70	41.25	41.29
53	42.91	42.62	42.63	42.14	42.18
54	43.87	43.54	43.55	43.03	43.08
55	44.82	44.46	44.48	43.93	43.97
56	45.77	45.38	45.40	44.82	44.86
57	46.72	46.31	46.33	45.71	45.76
58	47.67	47.23	47.25	46.60	46.65
59	48.63	48.15	48.18	47.50	47.55
60	49.58	49.07	49.10	48.39	48.44

บทสรุปย่อ ๒

๑. วัตถุประสงค์และคำอธิบายจำแนกของแบบทดสอบทั้งสามฉบับ

ค่าความยากและค่าอำนาจจำแนกของแบบทดสอบทั้งสามฉบับ

ข้อ	ฉบับ X		ฉบับ Y		ฉบับ Z	
	P	r	P	r	P	r
1	.53	.42	.56	.42	.66	.46
2	.45	.51	.42	.42	.54	.42
3	.65	.71	.44	.56	.76	.62
4	.67	.64	.76	.35	.63	.37
5	.62	.43	.54	.41	.39	.24
6	.70	.48	.49	.48	.65	.36
7	.62	.71	.53	.35	.73	.28
8	.66	.39	.69	.33	.69	.41
9	.53	.61	.44	.23	.78	.40
10	.58	.51	.76	.36	.32	.27
11	.75	.41	.50	.32	.56	.32
12	.54	.61	.69	.39	.73	.20
13	.78	.41	.65	.33	.32	.25
14	.71	.26	.53	.24	.46	.27
15	.79	.34	.78	.57	.74	.23
16	.65	.60	.72	.66	.66	.35
17	.74	.43	.47	.33	.80	.45
18	.31	.30	.66	.61	.43	.26
19	.71	.72	.54	.38	.77	.26
20	.61	.32	.78	.39	.42	.44
21	.58	.33	.73	.42	.78	.70
22	.51	.43	.67	.36	.48	.30
23	.56	.29	.54	.55	.39	.27
24	.40	.41	.63	.47	.77	.58

ចំនួន	ទំនាក់ទំនង X		ទំនាក់ទំនង Y		ទំនាក់ទំនង Z	
	P	r	P	r	P	r
25	.68	.32	.58	.33	.53	.39
26	.60	.36	.64	.29	.79	.54
27	.58	.38	.73	.30	.78	.46
28	.70	.50	.42	.30	.76	.69
29	.63	.33	.49	.29	.58	.44
30	.55	.43	.78	.42	.74	.32
31	.59	.40	.72	.34		
32	.77	.43	.77	.26		
33	.68	.40	.31	.22		
34	.79	.45	.79	.35		
35	.79	.50	.45	.30		
36	.77	.48	.33	.22		
37	.79	.41	.47	.41		
38	.78	.32	.71	.32		
39	.68	.39	.40	.24		
40	.80	.38	.80	.44		
41	.78	.25	.54	.31		
42	.78	.52	.65	.47		
43	.58	.45	.32	.27		
44	.78	.31	.79	.56		
45	.75	.42	.63	.35		
46	.75	.34	.69	.44		
47	.79	.53	.76	.52		
48	.78	.55	.60	.33		

ข้อ	ฉบับ X		ฉบับ Y		ฉบับ Z	
	P	r	P	r	P	r
49	.79	.51	.57	.34		
50	.74	.44	.63	.43		
51	.49	.32	.56	.46		
52	.79	.64	.72	.49		
53	.73	.42	.69	.56		
54	.53	.29	.59	.29		
55	.63	.43	.67	.28		
56	.75	.49	.78	.47		
57	.45	.28	.64	.24		
58	.38	.33	.52	.27		
59	.45	.24	.55	.39		
60	.59	.31	.75	.42		

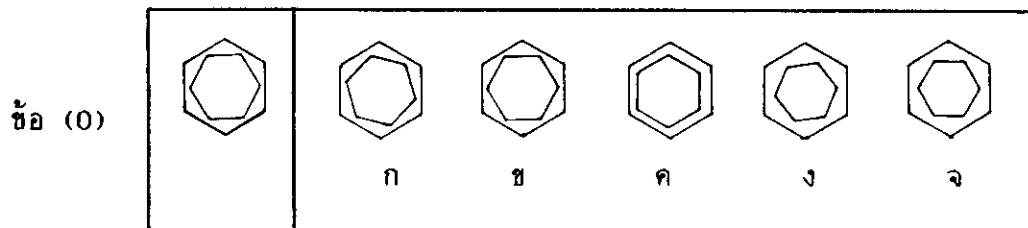
ภาคผนวก ค
แบบทดสอบที่สามฉบับ

แบบทดสอบฉบับที่ 1

คำอธิบาย

1. แบบทดสอบฉบับนี้ มี 2 ตอน ตอนละ 30 ข้อ รวม 60 ข้อ ให้อเวลาทำ 6 นาที ดังนั้นนักเรียนต้องทำอย่างรวดเร็ว และครบทุกข้อ จึงจะได้คะแนนดี
2. แบบทดสอบตอนที่ 1 เป็นแบบทดสอบวัดการรับรู้ โดยให้นักเรียนหาว่า ภาพในข้อ ก, ข, ค, ง หรือ จ เป็นภาพที่ เหมือนกับภาพที่กำหนดให้ทางซ้ายมือ ในข้อเดียวกัน

