

การแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของคะแนนการประเมินทักษะกระบวนการทาง
วิทยาศาสตร์โดยใช้ทฤษฎีมาตรฐานเดิมและทฤษฎีการตอบข้อคำถาม
ตามโมเดลของราสส์แบบหลายองค์ประกอบ

ปริญญานิพนธ์
ของ
ชาญวิทย์ จรัสสุทธิอิสร์

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาคุณวุฒิบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา
ตุลาคม 2550

การแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของคะแนนการประเมินทักษะกระบวนการทาง
วิทยาศาสตร์โดยใช้ทฤษฎีมาตรฐานเดิมและทฤษฎีการตอบข้อคำถาม
ตามโมเดลของราล์ชแบบหลายองค์ประกอบ

ปริญญานิพนธ์
ของ
ชาญวิทย์ จรัสสุทธิอิสร์

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุขฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา

ตุลาคม 2550

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้สำเร็จสมบูรณ์ลุล่วงได้เป็นอย่างดี ด้วยความกรุณาอย่างยิ่งจาก รองศาสตราจารย์ ดร. บุญเชิด ภิญโญนนันทพงษ์ ประธานคณะกรรมการควบคุมปริญญานิพนธ์ อาจารย์ ดร. สุวพร เข้มเฮง และผู้ช่วยศาสตราจารย์ ดร. สุรัชย์ มีชาญ กรรมการควบคุม ปริญญานิพนธ์ ที่ได้เสียสละเวลาอันมีค่าในการให้คำปรึกษาแนะนำการทำงานวิจัยนี้ทุกขั้นตอน จนเสร็จสมบูรณ์ รวมทั้งทำให้ผู้วิจัยได้รับประสบการณ์ในการทำงานวิจัย และรู้ถึงคุณค่าของ งานวิจัยที่จะช่วยให้การทำงานในด้านการวัดและประเมินผลได้พัฒนาไปอย่างมีคุณค่ามากขึ้น และท่านทั้งสามยังเป็นแบบฉบับของอาจารย์ที่ทุ่มเทให้กับศิษย์และงานด้านวิชาการอย่างไม่รู้จัก เหน็ดเหนื่อย ผู้วิจัยรู้สึกทราบบ้างถึงความกรุณาดังกล่าว และขอกราบขอบพระคุณไว้เป็นอย่าง สูงไว้ ณ โอกาสนี้

ขอกราบขอบพระคุณผู้ช่วยศาสตราจารย์ ดร. งามา นัยพัฒน์ และรองศาสตราจารย์ ดร. สิริรัตน์ วิภาสศิลป์ กรรมการที่แต่งตั้งเพิ่มเติมในการสอบปริญญานิพนธ์ที่ได้ให้ข้อเสนอแนะ เพิ่มเติมอันเป็นประโยชน์ ทำให้ปริญญานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น ผู้วิจัยขอกราบขอบพระคุณ ท่านเป็นอย่างสูง

ขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. บุญเชิด ภิญโญนนันทพงษ์ และอาจารย์ ดร. สุวพร เข้มเฮง อาจารย์ที่ปรึกษานิพนธ์ ซึ่งได้ให้การดูแลด้วยความห่วงใยและให้คำแนะนำ อย่างดียิ่ง ตลอดระยะเวลาที่ศึกษาอยู่ภายในมหาวิทยาลัย

ขอขอบพระคุณ Dr. John M Linacre ผู้อำนวยการศูนย์ทดสอบทางการศึกษาและ จิตวิทยา มหาวิทยาลัยชิคาโก ที่ได้กรุณาอนุเคราะห์โปรแกรมคอมพิวเตอร์ FACETS ที่ใช้ในการ วิเคราะห์โมเดลของราล์ชแบบหลายองค์ประกอบ

ขอขอบคุณ เขตพื้นที่การศึกษานครราชสีมา เขต 1 เพื่อนร่วมงานและครอบครัวที่ให้ กำลังใจ และช่วยเหลือเป็นอย่างดีตลอดระยะเวลาที่ทำการศึกษาและวิจัย ขอขอบคุณเพื่อน ๆ นิสิตปริญญาเอก สาขาการทดสอบและวัดผลการศึกษา ภาคปกติ รุ่นที่ 6 ทุกคนที่ให้กำลังใจ ร่วมมือกันคว้าข้อมูลด้วยกันตลอดมา

ท้ายสุดนี้คุณประโยชน์และความดีที่ได้รับจากปริญญานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบแด่ บิดา มารดา ครูอาจารย์ และผู้มีอุปการะคุณทุกท่านที่คอยเป็นกำลังใจ ให้ความช่วยเหลือ อบรม สั่งสอนผู้วิจัยจนประสบความสำเร็จในการศึกษาเล่าเรียน

ชาญวิทย์ จรัสสุทธิธร

สารบัญ

บทที่

หน้า

1 บทนำ	1
ภูมิหลัง	1
ความมุ่งหมายของการวิจัย	12
ความสำคัญของการวิจัย	13
ขอบเขตของการวิจัย.....	14
ประชากรและกลุ่มตัวอย่างที่ใช้ในการวิจัย	14
ตัวแปรที่ศึกษา.....	14
นิยามศัพท์เฉพาะ.....	15
สมมติฐานและโมเดลในการวิจัย	19
 2 เอกสารและงานวิจัยที่เกี่ยวข้อง	21
ทักษะกระบวนการทางวิทยาศาสตร์	21
ความหมายของทักษะกระบวนการทางวิทยาศาสตร์	21
ประเภทของทักษะกระบวนการทางวิทยาศาสตร์	23
การประเมินผลทักษะกระบวนการทางวิทยาศาสตร์	27
หลักฐานแสดงความเที่ยงตรง.....	28
ความหมายของความเที่ยงตรง.....	28
ประเภทของหลักฐานแสดงความเที่ยงตรง	31
การวิเคราะห์องค์ประกอบ.....	40
การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง.....	42
การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง	51
การวิเคราะห์กลุ่มพหุ	52
เมตริกซ์หลายคุณลักษณะหลายวิธี.....	55
โมเดลของราส์ชแบบหลายองค์ประกอบ	60
งานวิจัยที่เกี่ยวข้อง	67

3 วิธีดำเนินการวิจัย.....	71
การกำหนดประชากรและการเลือกกลุ่มตัวอย่าง.....	71
การสร้างเครื่องมือที่ใช้ในการวิจัย.....	72
ตัวอย่างเครื่องมือที่ใช้ในการวิจัย.....	74
การเก็บรวบรวมข้อมูล.....	76
การจัดกระทำและการวิเคราะห์ข้อมูล.....	77
4 ผลการวิเคราะห์ข้อมูล.....	82
5 สรุปผล อภิปราย และข้อเสนอแนะ.....	104
ความมุ่งหมายของการวิจัย.....	104
สมมติฐานของการวิจัย.....	105
วิธีดำเนินการวิจัย.....	105
สรุปผลการวิจัย.....	107
อภิปรายผล.....	112
ข้อเสนอแนะ.....	119
บรรณานุกรม.....	120
ภาคผนวก.....	129
ประวัติย่อผู้วิจัย.....	160

บัญชีตาราง

ตาราง		หน้า
1	เปรียบเทียบความเที่ยงตรงก่อนและหลัง ค.ศ. 1985	31
2	ผลการวิเคราะห์พฤติกรรมที่บ่งชี้ทักษะกระบวนการทางวิทยาศาสตร์.....	73
3	ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าความแปรปรวนและค่าความเอนเอียงของ ข้อมูลจำแนกตามตัวแปร.....	83
4	ผลการประมาณค่าการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง.....	85
5	ผลการประมาณค่าการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง	88
6	เมทริกซ์แสดงความสัมพันธ์ระหว่างตัวแปรที่ใช้ในการประเมินทักษะ กระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ.....	90
7	ผลการวิเคราะห์ค่าความยากของทักษะกระบวนการทางวิทยาศาสตร์ ตามโมเดลของราล์ฟ แบบ 3 องค์ประกอบ (จำแนกตามทักษะ).....	91
8	ผลการวิเคราะห์ความสามารถของผู้ตรวจให้คะแนนทักษะกระบวนการ ทางวิทยาศาสตร์ ตามโมเดลของราล์ฟ แบบ 3 องค์ประกอบ (จำแนกตาม ผู้ตรวจ).....	93
9	ผลการวิเคราะห์ความสามารถของนักเรียนตามโมเดลของราล์ฟ แบบ 3 องค์ประกอบ (จำแนกตามนักเรียน).....	95
10	ค่าความเชื่อมั่นของแบบทดสอบวัดทักษะกระบวนการทางวิทยาศาสตร์ ขั้นบูรณาการ	100
11	ความสอดคล้องระหว่างคู่ของผู้ตรวจให้คะแนนคำนวณโดยสัมประสิทธิ์แคปปา	101
12	ค่าน้ำหนักองค์ประกอบและความคลาดเคลื่อนของโมเดลการวิเคราะห์ องค์ประกอบเชิงยืนยันอันดับหนึ่ง.....	102
13	ค่าน้ำหนักองค์ประกอบและความคลาดเคลื่อนของโมเดลการวิเคราะห์ องค์ประกอบเชิงยืนยันอันดับสอง	103
14	ดัชนีความสอดคล้องของการประเมินจากผู้เชี่ยวชาญในการ ตรวจสอบคุณภาพของเครื่องมือ (IOC).....	147
15	ค่าความยาก (p) ค่าอำนาจจำแนก (r_{xy}) ของแบบวัดทักษะกระบวนการ ทางวิทยาศาสตร์ขั้นบูรณาการ.....	148
16	ผลการวิเคราะห์ความสามารถของนักเรียนตามโมเดลของราล์ฟ แบบ 3 องค์ประกอบ (จัดเรียงตามความสามารถของนักเรียน)	149

บัญชีภาพประกอบ

ภาพประกอบ

หน้า

1	โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง	19
2	โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง	20
3	ลักษณะของเมตริกซ์หลายคุณลักษณะหลายวิธี	56
4	เมตริกซ์สหสัมพันธ์ของ 3 คุณลักษณะและ 3 วิธีการวัด	58
5	แสดงการจับคู่ในการตรวจให้คะแนนระหว่างนักเรียนและผู้ตรวจ	77
6	ผลการประมาณค่าสัมประสิทธิ์องค์ประกอบโมเดลการวิเคราะห์ องค์ประกอบเชิงยืนยันอันดับหนึ่ง	87
7	ผลการประมาณค่าสัมประสิทธิ์องค์ประกอบโมเดลการวิเคราะห์ องค์ประกอบเชิงยืนยันอันดับสอง	89
8	ค่าความยากของทักษะกระบวนการ และการเปรียบเทียบความเหมาะสม ของข้อมูลกับโมเดล (เฉพาะองค์ประกอบของทักษะกระบวนการ)	92
9	ค่าความสามารถของผู้ตรวจ และการเปรียบเทียบความเหมาะสม ของข้อมูลกับโมเดล (เฉพาะองค์ประกอบของผู้ตรวจ)	94
10	ค่าความสามารถของนักเรียนและการเปรียบเทียบความเหมาะสม ของข้อมูลกับโมเดล (เฉพาะองค์ประกอบของนักเรียน)	97
11	ค่าพารามิเตอร์ของแต่ละองค์ประกอบบน Logit Scale เดียวกัน	98
12	โค้งลักษณะลำดับขั้นของทักษะในการประเมินทักษะกระบวนการ ทางวิทยาศาสตร์ ที่มี 4 ระดับ ของฟาเซ็ทโมเดล	99
13	โค้งความน่าจะเป็นในการประเมินทักษะกระบวนการทางวิทยาศาสตร์ ของฟาเซ็ทโมเดล	99

การแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของคะแนนการประเมินทักษะกระบวนการทาง
วิทยาศาสตร์โดยใช้ทฤษฎีมาตรฐานเดิมและทฤษฎีการตอบข้อคำถาม
ตามโมเดลของราล์ฟแบบหลายองค์ประกอบ

บทคัดย่อ
ของ
ชาญวิทย์ จรัสสุทธิอิสร์

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาคุณวุฒิบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา
ตุลาคม 2550

ชาญวิทย จรัสสุทธิอิศร. (2550), การแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของคะแนนการประเมินทักษะกระบวนการทางวิทยาศาสตร์โดยใช้ทฤษฎีมาตรฐานเดิมและทฤษฎีการตอบข้อคำถามตามโมเดลของราล์ชแบบหลายองค์ประกอบ. ปรินซ์นิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. คณะกรรมการควบคุม: รองศาสตราจารย์ ดร.บุญเชิด ภิญโญนนตพงษ์, อาจารย์ ดร.สุพร เข้มเฮง, ผู้ช่วยศาสตราจารย์ ดร.สุรัช มีชาญ.

การศึกษานี้มีจุดมุ่งหมายสำคัญ เพื่อแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของคะแนนการประเมินทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการ 5 ทักษะประกอบด้วย การตั้งสมมติฐาน การกำหนดนิยามเชิงปฏิบัติการ การกำหนดและควบคุมตัวแปร การทดลอง และการตีความหมายข้อมูลและลงข้อสรุป การวิเคราะห์ตามทฤษฎีมาตรฐานเดิมทำการวิเคราะห์องค์ประกอบเชิงยืนยันด้วยโปรแกรม LISREL 8.54 ของ Jöreskog & Sörbom การวิเคราะห์ความเที่ยงตรงรวมและความเที่ยงตรงแยกด้วยโปรแกรม SPSS 10.0.1 และการวิเคราะห์ทฤษฎีการตอบข้อคำถาม ทำการวิเคราะห์โมเดลของราล์ชแบบหลายองค์ประกอบด้วยโปรแกรม FACETS ของ Linacre พร้อมทั้งแสดงหลักฐานความเชื่อมั่นของการตรวจให้คะแนน

กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 สังกัดเขตพื้นที่การศึกษานครราชสีมา เขต 3 จำนวน 280 คน สุ่มแบบสองขั้นตอน และผู้ตรวจให้คะแนนจำนวน 14 คน ที่สุ่มแบบเจาะจง

ผลการวิจัยสรุปได้ดังนี้

การวิเคราะห์ข้อมูลตามทฤษฎีมาตรฐานเดิม โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งและอันดับสองมีความเหมาะสมกับข้อมูลเชิงประจักษ์ทั้งสองกรณี (อันดับหนึ่ง $\chi^2 = 205.65$, $df = 199$, $P = 0.36$ และอันดับสอง $\chi^2 = 209.90$, $df = 204$, $P = 0.37$) และมีความเที่ยงตรงรวมซึ่งได้จากค่าสหสัมพันธ์ภายในทักษะสูงกว่าความเที่ยงตรงแยกซึ่งได้จากค่าสหสัมพันธ์ระหว่างทักษะ ส่วนทฤษฎีการตอบข้อคำถาม ค่าพารามิเตอร์ทั้งสามองค์ประกอบอยู่ในเกณฑ์ที่ยอมรับได้และโมเดลการวิจัยมีความเหมาะสมกับข้อมูลเชิงประจักษ์ทุกองค์ประกอบสำหรับความเชื่อมั่นของการให้คะแนน ค่าความเชื่อมั่นของผู้ตรวจทุกคนที่คำนวณจากสูตรของเฟลด์-ราซ มีค่าเท่ากับ .8921 ค่าความเชื่อมั่นระหว่างผู้ตรวจแต่ละคู่ที่คำนวณจากสูตรแคปปา มีค่าเฉลี่ยเท่ากับ .7785 และค่าความเชื่อมั่นเชิงโครงสร้างของการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งและอันดับสองเท่ากับ .9055 และ .9361 ตามลำดับ

ข้อค้นพบทั้งสองทฤษฎีดังกล่าวเป็นหลักฐานแสดงว่า คะแนนการประเมินทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการมีความเที่ยงตรงตามโครงสร้าง

Construct Validity Evidence of Scores from Assessment in Science Process Skill
Using Classical Theory and Many-Facet Rasch Model

AN ABSTRACT
BY
CHARNWIT JARUSSOOTTI-ISORN

Presented in Partial Fulfillment of the Requirements for the
Doctor of Education Degree in Testing and Measurement
at Srinakharinwirot University
October 2007

Charnwit Jarussootti-Isorn. (2007). *Construct Validity Evidence of Scores from Assessment in Science Process Skill Using Classical Theory and Many-Facet Rasch Model*. Dissertation, Ed.D. (Testing and Measurement). Bangkok: Graduate School, Srinakharinwirot University. Advisor Committee: Assoc. Prof. Dr. Boonchird Pinyoanuntapong, Dr. Suwaporn Semheng, Asst. Prof. Dr. Surachai Meechan.

The main purpose of this study was to display the construct validity evidence of scores from the assessment of five integrated science process skills comprising establishing hypothesis, determining practical definition, determining and controlling experimental variables, interpreting data, and making conclusion. With the analysis of classical test theory, the confirmatory factors were analyzed by using Jöreskog & Sörbom's LISREL 8.54 program. The convergent validity and discriminant validity were analyzed by SPSS 10.0.1 program. For the analysis of item response theory, many-facet Rasch model was analyzed by Linacre's FACETS program together with the display of reliability evidence of scoring.

The samples were 280 Mathayomsuksa 4 students under the jurisdiction of Nakhonratchasima Educational Service Zone 3 obtained by two-stage random sampling, and 14 raters obtained by purposive sampling.

The results of research were as follows.

Regarding the analysis of classical test theory, the first and second models of confirmatory factor analysis fitted with the empirical data in both cases ($\chi^2 = 205.65$, $df = 199$, and $P = 0.36$ for the first model, and $\chi^2 = 209.90$, $df = 204$, and $P = 0.37$ for the second model). The mutual validity from the correlation within skills was higher than the separate validity from the correlation among skills. With the item response theory, the parameters of all three factors met the acceptable criteria, and the research model fitted with the empirical data in all factors. For the reliability of scoring, the reliability of all raters calculated by Feldt-Raju formula was .8921. The reliability between each pair of raters calculated by kappa formula was .7785. The construct reliabilities of the first and second confirmatory factor analysis were .9055 and .9361 respectively.

The findings of both theories showed the construct validity evidence of scores from the assessment of integrated science process skills.

บทที่ 1

บทนำ

ภูมิหลัง

วิทยาศาสตร์ทำให้คนได้พัฒนาวิธีคิดที่เป็นเหตุผล มีความคิดสร้างสรรค์ คิดวิเคราะห์ วิจัย มีทักษะที่สำคัญในการค้นคว้าหาความรู้ มีความสามารถในการแก้ปัญหาอย่างเป็นระบบ ตัดสินใจบนข้อมูลหลากหลายและประจักษ์พยานที่สามารถตรวจสอบได้ วิทยาศาสตร์เป็นวัฒนธรรมของโลกสมัยใหม่ซึ่งเป็นสังคมแห่งความรู้ (Knowledge based society) ทุกคนจึงจำเป็นต้องได้รับการพัฒนาให้รู้วิทยาศาสตร์ (Scientific literacy for all) เพื่อจะได้มีความรู้ความเข้าใจโลกธรรมชาติและเทคโนโลยีที่มนุษย์สร้างขึ้นและนำความรู้ไปใช้อย่างมีเหตุผลสร้างสรรค์ มีคุณธรรม ความรู้วิทยาศาสตร์ไม่เพียงแต่นำมาใช้ในการพัฒนาคุณภาพชีวิตที่ดี แต่ยังช่วยให้คนมีความรู้ความเข้าใจที่ถูกต้องเกี่ยวกับการใช้ประโยชน์ การดูแลรักษา ตลอดจนการพัฒนาสิ่งแวดล้อมและทรัพยากรธรรมชาติอย่างสมดุลและยั่งยืน และที่สำคัญก็คือ ความรู้วิทยาศาสตร์ช่วยเพิ่มขีดความสามารถในการพัฒนาเศรษฐกิจ สามารถแข่งขันกับนานาประเทศ และดำเนินชีวิตอยู่ร่วมกันในสังคมโลกได้อย่างมีความสุข (กรมวิชาการ. 2545: 1)

หลักสูตรการศึกษาขั้นพื้นฐาน ตามพระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 มุ่งให้ผู้เรียนได้เรียนรู้วิทยาศาสตร์ที่เน้นกระบวนการนำไปสู่การสร้างองค์ความรู้ โดยให้ผู้เรียนมีส่วนร่วมในการเรียนทุกขั้นตอน ผู้เรียนได้ทำกิจกรรมที่หลากหลายทั้งเป็นกลุ่มและเป็นรายบุคคล ในการสังเกตสิ่งต่าง ๆ รอบตัว ตั้งคำถามหรือปัญหาเกี่ยวกับสิ่งที่จะศึกษา พัฒนาการกระบวนการคิดขั้นสูง มีการคิดวางแผนและลงมือปฏิบัติการสำรวจตรวจสอบด้วยกระบวนการที่หลากหลายจากแหล่งเรียนรู้ทั้งส่วนที่เป็นสากลและท้องถิ่น คิดและตัดสินใจเลือกข้อมูลที่เป็นประโยชน์เพื่อนำไปใช้ในการตอบคำถาม หรือแก้ปัญหา ซึ่งจะนำไปสู่องค์ความรู้แนวคิดหลักทางวิทยาศาสตร์แล้วสื่อสารสิ่งที่เรียนรู้หรือองค์ความรู้ในรูปแบบต่างๆ ให้ผู้อื่นรับรู้กระบวนการเรียนรู้ดังกล่าว ทำให้ผู้เรียนได้สร้างองค์ความรู้และเกิดการพัฒนาเจตคติ คุณธรรม จริยธรรม และค่านิยมที่ดีต่อวิชาวิทยาศาสตร์ (กรมวิชาการ. 2545: 4)

เนื่องจากวิชาวิทยาศาสตร์เป็นเรื่องของการเรียนรู้เกี่ยวกับธรรมชาติ โดยมนุษย์ใช้กระบวนการสังเกต สำรวจตรวจสอบ และการทดลองเกี่ยวกับปรากฏการณ์ทางธรรมชาติและนำผลมาจัดระบบ หลักการ แนวคิดและทฤษฎี ดังนั้นการเรียนการสอนวิทยาศาสตร์จึงมุ่งเน้นให้ผู้เรียนได้เป็นผู้เรียนรู้และค้นพบด้วยตนเองมากที่สุด (กรมวิชาการ. 2545: 14) มุ่งพัฒนาให้นักเรียนคิดเป็น ทำเป็นและแก้ปัญหาได้เป็นสำคัญ การที่ผู้เรียนจะเกิดคุณภาพได้ตามที่กล่าวอ้างนั้น จำเป็นจะต้องมีทักษะพื้นฐานสำคัญที่เรียกว่า ทักษะกระบวนการทางวิทยาศาสตร์ ที่ครอบคลุมความสามารถในการปฏิบัติ มีความรู้ความเข้าใจ มีความคิดพื้นฐานทางวิทยาศาสตร์

และมีเจตคติทางวิทยาศาสตร์ (ยุพา ตันติเจริญ. 2531: 41-42) การปลูกฝังเพื่อให้ผู้เรียนเกิดกระบวนการทางวิทยาศาสตร์จึงมีความจำเป็น เพราะทักษะกระบวนการทางวิทยาศาสตร์เป็นวิธีการที่จะทำให้บุคคลแสวงหาความรู้ใหม่ ๆ อยู่เสมอ ดังที่นักวิทยาศาสตร์หลายท่านได้กล่าวไว้ว่า การเรียนการสอนวิทยาศาสตร์ในปัจจุบันคือการใช้กระบวนการทางวิทยาศาสตร์ (The Process of Science) ค้นหาความรู้ทางวิทยาศาสตร์ได้ ด้วยตนเอง (The Body of Knowledge) ในการแก้ปัญหาต้องอาศัยวิธีการทางวิทยาศาสตร์ (Scientific Method) ทักษะกระบวนการทางวิทยาศาสตร์ (Scientific Process Skill) และเกิดเจตคติทางวิทยาศาสตร์ (Scientific Attitude) (วรรณิพา รอดแรงคำ; และ พิมพ์พันธ์ เตชะคุปต์. 2532: คำนำ) อนันต์ จันทะกู (2525: 4-5) กล่าวว่า ทักษะกระบวนการทางวิทยาศาสตร์ถือเป็นองค์ประกอบร่วมของการสืบเสาะหาความรู้ทางวิทยาศาสตร์ ในขณะเดียวกันก็สามารถนำมาใช้ในวิชาอื่น ๆ ได้อย่างกว้างขวาง และสามารถนำไปใช้ได้ภายหลังที่จบจากโรงเรียนไปแล้ว ดังนั้น ผู้เรียนในทุกระดับชั้น ควรได้รับการฝึกทักษะกระบวนการทางวิทยาศาสตร์ ให้เกิดการเรียนรู้จนกลายเป็นทักษะทางสติปัญญา (Intellectual Skills) ซึ่งเป็นสมรรถภาพพื้นฐานของการเรียนรู้ (จำนง พรายแย้มแซ. 2529: 39) จากผลการวิจัยที่ผ่านมา มีข้อสรุปที่ยืนยันได้ว่า การพัฒนาทักษะกระบวนการทางวิทยาศาสตร์ให้เกิดขึ้นกับตัวผู้เรียนเป็นสิ่งจำเป็นมาก ผู้เรียนสามารถนำทักษะกระบวนการทางวิทยาศาสตร์ไปใช้ได้เหมาะสมกับสิ่งที่ต้องการศึกษาหรือปัญหาที่ต้องการแก้ไข รวมทั้งแสวงหาความรู้ในสาขาวิชาอื่น ๆ อีกด้วย

ถึงแม้ทักษะกระบวนการทางวิทยาศาสตร์จะมีความจำเป็นมากเพียงใด แต่การจัดการเรียนการสอนที่ผ่านมาถือว่ายังไม่ค่อยจะสอดคล้องกับลักษณะของวิชาวิทยาศาสตร์เท่าใดนัก ครูจะมุ่งสอนให้ผู้เรียนจดจำเนื้อหาความรู้ ละเลยในส่วนของกระบวนการทางวิทยาศาสตร์ การให้นักเรียนได้มีส่วนร่วมในการเรียนการสอน หรือการศึกษา ค้นคว้าและลงมือปฏิบัติจริง เพื่อค้นหาคำตอบด้วยตนเอง จากรายงานผลการประเมินคุณภาพการศึกษาที่ผ่านมา พบว่า การประเมินผลด้านทักษะกระบวนการทางวิทยาศาสตร์ส่วนใหญ่ครูจะประเมินด้านความรู้มากกว่าการปฏิบัติ ซึ่งไม่สอดคล้องกับธรรมชาติของวิทยาศาสตร์ซึ่งเป็นวิชาที่ว่าด้วยกระบวนการสืบเสาะหาความรู้อันเกิดจากการฝึกฝนและปฏิบัติจริงโดยใช้กระบวนการทางวิทยาศาสตร์อย่างเป็นระบบ เป็นเหตุให้ไม่สามารถตัดสินได้ว่า “ผู้เรียนมีทักษะกระบวนการทางวิทยาศาสตร์อยู่ในระดับใดควรต้องปรับปรุงข้อบกพร่องของนักเรียนในทักษะใดบ้าง” (วัฒนชัย จันทะกู. 2538: 3; อ้างอิงจาก นิเชต สุนทรพิทักษ์. 2533: 31) การวัดผลของครูโดยทั่วไปนิยมวัดผลสัมฤทธิ์ทางการเรียนและตัดสินผลโดยใช้แบบทดสอบเลือกตอบ ซึ่งถือว่าไม่สอดคล้องกับธรรมชาติของวิทยาศาสตร์เป็นอย่างยิ่ง การเรียนการสอนวิชาวิทยาศาสตร์ควรเน้นทางด้านการปฏิบัติ เพื่อต้องการให้ผู้เรียนสามารถนำความรู้ไปใช้ในชีวิตประจำวันมากกว่าการท่องจำได้หรือเข้าใจตามทฤษฎีต่าง ๆ เท่านั้น ในการสอบวัดก็เช่นเดียวกัน ควรจะให้ครอบคลุมพฤติกรรมในด้านต่าง ๆ ของวิทยาศาสตร์ทั้งความรู้ ความคิด ความรู้สึก และการปฏิบัติ (วัฒนชัย จันทะกู. 2538: 3;

อ้างอิงจาก อนันต์ ศรีโสภา. 2520: 1) มิใช่วัดแต่เฉพาะพฤติกรรมด้านความรู้ ความคิดเท่านั้น แบบทดสอบเลือกตอบที่ใช้กันอยู่เป็นส่วนมากก็สร้างได้ง่ายกว่าและไม่ต้องผ่านกระบวนการที่ยุ่งยากและซับซ้อนมากนัก ความเชื่อที่ว่าผู้มีความรู้ในเรื่องนั้นแล้วก็จะสามารถปฏิบัติในเรื่องนั้น ๆ ได้ถือเป็นสิ่งที่ไม่ถูกต้อง (วัณชัย จันทวิมลกุล. 2538: 3; อ้างอิงจาก กมล สุตประเสริฐ. 2528: 41) การวัดภาคปฏิบัติจึงเป็นการวัดอีกแนวทางหนึ่งที่สำคัญสำหรับวิชาวิทยาศาสตร์ เพราะการเรียนรู้ในรายวิชานี้จำเป็นต้องเน้นให้มีการปฏิบัติ และต้องปฏิบัติจนเกิดความชำนาญ ดังนั้น การวัดผลและประเมินผลในวิชาวิทยาศาสตร์จึงควรยึดแนวทางการประเมินการเรียนรู้ที่เน้น ผู้เรียนเป็นสำคัญ ตามพระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 รูปแบบที่เหมาะสม คือ รูปแบบการประเมินการปฏิบัติ (Performance Assessment) และการประเมินตามสภาพจริง (บุญเชิด ภิญโญอนันตพงษ์. 2544: 85, 125) การประเมินการปฏิบัติหากประเมินโดยการทดสอบมีหลายรูปแบบที่ใช้ ได้แก่ การเขียนตอบ (Constructed-Response) ความเรียง การสาธิต การนำเสนอปากเปล่า บทบาทสมมติ นิทรรศการ การสังเกตโดยตรงและพอดโฟลีโอ การประเมินการปฏิบัติส่วนใหญ่ผู้เรียนจะเป็นผู้กำหนดคำตอบ และสามารถให้คะแนนได้หลายระดับ ซึ่งตรงกันข้ามกับการให้ผู้สอบตอบถูกหรือผิด (Muraki; Hombo & Lee. 2000: 325; citing Messick. 1996) ดังนั้น แบบทดสอบเขียนตอบจึงนับเป็นอีกแนวทางหนึ่งที่เหมาะสมนำมาใช้เป็นเครื่องมือในการประเมินภาคปฏิบัติ (บุญเชิด ภิญโญอนันตพงษ์. 2544: 87-88) เวียสม่า และเจอร์ส (Wiersma; & Jurs. 1985: 165) กล่าวว่า แบบทดสอบเขียนตอบช่วยให้นักเรียนสามารถแสดงความต้องการที่จะตอบได้อย่างมีโครงสร้างใช้ข้อมูลทีละเอียดยาวมากขึ้น ฮอบกินส์ และแอนเตส (Hopkins; & Antes. 1990: 231) กล่าวว่า แบบทดสอบเขียนตอบเป็นข้อสอบประเภทขยายความที่สมบูรณ์ที่สุด เป็นข้อสอบที่ต้องการเห็นพัฒนาการในการตอบคำถามของนักเรียน กายย์ (Gay. 1992: 224) ให้ความเห็นว่า แบบทดสอบเขียนตอบประกอบด้วยหัวข้อที่ต้องการให้นักเรียนเขียนตอบ ซึ่งโดยปกติแล้วจะต้องเป็นคำตอบที่ยาวจริง ๆ และแบบทดสอบเขียนตอบไม่เหมาะสมสำหรับวัดผลในระดับความรู้ความจำ แต่เหมาะสำหรับวัดผลในระดับสูง เพราะแบบทดสอบลักษณะนี้จะต้องมีการสังเคราะห์และการประเมินค่าเวอร์ต และ แอนนี่ (Ward; & Annie. 1999: 129) สรุปว่า แบบทดสอบเขียนตอบเป็นแบบทดสอบที่วัดทักษะขั้นสูงที่ยากที่จะสร้างด้วยแบบทดสอบแบบปรนัยได้ แบบทดสอบประเภทนี้ต้องการให้นักเรียนแสดงความรู้ กระบวนการเรียนรู้ การรวบรวมข้อมูล และนำเสนอความรู้ในรูปของเหตุผลที่ประติดประต่อกันได้ จากเหตุผลดังกล่าว ผู้วิจัยจึงได้สร้างและพัฒนาแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ที่มีลักษณะเป็นแบบเขียนตอบ เพื่อต้องการวัดผลสัมฤทธิ์ทางการเรียนวิชาวิทยาศาสตร์ในพฤติกรรมการเรียนรู้ขั้นสูง โดยให้นักเรียนได้แสดงความคิดของตนเองออกมาด้วยภาษาและถ้อยคำที่ต้องอาศัยการเรียบเรียงและรวบรวมความรู้มาตอบด้วยตนเอง เพื่อบ่งชี้ระดับคุณภาพของทักษะกระบวนการทางวิทยาศาสตร์ของนักเรียน และศึกษาคุณภาพของการตรวจให้คะแนนแบบวัดโดยมุ่งเน้นไปที่การแสดงหลักฐาน

ความเที่ยงตรงตามโครงสร้างและความเชื่อมั่นซึ่งถือเป็นส่วนที่สำคัญที่สุดของการพัฒนาคุณภาพของเครื่องมือทางการวัดผล

ความเที่ยงตรง (Validity) เป็นคุณลักษณะประการหนึ่งที่สำคัญอย่างยิ่งในการสร้างและพัฒนาคุณภาพของแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ ความเที่ยงตรงของการวัดเป็นลักษณะของการตัดสินใจแบบบูรณาการในระดับที่ต้องมีประจักษ์พยาน เหตุผลที่เพียงพอและเหมาะสม ความเที่ยงตรงจึงเป็นข้อสรุปที่ได้จากการคิดแบบอุปมาน (inductive) ตามหลักฐานที่มีอยู่ สิ่งที่มีอิทธิพลที่จะทำให้สรุปได้ว่ามีความเที่ยงตรงเกิดขึ้นหรือไม่นั้น มิใช่มีเพียงคุณสมบัติของแบบวัดหรือการสังเกตเท่านั้น เพราะประโยชน์ของแบบวัดอยู่ที่ความสามารถในการอ้างอิงคะแนนจากการทดสอบหรือการวัดอื่น ๆ ไปสู่คุณลักษณะอย่างแท้จริงของบุคคลนั้นได้ และการอ้างอิงขึ้นอยู่กับจุดมุ่งหมายของการวัด (Walsh; & Betz. 1990: 59) การศึกษาเกี่ยวกับความเที่ยงตรงมีสองแนวคิด คือ แนวคิดเดิมกับแนวคิดปัจจุบัน แนวคิดเดิมได้อธิบายความเที่ยงตรงว่า เป็นคุณสมบัติของเครื่องมือวัด ถ้าเครื่องมือวัดสามารถให้ข้อมูลหรือคะแนนได้ตรงจุดมุ่งหมายของการวัด แสดงว่า เครื่องมือวัดนั้นมีความเที่ยงตรง เครื่องมือวัดที่มีความเที่ยงตรงสูง จะสามารถวัดในสิ่งที่ต้องการวัดได้ถูกต้องครอบคลุมมากกว่าเครื่องมือวัดที่มีความเที่ยงตรงต่ำ ความเที่ยงตรงของเครื่องมือวัดจึงเป็นระดับของความสอดคล้องหรือความสัมพันธ์ระหว่างคะแนนที่ได้จากการวัดกับสิ่งที่ต้องการวัดนั่นเอง “สิ่งที่ต้องการวัด” ก็คือ “ตัวเกณฑ์” (Criterion) ดังนั้น “ความเที่ยงตรง” จึงเป็นความสัมพันธ์ระหว่างคะแนนที่ได้จากการทดสอบกับคะแนนที่ได้จากตัวแปรเกณฑ์ (บุญเชิด ภิญญอนันตพงษ์. 2545: 171)

เนื่องจากตัวแปรเกณฑ์เป็นตัวแปรที่ผู้สร้างเครื่องมือวัดต้องการวัดคุณลักษณะนั้น หรือต้องการพยากรณ์ไปถึง เช่น เนื้อหาวิชาที่กำหนดให้, ความสามารถในการปฏิบัติงานจริง, ความสามารถในการเรียนในอนาคตและความสำเร็จในชีวิต คุณลักษณะที่ต้องการวัดในปัจจุบันหรืออนาคตต่างก็ใช้เป็นตัวแปรเกณฑ์ได้ทั้งสิ้น ความสัมพันธ์ระหว่างค่าที่วัดได้จากเครื่องมือวัดที่สร้างขึ้นกับตัวแปรเกณฑ์ต่าง ๆ กันเหล่านี้ จึงมีความเที่ยงตรงหลายชนิด เช่น เมื่อใช้เนื้อหาสาระวิชาเป็นเกณฑ์ก็เรียกว่า ความเที่ยงตรงตามเนื้อหา (Content Validity) เมื่อนำค่าที่วัดได้มาสัมพันธ์กับผลสัมฤทธิ์ในอนาคตก็เรียกว่า ความเที่ยงตรงเชิงพยากรณ์ (Predictive Validity) เมื่อนำไปเปรียบเทียบกับโครงสร้างหรือลักษณะทางจิตวิทยาที่ต้องการวัด เช่น ความมีเหตุผล ก็เรียกว่า ความเที่ยงตรงตามโครงสร้าง (Construct Validity) ดังนั้น สรุปได้ว่า ความเที่ยงตรงทุกชนิดเป็นความเที่ยงตรงที่เกี่ยวข้องกับเกณฑ์ทั้งสิ้นแต่เรียกชื่อต่างกันไปตามเกณฑ์

เครื่องมือวัดที่มีความเที่ยงตรง หมายความว่า เครื่องมือวัดนั้นสามารถวัดในสิ่งที่ผู้วัดต้องการวัดได้อย่างถูกต้อง ดังนั้น แนวคิดเดิมจึงถือว่าความเที่ยงตรงเป็นคุณภาพของเครื่องมือวัด ซึ่งเป็นคุณสมบัติที่สำคัญของเครื่องมือทุกชนิด ไม่ว่าจะเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน แบบทดสอบวัดความถนัด มาตราวัดทัศนคติ จริยธรรม และบุคลิกภาพ เครื่องมือวัดทุกฉบับจะต้องมีคุณภาพด้านความเที่ยงตรง จึงจะเชื่อได้ว่าเป็นเครื่องมือวัดที่มีคุณภาพและ

ผลที่ได้จากการวัดจะถูกต้องตรงตามความต้องการ การวัดใด ๆ จะไร้ประโยชน์เมื่อเครื่องมือวัดนั้นขาดความเที่ยงตรง และความเที่ยงตรงของเครื่องมือวัด ยังแบ่งเป็นหลายชนิด ขึ้นอยู่กับตัวแปรเกณฑ์ที่ใช้ในการเปรียบเทียบ (บุญเชิด ภิญโญนนตพงษ์. 2545: 172)

ส่วนแนวคิดปัจจุบัน ได้ให้ความหมายของความเที่ยงตรงตามมาตรฐานทางเทคนิคสำหรับประเมินคุณภาพของเครื่องมือวัด โดยคณะกรรมการร่วมที่ประกอบด้วยสมาชิกของสามสมาคมคือ สมาคมวิจัยการศึกษาอเมริกัน (American Educational Research Association: AERA) สมาคมจิตวิทยาอเมริกัน (American Psychological Association: APA) และสภาการวัดผลการศึกษาแห่งชาติของอเมริกัน (National Council American on Measurement in Educational: NCME) ได้เสนอไว้ในหนังสือ “Standards for Educational and Psychological Testing” ที่พิมพ์ในปี 1985 ให้ความหมายความเที่ยงตรงไว้ว่า ความเที่ยงตรงเป็นลักษณะสำคัญที่สุดที่ใช้ในการประเมินเครื่องมือวัด ซึ่งเป็นเรื่องเกี่ยวกับการลงความเห็น (Inference) จากคะแนนที่วัดได้อย่างเหมาะสม อย่างมีความหมาย และให้ประโยชน์ ส่วนการตรวจสอบความเที่ยงตรง (Validation) เป็นกระบวนการในการรวบรวมพยานหลักฐาน (Evidence) เพื่อสนับสนุนการลงความเห็นดังกล่าว และเก็บรวบรวมหลักฐานได้หลายวิธี ความเที่ยงตรงจึงขึ้นอยู่กับปริมาณและชนิดของหลักฐานที่สามารถสนับสนุนการลงความเห็นของผู้วัดจากข้อมูลที่วัดได้ (บุญเชิด ภิญโญนนตพงษ์. 2545: 172) เมสสิก (Popham. 1990 : 94; citing Messick. 1989) ได้ให้ความเห็นว่า ความเที่ยงตรงของแบบทดสอบมิได้เกิดจากตัวแบบทดสอบ หากแต่เกิดขึ้นจากการแปลความหมายของข้อมูลที่ได้จากวิธีการเฉพาะอันหนึ่ง สรุปได้ว่า ความหมายของความเที่ยงตรงในระยะหลังได้เน้นไปที่การขยายการอ้างอิงไปสู่สิ่งที่แบบวัดต้องการจะวัด โดยผลที่ได้จากการวัดก็คือคะแนนที่เป็นพื้นฐานสำคัญของการอ้างอิง (จินตนา ธนวิบูลย์ชัย. 2535: 65)

กรอนลันและลินน์ (Gronlund; & Linn. 1990: 48-49) กล่าวถึงธรรมชาติของความเที่ยงตรงว่า เมื่อใช้คำว่าความเที่ยงตรงในแง่ของการทดสอบและการประเมินจะต้องมีข้อควรคำนึงดังต่อไปนี้คือ (1) ความเที่ยงตรง หมายถึง ความเหมาะสมของการแปลความหมายของผลการทดสอบหรือการประเมินเครื่องมือสำหรับกลุ่มใดกลุ่มหนึ่งไม่ได้เป็นการประเมินตัวเครื่องมือ บางครั้งเราอาจกล่าวว่า “ความเที่ยงตรงของแบบทดสอบ” ซึ่งที่จริงควรกล่าวว่า “ความเที่ยงตรงของการแปลความหมายที่เกิดขึ้นจากผลการประเมิน” จะถูกต้องมากกว่า (2) ความเที่ยงตรงเป็นขนาดของความเข้มข้น (Degree) ซึ่งไม่ได้ตอบคำถามว่ามีหรือไม่มี เราจึงไม่ควรจะกล่าวว่า เครื่องมือนี้มีความเที่ยงตรงหรือไม่มีความเที่ยงตรง แต่ควรกล่าวถึงในแง่ความหมายของระดับความเข้มข้น เช่น มีความเที่ยงตรงสูง มีความเที่ยงตรงปานกลางหรือมีความเที่ยงตรงต่ำ (3) ความเที่ยงตรงจะมีลักษณะเฉพาะเจาะจงในการแปลความหมาย ไม่มีแบบทดสอบใดที่มีความเที่ยงตรงในทุกวัตถุประสงค์ (4) การแสดงหลักฐานความเที่ยงตรงของแบบทดสอบมีหลายวิธี เนื่องจากความเที่ยงตรงมีหลายชนิด (5) ในปี 1995 ลินน์และกรอนลัน (Linn; & Gronlund.