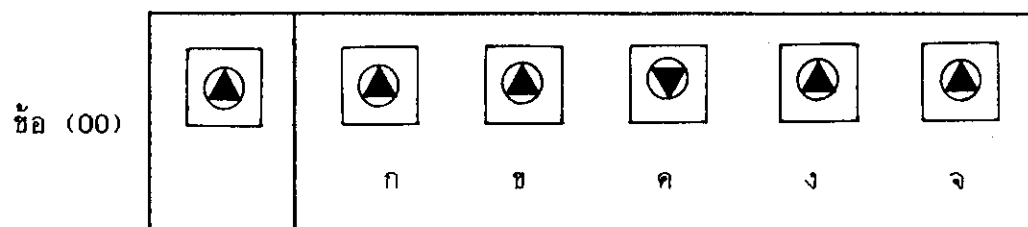
ตัวอย่าง



คำตอบคือ ข้อ ข

3. แบบทดสอบตอนที่ 2 เป็นแบบทดสอบวัดการรับรู้ โดยให้นักเรียนหาว่า ภาพในข้อ ก, ข, ค, ง หรือ จ เป็นภาพที่ ต่างจากภาพที่กำหนดให้ทางซ้ายมือ ในข้อเดียวกัน

ตัวอย่าง



คำตอบคือ ข้อ ค

4. เมื่อนักเรียนเลือกคำตอบได้แล้ว นำไปกากบาทลงในกระดาษคำตอบ ดังนี้

	ก	ข	ค	ง	จ
ข้อ (0)	☒	☐	☐	☐	☐
ข้อ (00)	☐	☐	☐	☐	☐

5. ถ้านักเรียนต้องการเปลี่ยนคำตอบ ให้ขีดเครื่องหมาย = ทับคำตอบเดิม แล้วจึงกากบาทในคำตอบใหม่ เช่น ต้องการเปลี่ยนคำตอบในข้อ (0) จากตัวเลือก ก ไปเป็นตัวเลือก ค ให้ทำดังนี้

	ก	ข	ค	ง	จ
ข้อ (0)	☒	☐	☒	☐	☐
ข้อ (00)	☐	☐	☐	☐	☐

6. ถ้าพบว่าข้อใดยาก อย่าท้อใจ ให้ข้ามไปทำข้ออื่นก่อน แล้วจึงกลับมาทำข้อนั้นภายหลังก่อนหมดเวลา

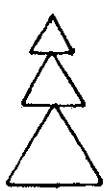
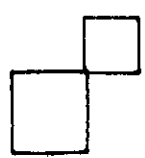
7. อย่าขีดเขียนเครื่องหมายใด ๆ ลงบนแบบทดสอบนี้

8. ถ้านักเรียนทำเสร็จ หรือหมดเวลา ให้นักเรียนส่งแบบทดสอบ และกระดาษคำตอบทันที

ขอให้นักเรียนทุกคนโชคดี

ตอนที่ ๑

ให้นักเรียนเลือกรูปที่เหมือนกับรูปที่กำหนดให้ ทางซ้ายมือ

๑	 ก ข ค ง จ
๒	 ก ข ค ง จ
๓	 ก ข ค ง จ
๔	 ก ข ค ง จ
๕	 ก ข ค ง จ

b



၁



၂



၃

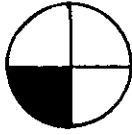
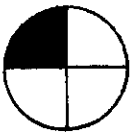


၄

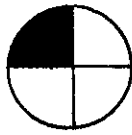


၅

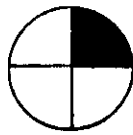
c



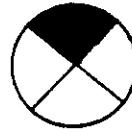
၁



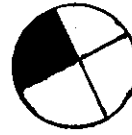
၂



၃



၄

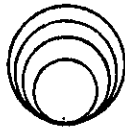


၅

d



၁



၂



၃



၄

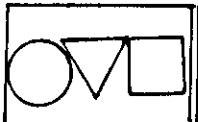


၅

e



၁



၂



၃

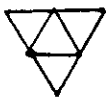
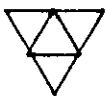


၄



၅

f



၁



၂



၃

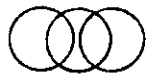


၄

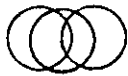


၅

၁၁



၈



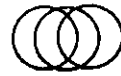
၉



၁၀



၁၁



၁၂

၁၂



၈



၉



၁၀



၁၁



၁၂

၁၃



၈



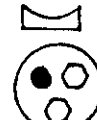
၉



၁၀

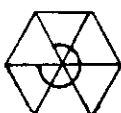


၁၁



၁၂

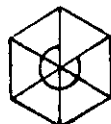
၁၄



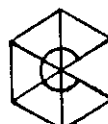
၈



၉



၁၀



၁၁



၁၂

၁၅



၈



၉



၁၀

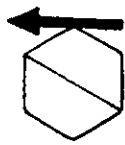
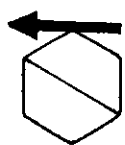


၁၁

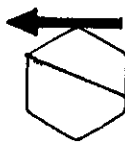


၁၂

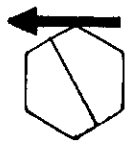
၁၆



၈



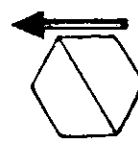
၉



၁၀



၁၁



၁၂

၁၇



၈



၉



၁၀



၁၁



၁၂

၁၈



၈



၉



၁၀

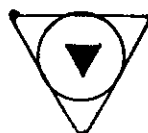
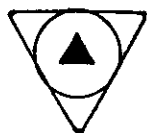