1995: 49) เพิ่มข้อคำนึงเกี่ยวกับธรรมชาติของความเที่ยงตรงอีก คือ ความเที่ยงตรงเกี่ยวข้องกับการตัดสินใจเชิงประเมินทั้งหมด (Overall Evaluative Judgment) กล่าวคือ เป็นความต้องการประเมินในระดับของความเข้มข้นในการตีความ (Interpretation) และใช้ผลการประเมิน (Assessment) ว่า ได้รับการพิสูจน์ความถูกต้องจากการแสดงหลักฐานความเที่ยงตรง และหลักเหตุผลของการตีความ และการใช้งานดังกล่าว

ดังนั้นสรุปได้ว่า ความเที่ยงตรง หมายถึง ระดับของหลักฐานที่สามารถสนับสนุน การลงความเห็นจากข้อมูลที่วัดจากเครื่องมือวัด ซึ่งเป็นการลงความเห็นจากข้อมูลที่เกี่ยวข้องกับการใช้เครื่องมือวัดเฉพาะครั้งนั้น ความเที่ยงตรงจึงมีความหมายเดียว (Unitary Concept) แต่การลงความเห็นและการตรวจสอบความเที่ยงตรงอาจมีได้หลายชนิด แบ่งตามหลักฐานพยานอาจแบ่งได้เป็นสามชนิดตามแนวคิดแบบเก่า คือ หลักฐานความเที่ยงตรงตามเนื้อหา (Content Validity Evidence) หลักฐานแสดงความเที่ยงตรงตามเกณฑ์ (Criterion Validity Evidence) หลักฐานความเที่ยงตรงตามโครงสร้าง (Construct Validity Evidence) เป็นต้น

หลักฐานแสดงความเที่ยงตรงตามเนื้อหา หมายถึง การลงความเห็นเกี่ยวกับความสอดคล้องระหว่างเนื้อหาของเครื่องมือวัดกับเนื้อหาของเกณฑ์ ถ้ามีความสอดคล้องกันมาก แสดงว่า มีหลักฐานแสดงความเที่ยงตรงตามเนื้อหาสูง เนื้อหาของเครื่องมือวัดจะครอบคลุมเนื้อหาที่เป็นตัวเกณฑ์ เครื่องมือวัดที่แสดงความเที่ยงตรงตามเนื้อหาสูง เครื่องมือวัดนั้นจะต้องมีเนื้อหา ข้อคำถามหรือเนื้อหาที่นำมาสร้างเป็นข้อคำถามแต่ละข้อครอบคลุมเนื้อหาอย่างครบถ้วน ดังนั้น จึงกล่าวได้ว่า หลักฐานแสดงความเที่ยงตรงตามเนื้อหาเป็นหลักฐานที่แสดงให้เห็นการสุมเนื้อหามาสร้างเครื่องมือวัดว่า เนื้อหาของเครื่องมือวัดที่สุมเลือกมานั้นเป็นตัวอย่างที่เป็นตัวแทนที่ดีของตัวเกณฑ์ หรือกล่าวอีกอย่างหนึ่งได้ว่า เนื้อหาของเครื่องมือวัดเป็นตัวแทนของมวลประชากรของเนื้อหาหรือประชากรพฤติกรรมที่ต้องการวัดเพียงไร ทั้งนี้เพราะเครื่องมือวัดแต่ละฉบับไม่สามารถวัดประชากรความรู้ตามเนื้อหาได้ทั้งหมด ตัวเครื่องมือวัดเป็นเพียงกลุ่มตัวอย่างของมวลประชากรความรู้เท่านั้น วิธีการแสดงหลักฐานความเที่ยงตรงตามเนื้อหาเป็นการแสดงหรือหาว่าเครื่องมือวัดนั้นสามารถวัดได้ตรงและครอบคลุมเนื้อหาวิชาได้มากหรือน้อยเพียงใด การแสดงหลักฐานความเที่ยงตรงตามเนื้อหานี้จะใช้ดุลยพินิจของผู้เชี่ยวชาญตรวจสอบความเที่ยงตรงว่า ข้อสอบมีความเที่ยงตรงมากน้อยเพียงใด โดยการเทียบกับตารางวิเคราะห์หลักสูตร หรือตารางกำหนดข้อสอบ (Test Blueprint) ซึ่งกำหนดตัวอย่างหัวข้อเนื้อหาสาระวิชาและพฤติกรรมจากเนื้อหาสาระวิชาทั้งหมด จินตนา ธนวิบูลย์ชัย (2535: 71) กล่าวว่า การพิจารณาโดยอาศัยหลักฐานผลการตัดสินของผู้เชี่ยวชาญในคุณลักษณะที่ต้องการจะวัดนั้น ถือได้ว่าเป็นหลักฐานที่มีความเป็นอัตนัยมาก การสร้างและคัดเลือกสิ่งบ่งชี้หรือข้อคำถามต่าง ๆ จะต้องมีวิธีการที่กำหนดขึ้นโดยเฉพาะซึ่งมีความเป็นระบบ มีความชัดเจน มีการวางแผนอย่างละเอียด และใช้ความคิดอย่างรอบคอบ ตลอดจนมีข้อมูลเชิงประจักษ์ที่ช่วยสนับสนุนให้เห็นว่า สิ่งบ่งชี้เหล่านี้อยู่ภายใต้สังกัดที่นิยามเอาไว้ อย่างไรก็ตามหลักฐานที่จำเป็นอย่างยิ่งสำหรับ

ความเที่ยงตรงประเภทนี้คือ ความรอบคอบในการสร้างและคัดเลือกสิ่งบ่งชี้หรือข้อคำถาม ความรอบคอบในการตัดสินใจของผู้เชี่ยวชาญ และวิธีการที่ใช้ในการตรวจสอบโครงสร้างทางจิตวิทยาที่กล่าวมาแล้ว (Walsh; & Betz. 1990: 61) สิ่งที่น่าพึงระวังไว้เสมอก็คือข้อสรุปที่ได้จากการตรวจสอบความเที่ยงตรงนี้จะมีลักษณะเป็นเพียงว่าแบบทดสอบที่สร้างขึ้นนั้นดูเหมือนว่าจะวัดคุณลักษณะที่ต้องการวัดเท่านั้น ยังสรุปไม่ได้ว่าสามารถวัดคุณลักษณะนั้นได้อย่างแท้จริง แต่การแสดงผลหลักฐานความเที่ยงตรงของหลักฐานตามเนื้อหานี้นับได้ว่าควรเป็นขั้นตอนแรกที่สำคัญของการหาความเที่ยงตรงชนิดอื่น ๆ

หลักฐานแสดงความเที่ยงตรงตามเกณฑ์ของเครื่องมือวัด เป็นการหาความสัมพันธ์ระหว่างคะแนนผลการวัดจากเครื่องมือที่สร้างกับค่าที่วัดได้จากเกณฑ์ภายนอก ถ้าเกณฑ์ภายนอกนั้นเป็นเกณฑ์ที่วัดได้ในปัจจุบัน ก็จะเรียกว่า หลักฐานแสดงความเที่ยงตรงตามสภาพ (Concurrent Validity Evidence) ถ้าเป็นเกณฑ์ที่เก็บรวบรวมได้ในอนาคต ก็เรียกว่า หลักฐานแสดงความเที่ยงตรงตามพยากรณ์ (Predictive Validity Evidence) การแสดงผลหลักฐานความเที่ยงตรงทั้งสองชนิดนี้ เป็นการหาความสัมพันธ์ระหว่างคะแนนของเครื่องมือวัดที่สร้างขึ้นกับค่าที่วัดได้จากตัวเกณฑ์ (บุญเชิด ภิญโญนันตพงษ์. 2545: 174) วิธีการแสดงผลหลักฐานความเที่ยงตรงตามเกณฑ์ ขึ้นอยู่กับข้อมูลเชิงประจักษ์ที่ได้จากแนวทางใหญ่ ๆ 2 แนวทาง คือ การพิจารณาระดับความสัมพันธ์ระหว่างคะแนนที่ได้จากการวัดคุณลักษณะที่สนใจกับคะแนนเกณฑ์ ส่วนอีกแนวทางหนึ่งนั้นพิจารณาจากความแตกต่างระหว่างกลุ่มของข้อมูล ข้อมูลเชิงประจักษ์ที่ได้มาเช่นนี้ความสนใจไม่ได้อยู่ที่ความสัมพันธ์ระหว่างคะแนนสอบกับผลการวัดต่าง ๆ หลาย ๆ ตัวซึ่งเป็นโครงสร้างภายนอกที่เกี่ยวข้องกับคุณลักษณะที่วัด แต่ค่อนข้างจะสนใจแค่ว่าความสัมพันธ์ระหว่างคะแนนสอบกับเกณฑ์บางอย่างที่มีจุดมุ่งหมายในการนำไปใช้เฉพาะด้าน (จินตนา ธนวิบูลย์ชัย. 2535: 77) การหาหลักฐานข้อมูลเชิงประจักษ์ทั้ง 2 แนวทางดังกล่าวได้แก่ (1) การหาความสัมพันธ์ เป็นการหาค่าความสัมพันธ์ของค่าที่วัดได้จากเครื่องมือวัดที่ต้องการกับค่าที่วัดได้จากเกณฑ์ ส่วนมากนิยมใช้สหสัมพันธ์ของเพียร์สัน (Pearson Product Moment Correlation) วิเคราะห์ข้อมูลในกรณีที่ระหว่างตัวแปรสองตัวหากความสัมพันธ์นั้นมีข้อตกลงเบื้องต้นว่าสัมพันธ์กันเป็นเส้นตรง แต่ถ้าเมื่อใดมีความสัมพันธ์เป็นเส้นโค้ง ก็ควรใช้สัมประสิทธิ์ เอต้า (Eta Coefficient) เป็นตัวบ่งชี้ความสัมพันธ์ (2) การพิจารณาความแตกต่างระหว่างกลุ่มข้อมูล ในลักษณะนี้จะพิจารณาว่าคะแนนสอบที่ได้นั้นสามารถแยกแยะกลุ่มออกได้ตามเกณฑ์ที่กำหนดหรือไม่ กล่าวคือ ถ้าจะใช้คะแนนสอบในการพยากรณ์ความสำเร็จหรือความล้มเหลวในการศึกษาของบุคคล แบบทดสอบที่มีความเที่ยงตรงตามเกณฑ์จะต้องแยกกลุ่มคนที่สามารถสำเร็จการศึกษาและคนที่ล้มเหลวได้ตามที่พยากรณ์ไว้ โดยทั่วไปในการคำนวณหาค่าเฉลี่ยของคะแนนแบบทดสอบภายในกลุ่มต่าง ๆ แล้วนำค่าเฉลี่ยมาเปรียบเทียบกันนั้น จะใช้ t-test หรือ ANOVA ในการสรุปการเปรียบเทียบความแตกต่างอย่างมีนัยสำคัญ ข้อมูลที่ได้จากการเปรียบเทียบความแตกต่างนี้สามารถนำมาใช้ในการหาความเที่ยงตรงเชิงสภาพและเชิง

ทำนายได้เช่นเดียวกับการหาความสัมพันธ์ บุญเชิด ภิญญอนันตพงษ์ (2545: 176) กล่าวว่า มักมีข้อโต้แย้งอยู่เสมอในการเลือกใช้เกณฑ์เพื่อพิจารณาความสำเร็จเนื่องจากตัวพยากรณ์ แต่ละตัวมีสหสัมพันธ์กับคะแนนเกณฑ์แตกต่างกัน ทำให้การแสดงผลฐานความเที่ยงตรงตามพยากรณ์มีความแตกต่างกันและไม่มีการแสดงผลฐานความเที่ยงตรงตามพยากรณ์ได้ดีที่สุด ดังนั้นจำเป็นต้องใช้การสะสมการแสดงผลฐานความเที่ยงตรงสูง จึงจะแปลความหมายได้ด้วย ความมั่นใจและนำไปสู่การตัดสินใจที่มีประสิทธิภาพสูง ส่วนการแสดงผลฐานความเที่ยงตรงตามสภาพ ก็มักมีปัญหาในแง่การแปลความหมาย เพราะดัชนีที่ใช้วัดระดับความสัมพันธ์เป็นค่าสหสัมพันธ์ระหว่างตัวแปรที่ใช้วัดกับตัวเกณฑ์ ถ้าตัวเกณฑ์ที่วัดได้นั้นเป็น “ตัวเกณฑ์ที่แท้จริง” ค่าสหสัมพันธ์ที่ได้จะแปลความหมายได้มั่นใจว่าตรงกับเกณฑ์นั้นเพียงไร แต่ถ้าค่าวัดที่แทนตัวเกณฑ์นั้นมีความเที่ยงตรงต่ำกล่าวคือ ค่าวัดตัวเกณฑ์ที่ได้มาไม่สามารถเป็นตัวแทนที่ดีของตัวเกณฑ์แล้ว ก็ยากที่จะแปลความหมายของความเที่ยงตรงนี้ได้ ค่าความเที่ยงตรงตามสภาพนี้ จะมีความหมายดีก็ต่อเมื่อ ค่าวัดทั้งคู่เป็นตัวแทนที่ดีของตัวแปรนั้น

หลักฐานแสดงความเที่ยงตรงตามโครงสร้าง เป็นการแสดงผลฐานความเที่ยงตรงว่าเครื่องมือวัดนั้นสามารถวัดภายใต้ขอบเขต ความหมาย หรือคุณลักษณะประจำตามโครงสร้างทางทฤษฎีที่สมมติขึ้นนั้นได้เพียงใด คำว่า “โครงสร้าง” (Construct) ในที่นี้หมายถึง ตัวประกอบ หรือองค์ประกอบ (Factor) จึงเรียกความเที่ยงตรงตามโครงสร้างนี้อีกอย่างหนึ่งว่า “หลักฐานแสดงความเที่ยงตรงตามองค์ประกอบ” แต่ละโครงสร้างต้องเกี่ยวกับทฤษฎี ซึ่งอธิบายและพยากรณ์พฤติกรรมของมนุษย์ การแสดงผลฐานความเที่ยงตรงตามโครงสร้าง จึงเป็นการหาความสอดคล้องระหว่างองค์ประกอบที่วัดในเครื่องมือวัด และองค์ประกอบที่ต้องการวัดว่า วัดในองค์ประกอบเดียวกันหรือไม่ นั่นคือ วัดลักษณะทางจิตวิทยาหรือคุณสมบัติตามที่ต้องการหรือไม่ และปริมาณที่วัดแต่ละองค์ประกอบเป็นสัดส่วนสอดคล้องกับที่ต้องการหรือไม่ (บุญเชิด ภิญญอนันตพงษ์. 2545: 177) ความเที่ยงตรงชนิดนี้จะเกี่ยวกับความสัมพันธ์ระหว่างคะแนนสอบกับการทำนายคุณลักษณะตามทฤษฎีบางอย่าง ที่ตอบคำถามว่าแบบทดสอบนั้นๆ สามารถวัดได้ในสิ่งที่ต้องการจะวัดหรือไม่นั้น ความเที่ยงตรงตามโครงสร้างจะสามารถตอบคำถามนี้ได้มากกว่าความเที่ยงตรง 2 ชนิดแรกที่กล่าวมาแล้ว การทำความเข้าใจเกี่ยวกับหลักฐานความเที่ยงตรงตามโครงสร้างนี้มีความจำเป็นอย่างยิ่งที่จะต้องระลึกไว้เสมอว่า ตัวแบบทดสอบไม่ใช่สิ่งที่เราสนใจ แต่ความสนใจจะอยู่ที่แบบทดสอบได้สะท้อนปรากฏการณ์ในโลกของความเป็นจริงบางประการ หรือคุณสมบัติหรือพฤติกรรมบางอย่างที่มีความหมายและมีความสำคัญต่อปรากฏการณ์หรือคุณสมบัติหรือพฤติกรรมอื่น ๆ นั้นได้หรือไม่ เพราะปรากฏการณ์หรือคุณสมบัติหรือพฤติกรรมต่าง ๆ มีลักษณะที่เป็นระเบียบแบบแผนและมีความเป็นนามธรรม ไม่สามารถสังเกตได้โดยตรงด้วยประสาทสัมผัสทั้ง 5 แต่ต้องอาศัยการอ้างอิงหรือสันนิษฐานเอาจากปรากฏการณ์อื่น ๆ ที่สังเกตได้ สิ่งนี้เองเราเรียกว่า Construct การวัด Construct ทางจิตวิทยานับเป็นเรื่องที่ยาก เพราะ Construct เกิดขึ้นจากกระบวนการทางจิตซึ่งอยู่ภายในตัวบุคคล การวัด Construct จึง

กระทำโดยอ้อมเพื่ออธิบายบุคคลโดยยึดตามทฤษฎีคุณลักษณะนั้น(จินตนา ธนวิบูลย์ชัย. 2535: 80) บุญเชิด ภิญโญนันตพงษ์. (2545: 178) กล่าวว่า หลักฐานความเที่ยงตรงตามโครงสร้างมีลักษณะคล้ายกับหลักฐานความเที่ยงตรงตามเนื้อหาและหลักฐานความเที่ยงตรงตามเกณฑ์ แต่แตกต่างกันตรงที่ หลักฐานความเที่ยงตรงตามโครงสร้างจะต้องสะสมรายละเอียดจากหลาย ๆ แหล่งที่แตกต่างกันมากกว่า เพื่อนำมารวมเป็นคำบรรยายพฤติกรรมได้อย่างกว้างขวางกว่า คงทนกว่า และมีลักษณะความเป็นนามธรรมมากกว่า การกำหนดเกณฑ์เพียงเกณฑ์เดียว จึงไม่เพียงพอที่จะอธิบายคุณลักษณะของโครงสร้างได้ จึงต้องอาศัยรายละเอียดที่มากกว่าหรือต้องอาศัยเกณฑ์หลายเกณฑ์ ถ้าต้องการสร้างเครื่องมือวัดให้มีหลักฐานแสดงความเที่ยงตรงตามโครงสร้าง จะต้องให้คำนิยามปฏิบัติการ (Operational Definition) ของคำนั้นโดยอาศัยทฤษฎีใด ทฤษฎีหนึ่งเป็นหลัก แล้วจึงสร้างเครื่องมือวัดสิ่งนั้นตามคำนิยามปฏิบัติการ ความหมายของ คำว่า นิยามปฏิบัติการ ประกอบด้วยสามส่วน คือ คุณลักษณะ (Trait) ที่ต้องการวัด สิ่งเร้าหรือสถานการณ์ที่จะกระตุ้นให้คุณลักษณะนั้นแสดงออกมา และการตอบสนองแสดงออกที่สามารถสังเกตเห็นได้ เมื่อสามารถให้คำนิยามปฏิบัติการในสิ่งที่ต้องการจะทดสอบได้แล้ว จึงสร้างข้อคำถามขึ้นตามนิยามปฏิบัติการที่ให้มานั้น และนำผลจากการทดสอบไปคำนวณความเที่ยงตรง เพื่อแสดงหลักฐานความเที่ยงตรงตามโครงสร้างต่อไป