၁၁



၁၂

၁၉



၈



၉



၁၀



၁၁



၁၂

၂၀



၈



၉



၁၀



၁၁



၁၂

1. 1



1

2

3

4

5

2. 1



1

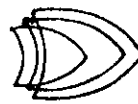
2

3

4

5

3. 1



1

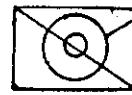
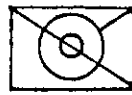
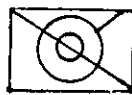
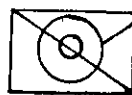
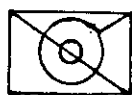
2

3

4

5

4. 1



1

2

3

4

5

5. 1



1

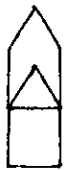
2

3

4

5

๔๖



ก



ข



ค



ง



จ

๔๗



ก



ข



ค



ง



จ

๔๘



ก



ข



ค

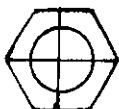


ง



จ

๔๙



ก



ข



ค



ง



จ

๕๐



ก



ข



ค

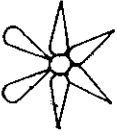
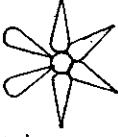
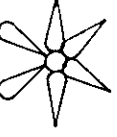
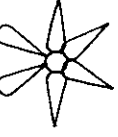
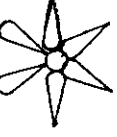


ง



จ

ตอนที่ ๒ ให้นักเรียนเลือกรูปที่ต่างจากรูปที่กำหนดให้ ทางซ้ายมือ

๓๑						
		ก	ข	ค	ง	จ
๓๒						
		ก	ข	ค	ง	จ
๓๓						
		ก	ข	ค	ง	จ
๓๔						
		ก	ข	ค	ง	จ
๓๕						
		ก	ข	ค	ง	จ

៣៦



ក



ខ



គ



ឃ



ង

៣៧



ក



ខ



គ



ឃ



ង

៣៨



ក



ខ



គ



ឃ



ង

៣៩



ក



ខ



គ



ឃ



ង

៤០



ក



ខ



គ

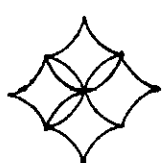
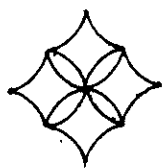


ឃ

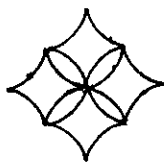


ង

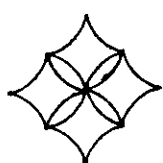
៤១



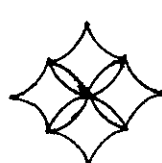
១



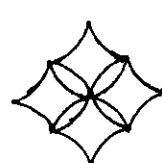
២



៣



៤



៥

៤២



១



២



៣

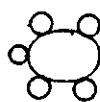
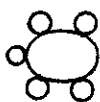


៤

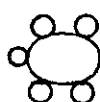


៥

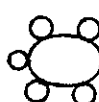
៤៣



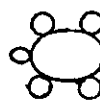
១



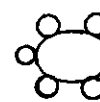
២



៣



៤



៥

៤៤



១



២



៣



៤



៥

៤៥



១



២



៣



៤



៥

၆၀



၈

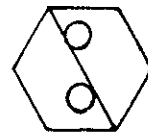
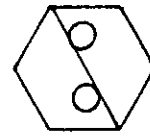
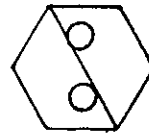
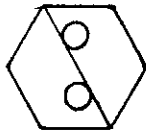
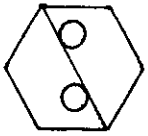
၉

၁၀

၁၁

၁၂

၆၁



၈

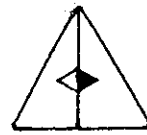
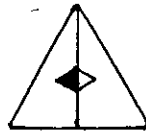
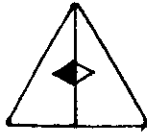
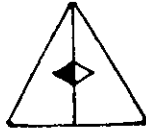
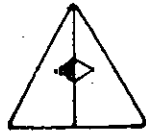
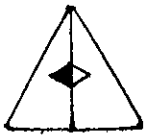
၉

၁၀

၁၁

၁၂

၆၂



၈

၉

၁၀

၁၁

၁၂

၆၃



၈

၉

၁၀

၁၁

၁၂

၆၄



၈

၉

၁၀

၁၁

၁၂

๕๑



ก



ข



ค

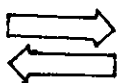
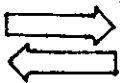


ง

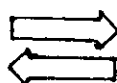


จ

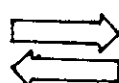
๕๒



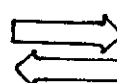
ก



ข



ค



ง



จ

๕๓



ก



ข



ค



ง



จ

๕๔



ก



ข



ค



ง



จ

๕๕



ก



ข



ค



ง



จ

๕๖



ก



ข



ค

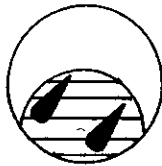


ง

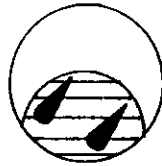


จ

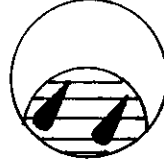
๕๗



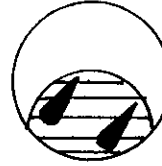
ก



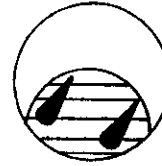
ข



ค



ง

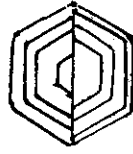


จ

๕๘



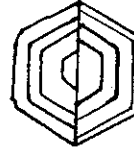
ก



ข



ค

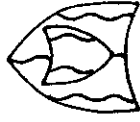
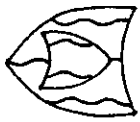