การแสดงผลฐานความเที่ยงตรงตามโครงสร้าง เป็นการแสดงผลฐานความเที่ยงตรงว่า เครื่องมือวัดนั้นสามารถวัดขอบเขต ความหมาย หรือคุณลักษณะประจำตามโครงสร้างทางทฤษฎีที่สมมติขึ้นนั้นได้เพียงใด การแสดงผลฐานความเที่ยงตรงตามโครงสร้างสามารถทำได้หลายวิธีการ ได้แก่ วิธีพิจารณาเทียบกับโครงสร้างที่กำหนด วิธีเปรียบเทียบจากกลุ่มที่ต่างกัน วิธีเทียบกับเครื่องมือมาตรฐานที่วัดคุณลักษณะเดียวกัน วิธีหาค่าความสอดคล้องภายในเครื่องมือวัด วิธีวิเคราะห์องค์ประกอบ (Factor Analysis) และวิธีวิเคราะห์หลายลักษณะหลายวิธี (Multitrait-Multimethods) เป็นต้น

การแสดงผลฐานความเที่ยงตรงที่ดีที่สุดควรต้องแสดงให้เห็นถึงความเที่ยงตรงทั้ง 3 ประเภท อย่างไรก็ตาม ในแง่ของการปฏิบัติแล้วไม่มีความจำเป็นที่จะต้องแสดงผลฐานให้ครบทั้ง 3 ประเภท ขึ้นกับคุณลักษณะของแบบทดสอบที่ต้องการหลักฐานมายืนยันความเที่ยงตรงโดยทั่วไปแล้วความเที่ยงตรงเชิงเนื้อหาจะยืนยันขึ้น สำหรับการแสดงผลฐานความเที่ยงตรงตามโครงสร้างจะแสดงเพื่อตอบคำถามเกี่ยวกับความสามารถของแบบทดสอบในการแปลความหมายของคุณลักษณะที่วัดได้ดีเพียงใด ความเที่ยงตรงตามเกณฑ์จะใช้เมื่อแบบทดสอบนั้นต้องการหลักฐานแสดงว่า แบบทดสอบสามารถพยากรณ์ความสามารถของการทำงานในอนาคตได้ดีเพียงใดหรือประมาณค่าความสามารถในปัจจุบันจากการวัดแบบอื่น ๆ ที่เป็นเกณฑ์ได้ดีเพียงใด (Linn; & Gronlund. 1995: 50-51)

เนื่องจากธรรมชาติของความเที่ยงตรงมีความเฉพาะเจาะจงในการแปลความหมายผลการวัด ไม่มีแบบทดสอบใดมีความเที่ยงตรงในทุกวัตถุประสงค์ (Gronlund; & Linn. 1990) ในการศึกษาครั้งนี้ ผู้วิจัยจึงเลือกศึกษาเฉพาะกรณีการแสดงผลหลักฐานความเที่ยงตรงตามโครงสร้างเป็นประเด็นหลักในการศึกษา โดยการยืนยันองค์ประกอบทักษะกระบวนการทางวิทยาศาสตร์ที่ประยุกต์ใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน และการวิเคราะห์ความเที่ยงตรงรวมและความเที่ยงตรงแยกตามแนวทางของทฤษฎีมาตรฐานเดิม (Classical) ซึ่งทั้งสองวิธีนี้จัดเป็นวิธีการแสดงผลหลักฐานความเที่ยงตรงตามโครงสร้างที่มีความน่าเชื่อถือในการแสดงผลหลักฐาน และมี ผู้ทำการวิจัยเป็นจำนวนมาก

ด้วยเหตุที่ความเที่ยงตรงเป็นระบบหนึ่งของเรื่องราวหรือเหตุการณ์ที่เป็นหลักฐาน แต่ไม่ใช่เหตุการณ์ทั้งหมด นอกจากนี้ความเที่ยงตรงที่ได้จากพยานหลักฐานนั้น ๆ จะไม่มีความคงทนเพียงพอ เมื่อเวลาผ่านไปเหตุการณ์ที่เป็นพยานหลักฐานก็สามารถเปลี่ยนแปลงไปตามเงื่อนไขทางสังคมที่เปลี่ยนแปลงไป หรือเมื่อมีข้อค้นพบใหม่ ๆ เกิดขึ้น ความเที่ยงตรงจึงเป็นคุณสมบัติที่มีการพัฒนาและเป็นกระบวนการต่อเนื่อง ด้วยเหตุที่หลักฐานที่มีอยู่มากไม่ค่อยมีความสมบูรณ์ จึงมีความสำคัญยิ่งที่ในการหาความเที่ยงตรงนั้นจะต้องพยายามทำให้เกิดความเป็นเหตุเป็นผลให้มากที่สุด เพียงพอที่จะนำมาใช้ได้อย่างเป็นปัจจุบันทั้งในการทดสอบและการทำวิจัยต่าง ๆ (Messick. 1989: 13) ความเป็นเหตุเป็นผลนั้นจะเป็นสิ่งที่เชื่อถือได้หรือไม่ จะขึ้นอยู่กับว่าได้พยายามใช้วิธีการทางวิทยาศาสตร์เข้ามาช่วยในการแปลความหมายของคะแนนสามารถอ้างอิงไปสู่จุดมุ่งหมายที่ต้องการได้มากน้อยเพียงใด จากผลการตรวจสอบความเป็นเหตุเป็นผลโดยใช้หลักฐานอ้างอิงในการทดสอบแบบต่าง ๆ อาจต้องการหลักฐานที่แตกต่างกัน หลักฐานนี้ย่อมหาได้จากข้อมูลหรือข้อเท็จจริงต่าง ๆ ที่แสดงความเชื่อมโยงขององค์ประกอบต่าง ๆ ที่จะช่วยให้ข้อสันเทศในการพิจารณาตัดสินใจนำไปใช้ในการอ้างอิง สังเกตได้ว่าข้อมูลนี้มีข้อสันเทศแต่เป็นสิ่งที่ทำให้เกิดข้อสันเทศ เพราะข้อสันเทศ ได้จากการแปลความหมายของข้อมูลนั่นเอง (จินตนา ธนวิบูลย์ชัย. 2535: 64) สอดคล้องกับบุญเชิด ภิญโญอนันตพงษ์. (2545: 178) ที่กล่าวว่า หลักฐานความเที่ยงตรงตามโครงสร้างจะต้องสะสมรายละเอียดจากหลาย ๆ แหล่งที่แตกต่างกันมากกว่า เพื่อนำมารวมเป็นคำบรรยายพฤติกรรมได้อย่างกว้างขวางกว่าคงทนกว่า และมีลักษณะความเป็นนามธรรมมากกว่า การกำหนดเกณฑ์เพียงเกณฑ์เดียว จึงไม่เพียงพอที่จะอธิบายคุณลักษณะของโครงสร้างได้ จึงต้องอาศัยรายละเอียดที่มากกว่าหรือต้องอาศัยเกณฑ์หลายเกณฑ์

เพื่อเป็นการสะสมรายละเอียดจากหลาย ๆ แหล่งที่แตกต่างกันที่จะนำมาใช้ในการแสดงผลหลักฐานความเที่ยงตรงตามโครงสร้างให้เพียงพอที่จะอธิบายคุณลักษณะของโครงสร้างได้ จึงควรมีการศึกษาเพิ่มเติมในอีกทฤษฎีที่มีความแตกต่างกัน นอกเหนือจากทฤษฎีมาตรฐานเดิม (Classical) ผู้วิจัยจึงสนใจที่จะเลือกทฤษฎีการตอบข้อคำถาม (Item Respond Theory; IRT) เป็นแนวทางในการศึกษาเพื่อแสดงผลหลักฐานความเที่ยงตรงตามโครงสร้างของคะแนนในการ

ประเมินทักษะกระบวนการทางวิทยาศาสตร์เป็นกรณีศึกษาเพิ่มเติม จุดเด่นของทฤษฎีนี้ก็คือสามารถประมาณค่าความสามารถหรือคุณลักษณะของผู้ตอบเป็นอิสระจากชุดของข้อคำถามหรือข้อความที่เลือกใช้ และไม่ว่าจะเลือกกลุ่มตัวอย่างย่อย ๆ กลุ่มใดมาใช้ในการประมาณค่าพารามิเตอร์ของข้อความ ค่าประมาณเหล่านั้นจะมีค่าคงที่ ซึ่งคุณสมบัติข้อนี้เป็นประโยชน์อย่างมากสำหรับการนำมาใช้ในการสร้างและพัฒนาแบบทดสอบในปัจจุบัน (ผจงจิต อินทสุวรรณ. 2525: 66-67) ทฤษฎีการตอบข้อคำถามเป็นทฤษฎีที่เกี่ยวกับการใช้โมเดลทางคณิตศาสตร์เพื่ออธิบายความสัมพันธ์ระหว่างระดับความสามารถ (Ability) หรือคุณลักษณะภายในตัวบุคคล (Trait) กับพฤติกรรมการตอบสนองต่อข้อคำถามหรือข้อความ (Performance) โมเดลที่อธิบายความสัมพันธ์ของทฤษฎีนี้มีอยู่หลายโมเดลด้วยกัน แต่โมเดลที่มีผู้นิยมนำมาใช้มากในปัจจุบัน ได้แก่ ราสซัส โมเดล (Rasch Model; Rasch. 1960) เพราะโมเดลนี้มีจำนวนพารามิเตอร์ที่ต้องประมาณค่าน้อยกว่าโมเดลอื่น ๆ ทำให้การวิเคราะห์ง่าย ไม่ยุ่งยากซับซ้อน (Hambleton. 1989: 157) โดยเฉพาะอย่างยิ่ง ถ้ากลุ่มตัวอย่างที่ใช้ในการวิเคราะห์มีจำนวนไม่มากนัก สามารถใช้เครื่องคิดเลขคำนวณด้วยมือ (Wright; & Stone. 1979: 28-45) นอกจากนี้ ราสซัส โมเดล ยังสามารถประมาณค่าพารามิเตอร์ให้ค่าถูกต้องกว่าโมเดลแบบอื่น ๆ เมื่อนำไปใช้กับกลุ่มตัวอย่างขนาดเล็ก (van de Vijver. 1986: 48; & Lord. 1980: 190) ด้วยเหตุผลดังกล่าว จึงทำให้มีการนำโมเดลนี้มาใช้ในการประมาณค่าพารามิเตอร์ได้อย่างมีประสิทธิภาพ ประหยัดเวลาและค่าใช้จ่ายในการวิเคราะห์ข้อมูล ในระยะที่ผ่านมาได้มีการนำราสซัส โมเดลมาประยุกต์ใช้กับการวัดพฤติกรรมทางด้านพุทธิพิสัย ในแบบทดสอบที่มีการให้คะแนนเป็นสองค่า (Dichotomous) ทั้งที่การวัดทางพฤติกรรมศาสตร์มิได้จำกัดอยู่เฉพาะการวัดความสามารถทางสมองประการเดียวเท่านั้น แต่ยังรวมไปถึงการวัดเจตคติ การวัดบุคลิกภาพ หรือการวัดพฤติกรรมด้านอื่น ๆ ที่ให้คะแนนเป็นแบบหลายค่า (Polytomous) ด้วยเหตุนี้จึงทำให้ในระยะต่อมา นักวัดผลได้ปรับขยายโมเดลต่าง ๆ ที่พัฒนาขึ้นในระยะแรก เพื่อนำไปใช้วัดพฤติกรรมทางด้านคุณลักษณะและทักษะการปฏิบัติ (Performance) เพิ่มขึ้นด้วย อาทิ โมเดลที่พัฒนาขึ้นจากราสซัส โมเดล เช่น เรทติ้ง สเกล โมเดล (Rating Scale Model) ของแอนดริช (Andrich. 1978) พาเชียล เครดิต โมเดล (Partial Credit Model) ของมาสเตอร์ (Masters. 1980) และนอกจากนี้ยังมี เกรด เรสพอน โมเดล (Grade Response Model) ของซามิไจมา (Samejima. 1969) ซึ่งพัฒนามาจากโลจิสติก โมเดล ที่ใช้พารามิเตอร์สองตัว (Two – Logistic Model ;2PL) ของเบิร์นบอม (Birnbaum. 1968) เป็นต้น

ต่อมา ลินาเคอร์ (Linacre. 1989) ได้พัฒนาโมเดลที่มีความแตกต่างจากโมเดลของทฤษฎีการตอบข้อคำถามแบบทั่วไป คือมีการขยายแนวคิดโมเดลของทฤษฎีการตอบข้อคำถามจากที่ให้ค่าพารามิเตอร์แค่ 2 ตัว คือ พารามิเตอร์ของข้อคำถาม (Item Parameter) ที่ประกอบด้วย ค่าความยาก ค่าอำนาจจำแนก และค่าการเดา กับ พารามิเตอร์ของผู้เข้าสอบ (Examinee Parameter) ซึ่งได้แก่ ค่าความสามารถของผู้ตอบ โมเดลที่ลินาเคอร์พัฒนาขึ้นคือ

โมเดลของราสช์แบบหลายองค์ประกอบ (Many-Facets Rasch Model: FACETS) ใช้ในการตรวจสอบ คุณภาพของการให้คะแนนการตอบที่เป็นลำดับขั้น (Order Response Categories) ลินาเคอร์ ได้พัฒนาขึ้นมาจากแนวคิดของราสช์โมเดล (Engelhard. 1994: 94; citing Andrich. 1988; Rasch. 1980; Wright; & Masters. 1982) โดยการเพิ่มค่าพารามิเตอร์ที่สนใจให้ศึกษาได้มากกว่า 2 พารามิเตอร์ เช่น พารามิเตอร์ของผู้ตรวจให้คะแนนในกรณีที่มีการใช้ผู้ตรวจให้คะแนนมากกว่า 2 คนขึ้นไป เป็นต้น โมเดลของราสช์แบบหลายองค์ประกอบใช้การตรวจสอบความเหมาะสมของข้อมูลกับโมเดลในการแสดงหลักฐานความเที่ยงตรงตามโครงสร้าง นอกจากนั้น โมเดลยังให้สารสนเทศเกี่ยวกับพฤติกรรมการตรวจให้คะแนนของผู้ตรวจ (Rater) เพิ่มเติมอีกด้วย จากคุณสมบัติดังกล่าวข้างต้น ผู้วิจัยจึงสนใจที่จะเลือกใช้โมเดลนี้ในการศึกษาเพิ่มเติมเกี่ยวกับการแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของคะแนนประเมินทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ ซึ่งมีความจำเป็นที่ต้องสะสมรายละเอียดจากหลายๆ แหล่งที่แตกต่างกัน เพื่อนำมารวมเป็นคำบรรยายพฤติกรรมให้กว้างขวาง คงทน และมีความเพียงพอที่จะอธิบายคุณลักษณะโครงสร้างของทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการได้

ความมุ่งหมายของการวิจัย

การศึกษาในครั้งนี้ มีความมุ่งหมายเพื่อแสดงหลักฐานความเที่ยงตรงตามโครงสร้างคะแนนการประเมินทักษะกระบวนการทางวิทยาศาสตร์ ใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน การวิเคราะห์ความเที่ยงตรงรวมและความเที่ยงตรงแยกตามแนวทางทฤษฎีมาตรฐานเดิม และจากการวิเคราะห์โมเดลของราสช์แบบหลายองค์ประกอบตามแนวทางของทฤษฎีการตอบข้อคำถาม พร้อมทั้งการแสดงหลักฐานความเชื่อมั่นในการประเมินจากการให้คะแนนของผู้ตรวจมากกว่าหนึ่งคนในทั้งสองทฤษฎี แยกเป็นความมุ่งหมายเฉพาะดังนี้

1. เพื่อแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของคะแนนการประเมินทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการโดยใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยันตามแนวทางของทฤษฎีมาตรฐานเดิม ดังนี้

- 1.1 เพื่อศึกษาโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งของแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการที่มี 5 องค์ประกอบ ได้แก่ องค์ประกอบของทักษะการตั้งสมมติฐาน ทักษะการกำหนดและควบคุมตัวแปร ทักษะการกำหนดนิยามเชิงปฏิบัติการ ทักษะการทดลอง และทักษะการตีความหมายข้อมูลและลงข้อสรุป

- 1.2 เพื่อศึกษาโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองของแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ที่มี 1 องค์ประกอบ คือ องค์ประกอบของทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ

2. เพื่อแสดงหลักฐานความเที่ยงตรงรวมและความเที่ยงตรงแยกของคะแนนการประเมินทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการตามแนวทางของทฤษฎีมาตรฐานเดิม
3. เพื่อแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของคะแนนการประเมินทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการโดยการวิเคราะห์โมเดลของราส์ชแบบหลายองค์ประกอบตามแนวทางของทฤษฎีการตอบข้อคำถาม
4. เพื่อแสดงหลักฐานความเชื่อมั่นในการให้คะแนนการวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการที่ใช้ผู้ตรวจให้คะแนนมากกว่าหนึ่งคน โดยใช้สถิติดังนี้
 - 4.1 ความเชื่อมั่นของผลการตรวจให้คะแนน (Reliability of rating) แบบความสอดคล้องภายใน (Internal consistency) โดยใช้สูตร r_{F-R} ของ Feldt – Raju
 - 4.2 ดัชนีความสอดคล้องของผู้ตรวจให้คะแนน โดยใช้สัมประสิทธิ์ Kappa
 - 4.3 ความเชื่อมั่นเชิงโครงสร้าง (Construct Reliability) และค่า Variance

Extracted

ความสำคัญของการวิจัย

การวิจัยในครั้งนี้มีเป้าหมายในการแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของคะแนนการประเมินด้านทักษะกระบวนการทางวิทยาศาสตร์ โดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยันและการวิเคราะห์ความเที่ยงตรงรวมและความเที่ยงตรงแยกตามแนวทางของทฤษฎีมาตรฐานเดิม จากการวิเคราะห์ทั้งสองวิธีนั้นจะให้ผลในการแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของคะแนนการประเมินทักษะกระบวนการทางวิทยาศาสตร์ที่สอดคล้องหรือขัดแย้งกันอย่างไร ถ้าวิเคราะห์เพิ่มเติมโดยใช้การวิเคราะห์โมเดลของราส์ชแบบหลายองค์ประกอบตามแนวทางของทฤษฎีการตอบข้อคำถาม จะให้ผลการแสดงหลักฐานความเที่ยงตรงตามโครงสร้างในแนวทางเดียวกันกับทฤษฎีมาตรฐานเดิมหรือไม่ การแสดงหลักฐานความเชื่อมั่นการประเมินจากการให้คะแนนของผู้ประเมินมากกว่าหนึ่งคนในทั้งสองทฤษฎี จะช่วยสะสมรายละเอียดจากหลาย ๆ แหล่งที่แตกต่างกัน เพื่อนำมารวมเป็นคำบรรยายพฤติกรรมได้อย่างกว้างขวาง คงทน และให้เพียงพอที่จะอธิบายคุณลักษณะโครงสร้างของทักษะกระบวนการทางวิทยาศาสตร์ได้หรือไม่