ง

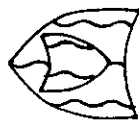


จ

๕๙



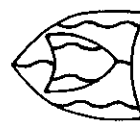
ก



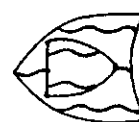
ข



ค



ง



จ

๖๐



ก



ข



ค



ง



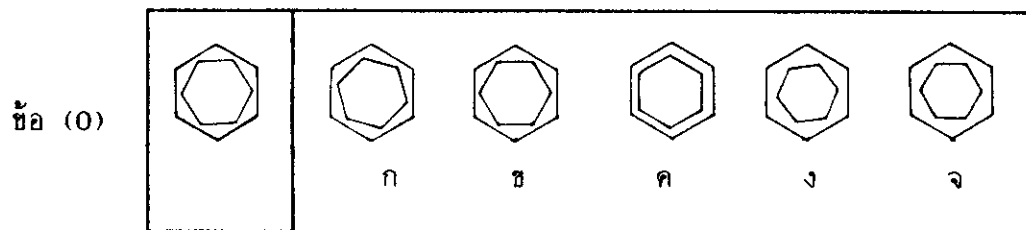
จ

แบบทดสอบฉบับที่ 2

คำอธิบาย

1. แบบทดสอบฉบับนี้ มี 2 ตอน ตอนละ 30 ข้อ รวม 60 ข้อ ให้อเวลาทำ 6 นาที ดังนั้นนักเรียนต้องทำอย่างรวดเร็ว และครบทุกข้อ จึงจะได้คะแนนดี
2. แบบทดสอบตอนที่ 1 เป็นแบบทดสอบวัดการรับรู้ โดยให้นักเรียนหาว่า ภาพในข้อ ก, ข, ค, ง หรือ จ เป็นภาพที่ เหมือนกับภาพที่กำหนดให้ทางซ้ายมือ ในข้อเดียวกัน

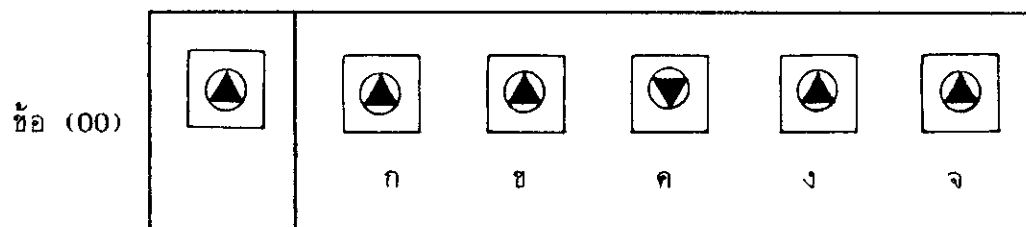
ตัวอย่าง



คำตอบคือ ข้อ ข

3. แบบทดสอบตอนที่ 2 เป็นแบบทดสอบวัดการรับรู้ โดยให้นักเรียนหาว่า ภาพในข้อ ก, ข, ค, ง หรือ จ เป็นภาพที่ ต่างจากภาพที่กำหนดให้ทางซ้ายมือ ในข้อเดียวกัน

ตัวอย่าง



คำตอบคือ ข้อ ค

4. เมื่อนักเรียนเลือกคำตอบได้แล้ว นำไปกากบาทลงในกระดาษคำตอบ ดังนี้

	ก	ข	ค	ง	จ
ข้อ (0)	X				
ข้อ (00)					

5. ถ้านักเรียนต้องการเปลี่ยนคำตอบ ให้ขีดเครื่องหมาย = ทับคำตอบเดิม แล้วจึงกากบาทในคำตอบใหม่ เช่น ต้องการเปลี่ยนคำตอบในข้อ (0) จากตัวเลือก ก ไปเป็นตัวเลือก ค ให้ทำดังนี้

	ก	ข	ค	ง	จ
ข้อ (0)	X		X		
ข้อ (00)					

6. ถ้าพบว่าข้อใดยาก อย่าท้อใจ ให้ข้ามไปทำข้ออื่นก่อน แล้วจึงกลับมาทำข้อนั้นภายหลังก่อนหมดเวลา

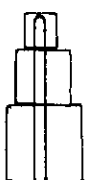
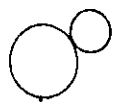
7. อย่าขีดเขียนเครื่องหมายใด ๆ ลงบนแบบทดสอบนี้

8. ถ้านักเรียนทำเสร็จ หรือหมดเวลา ให้นักเรียนส่งแบบทดสอบ และกระดาษคำตอบทันที

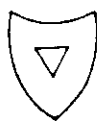
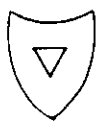
ขอให้นักเรียนทุกคนโชคดี

ตอนที่ ๑

ให้นักเรียนเลือกรูปที่เหมือนกับรูปที่กำหนดให้ ทางซ้ายมือ

๑	 ก ข ค ง จ
๒	 ก ข ค ง จ
๓	 ก ข ค ง จ
๔	 ก ข ค ง จ
๕	 ก ข ค ง จ

၁၆



၀



၁



၂



၃



၄

၁၇



၀



၁



၂



၃



၄

၁၈



၀



၁



၂



၃



၄

၁၉



၀



၁



၂



၃



၄

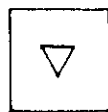
၂၀



၀



၁



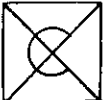
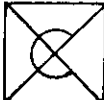
၂