ผลการศึกษาในครั้งนี้ จะทำให้ได้หลักฐานอ้างอิงในการทดสอบแบบต่าง ๆ ที่ต้องการหลักฐานที่แตกต่างกัน หลักฐานนี้หาได้มาจากข้อมูลหรือข้อเท็จจริงที่แสดงความเชื่อมโยงขององค์ประกอบต่างๆ ของทักษะกระบวนการทางวิทยาศาสตร์ที่จะช่วยให้ได้ข้อสังเกตในการพิจารณาตัดสินใจนำไปใช้ในการอ้างอิงการแปลความหมายของข้อมูล ที่จะอธิบายคุณภาพของการประเมินทักษะกระบวนการทางวิทยาศาสตร์โดยใช้เครื่องมือชุดนี้ อีกทั้งยังช่วยให้ผู้วิจัยได้มีแนวทางใหม่ๆ ในการตรวจสอบความเที่ยงตรงและความเชื่อมั่นของเครื่องมือทางการวัดผลต่อไป

ขอบเขตของการวิจัย

ประชากรที่ใช้ในการวิจัย

ประชากรที่ใช้ในการวิจัยเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 โรงเรียนมัธยมศึกษาสังกัดเขตพื้นที่การศึกษานครราชสีมา เขต 3 ปีการศึกษา 2550 จำนวน 30,560 คน

กลุ่มตัวอย่างที่ใช้ในการวิจัย

กลุ่มตัวอย่างที่ใช้ในการวิจัย แบ่งออกเป็น 2 กลุ่ม ได้แก่ (1) กลุ่มตัวอย่างที่ใช้ในการเก็บข้อมูลเพื่อการแสดงหลักฐานความเที่ยงตรงตามโครงสร้างและความเชื่อมั่นของการประเมินทักษะกระบวนการทางวิทยาศาสตร์ โดยใช้ทฤษฎีมาตรฐานเดิม (Classical) และทฤษฎีการตอบข้อคำถามตามโมเดลของราสช์แบบหลายองค์ประกอบ (Many-Facet Rasch Model) โดยมีตัวแปรที่เกี่ยวข้อง 3 ตัวแปร คือ ทักษะ (Task), นักเรียนหรือผู้เข้าสอบ (Examinee) และผู้ตรวจให้คะแนน (Rater) กลุ่มตัวอย่างทั้งหมดเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 โรงเรียนมัธยมศึกษา สังกัดเขตพื้นที่การศึกษานครราชสีมา เขต 3 ปีการศึกษา 2550 จำนวน 330 คน ซึ่งเลือกมาโดยการสุ่มแบบสองขั้นตอน (Two - stage random sampling) (2) กลุ่มตัวอย่างที่เป็นผู้ตรวจให้คะแนน (Rater) แบบวัดทักษะกระบวนการทางวิทยาศาสตร์ โดยการเลือกจากครูผู้สอนวิทยาศาสตร์ที่มีประสบการณ์ทางวิทยาศาสตร์ เป็นต้น โดยคัดเลือกมาจำนวน 14 คน เพื่อทำการตรวจคะแนนแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ ให้กับนักเรียนจำนวน 280 คน โดยการเลือกแบบเจาะจง (Purposive sampling)

ตัวแปรที่ศึกษา

การแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ ที่ได้จากการวิเคราะห์องค์ประกอบเชิงยืนยัน มีตัวแปรที่ต้องการศึกษาดังต่อไปนี้

1. ตัวแปรสังเกตในองค์ประกอบอันดับหนึ่งและอันดับสอง จำนวน 22 ตัวแปร
2. ตัวแปรแฝงในองค์ประกอบอันดับหนึ่ง จำนวน 5 องค์ประกอบ ได้แก่ องค์ประกอบของทักษะการตั้งสมมติฐาน ทักษะการกำหนดนิยามเชิงปฏิบัติการ ทักษะการกำหนดและควบคุมตัวแปร ทักษะการทดลอง และทักษะการตีความหมายข้อมูลและลงข้อสรุป
3. ตัวแปรแฝงในองค์ประกอบอันดับสอง จำนวน 1 องค์ประกอบ ได้แก่ องค์ประกอบของทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ

นิยามศัพท์เฉพาะ

1. ทักษะกระบวนการทางวิทยาศาสตร์ (Science Process Skills) หมายถึง พฤติกรรมที่เกิดจากการปฏิบัติและฝึกฝนความนึกคิดอย่างเป็นระบบ เป็นความสามารถในการเสาะแสวงหาความรู้ทางวิทยาศาสตร์ มี 2 ประเภท คือ ทักษะกระบวนการทางวิทยาศาสตร์ขั้นพื้นฐาน (Basic Science Process Skills) และทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ (Integrated Science Process Skills) แต่ละประเภทมีลักษณะดังนี้

1.1 ทักษะกระบวนการทางวิทยาศาสตร์ขั้นพื้นฐาน (Basic Science Process Skills) หมายถึง ปฏิบัติการหรือวิธีการเสาะแสวงหาความรู้ทางวิทยาศาสตร์ก่อให้เกิดพัฒนาการทางด้านสติปัญญา ประกอบด้วย 8 ทักษะ ได้แก่ ทักษะการสังเกต ทักษะการจำแนกประเภท ทักษะการวัด ทักษะการคำนวณ ทักษะการหาความสัมพันธ์ระหว่างสเปสกับสเปสและสเปสกับเวลา ทักษะการจัดกระทำข้อมูลและสื่อความหมาย ทักษะการลงความเห็นจากข้อมูล และทักษะการพยากรณ์ มีรายละเอียดดังนี้

1.1.1 ทักษะการสังเกต (Observing) หมายถึง ความสามารถในการใช้ประสาทสัมผัสอย่างใดอย่างหนึ่งหรือหลายอย่างรวมกัน ซึ่งได้แก่ ตา หู จมูก ลิ้น และกายสัมผัส เข้าสัมผัสโดยตรงกับวัตถุ

1.1.2 ทักษะการจำแนกประเภท (Classifying) หมายถึงความสามารถในการจัดแบ่งหรือเรียงลำดับวัตถุหรือสิ่งที่อยู่ในปรากฏการณ์ต่างๆ ออกเป็นพวกๆ โดยมีเกณฑ์ในการจัดแบ่ง

1.1.3 ทักษะการวัด (Measuring) หมายถึง ความสามารถในการใช้เครื่องมือวัดปริมาณของสิ่งต่าง ๆ ได้อย่างถูกต้อง โดยมีหน่วยกำกับเสมอ รวมไปถึงการเลือกใช้เครื่องมือวัดได้อย่างถูกต้องเหมาะสมต่อสิ่งที่ต้องการวัด

1.1.4 ทักษะการคำนวณ (Using Numbers) หมายถึง ความสามารถในการบวก ลบ คูณและหารตัวเลขที่แสดงค่าปริมาณของสิ่งใดสิ่งหนึ่งซึ่งได้จากการสังเกต การวัด การทดลองโดยตรง หรือจากแหล่งอื่น ๆ อีกทอดหนึ่ง

1.1.5 ทักษะการหาความสัมพันธ์ระหว่างสเปสกับสเปสและสเปสกับเวลา (Space/Space and Space/Time Relationships) หมายถึง ความสามารถในการหาความสัมพันธ์ระหว่าง 3 มิติกับ 2 มิติ ระหว่างตำแหน่งของวัตถุหนึ่งกับอีกวัตถุหนึ่ง ระหว่างสเปสของวัตถุกับเวลา

1.1.6 ทักษะการจัดกระทำข้อมูลและสื่อความหมาย (Organizing Data and Communicating) หมายถึง ความสามารถในการนำข้อมูลดิบที่ได้จากการสังเกต การทดลอง หรือจากแหล่งอื่นที่มีข้อมูลดิบอยู่แล้วมาจัดกระทำเสียใหม่ โดยอาศัยวิธีการต่าง ๆ เช่น การจัดเรียงลำดับ การจัดแยกประเภท การหาค่าเฉลี่ย เป็นต้น

1.1.7 ทักษะการลงความเห็นจากข้อมูล (Inferring) หมายถึง ความสามารถในการนำข้อมูลที่ได้จากการสังเกตวัตถุหรือปรากฏการณ์ไปสัมพันธ์กับความรู้ หรือประสบการณ์เดิมเพื่อลงข้อสรุป หรืออธิบายปรากฏการณ์หรือวัตถุนั้น

1.1.8 ทักษะการพยากรณ์ (predicting) หมายถึง ความสามารถในการทำนายหรือคาดคะเนสิ่งที่จะเกิดขึ้นล่วงหน้า โดยอาศัยการสังเกตปรากฏการณ์ที่เกิดขึ้นซ้ำ ๆ หรือความรู้ที่เป็นหลักการ กฎหรือทฤษฎีในเรื่องนั้นมาช่วยในการทำนาย

1.2 ทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ (Integrated Science Process Skills) หมายถึง วิธีการแสวงหาความรู้ทางวิทยาศาสตร์ เป็นพฤติกรรมที่เกิดจากการปฏิบัติและฝึกฝนความคิดอย่างเป็นระบบ โดยอาศัยการผสมผสานจากทักษะกระบวนการ ขั้นพื้นฐาน ประกอบด้วย 5 ทักษะ ได้แก่ ทักษะการตั้งสมมติฐาน ทักษะการกำหนดและควบคุมตัวแปร ทักษะการกำหนดนิยามเชิงปฏิบัติการ ทักษะการตีความหมายข้อมูลและลงข้อสรุป และทักษะการทดลอง มีรายละเอียดดังนี้

1.2.1 ทักษะการตั้งสมมติฐาน (Formulating Hypothesis) หมายถึง ความสามารถในการให้คำอธิบายซึ่งเป็นคำตอบล่วงหน้าก่อนที่จะดำเนินการทดลองเพื่อตรวจสอบความถูกต้องเป็นจริงในเรื่องนั้น ๆ

1.2.2 ทักษะการกำหนดและควบคุมตัวแปร (Identifying and Controlling) หมายถึง ความสามารถในการบ่งชี้และกำหนดตัวแปรต้น ตัวแปรตาม และควบคุมตัวแปรที่เป็นสาเหตุอื่น ๆ ในปรากฏการณ์หนึ่ง ๆ ที่ต้องการศึกษา

1.2.3 ทักษะการกำหนดนิยามเชิงปฏิบัติการ (Defining Operationally) หมายถึง ความสามารถในการกำหนดความหมายและขอบเขตของคำหรือตัวแปรต่าง ๆ ในปฏิบัติการทดลองให้สามารถทดสอบหรือวัดได้

1.2.4 ทักษะการทดลอง (Experimenting) หมายถึง ความสามารถในการดำเนินการตรวจสอบสมมติฐานโดยการทดลอง ซึ่งเริ่มตั้งแต่การออกแบบการทดลอง การปฏิบัติการทดลองตามขั้นตอนที่ออกแบบไว้ ตลอดจนการใช้วัสดุอุปกรณ์ได้อย่างถูกต้อง

1.2.5 ทักษะการตีความหมายข้อมูลและลงข้อสรุป (Interpreting Data and Making Conclusion) หมายถึง ความสามารถในการแปลความหมายหรือบรรยายลักษณะ และคุณสมบัติของข้อมูลพร้อมทั้งสรุปความสัมพันธ์ของตัวแปรที่อยู่ในข้อมูลที่ต้องการศึกษา

2. แบบวัดทักษะกระบวนการทางวิทยาศาสตร์ หมายถึง เครื่องมือที่ใช้ในการวัดความสามารถของนักเรียนด้านทักษะกระบวนการทางวิทยาศาสตร์ สำหรับนักเรียนระดับชั้นมัธยมศึกษาตอนปลาย ที่ประกอบด้วยองค์ประกอบของทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ

3. คะแนนการประเมินทักษะกระบวนการทางวิทยาศาสตร์ หมายถึง ผลที่ได้จากการประเมินระดับความสามารถของนักเรียนจำนวน 280 คน ในด้านทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการที่ได้จากการตอบแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ ซึ่งถูกประเมินโดยผู้ตรวจจำนวน 14 คน

4. กฎเกณฑ์การให้คะแนน (Scoring Rubric) หมายถึง แนวทางที่กำหนดขึ้นเพื่อชั่งบอกระดับของทักษะกระบวนการทางวิทยาศาสตร์ในแต่ละด้าน ข้อมูลได้มาจากการพิจารณาผลของผู้ตรวจให้คะแนนเพื่อให้เกิดความเข้าใจตรงกันการวิจัยในครั้งนี้ผู้วิจัยได้พัฒนากฎเกณฑ์การให้คะแนน (Scoring Rubric) โดยใช้รูปแบบเฉพาะ (Specific Rubric) ที่กำหนดคุณภาพเป็น 4 ระดับ

5. ความเที่ยงตรง (Validity) หมายถึง ระดับของหลักฐานที่สามารถสนับสนุนการลงความเห็น (Inference) จากคะแนนที่วัดได้อย่างเหมาะสม มีความหมาย และให้ประโยชน์ การตรวจสอบความเที่ยงตรงจึงต้องรวบรวมพยานหลักฐาน (Evidence) เพื่อสนับสนุนการลงความเห็นดังกล่าว และเก็บรวบรวมหลักฐานได้หลายวิธี ความเที่ยงตรงจึงขึ้นอยู่กับปริมาณและชนิดของหลักฐานที่สนับสนุนการลงความเห็นของผู้วัดจากข้อมูลที่วัดได้

6. การแสดงหลักฐานความเที่ยงตรงตามโครงสร้าง หมายถึง การเก็บรวบรวมข้อมูลมาสนับสนุนว่า แบบวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ สามารถสะท้อนโครงสร้างทางสติปัญญาและแปลความหมายของคุณลักษณะที่วัดตามลักษณะองค์ประกอบที่กำหนดไว้ โดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน การวิเคราะห์หลายลักษณะหลายวิธี และการวิเคราะห์โมเดลของราส์ชแบบหลายองค์ประกอบ

6.1 การวิเคราะห์องค์ประกอบเชิงยืนยัน หมายถึง โครงสร้างองค์ประกอบตามทฤษฎีหรืองานวิจัยที่ต้องการตรวจสอบยืนยันด้วยข้อมูลเชิงประจักษ์ที่รวบรวมได้จากการวิจัยเชิงประจักษ์ว่ามีความเหมาะสมกันหรือไม่ ประกอบด้วยโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง และโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง

6.1.1 โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง หมายถึง โครงสร้างองค์ประกอบที่มีองค์ประกอบ/ตัวแปรแฝง 5 ตัวแปร ซึ่งเกิดจากตัวแปรสังเกต 22 ตัวแปร ตามโครงสร้างองค์ประกอบของตัวแปรการประเมินทักษะกระบวนการทางวิทยาศาสตร์โดยผู้เรียน ที่ใช้ในการตรวจสอบความเหมาะสมกับข้อมูลเชิงประจักษ์

6.1.2 โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง หมายถึง โครงสร้างองค์ประกอบที่มีองค์ประกอบ/ตัวแปรแฝงสองระดับโดยองค์ประกอบที่มีองค์ประกอบ/ตัวแปรแฝงอันดับสองเกิดจากการวิเคราะห์องค์ประกอบอันดับหนึ่ง ที่ใช้ในการตรวจสอบความเหมาะสมกับข้อมูลเชิงประจักษ์กำหนดไว้เพียงรูปแบบเดียวคือ ทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ

6.2 การวิเคราะห์หลายลักษณะหลายวิธี หมายถึง กระบวนการตรวจสอบความเที่ยงตรงตามโครงสร้างของผลการประเมินทักษะกระบวนการทางวิทยาศาสตร์ ตามแนวทางของทฤษฎีมาตรฐานเดิม โดยอาศัยสมมติฐานที่ว่าลักษณะร่วมกันภายในทักษะกระบวนการจะมีค่าสหสัมพันธ์กันสูง ลักษณะที่แตกต่างกันระหว่างทักษะกระบวนการจะมีค่าสหสัมพันธ์กันต่ำ มีสองรูปแบบ คือ ความเที่ยงตรงร่วม (Convergent Validity) และความเที่ยงตรงแยก (Discriminant Validity)

6.2.1 ความเที่ยงตรงร่วม หมายถึง ระดับความคล้ายคลึงกันภายในทักษะกระบวนการทางวิทยาศาสตร์ ที่ควรจะคล้ายคลึงกันตามทฤษฎีและมีความสัมพันธ์กันสูง

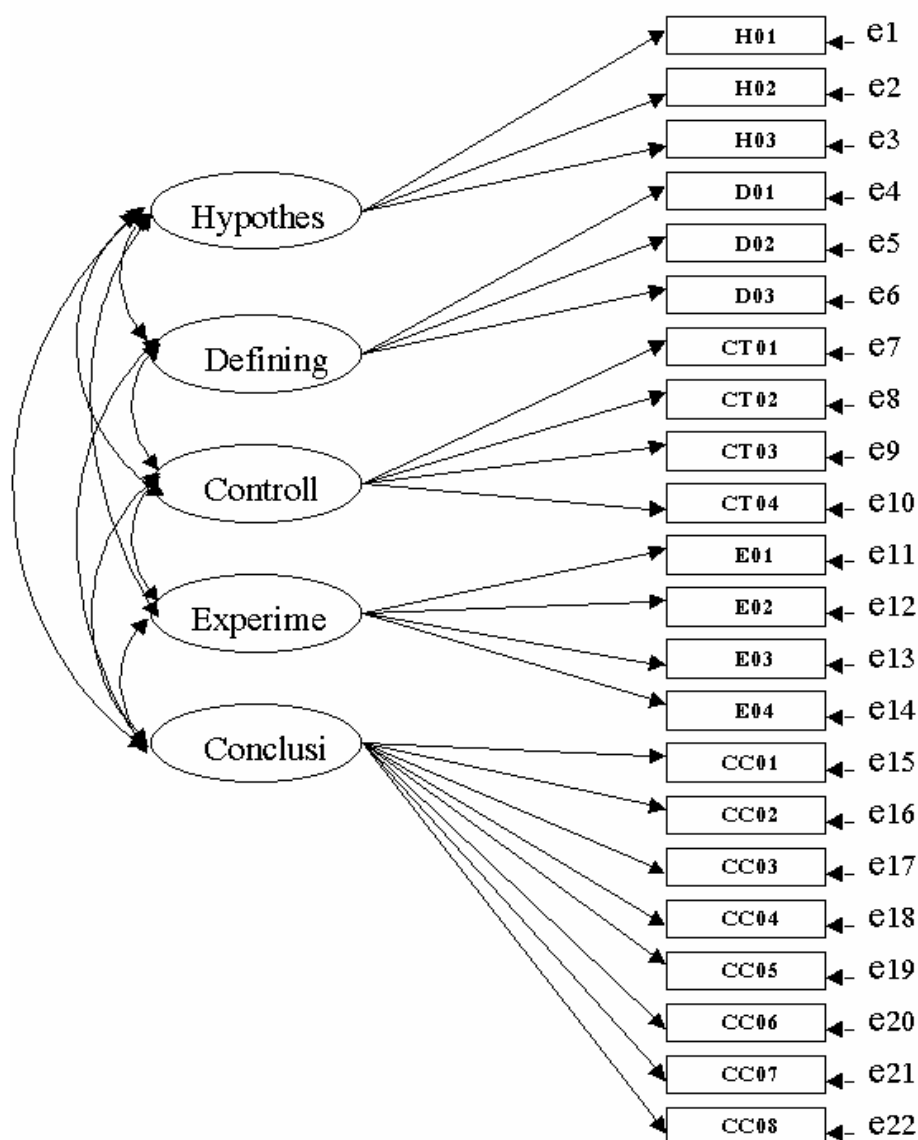
6.2.2 ความเที่ยงตรงแยก หมายถึง ระดับความแตกต่างกันระหว่างทักษะกระบวนการทางวิทยาศาสตร์ ที่ควรจะแตกต่างกันตามทฤษฎีและมีความสัมพันธ์กันต่ำ

6.3 โมเดลของราส์ชแบบหลายองค์ประกอบ (Many-Facet Rasch Model) หมายถึง โมเดลของทฤษฎีการตอบข้อคำถามที่ใช้ในการตรวจสอบคุณภาพของการให้คะแนนการตอบที่เป็นลำดับขั้น (Order Response Categories) ที่พัฒนาขึ้นมาจากแนวคิดของราส์ช โมเดล ผลของการตรวจสอบความเหมาะสมของข้อมูลกับโมเดลจะแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของการให้คะแนน

สมมติฐานและโมเดลในการวิจัย

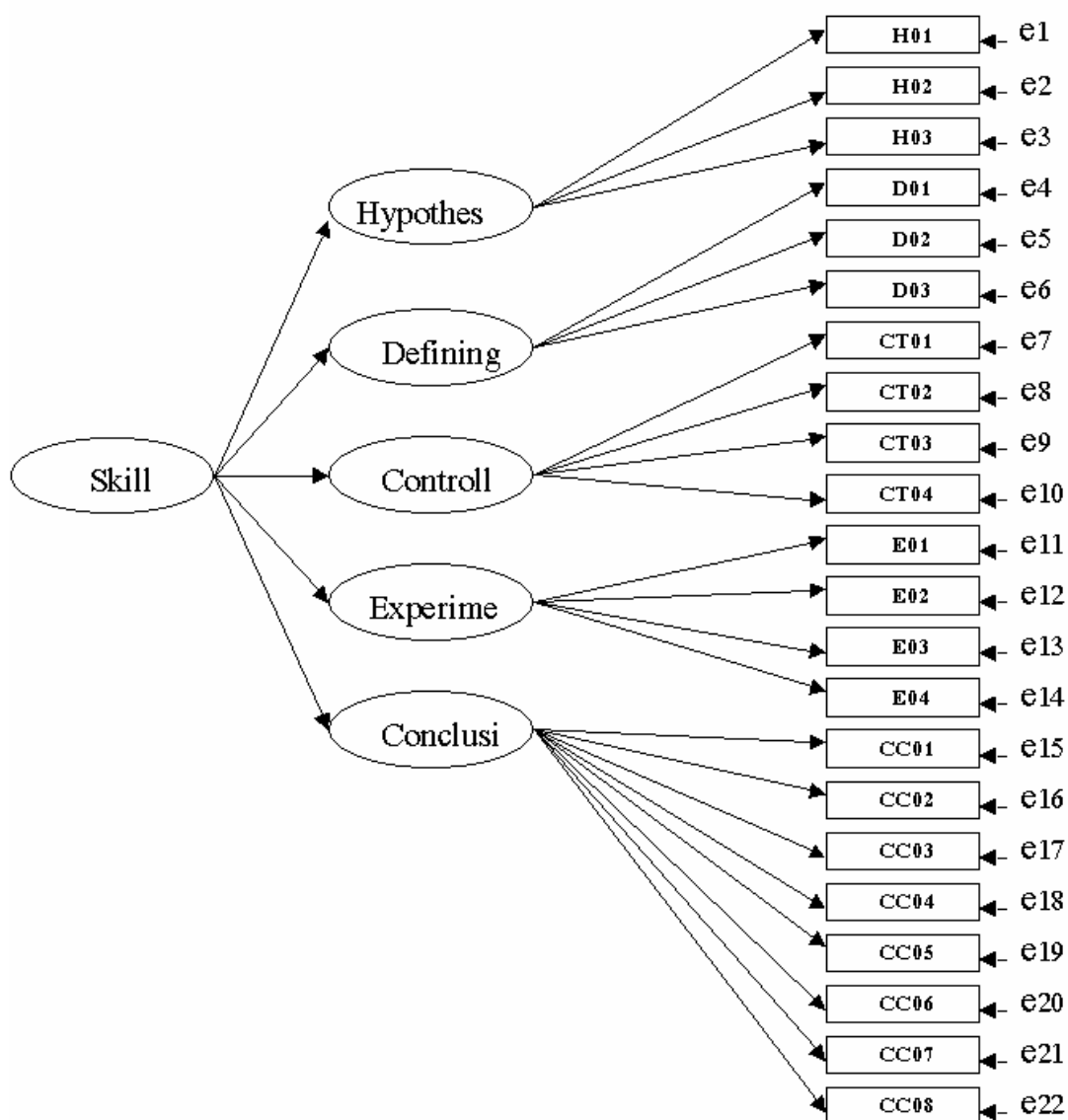
การวิจัยครั้งนี้ตั้งสมมติฐานไว้ว่า ผลการตรวจให้คะแนนแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการมีความเที่ยงตรงตามโครงสร้าง โดยมีสมมติฐานย่อยดังนี้

1. องค์ประกอบตัวแปรแฝง 5 องค์ประกอบของโมเดล ได้แก่ องค์ประกอบของการตั้งสมมติฐาน การกำหนดนิยามเชิงปฏิบัติการ การกำหนดและควบคุมตัวแปร การทดลอง และการตีความหมายข้อมูลและลงข้อสรุป มีความเหมาะสมกับข้อมูลเชิงประจักษ์ โดยลักษณะของโมเดลที่ใช้ในการศึกษาองค์ประกอบเชิงยืนยันอันดับหนึ่งมีลักษณะดังนี้



ภาพประกอบ 1 โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง

องค์ประกอบเชิงยืนยันอันดับหนึ่ง สามารถอธิบายองค์ประกอบใหม่เป็นองค์ประกอบอันดับสองของตัวแปรแฝง 1 องค์ประกอบ คือ องค์ประกอบทักษะกระบวนการทางวิทยาศาสตร์ ขั้นบูรณาการโดยลักษณะของโมเดลที่ใช้ในการศึกษาองค์ประกอบอันดับสองมีลักษณะดังนี้



ภาพประกอบ 2 โมเดลการวิเคราะห์หึ่งค์ประกอบเชิงยืนยันอันดับสอง

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยในครั้งนี้ มีความมุ่งหมายในการแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของคะแนนการประเมินทักษะกระบวนการทางวิทยาศาสตร์ โดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยันและการวิเคราะห์ความเที่ยงตรงรวมและความเที่ยงตรงแยกตามทฤษฎีมาตรฐานเดิม และการวิเคราะห์โมเดลของราส์ชแบบหลายองค์ประกอบตามทฤษฎีการตอบข้อคำถาม พร้อมทั้งแสดงหลักฐานความเชื่อมั่นในการประเมินทักษะกระบวนการทางวิทยาศาสตร์ทั้งสองทฤษฎีในการศึกษาดังกล่าว ผู้วิจัยได้ศึกษากรอบแนวคิด ทฤษฎี หลักการทางด้านทักษะกระบวนการทางวิทยาศาสตร์ ความเที่ยงตรงตามโครงสร้างและความเชื่อมั่นในการประเมินตลอดจนงานวิจัยต่าง ๆ ที่เกี่ยวข้อง เพื่อนำองค์ความรู้ที่ได้จากการศึกษาค้นคว้ามาประยุกต์ใช้กับการวิจัยในครั้งนี้ แนวทางของการศึกษาค้นคว้า สามารถกำหนดเป็นกรอบการศึกษาใน 5 ประเด็นดังนี้

1. ทักษะกระบวนการทางวิทยาศาสตร์
2. หลักฐานแสดงความเที่ยงตรง
3. การวิเคราะห์องค์ประกอบเชิงยืนยัน
4. การวิเคราะห์หลายลักษณะหลายวิธี
5. การวิเคราะห์โมเดลของราส์ชแบบหลายองค์ประกอบ

1. ทักษะกระบวนการทางวิทยาศาสตร์

1.1 ความหมายทักษะกระบวนการทางวิทยาศาสตร์

วิทยาศาสตร์มีบทบาทสำคัญในสังคมโลกปัจจุบันและอนาคต วิทยาศาสตร์เกี่ยวข้องกับชีวิตของทุกคนทั้งในการดำรงชีวิตประจำวัน ในงานอาชีพต่าง ๆ เครื่องมือเครื่องใช้ตลอดจนผลผลิตต่าง ๆ ที่เราใช้อำนวยความสะดวกในชีวิตและในการทำงาน ผลของความรู้วิทยาศาสตร์ที่ผสมผสานกับความคิดสร้างสรรค์และศาสตร์แขนงอื่น ๆ จึงช่วยให้เกิดการพัฒนาเทคโนโลยีอย่างมากมาย และในทางกลับกัน เทคโนโลยีก็มีส่วนสำคัญที่จะให้มีการศึกษาค้นคว้าความรู้ทางวิทยาศาสตร์เพิ่มขึ้นอย่างไม่หยุดยั้ง โดยเฉพาะหลักสูตรการศึกษาขั้นพื้นฐาน ตามพระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 ในปัจจุบันที่มุ่งให้ผู้เรียนได้เรียนรู้วิทยาศาสตร์ที่เน้นกระบวนการนำไปสู่การสร้างองค์ความรู้ โดยผู้เรียนมีส่วนร่วมในการเรียนทุกขั้นตอน ผู้เรียนได้ทำกิจกรรมที่หลากหลายทั้งเป็นกลุ่มและเป็นรายบุคคล ในการสังเกตสิ่งต่าง ๆ รอบตัว ตั้งคำถามหรือปัญหาเกี่ยวกับสิ่งที่จะศึกษา พัฒนาระบวนการคิดขั้นสูง มีการคิดวางแผน

และลงมือปฏิบัติการสำรวจตรวจสอบด้วยกระบวนการที่หลากหลายจากแหล่งเรียนรู้ทั้งส่วนที่เป็นสากลและท้องถิ่น คิดและตัดสินใจเลือกข้อมูลที่เป็นประโยชน์เพื่อนำไปใช้ในการตอบคำถามหรือแก้ปัญหา ซึ่งจะนำไปสู่องค์ความรู้แนวคิดหลักทางวิทยาศาสตร์ แล้วสื่อสารสิ่งที่เรียนรู้หรือองค์ความรู้ในรูปแบบต่าง ๆ ให้ผู้อื่นรับรู้ กระบวนการเรียนรู้ดังกล่าว ทำให้ผู้เรียนได้สร้างองค์ความรู้และเกิดการพัฒนาเจตคติ คุณธรรม จริยธรรมและค่านิยมที่ดีต่อวิชาวิทยาศาสตร์ (กรมวิชาการ. 2545: 4) ดังนั้น วิทยาศาสตร์จึงมุ่งเน้นให้ผู้เรียนได้เป็นผู้เรียนรู้และค้นพบด้วยตนเองมากที่สุด (กรมวิชาการ. 2545: 14) มุ่งพัฒนาให้นักเรียนคิดเป็น ทำเป็นและแก้ปัญหาได้เป็นสำคัญ การที่ผู้เรียนจะเกิดคุณภาพได้ตามที่กล่าวอ้างนั้น จำเป็นต้องมีทักษะพื้นฐานสำคัญที่เรียกว่า **ทักษะกระบวนการทางวิทยาศาสตร์** ที่ครอบคลุมความสามารถในการปฏิบัติ มีความรู้ความเข้าใจ มีความคิดพื้นฐานทางวิทยาศาสตร์ และมีเจตคติทางวิทยาศาสตร์ (ยุพาทันติเจริญ. 2531: 41-42) คำว่า ทักษะกระบวนการทางวิทยาศาสตร์ (Science Process Skills) ได้มีนักการศึกษาให้ความหมายไว้หลายทัศนะดังนี้

นิคม ทาแดง และสุจินต์ วิสุทธิรานนท์ (2529: 48) กล่าวว่า “ทักษะกระบวนการทางวิทยาศาสตร์เป็นองค์ประกอบที่สำคัญประการหนึ่งของการแสวงหาความรู้ทางวิทยาศาสตร์ เพราะการทำงานตามขั้นตอนของวิธีการทางวิทยาศาสตร์แต่ละขั้นตอนจะประสบผลสำเร็จหรือล้มเหลว ขึ้นอยู่กับความสามารถและทักษะกระบวนการทางวิทยาศาสตร์ของแต่ละคน ทักษะกระบวนการทางวิทยาศาสตร์สามารถฝึกให้มีขึ้นได้” สอดคล้องกับคลอฟเฟอร์ (Klopfers. 1971 : 568-573) ที่กล่าวว่า “ทักษะกระบวนการทางวิทยาศาสตร์เป็นกระบวนการที่ใช้ในการแสวงหาความรู้ทางวิทยาศาสตร์” และปีเตอร์สัน (Peterson. 1978: 153) อธิบายว่า ทักษะกระบวนการทางวิทยาศาสตร์เป็นปฏิบัติการสืบเสาะหาความรู้ทางวิทยาศาสตร์ที่ประกอบด้วย การสังเกต การตั้งคำถาม การทดลอง การเปรียบเทียบ การสรุปพาดพิง การสรุปหลักเกณฑ์ การสื่อความหมายและการนำไปใช้ แอนเดอร์สัน (Anderson. 1978: 15) ได้ให้ความหมายของทักษะกระบวนการทางวิทยาศาสตร์ ไว้ว่า เป็นวิธีการที่นักวิทยาศาสตร์ใช้ในการเสาะแสวงหาความรู้ความหมายที่สำคัญของกระบวนการ คือ เป็นวิถีทางของกระบวนการในการเสาะแสวงหาความรู้ทำให้เกิดพัฒนาการทางด้านสติปัญญา ส่วนวรรณทิพา รอดแรงคำ และพิมพ์พันธ์ เตชะคุปต์ (2532: V) ให้ความหมายของทักษะกระบวนการทางวิทยาศาสตร์ว่า เป็นทักษะทางสติปัญญาที่นักวิทยาศาสตร์และผู้ที่เกี่ยวข้องนำเอาวิธีการทางวิทยาศาสตร์มาแก้ปัญหา ใช้ในการศึกษาค้นคว้าหาความรู้และแก้ปัญหาต่าง ๆ หน่วยงานที่รับผิดชอบการจัดการเรียนการสอนด้านวิทยาศาสตร์ คือ สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (2524: 1) ให้ความหมายของทักษะกระบวนการทางวิทยาศาสตร์ไว้ว่า “เป็นพฤติกรรมที่เกิดจากการปฏิบัติและการฝึกฝนความคิดอย่างมีระบบ ซึ่งก่อให้เกิดความมั่งคั่งทางสติปัญญา”

จากแนวคิดด้านทักษะกระบวนการทางวิทยาศาสตร์ที่นักการศึกษาได้อธิบายดังกล่าว จึงสรุปได้ว่า ทักษะกระบวนการทางวิทยาศาสตร์ หมายถึง ความสามารถในการใช้กระบวนการทางวิทยาศาสตร์ต่าง ๆ ได้แก่ การสังเกต การจำแนกประเภท การวัด การคำนวณ การหาความสัมพันธ์ระหว่างสเปสและเวลา การจัดกระทำและสื่อความหมายข้อมูล การลงความเห็นจากข้อมูล การพยากรณ์ การตั้งสมมติฐาน การกำหนดและควบคุมตัวแปร การกำหนดนิยามเชิงปฏิบัติการ การตีความหมายข้อมูลและลงข้อสรุป และการทดลองได้อย่างคล่องแคล่วถูกต้อง และแม่นยำ เพื่อการเสาะแสวงหาความรู้ หรือแก้ปัญหาอันเกิดจากการปฏิบัติและฝึกฝนความคิดอย่างเป็นระบบ

1.2 ประเภทของทักษะกระบวนการทางวิทยาศาสตร์

หน่วยงานทางการศึกษาและนักการศึกษาที่รับผิดชอบเกี่ยวกับการจัดการเรียนการสอนวิทยาศาสตร์ทั้งในประเทศและต่างประเทศได้จัดประเภทของทักษะกระบวนการทางวิทยาศาสตร์ไว้ทั้งที่เหมือนกันและแตกต่างกัน ดังนี้

คลอปเฟอร์ (Klopfer. 1971: 568-573) ได้จัดประเภทของทักษะกระบวนการทางวิทยาศาสตร์ไว้ 4 ทักษะ คือ ทักษะการสังเกตและการวัด (Observing and Measuring), ทักษะการมองเห็นปัญหาและวิธีการหาคำตอบ (Seeing a Problem and Seeking Ways to Solve It), ทักษะการแปลความหมายข้อมูลและการสรุป (Interpreting Data and Formulating Generalizations) และทักษะการสร้างทฤษฎี การตรวจสอบและการปรับปรุงทฤษฎีที่สร้างขึ้น (Building, Testing and Revising a Theoretical Model) การ์แลนด์ และคณะ (Garland; & Others. 1973: 127) ได้จัดประเภทของทักษะกระบวนการทางวิทยาศาสตร์ออกเป็น 9 ทักษะที่คล้ายกับสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (2519: 8-9) แต่เรียกชื่อแตกต่างกัน ประเภทของทักษะกระบวนการทางวิทยาศาสตร์ของการ์แลนด์และคณะ ประกอบด้วย ทักษะการสังเกต (Observing), ทักษะการจัดกระทำข้อมูล (Data Treatment), ทักษะการพยากรณ์และตั้งสมมติฐาน (Predicting and Formulating Hypothesis), ทักษะการจำแนกประเภท (Classifying), ทักษะการบ่งชี้ (Identifying), ทักษะการวัด (Measuring), ทักษะการพัฒนาเทคนิควิธีการปฏิบัติในห้องทดลอง (Development of Acceptable Laboratory Technique), ทักษะการวิเคราะห์และสังเคราะห์ (Analysis and Synthesis) และทักษะการสื่อความหมาย (Communicating) ส่วนทักษะกระบวนการทางวิทยาศาสตร์ของสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยีประกอบด้วย ทักษะการสังเกต, ทักษะในการเลือกใช้เครื่องมือ, ทักษะในการบันทึกข้อมูลและสื่อความหมาย, ทักษะในการจัดกระทำข้อมูล, ทักษะในการแปลความหมายข้อมูลและลงข้อสรุป, ทักษะในการสร้างสมมติฐาน, ทักษะในการออกแบบการทดลอง, ทักษะในการคิดคำนวณ และทักษะในการหาความสัมพันธ์ระหว่างมิติ โอเคย์ และฟิลด์ (Okey; & Field. 1973: 1-10), ลันสตรอม และโลเวอร์รี่ (Lunstrom; & Lowery. 1978: 209-

210) และสำนักประเมินผลการศึกษาเพื่อความก้าวหน้าแห่งชาติ (Duren. 1978: 20; citing The National Assessment of Education Progress : NAEP. 1978) ได้จัดประเภทของทักษะกระบวนการทางวิทยาศาสตร์เป็น 10 ทักษะ ทั้งที่เหมือนกันและแตกต่างกัน ทักษะกระบวนการทางวิทยาศาสตร์ของโอเคย์และฟิล ประกอบด้วย ทักษะการกำหนดตัวแปร (Identifying Variables), ทักษะการสร้างตารางเสนอข้อมูล (Constructing a Table of Data), ทักษะการเขียนกราฟ (Constructing a Graph), ทักษะการอธิบายความสัมพันธ์ของตัวแปร (Describing Relationships between Variables), ทักษะการรวบรวมและจัดกระทำข้อมูล (Acquiring and Processing Your Own Data), ทักษะการตั้งสมมติฐาน (Constructing Hypothesis), ทักษะการวิเคราะห์กระบวนการทดลอง (Analyzing Investigations), ทักษะการให้นิยามเชิงปฏิบัติการ (Defining Variables Operationally), ทักษะการออกแบบการทดลอง (Designing Investigations) และทักษะการดำเนินการทดลอง (Experimenting) ส่วนทักษะกระบวนการทางวิทยาศาสตร์ของลินสตรอมและโลเวอร์ ประกอบด้วย ทักษะการสังเกตและการบรรยาย (Observation and Description), ทักษะการเปรียบเทียบ (Comparison), ทักษะการจำแนก (Classification), ทักษะการใช้เทคนิคทางคณิตศาสตร์ (Mathematical Techniques), ทักษะการปฏิบัติการทดลอง (Experimentation), ทักษะการสรุปอ้างอิง (Inferences), ทักษะการแสดงความหมาย (Implications), ทักษะการทดลองเพื่อทดสอบสมมติฐาน (Verification in Independent Study), ทักษะการศึกษาลึกซึ้ง (Depth Study) และทักษะการใช้เครื่องมือและความสามารถเฉพาะ (Instrumentation and Other Specialization) ทักษะกระบวนการทางวิทยาศาสตร์ของสำนักประเมินผลการศึกษาเพื่อความก้าวหน้าแห่งชาติ ประกอบด้วย ทักษะการให้นิยามทางวิทยาศาสตร์ (Define a Scientific Problem), ทักษะการเสนอแนะหรือการรู้สมมติฐานทางวิทยาศาสตร์ (Suggest or Recognize a Scientific Hypothesis), ทักษะการเสนอหรือเลือกวิธีการที่เหมาะสมในด้านเหตุผลและการปฏิบัติ (Propose or Select Validating Procedures), ทักษะการเก็บข้อมูล (Obtain Requisite Data), ทักษะการตีความหมายข้อมูล (Interpret Data), ทักษะการตรวจสอบความถูกต้องอย่างมีเหตุผลของสมมติฐานกับข้อเท็จจริง การสังเกตหรือการทดลอง (Check the Logical Consistency of Hypothesis with Relevant Laws, Fact, Observations or Experiments), ทักษะการให้เหตุผลทางด้านปริมาณและสัญลักษณ์ (Reason Quantitative and Symbolically), ทักษะการจำแนกความแตกต่างระหว่างข้อเท็จจริง สมมติฐานและความคิดเห็น สิ่งที่เกี่ยวข้องกับสิ่งที่ไม่เกี่ยวข้องและรูปแบบจากการสังเกต (Distinguish Between Fact, Hypothesis and Opinions the Relevant from the Irrelevant and the Model from the Observations), ทักษะการวิเคราะห์และวิจารณ์เอกสารทางวิทยาศาสตร์ (Read Scientific Materials Critically) และทักษะการใช้กฎและหลักการทางวิทยาศาสตร์ ทั้งในสถานที่คุ้นเคยและไม่คุ้นเคย (Employ Scientific Laws and Principles in Familiar and Unfamiliar Situations) โชวาลเตอร์ (Showalter. 1974: 1-8) ได้