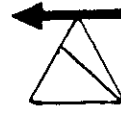
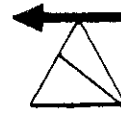
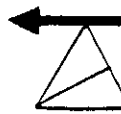
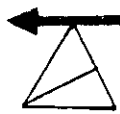
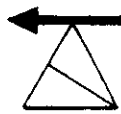
၃



၄

၁၁						
	၀	၁	၂	၃	၄	၅
၁၂						
	၀	၁	၂	၃	၄	၅
၁၃						
	၀	၁	၂	၃	၄	၅
၁၄						
	၀	၁	၂	၃	၄	၅
၁၅						
	၀	၁	၂	၃	၄	၅

၁၆



၀

၁

၂

၃

၄

၁၇



၀

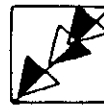
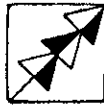
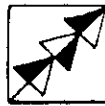
၁

၂

၃

၄

၁၈



၀

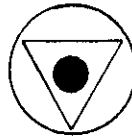
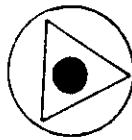
၁

၂

၃

၄

၁၉



၀

၁

၂

၃

၄

၂၀



၀

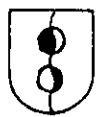
၁

၂

၃

၄

၁၃၁



၀

၁

၂

၃

၄

၁၃၂



၀

၁

၂

၃

၄

၁၃၃



၀

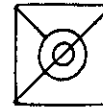
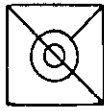
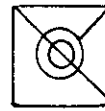
၁

၂

၃

၄

၁၃၄



၀

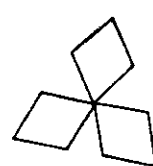
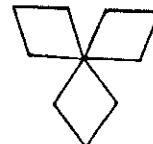
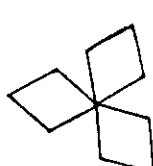
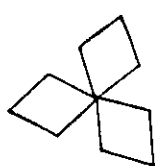
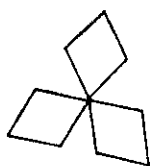
၁

၂

၃

၄

၁၃၅



၀

၁

၂

၃

၄

១២



ក

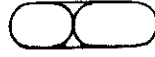
ខ

គ

ឃ

ង

១៣



ក

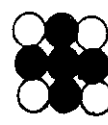
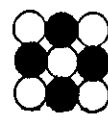
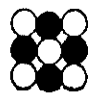
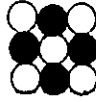
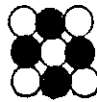
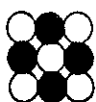
ខ

គ

ឃ

ង

១៤



ក

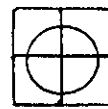
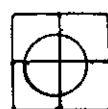
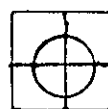
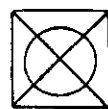
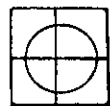
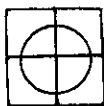
ខ

គ

ឃ

ង

១៥



ក

ខ

គ

ឃ

ង

១៦



ក

ខ

គ

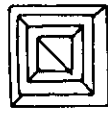
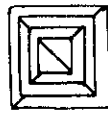
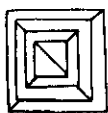
ឃ

ង

ตอนที่ ๒ ให้นักเรียนเลือกรูปที่ต่างจากรูปที่กำหนดให้ทางซ้ายมือ

๓๑						
		ก	ข	ค	ง	จ
๓๒						
		ก	ข	ค	ง	จ
๓๓						
		ก	ข	ค	ง	จ
๓๔						
		ก	ข	ค	ง	จ
๓๕						
		ก	ข	ค	ง	จ

၈၆



၈

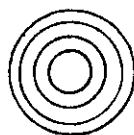
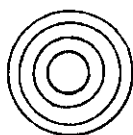
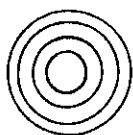
၉

၁၀

၁၁

၁၂

၈၇



၈

၉

၁၀

၁၁

၁၂

၈၈



၈

၉

၁၀

၁၁

၁၂

၈၉



၈

၉

၁၀

၁၁

၁၂

၉၀



၈

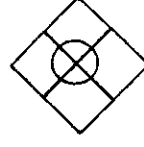
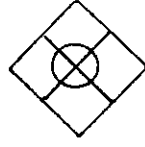
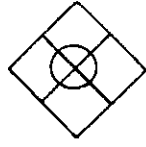
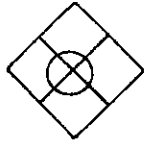
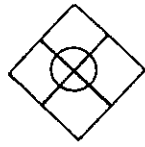
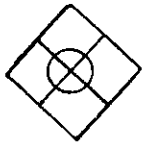
၉

၁၀

၁၁

၁၂

៤៦



១

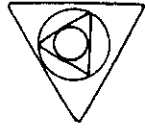
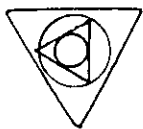
២

៣

៤

៥

៤៧



១

២

៣

៤

៥

៤៨



១

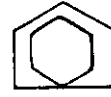
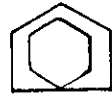
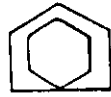
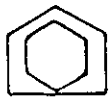
២

៣

៤

៥

៤៩



១

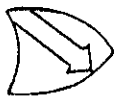
២

៣

៤

៥

៥០



១

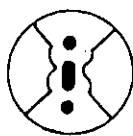
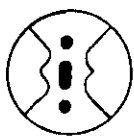
២