จัดแบ่งประเภทของทักษะกระบวนการทางวิทยาศาสตร์ต่าง ๆ ที่สำคัญ โดยอาศัยแนวคิดของ คณะกรรมการการศึกษาวิทยาศาสตร์ของสมาคมอเมริกันเพื่อความก้าวหน้าทางวิทยาศาสตร์ (AAAS) ออกเป็น 15 ทักษะ คือ ทักษะการจำแนก (Classifying), ทักษะการสื่อความหมาย (Communicating), ทักษะการควบคุมตัวแปร (Controlling Variables), ทักษะการให้คำนิยาม (Defining Operationally), ทักษะการออกแบบการทดลอง (Designing Experimenting), ทักษะการสร้างแบบจำลอง (Formulating Models), ทักษะการลงความคิดเห็น (Inferring), ทักษะการตั้งสมมติฐาน (Formulating Hypothesis), ทักษะการตีความหมายข้อมูล (Interpreting Data), ทักษะการวัด (Measuring), ทักษะการสังเกต (Observing), ทักษะการทำนาย (Predicting), ทักษะการตั้งคำถาม (Questioning), ทักษะการใช้ตัวเลข (Using Numbers) และทักษะการใช้ความสัมพันธ์เกี่ยวกับช่วงเวลา (Using Spacetime Relationships) ส่วนทบวงมหาวิทยาลัย (2525: 59) ได้จัดประเภทของทักษะกระบวนการทางวิทยาศาสตร์โดยใช้แนวทางของโซวาลเตอร์ป รับลดให้เหลือเพียง 13 ทักษะ ตัดทักษะของโซวาลเตอร์ออก 2 ทักษะ ทักษะที่ถูกตัดออกคือ ทักษะการสร้างแบบจำลอง (Formulating Models) และทักษะการตั้งคำถาม (Questioning) เจคอบสัน และบาร์รี (Jacobson; & Barry. 1980: 450-451) ได้จัดประเภททักษะกระบวนการทางวิทยาศาสตร์ที่ควรฝึกให้กับนักเรียน เพื่อให้นักเรียนได้มีโอกาสพัฒนาความสามารถด้านนี้ และสามารถนำเอาทักษะด้านนี้ไปใช้ในชีวิตประจำวันได้ โดยการแบ่งทักษะกระบวนการทางวิทยาศาสตร์ออกเป็น 13 ทักษะ คล้ายกับทบวงมหาวิทยาลัย แต่มีบางทักษะที่ไม่เหมือนกัน คือ ทักษะการกำหนดปัญหา (Defining Problems), ทักษะการวางแผนทางในการค้นคว้า (Designing Investigations), ทักษะการค้นคว้าเอกสารที่เกี่ยวข้อง (Literature Search), ทักษะการสร้างความสัมพันธ์กับทฤษฎี (Relating to Theory) และทักษะการลงข้อสรุป (Drawing Conclusions)

สมาคมอเมริกันเพื่อความก้าวหน้าทางวิทยาศาสตร์ (American Association for the Advancement of Science: AAAS. 1970: 11-176) ได้วิเคราะห์การทำงานของนักวิทยาศาสตร์ และจัดประเภทของทักษะกระบวนการทางวิทยาศาสตร์ได้สอดคล้องกับสถาบันส่งเสริมการสอน วิทยาศาสตร์และเทคโนโลยี (สสวท.) (2524: 1-16) โดยจัดประเภทของทักษะกระบวนการทาง วิทยาศาสตร์ออกเป็น 2 ประเภท 13 ทักษะ คือ

1. ทักษะกระบวนการขั้นพื้นฐาน จำนวน 8 ทักษะ ประกอบด้วย
 - 1.1 ทักษะการสังเกต (Observing)
 - 1.2 ทักษะการใช้ความสัมพันธ์ระหว่างสเปสและเวลา (Using Space and Relationships)
 - 1.3 ทักษะการจัดจำพวก (Classifying)
 - 1.4 ทักษะการใช้ตัวเลข (Using Numbers)
 - 1.5 ทักษะการวัด (Measuring)

- 1.6 ทักษะการสื่อความหมาย (Communicating)
- 1.7 ทักษะการทำนาย (Predicting)
- 1.8 ทักษะการสรุปอ้างอิง (Inferring)
2. ทักษะกระบวนการขั้นบูรณาการหรือขั้นผสม จำนวน 5 ทักษะ ประกอบด้วย
 - 2.1 ทักษะการกำหนดและควบคุมตัวแปร (Identifying controlling and Manipulating Variables)
 - 2.2 ทักษะการประมวลผลและตีความหมายข้อมูล (Data Processing and Interpreting)
 - 2.3 ทักษะการตั้งสมมติฐาน (Formulating Hypothesis)
 - 2.4 ทักษะการให้นิยามเชิงปฏิบัติการ (Defining Operationally)
 - 2.5 ทักษะการออกแบบและดำเนินการทดลอง (Designing Investigations and Experimenting)

จากแนวคิดการจัดประเภทของทักษะกระบวนการทางวิทยาศาสตร์ที่กล่าวมา ผู้วิจัย จึงได้จัดแบ่งประเภทของทักษะกระบวนการทางวิทยาศาสตร์ออกเป็น 2 ประเภท 13 ทักษะ ตามแนวทางของสมาคมอเมริกันเพื่อความก้าวหน้าทางวิทยาศาสตร์และสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.) ดังนี้

1. ทักษะกระบวนการขั้นพื้นฐาน (Basic Science Process Skills) เป็นทักษะกระบวนการที่ผู้เรียนควรฝึกฝนให้เกิดความชำนาญเป็นพื้นฐาน ก่อนไปฝึกทักษะกระบวนการขั้นผสม หรือขั้นบูรณาการ ทักษะกระบวนการขั้นพื้นฐานประกอบด้วย 8 ทักษะ คือ
 - 1.1 ทักษะการสังเกต (Observing)
 - 1.2 ทักษะการจำแนกประเภท (Classifying)
 - 1.3 ทักษะการวัด (Measuring)
 - 1.4 ทักษะการคำนวณ (Using Numbers)
 - 1.5 ทักษะการหาความสัมพันธ์ระหว่างสเปซกับสเปซและสเปซกับเวลา (Space/Space and Space/Time Relationships)
 - 1.6 ทักษะการจัดกระทำข้อมูลและสื่อความหมาย (Organizing Data and Communicating)
 - 1.7 ทักษะการลงความเห็นจากข้อมูล (Inferring)
 - 1.8 ทักษะการพยากรณ์ (predicting)

2. ทักษะกระบวนการขั้นผสมหรือขั้นบูรณาการ (Integrated Science Process Skills) เป็นทักษะกระบวนการที่ต้องอาศัยการบูรณาการจากทักษะกระบวนการขั้นพื้นฐาน ซึ่งผู้เรียนควรฝึกฝนให้เกิดความชำนาญมาก่อน จึงจะทำให้ทักษะกระบวนการขั้นผสมหรือขั้นบูรณาการมีประสิทธิภาพ ประกอบด้วย 5 ทักษะ คือ

2.1 ทักษะการตั้งสมมติฐาน (Formulating Hypothesis)

2.2 ทักษะการกำหนดนิยามเชิงปฏิบัติการ (Defining Operationally)

2.3 ทักษะการกำหนดและควบคุมตัวแปร (Identifying and Controlling Variables)

2.4 ทักษะการทดลอง (Experimenting)

2.5 ทักษะการตีความหมายข้อมูลและลงข้อสรุป (Interpreting Data and Making Conclusion)

1.3 การประเมินผลทักษะกระบวนการทางวิทยาศาสตร์

การประเมินผลทักษะกระบวนการทางวิทยาศาสตร์ ต้องศึกษาวัตถุประสงค์เชิงพฤติกรรมของแต่ละทักษะ เพื่อใช้เป็นแนวทางการประเมินผลว่า นักเรียนมีความสามารถในการใช้ทักษะกระบวนการทางวิทยาศาสตร์ในด้านต่าง ๆ หรือไม่ มากน้อยเพียงใด (Sund; & Picard. 1972: 31-25) ส่วนหน่วยทดสอบและประเมินผลสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี ได้เสนอแนวทางในการสร้างแบบทดสอบวัดทักษะกระบวนการทางวิทยาศาสตร์ตามขั้นตอน (สสวท. 2518 : 5) ดังนี้ (1) กำหนดความมุ่งหมายเชิงพฤติกรรม ซึ่งต้องแจกแจงให้ชัดเจน โดยครูต้องศึกษาจุดมุ่งหมายในแต่ละทักษะให้เข้าใจแล้วแจกแจงให้เป็นจุดมุ่งหมายเชิงพฤติกรรม ซึ่งจะมีทั้งภาคสถานการณ์ ภาคพฤติกรรมที่คาดหวัง และเกณฑ์ในการกำหนดพฤติกรรมนั้น ๆ (2) การเลือกเนื้อหาที่จะวัด หมายถึง การเลือกความมุ่งหมายเชิงพฤติกรรมกับเนื้อหาที่จำเป็นที่ขาดเสียมิได้ในบทหนึ่ง ๆ ควรจะกำหนดว่าทักษะใด เนื้อหาใดเป็นสิ่งที่ขาดไม่ได้ ทักษะนั้นและเนื้อหานั้นก็ควรจะปรากฏในข้อสอบ (3) การสร้างตารางเพื่อกำหนดเนื้อหาและพฤติกรรมทักษะ ซึ่งมีความมุ่งหมายอยู่ที่จะกำหนดว่าจะวัดทักษะหรือพฤติกรรมได้เท่าไรอย่างละกี่ข้อ จะได้ไม่บกพร่อง นอกจากนั้นผู้ออกข้อสอบยังจะต้องทราบต่อไปอีกว่า ข้อสอบวัดพฤติกรรมทักษะใดมีสัดส่วนมากน้อยเพียงใด (4) การเลือกแนวทางการออกข้อสอบ ควรจะถือหลักกว่าจะการใช้การสอบแบบใด จึงจะวัดพฤติกรรมนั้น ได้ตรงและถูกต้องเหมาะสมที่สุด ตลอดทั้งเหมาะสมกับวัยของเด็ก ประหยัดเวลาและง่ายต่อการปฏิบัติด้วย นอกจากนี้ ยังได้เสนอลักษณะข้อทดสอบเพื่อวัดความสามารถในการดำเนินการตามกระบวนการทางวิทยาศาสตร์ตามขั้นตอน (สสวท. 2519 : 10) ดังนี้ (1) การสร้างสถานการณ์ (1.1) สถานการณ์ที่สร้างขึ้น จะเป็นสถานการณ์สมมติ หรือนำมาจากเอกสารอื่นใดก็ตาม จะต้องมีความยากง่ายเหมาะสมกับระดับชั้นของนักเรียน (1.2) ใช้คำพูดที่เข้าใจง่าย ศัพท์เทคนิคต้องไม่นอกเหนือจากที่นักเรียน

เรียนรูมาแล้ว (1.3) สถานการณ์นั้น ๆ ต้องมีความเป็นไปได้ จะต้องเป็นจริงสมเหตุผล (1.4) ถ้าเป็นเรื่องที่มีหน่วยวัด จะต้องระบุให้ชัดเจนว่าเป็นหน่วยใด และ (1.5) สถานการณ์ที่ยกมาต้องสั้น กระชับรัด อ่านเข้าใจได้ง่าย แต่ละสถานการณ์ควรใช้สำหรับถามได้มากกว่า 1 ข้อ เพื่อให้มีให้นักเรียนเสียเวลาในการอ่านมากเกินไปจนความจำเป็น (2) การสร้างคำถาม คำถามที่จะให้ตอบตามสถานการณ์ที่ยกมาจะมีคุณสมบัติดังนี้ (2.1) ถามในเรื่องที่ต้องใช้ความสามารถในด้านกระบวนการวิทยาศาสตร์ ไม่ถามเรื่องที่เป็นความรู้-ความจำ (2.2) ไม่ถามถึงปัญหา หรือสมมติฐานที่เคยอภิปราย หรือสรุปกันมาแล้วเพราะจะกลายเป็นความจำทั้ง ๆ ที่คำถามเหมือนกับจะวัดทักษะกระบวนการทางวิทยาศาสตร์ (2.3) ใช้คำถามรัดกุม บังคับว่าจะให้ตอบเรื่องใด แม้ว่าบางคำถามจะให้เห็นความคิดเห็นได้แตกต่างกัน แต่ก็ต้องเป็นความคิดเห็นเกี่ยวกับเรื่องนั้น ๆ โดยเฉพาะ และ (2.4) ข้อความที่จะให้ตอบแต่ละคำถาม ควรเป็นตอนละเรื่อง และกำหนดระดับคะแนนให้เหมาะสม ถ้าเป็นไปได้ควรให้คะแนนเป็น 1 ถ้าตอบถูก และให้ 0 ถ้าตอบผิด (3) การตรวจให้คะแนน ถ้าเป็นข้อทดสอบให้ตอบสั้น ๆ แม้จะตั้งคำถามที่ผู้ถามคิดว่าจำเพาะเจาะจง คำตอบน่าจะแน่นอน แต่ในการตรวจจะต้องดูที่เหตุผลของนักเรียนบางคนที่ตอบแตกต่างไปจากเกณฑ์ที่ตั้งไว้ด้วย ถ้าเหตุผลถูกต้องก็ต้องยอมรับ

จากแนวคิดการประเมินผลทักษะกระบวนการทางวิทยาศาสตร์ดังกล่าว ผู้วิจัยจึงได้ใช้เป็นแนวทางในการสร้างแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ โดยกำหนดสถานการณ์ให้ตอบ คำตอบของผู้เรียนจะเป็นตัวบ่งชี้ถึงความสามารถด้านทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการที่เกิดขึ้นกับผู้เรียนในแต่ละทักษะ

2. หลักฐานแสดงความเที่ยงตรง

2.1 ความหมายของความเที่ยงตรง

ความเที่ยงตรงเป็นลักษณะที่สำคัญที่สุดของเครื่องมือวัดที่มีคุณภาพ การศึกษาเกี่ยวกับความเที่ยงตรงมีสองแนวคิด คือ แนวคิดเดิมกับแนวคิดปัจจุบัน แนวคิดเดิมได้อธิบายความเที่ยงตรงว่า เป็นคุณสมบัติของเครื่องมือวัด ถ้าเครื่องมือวัดสามารถให้ข้อมูลหรือคะแนนได้ตรงตามจุดมุ่งหมายของการวัด แสดงว่า เครื่องมือวัดนั้นมีความเที่ยงตรง เครื่องมือวัดที่มีความเที่ยงตรงสูง จะสามารถวัดในสิ่งที่ต้องการวัดได้ถูกต้องครอบคลุมมากกว่าเครื่องมือวัดที่มีความเที่ยงตรงต่ำ ความเที่ยงตรงของเครื่องมือวัดจึงเป็นระดับของความสอดคล้อง หรือความสัมพันธ์ระหว่างคะแนนที่ได้จากการวัดกับสิ่งที่ต้องการวัดนั่นเอง “สิ่งที่ต้องการวัด” ก็คือ “ตัวเกณฑ์” (Criterion) ดังนั้น “ความเที่ยงตรง” จึงเป็นความสัมพันธ์ระหว่างคะแนนที่ได้จากการทดสอบกับคะแนนที่ได้จากตัวแปรเกณฑ์ (บุญเชิด ภิญโญนนตพงษ์. 2545: 171)

ความเที่ยงตรงเป็นระบบหนึ่งของเรื่องราวหรือเหตุการณ์ที่เป็นหลักฐาน แต่ไม่ใช่เหตุการณ์ทั้งหมด นอกจากนี้ความเที่ยงตรงที่ได้จากพยานหลักฐานนั้น ๆ จะไม่มีความคงทนเพียงพอ เมื่อเวลาผ่านไปเหตุการณ์ที่เป็นพยานหลักฐานก็สามารถเปลี่ยนแปลงไปตามเงื่อนไขทางสังคมที่เปลี่ยนแปลงไป หรือเมื่อมีข้อค้นพบใหม่ ๆ เกิดขึ้น ความเที่ยงตรงจึงเป็นคุณสมบัติที่มีการพัฒนาและเป็นกระบวนการต่อเนื่อง ด้วยเหตุที่หลักฐานที่มีอยู่มากไม่ค่อยมีความสมบูรณ์ จึงมีความสำคัญยิ่งที่ในการหาความเที่ยงตรงนั้นจะต้องพยายามทำให้เกิดความเป็นเหตุเป็นผลให้มากที่สุด เพียงพอที่จะนำมาใช้ได้อย่างเป็นปัจจุบันทั้งในการทดสอบและการทำวิจัยต่าง ๆ (Messick, 1989: 13) ความเป็นเหตุเป็นผลนั้นจะเป็นสิ่งที่เชื่อถือได้หรือไม่ จะขึ้นอยู่กับว่าได้พยายามใช้วิธีการทางวิทยาศาสตร์เข้ามาช่วยในการแปลความหมายของคะแนนสามารถอ้างอิงไปสู่จุดมุ่งหมายที่ต้องการได้มากน้อยเพียงใด จากผลการตรวจสอบความเป็นเหตุเป็นผลโดยใช้หลักฐานอ้างอิงในการทดสอบแบบต่าง ๆ อาจต้องการหลักฐานที่แตกต่างกัน หลักฐานนี้ย่อมหาได้มาจากข้อมูลหรือข้อเท็จจริงต่าง ๆ ที่แสดงความเชื่อมโยงขององค์ประกอบต่าง ๆ ที่จะช่วยให้ข้อสนเทศในการพิจารณาตัดสินใจนำไปใช้ในการอ้างอิง สังเกตได้ว่าข้อมูลนี้ไม่ใช่ข้อสนเทศแต่เป็นสิ่งที่ทำให้เกิดข้อสนเทศ เพราะข้อสนเทศได้จากการแปลความหมายของข้อมูลนั่นเอง (จินตนา ธนวิบูลย์ชัย, 2535: 64)

การหาความเที่ยงตรงของเครื่องมือวัด จึงเป็นการตรวจหาความถูกต้อง แม่นยำของค่าวัดที่ใช้สำหรับพยากรณ์หรือลงความเห็นค่าวัดหรือข้อมูลที่ได้จากเครื่องมือวัด ซึ่งกล่าวให้กว้างขึ้นได้ว่า ในการหาความเที่ยงตรง เป็นการตรวจสอบความสมเหตุสมผลของการแปลความหมายคะแนนจากเครื่องมือวัด และการตรวจสอบดังกล่าว เป็นการเปรียบเทียบค่าที่วัดได้จากเครื่องมือวัดที่สร้างขึ้นกับตัวแปรที่ถือเป็นตัวแปรเกณฑ์ ดังนั้น เป้าหมายของการหาความเที่ยงตรง จึงทำเพื่อแปลความหมายของข้อมูลที่ได้จากการวัดว่าข้อมูลที่ได้นั้นเป็นตัวแทนที่ดีในการบรรยายคุณลักษณะและพยากรณ์คุณลักษณะอื่นได้ดีเพียงไร ความเที่ยงตรงจึงถือเป็นความสัมพันธ์ระหว่างค่าที่วัดได้จากเครื่องมือวัดกับค่าวัดของตัวแปรเกณฑ์ หรือระดับความถูกต้องแม่นยำที่เครื่องมือวัดจะวัดในสิ่งที่ต้องการวัด ถ้ามีความถูกต้องแม่นยำสูงก็แสดงว่าเครื่องมือวัดนั้นมีความเที่ยงตรงสูง (บุญเชิด ภิญญอนันตพงษ์, 2545: 171-172)