៣

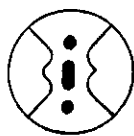
៤

៥

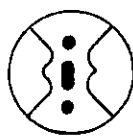
๔๖



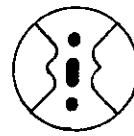
ก



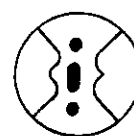
ข



ค

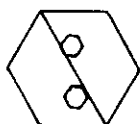


ง



จ

๔๗



ก



ข



ค



ง



จ

๔๘



ก



ข



ค

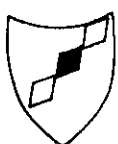


ง

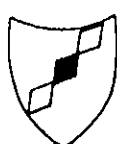


จ

๔๙



ก



ข



ค



ง



จ

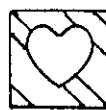
๕๐



ก



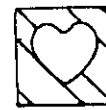
ข



ค

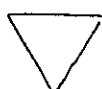
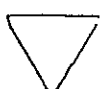
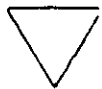


ง



จ

៤១



១

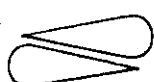
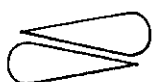
២

៣

៤

៥

៤២



១

២

៣

៤

៥

៤៣



១

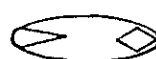
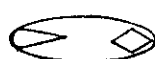
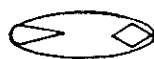
២

៣

៤

៥

៤៤



១

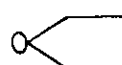
២

៣

៤

៥

៤៥



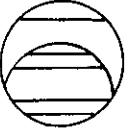
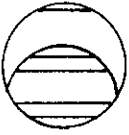
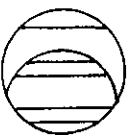
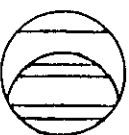
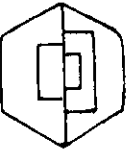
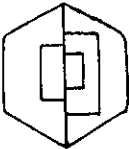
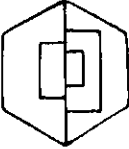
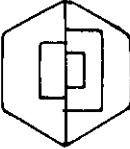
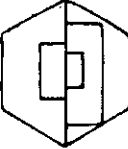
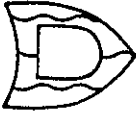
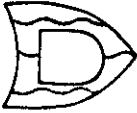
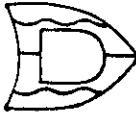
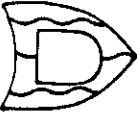
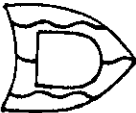
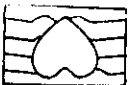
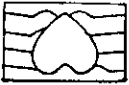
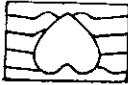
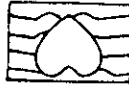
១

២

៣

៤

៥

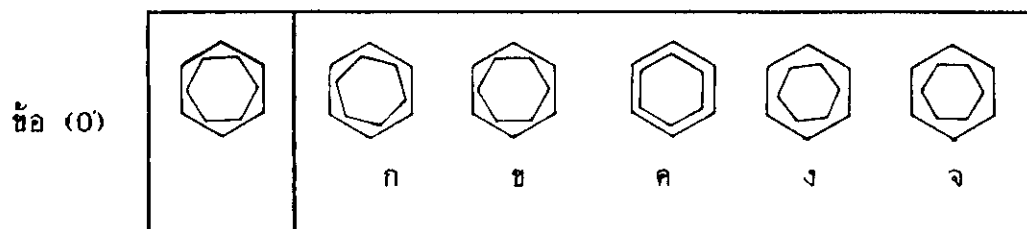
៤៦		 ក	 ខ	 គ	 ឌ	 ឍ
៤៧		 ក	 ខ	 គ	 ឌ	 ឍ
៤៨		 ក	 ខ	 គ	 ឌ	 ឍ
៤៩		 ក	 ខ	 គ	 ឌ	 ឍ
៥០		 ក	 ខ	 គ	 ឌ	 ឍ

แบบทดสอบฉบับที่ 3

คำอธิบาย

1. แบบทดสอบฉบับนี้ มี 2 ตอน ตอนละ 15 ข้อ รวม 30 ข้อ ให้เวลาทำ 3 นาที ดังนั้นนักเรียนต้องทำอย่างรวดเร็ว และครบทุกข้อ จึงจะได้คะแนนดี
2. แบบทดสอบตอนที่ 1 เป็นแบบทดสอบวัดการรับรู้ โดยให้นักเรียนหาว่า ภาพในข้อ ก, ข, ค, ง หรือ จ เป็นภาพที่ เหมือนกับภาพที่กำหนดให้ทางซ้ายมือ ในข้อเดียวกัน

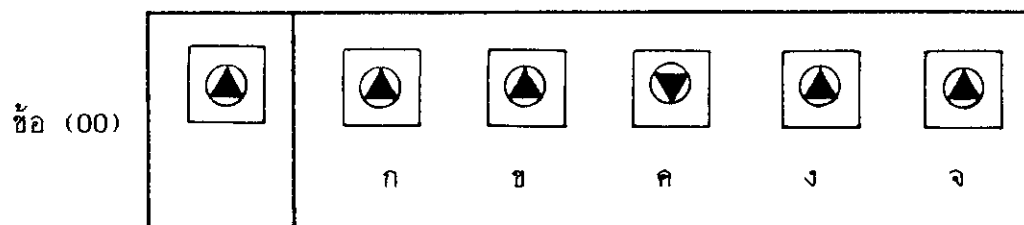
ตัวอย่าง



คำตอบคือ ข้อ ข

3. แบบทดสอบตอนที่ 2 เป็นแบบทดสอบวัดการรับรู้ โดยให้นักเรียนหาว่า ภาพในข้อ ก, ข, ค, ง หรือ จ เป็นภาพที่ ต่างจากภาพที่กำหนดให้ทางซ้ายมือ ในข้อเดียวกัน

ตัวอย่าง



คำตอบคือ ข้อ ค

4. เมื่อนักเรียนเลือกคำตอบได้แล้ว นำไปกากบาทลงในกระดาษคำตอบ ดังนี้

	ก	ข	ค	ง	จ
ข้อ (0)	☒	☐	☐	☐	☐
ข้อ (00)	☐	☐	☐	☐	☐

5. ถ้านักเรียนต้องการเปลี่ยนคำตอบ ให้ขีดเครื่องหมาย = ทับคำตอบเดิม แล้วจึงกากบาทในคำตอบใหม่ เช่น ต้องการเปลี่ยนคำตอบในข้อ (0) จากตัวเลือก ก ไปเป็นตัวเลือก ค ให้ทำดังนี้

	ก	ข	ค	ง	จ
ข้อ (0)	☒	☐	☒	☐	☐
ข้อ (00)	☐	☐	☐	☐	☐

6. ถ้าพบว่าข้อใดยาก อย่าท้อใจ ให้ข้ามไปทำข้ออื่นก่อน แล้วจึงกลับมาทำข้อนั้นภายหลังก่อนหมดเวลา

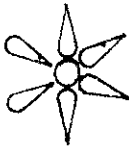
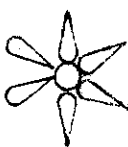
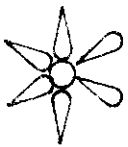
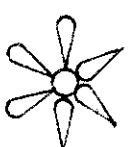
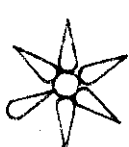
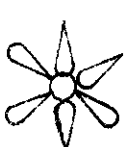
7. อย่าขีดเขียนเครื่องหมายใด ๆ ลงบนแบบทดสอบนี้

8. ถ้านักเรียนทำเสร็จ หรือหมดเวลา ให้นักเรียนส่งแบบทดสอบ และกระดาษคำตอบทันที

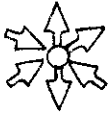
ขอให้นักเรียนทุกคนโชคดี

ตอนที่ ๑

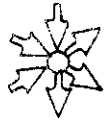
ให้นักเรียนเลือกรูปที่เหมือนกับรูปที่กำหนดให้ทางซ้ายมือ

๑	  ก  ข  ค  ง  จ
๒	  ก  ข  ค  ง  จ
๓	  ก  ข  ค  ง  จ
๔	  ก  ข  ค  ง  จ
๕	  ก  ข  ค  ง  จ

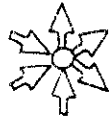
6



1



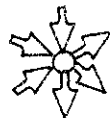
2



3



4



5

3



1



2



3



4



5

2



1



2



3



4



5

4



1



2



3



4



5

5



1



2



3

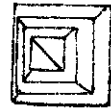
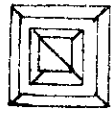


4



5

၅၅



ဂ

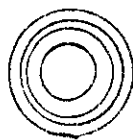
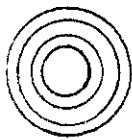
ဃ

ဂ

၁

၃

၅၆



ဂ

ဃ

ဂ

၁

၃

၅၇



ဂ

ဃ

ဂ

၁

၃

၅၈



ဂ

ဃ

ဂ

၁

၃

၅၉



ဂ

ဃ

ဂ

၁

၃

ตอนที่ ๒

ให้นักเรียนเลือกรูปที่ต่างจากรูปที่กำหนดให้ทางขวามือ

๑๖



ก



ข



ค



ง



จ

๑๗



ก



ข



ค



ง



จ

๑๘



ก



ข



ค



ง



จ

๑๙



ก



ข



ค



ง



จ

๒๐



ก



ข



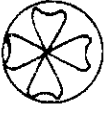
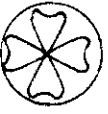
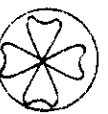
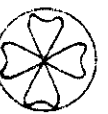
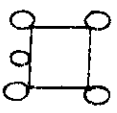
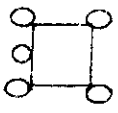
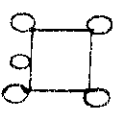
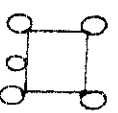
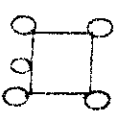
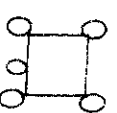
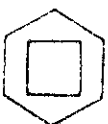
ค



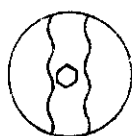
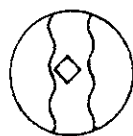
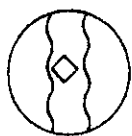
ง



จ

၁၁		 ၀	 ၁	 ၂	 ၃	 ၄
၁၂		 ၀	 ၁	 ၂	 ၃	 ၄
၁၃		 ၀	 ၁	 ၂	 ၃	 ၄
၁၄		 ၀	 ၁	 ၂	 ၃	 ၄
၁၅		 ၀	 ၁	 ၂	 ၃	 ၄

၁၀၀



၀

၁

၂

၃

၄

၁၀၁



၀

၁

၂

၃

၄

၁၀၂



၀

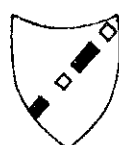
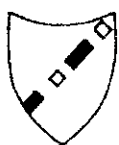
၁

၂

၃

၄

၁၀၃



၀

၁

၂

၃

၄

၁၀၄



၀

၁

၂

၃

၄

ประวัติย่อของผู้วิจัย

ชื่อ นางสาวนิภาพร	ชื่อสกุล ยอดเมือง
เกิดวันที่ 19 เดือน พฤษภาคม	พุทธศักราช 2498
สถานที่อยู่ปัจจุบัน	บ้านเลขที่ 122 หมู่ 4 ตำบล ภาชี อำเภอกาชี จังหวัดพระนครศรีอยุธยา
ตำแหน่งหน้าที่	อาจารย์ 2 ระดับ 6
สถานที่ทำงาน	โรงเรียนบางเขน (ไผ่สาละวันสรณ์) แขวงตลาดบางเขน เขตดอนเมือง กรุงเทพมหานคร

ประวัติการศึกษา

พ.ศ. 2525	ศษ.บ. (วิชาเอกสุขศึกษา) จากมหาวิทยาลัยเกษตรศาสตร์
พ.ศ. 2526	ศษ.บ. (วิชาเอกบริหารการศึกษา) จากมหาวิทยาลัยสุโขทัยธรรมมาธิราช
พ.ศ. 2530	ศษ.บ. (วิชาเอกคณิตศาสตร์) จากมหาวิทยาลัยสุโขทัยธรรมมาธิราช
พ.ศ. 2539	กศ.ม. (วิชาเอกการวัดผลการศึกษา) จากมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

ความคลาดเคลื่อน และความลำเอียงทางสถิติในการเทียบคะแนนเชิงเส้นตรง
แบบคะแนนจริงสัมพันธ์ในกลุ่มตัวอย่างขนาดเล็ก

บทคัดย่อ
ของ
นิภาพร ขอดเมือง

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา
กันยายน 2539

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อศึกษาว่าคะแนนเทียบเฉลี่ย ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน และความลำเอียงทางสถิติในการเทียบคะแนน ของกลุ่มตัวอย่างขนาด 30 คน 60 คน 120 คน และ 250 คนว่าแตกต่างกันหรือไม่ โดยศึกษากับนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2538 สังกัดกรุงเทพมหานคร เขตบางเขน เขตดอนเมือง เขตจตุจักร จำนวนนักเรียนทั้งสิ้น 1,826 คน ใช้แบบทดสอบวัดการรับรู้ 3 ฉบับ เป็นแบบทดสอบเทียบคะแนน 2 ฉบับ เป็นแบบทดสอบเชื่อมโยง 1 ฉบับ ประชากรที่ศึกษาแบ่งเป็น 2 กลุ่มกลุ่มที่ 1 จำนวน 913 คน ทำแบบทดสอบเทียบคะแนนฉบับที่หนึ่ง กับแบบทดสอบเชื่อมโยง กลุ่มที่ 2 จำนวน 913 คน ทำแบบทดสอบเทียบคะแนนฉบับที่สอง กับแบบทดสอบเชื่อมโยงฉบับเดิม

ผลการวิจัยสรุปดังนี้

1. คะแนนเทียบเฉลี่ยของกลุ่มตัวอย่างขนาด 30 คน 60 คน 120 คน และ 250 คนแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ
2. ค่าความลำเอียงทางสถิติในการเทียบคะแนน ในกลุ่มตัวอย่างขนาด 30 คน 60 คน 120 คน และ 250 คน แตกต่างกันน้อยมาก ค่าความลำเอียงทางสถิติในการเทียบคะแนนของกลุ่มตัวอย่างขนาด 60 คน และ 250 คน มีค่าเท่ากับ 0 เมื่อคะแนนสังเกตเข้าใกล้คะแนนเฉลี่ย
3. ค่าความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนในกลุ่มตัวอย่างขนาด 30 คน 60 คน 120 คน และ 250 คน แตกต่างกันเล็กน้อยเมื่อคะแนนสังเกตใกล้ค่าเฉลี่ย (38.8204)

ดังนั้นจากผลการวิจัยสรุปได้ว่า กลุ่มตัวอย่างที่สามารถนำไปใช้ในการเทียบคะแนนได้ คือ กลุ่มตัวอย่างทั้ง 4 ขนาด คือขนาด 30 คน 60 คน 120 คน และ 250 คน แต่ถ้าต้องการให้ค่าความคลาดเคลื่อนมาตรฐาน และความลำเอียงทางสถิติในการเทียบคะแนนมีค่าน้อย ก็สมควรใช้กลุ่มตัวอย่างขนาด 250 คน

EQUATING ERROR AND STATISTICAL BIAS IN SMALL SAMPLE
CONGENERIC LINEAR EQUATING

AN ABSTRACT

BY

NIPAPORN YODMUANG

Presented in partial fulfillment of the requirements for the
Master of Education degree in Educational Measurement
Srinakharinwirot University
September 1996

A study was conducted to find whether the equating mean score, the equating statistical bias and the equating standard errors of nonequivalent-group linear test equating in small samples of examinees were different. Sample sizes of 30, 60, 120 and 250 were examined. The samples were 1,826 Prathom 6 students of academic year 2538 in Bangkok, Donmuang and Jatujak in Bangkok. Three perception aptitude test, being two parallel tests and anchor test were used.

The students were divided into two groups. 913 students did the first parallel and the anchor test and the other 913 did the second parallel test and the anchor test.

The findings were as follows:

1. The equating mean score of the four sample sizes was not significantly different.

2. The equating statistical bias of the four sample sizes was different slightly, but that of the sample sizes of 60 and 250 was zero while the two scores were near the mean score.

3. The equating standard errors of the four sample sizes was different slightly while the score was near the mean score (38.8204).

The findings were concluded that the four sample sizes could be used to equate the score. It was suggested that the 250 sample size should be used if the equating statistical bias and the equating standard errors were low.