สรุปได้ว่า เครื่องมือวัดที่มีความเที่ยงตรง หมายถึง เครื่องมือวัดนั้นสามารถวัดในสิ่งที่ผู้วัดต้องการวัดได้อย่างถูกต้อง ดังนั้น ในแนวคิดเดิมจึงถือว่า ความเที่ยงตรง เป็นคุณภาพของเครื่องมือวัด ซึ่งเป็นคุณสมบัติที่สำคัญของเครื่องมือทุกชนิด ไม่ว่าจะเป็นแบบทดสอบผลสัมฤทธิ์ แบบทดสอบความถนัด มาตรฐานทัศนคติ จริยธรรม และบุคลิกภาพ เครื่องมือวัดทุกฉบับจะต้องมีคุณภาพด้านความเที่ยงตรง จึงจะเชื่อได้ว่าเป็นเครื่องมือวัดที่ดีและผลที่ได้จากการวัดจะถูกต้องตรงตามความต้องการ การวัดใด ๆ จะไร้ประโยชน์เมื่อเครื่องมือที่ใช้วัดนั้นขาดความเที่ยงตรง และความเที่ยงตรงของเครื่องมือวัด แบ่งเป็นหลายชนิด ขึ้นอยู่กับตัวแปรเกณฑ์ที่ใช้เทียบ (บุญเชิด ภิญญอนันตพงษ์, 2545: 172)

ส่วนแนวคิดปัจจุบัน ให้ความหมายของความเที่ยงตรงตามมาตรฐานทางเทคนิค สำหรับประเมินคุณภาพของเครื่องมือวัด โดยคณะกรรมการร่วมที่ประกอบด้วยสมาชิกของสมาคมคือ สมาคมวิจัยการศึกษาอเมริกัน (American Educational Research Association: AERA) สมาคมจิตวิทยาอเมริกัน (American Psychological Association: APA) และสภาการวัดผลการศึกษาแห่งชาติของอเมริกัน (National Council American on Measurement in Educational: NCME) ได้เสนอไว้ในหนังสือ “Standards for Educational and Psychological Testing” ที่พิมพ์ในปี 1985 ให้ความหมายความเที่ยงตรงไว้ว่า ความเที่ยงตรงเป็นลักษณะสำคัญที่สุดที่ใช้ในการประเมินเครื่องมือวัด ซึ่งเป็นเรื่องเกี่ยวกับการลงความเห็น (Inference) จากคะแนนที่วัดได้อย่างเหมาะสม อย่างมีความหมาย และให้ประโยชน์ ส่วนการตรวจสอบความเที่ยงตรง (Validation) เป็นกระบวนการในการรวบรวมพยานหลักฐาน (Evidence) เพื่อสนับสนุนการลงความเห็นดังกล่าว และเก็บรวบรวมหลักฐานได้หลายวิธี ความเที่ยงตรงจึงขึ้นอยู่กับปริมาณ และชนิดของหลักฐานที่สามารถสนับสนุนการลงความเห็นของผู้วัดจากข้อมูลที่วัดได้ (บุญเชิด ภิญโญนนตพงษ์. 2545: 172) เมสสิค (Popham. 1990 : 94; citing Messick. 1989) ได้ให้ความเห็นว่า ความเที่ยงตรงของแบบทดสอบมิได้เกิดจากตัวแบบทดสอบ หากแต่เกิดขึ้นจากการแปลความหมายของข้อมูลที่ได้จากวิธีการเฉพาะอันหนึ่ง จึงพิจารณาได้ว่า ความหมายของความเที่ยงตรงในระยะหลังได้เน้นไปที่การขยายการอ้างอิงไปสู่สิ่งที่แบบทดสอบต้องการจะวัด โดยผลที่ได้จากการวัดก็คือคะแนนที่เป็นพื้นฐานสำคัญของการอ้างอิง (จินตนา ธนวิบูลย์ชัย. 2535: 65)

กรอนลันและลินน์ (Gronlund; & Linn. 1990: 48-49) ได้กล่าวถึงธรรมชาติของความเที่ยงตรงว่า เมื่อใช้คำว่าความเที่ยงตรงในแง่ของการทดสอบและการประเมินจะต้องมีข้อควรคำนึงดังต่อไปนี้คือ (1) ความเที่ยงตรง หมายถึง ความเหมาะสมของการแปลความหมายของผลการทดสอบหรือการประเมินเครื่องมือสำหรับกลุ่มใดกลุ่มหนึ่ง ไม่ได้เป็นการประเมินที่ตัวเครื่องมือ บางครั้งเราอาจกล่าวว่า “ความเที่ยงตรงของแบบทดสอบ” ซึ่งที่จริงแล้วควรกล่าวว่า “ความเที่ยงตรงของการแปลความหมายที่เกิดขึ้นจากผลการประเมิน” จะถูกต้องมากกว่า (2) ความเที่ยงตรงเป็นขนาดของความเข้มข้น (Degree) ซึ่งไม่ได้ตอบคำถามว่ามีหรือไม่มี เราจึงไม่ควรจะกล่าวว่า เครื่องมือนี้มีความเที่ยงตรงหรือไม่มีความเที่ยงตรง แต่ควรกล่าวถึงในแง่ของความหมายของระดับความเข้มข้น เช่น มีความเที่ยงตรงสูง มีความเที่ยงตรงปานกลาง หรือมีความเที่ยงตรงต่ำ (3) ความเที่ยงตรงจะมีลักษณะเฉพาะเจาะจงในการแปลความหมาย ไม่มีแบบทดสอบใดที่มีความเที่ยงตรงในทุกวัตถุประสงค์ (4) การแสดงหลักฐานความเที่ยงตรงของแบบทดสอบมีหลายวิธี เนื่องจากความเที่ยงตรงมีหลายชนิด (5) ในปี 1995 ลินน์และกรอนลัน (Linn; & Gronlund. 1995: 49) ได้เพิ่มข้อคำนึงเกี่ยวกับธรรมชาติของความเที่ยงตรงอีก คือ ความเที่ยงตรงเกี่ยวข้องกับการตัดสินใจเชิงประเมินทั้งหมด (Overall Evaluative Judgment) กล่าวคือ เป็นความต้องการประเมินในระดับของความเข้มข้นในการตีความ (Interpretation)

และใช้ผลการประเมิน (Assessment) ว่า ได้รับการพิสูจน์ความถูกต้องจากการแสดงหลักฐานความเที่ยงตรง หลักเหตุผลของการตีความ และการใช้งานดังกล่าว

ดังนั้นสรุปได้ว่า ความเที่ยงตรง หมายถึง ระดับของหลักฐานที่สามารถสนับสนุนการลงความเห็นจากข้อมูลที่วัดจากเครื่องมือนั้น ซึ่งเป็นการลงความเห็นจากข้อมูลที่เกี่ยวข้องกับการใช้เครื่องมือวัดเฉพาะครั้งนั้น ความเที่ยงตรงจึงมีความหมายเดียว (Unitary Concept) แต่การลงความเห็นและการตรวจสอบความเที่ยงตรงอาจมีได้หลายชนิด แบ่งตามหลักฐานพยาน อาจแบ่งได้เป็นสามชนิดตามแนวคิดแบบเก่า คือ หลักฐานความเที่ยงตรงตามเนื้อหา (Content Validity Evidence) หลักฐานแสดงความเที่ยงตรงตามเกณฑ์ (Criterion Validity Evidence) และหลักฐานความเที่ยงตรงตามโครงสร้าง (Construct Validity Evidence)

2.2 ประเภทของหลักฐานแสดงความเที่ยงตรง

ในระยะแรกแนวคิดเกี่ยวกับความเที่ยงตรงนี้ได้มีการศึกษาอย่างมากมาโดยผู้ที่สนใจในเรื่องนี้ได้ศึกษากันไปตามความเชื่อของตนแบบต่างคนต่างหา จึงก่อให้เกิดความเที่ยงตรงประเภทต่างๆ ขึ้นอย่างหลากหลาย ดังนั้นเพื่อแก้ปัญหาความสับสนของความเที่ยงตรงดังกล่าว สมาคมวิชาชีพด้านการศึกษา 3 สมาคม ได้แก่ สมาคมวิจัยการศึกษาอเมริกัน (American Educational Research Association: AERA) สมาคมจิตวิทยาอเมริกัน (American Psychological Association: APA) และสภาการวัดผลการศึกษาแห่งชาติของอเมริกัน (National Council American on Measurement in Educational: NCME) จึงได้ร่วมมือกันจัดทำ Standards for Educational and Psychological Tests and Manuals ขึ้นในปี ค.ศ. 1966 และ 1974 จนกระทั่งเมื่อ ค.ศ. 1985 นี้ ได้มีการทบทวนใหม่อีกครั้งใช้ชื่อเรียกว่า Standards for Educational and Psychological Testing เป็นมาตรฐานที่ช่วยให้เกิดความชัดเจนในการสื่อสารและได้รับการยอมรับอย่างกว้างขวาง โดยได้แบ่งชนิดของความเที่ยงตรงออกเป็น 3 ประเภท เช่นเดียวกับที่เคยจัดแบ่งไว้ก่อนปี ค.ศ. 1985 แต่ได้มีการปรับเปลี่ยนถ้อยคำให้มีความแจ่มชัดในการอธิบายชนิดของความเที่ยงตรง ดังตาราง 1

ตาราง 1 เปรียบเทียบความเที่ยงตรงก่อนและหลัง ค.ศ. 1985

ก่อน ค.ศ. 1985	หลัง ค.ศ. 1985
1. ความเที่ยงตรงตามเนื้อหา (Content Validity)	1. หลักฐานแสดงความเที่ยงตรงตามเนื้อหา
2. ความเที่ยงตรงตามเกณฑ์ (Criterion – Related Validity)	(Content – Related Evidence of Validity)
3. ความเที่ยงตรงตามโครงสร้าง (Construct Validity)	2. หลักฐานแสดงความเที่ยงตรงตามเกณฑ์ (Criterion – Related Evidence of Validity)
	3. หลักฐานแสดงความเที่ยงตรงตามโครงสร้าง (Construct – Related Evidence of Validity)

ชื่อใหม่ของความเที่ยงตรงนี้ได้เน้นให้เห็นความจำเป็นของการรวบรวมหลักฐาน (evidence) ที่จะช่วยให้เราสามารถพิจารณาได้ว่า เราจะมีเชื่อมั่นในการอ้างอิงคะแนนได้มากน้อยเพียงใด ดังนั้น ผู้เชี่ยวชาญด้านการวัดผลส่วนมากในต่างประเทศจึงเชื่อว่า การปรับเปลี่ยนชื่อประเภทของความเที่ยงตรงเสียใหม่จะเป็นสิ่งที่ดีประการหนึ่ง (Popham. 1990: 113)

2.2.1 หลักฐานแสดงความเที่ยงตรงตามเนื้อหา (Content – Related Evidence of Validity)

หลักฐานแสดงความเที่ยงตรงตามเนื้อหาหมายถึง การลงความเห็นเกี่ยวกับความสอดคล้องระหว่างเนื้อหาของเครื่องมือวัดกับเนื้อหาของเกณฑ์ ถ้ามีความสอดคล้องกันมากแสดงว่า มีหลักฐานแสดงความเที่ยงตรงตามเนื้อหาสูง เพราะเนื้อหาของเครื่องมือวัดครอบคลุมเนื้อหาที่เป็นตัวเกณฑ์ เครื่องมือวัดที่แสดงความเที่ยงตรงตามเนื้อหาสูง เครื่องมือวัดนั้นจะต้องมีเนื้อหา ข้อคำถามหรือเนื้อหาที่นำมาสร้างเป็นข้อคำถามแต่ละข้อครอบคลุมเนื้อหาอย่างครบถ้วน ดังนั้น จึงกล่าวได้ว่า หลักฐานแสดงความเที่ยงตรงตามเนื้อหาเป็นหลักฐานที่แสดงให้เห็นการสุมเนื้อหามาสร้างเครื่องมือวัดว่า เนื้อหาของเครื่องมือวัดที่สุมเลือกมานั้นเป็นตัวอย่างที่เป็นตัวแทนที่ดีของตัวเกณฑ์ หรือกล่าวอีกอย่างหนึ่งได้ว่า เนื้อหาของเครื่องมือวัดเป็นตัวแทนของมวลประชากรของเนื้อหาหรือประชากรพฤติกรรมที่ต้องการวัดเพียงไร ทั้งนี้เพราะเครื่องมือวัดแต่ละฉบับไม่สามารถวัดประชากรความรู้ตามเนื้อหาได้ทั้งหมด ตัวเครื่องมือวัดเป็นเพียงกลุ่มตัวอย่างของมวลประชากรความรู้เท่านั้น

คำว่า “เนื้อหา” ซึ่งเป็นตัวเกณฑ์ของความเที่ยงตรงตามเนื้อหานี้ มิได้หมายถึงเฉพาะเนื้อหาวิชาเพียงเดียว แต่ยังหมายถึงพฤติกรรมแสดงออกตามเนื้อหานั้นด้วย หลักฐานแสดงความเที่ยงตรงตามเนื้อหาจึงเป็นการแสดงความเที่ยงตรงทั้งต่อรายละเอียดของเนื้อหาสาระวิชาและความเที่ยงตรงต่อพฤติกรรมที่คาดหวังในเนื้อหาวิชานั้นด้วย

วิธีการแสดงหลักฐานความเที่ยงตรงตามเนื้อหา เป็นการแสดงหรือหาว่าเครื่องมือวัดนั้นสามารถวัดได้ตรงและครอบคลุมเนื้อหาวิชา การแสดงหลักฐานความเที่ยงตรงตามเนื้อหาจะใช้ดุลยพินิจของผู้เชี่ยวชาญตรวจสอบความเที่ยงตรงว่า ข้อสอบมีความเที่ยงตรงมากน้อยเพียงใด โดยการเทียบกับตารางวิเคราะห์หลักสูตร หรือตารางกำหนดข้อสอบ (Test Blueprint) ซึ่งกำหนดตัวอย่างหัวข้อเนื้อหาสาระวิชาและพฤติกรรมจากเนื้อหาสาระวิชาทั้งหมด และถือได้ว่าเป็นตัวแทนที่ดี การพิจารณาโดยอาศัยหลักฐานผลการตัดสินของผู้เชี่ยวชาญในคุณลักษณะที่ต้องการจะวัดนั้น ถือได้ว่าเป็นหลักฐานที่มีความเป็นอัตนัยมาก (จินตนา ธนวิบูลย์ชัย. 2535: 71) อย่างไรก็ตามหลักฐานที่มีความจำเป็นสำหรับความเที่ยงตรงประเภทนี้คือ ความรอบคอบในการสร้างและคัดเลือกสิ่งบ่งชี้หรือข้อคำถาม ความรอบคอบในการตัดสินใจของผู้เชี่ยวชาญ และวิธีการที่ใช้ในการตรวจสอบโครงสร้างทางจิตวิทยาที่กล่าวมาแล้ว (Walsh; & Betz. 1990:

61) สิ่งที่ยังระลึกไว้เสมอก็คือข้อสรุปที่ได้จากการตรวจสอบความเที่ยงตรงนี้จะมีลักษณะเป็นเพียงว่าแบบทดสอบที่สร้างขึ้นนั้นดูเหมือนว่าจะวัดคุณลักษณะที่ต้องการวัดเท่านั้นยังสรุปไม่ได้ว่าสามารถวัดคุณลักษณะนั้นได้อย่างแท้จริง แต่การหาความเที่ยงตรงของหลักฐานตามเนื้อหาที่น่าับได้ว่าควรเป็นขั้นตอนที่สำคัญแรกของการหาความเที่ยงตรงชนิดอื่น ๆ

2.2.2 หลักฐานแสดงความเที่ยงตรงตามเกณฑ์ (Criterion – Related Evidence of Validity)

หลักฐานแสดงความเที่ยงตรงตามเกณฑ์ของเครื่องมือวัด เป็นการหาความสัมพันธ์ระหว่าง คะแนนผลการวัดจากเครื่องมือที่สร้างกับค่าที่วัดได้จากเกณฑ์ภายนอก ถ้าเกณฑ์ภายนอกนั้นเป็นเกณฑ์ที่วัดได้ในปัจจุบัน ก็เรียกว่า หลักฐานแสดงความเที่ยงตรงตามสภาพ (Concurrent Validity Evidence) แต่ถ้าเป็นเกณฑ์ที่เก็บรวบรวมได้ในอนาคต ก็เรียกว่า หลักฐานแสดงความเที่ยงตรงตามพยากรณ์ (Predictive Validity Evidence) การแสดงหลักฐานความเที่ยงตรงทั้งสองชนิดนี้ เป็นการหาความสัมพันธ์ระหว่างคะแนนของเครื่องมือวัดที่สร้างขึ้นกับค่าที่วัดได้จากตัวเกณฑ์ (บุญเชิด ภิญญอนันตพงษ์. 2545: 174) ขนาดของความสัมพันธ์ระหว่างคะแนนที่ได้จากการวัดคุณลักษณะที่เราสนใจกับพฤติกรรมที่เราสนใจจะพยากรณ์นี้มีอยู่ด้วยกัน 2 ลักษณะ (จินตนา ธนวิบูลย์ชัย. 2535: 73) คือ

2.2.2.1 หลักฐานแสดงความเที่ยงตรงตามสภาพ

หลักฐานแสดงความเที่ยงตรงตามสภาพเป็นการหาระดับความสัมพันธ์หรือระดับความสอดคล้องระหว่างคะแนนการวัดของเครื่องมือวัดในปัจจุบันกับตัวเกณฑ์ในปัจจุบัน หรือสภาพที่แท้จริงในปัจจุบันของนักเรียน (บุญเชิด ภิญญอนันตพงษ์. 2545: 174) ความเที่ยงตรงชนิดนี้จะเกิดขึ้นเมื่อทำการตรวจสอบความสัมพันธ์ระหว่างคะแนนที่ได้จากการวัดกับพฤติกรรมที่เราสนใจจะพยากรณ์ซึ่งเป็นพฤติกรรมที่เกิดขึ้นในปัจจุบัน

2.2.2.2 หลักฐานความเที่ยงตรงตามพยากรณ์

หลักฐานแสดงความเที่ยงตรงตามพยากรณ์ เป็นการหาความสัมพันธ์ระหว่างคะแนนของเครื่องมือวัดที่ต้องการหาความเที่ยงตรงกับตัวเกณฑ์ที่เกิดขึ้นในอนาคตว่ามีความสัมพันธ์กันเพียงไร เพื่อจะเอาผลการวัดจากเครื่องมือไปพยากรณ์ความสำเร็จในอนาคต ความเที่ยงตรงชนิดนี้จะเกิดขึ้นเมื่อทำการตรวจสอบความสัมพันธ์ระหว่างคะแนนที่ได้จากการวัดกับพฤติกรรมที่เราสนใจจะพยากรณ์ซึ่งเป็นพฤติกรรมที่จะเกิดขึ้นในอนาคต

หลักฐานความเที่ยงตรงตามเกณฑ์ที่ค่อนข้างเป็นปัญหามากคือ หลักฐานของความเที่ยงตรงตามพยากรณ์เพราะส่วนมากจะนิยมนำไปใช้ในการคัดเลือกผู้สมัครเข้าศึกษาต่อหรือผู้สมัครเข้าทำงาน ซึ่งการคัดเลือกในแต่ละกรณีย่อมมีความแตกต่างกันตามจุดมุ่งหมาย

ของการคัดเลือก แบบทดสอบที่นำมาใช้จึงมีความเที่ยงตรงตามพยากรณ์ที่แตกต่างกัน นอกจากนี้ โดยทั่วไปเราจะไม่สามารถสร้างความเที่ยงตรงของหลักฐานตามพยากรณ์ได้อย่างเต็มที่ ทั้งนี้ ก็เพราะว่าเครื่องมือหรือแบบทดสอบที่สร้างขึ้นนั้นมักไม่ได้ใช้ในการทำนายหากแต่แท้จริงแล้วนำไปใช้ในการคัดเลือกบุคคลว่าผู้ใดบ้างที่ควรจะได้รับและผู้ใดบ้างที่ไม่ควรจะได้รับเข้ามา (จินตนา ธนวิบูลย์ชัย. 2535: 73)

บุญเชิด ภิญโญนนตพงษ์ (2545: 176) กล่าวว่า มักมีข้อโต้แย้งอยู่เสมอในการเลือกใช้เกณฑ์เพื่อพิจารณาความสำเร็จเนื่องจากตัวพยากรณ์แต่ละตัวมีสหสัมพันธ์กับคะแนนเกณฑ์แต่ละตัวแตกต่างกัน ทำให้การแสดงผลหลักฐานค่าความเที่ยงตรงตามพยากรณ์มีความแตกต่างกันและไม่มีการแสดงผลหลักฐานความเที่ยงตรงตามพยากรณ์ได้ดีที่สุด ดังนั้น จำเป็นต้องใช้การสะสมการแสดงผลหลักฐานความเที่ยงตรงสูง จึงจะแปลความหมายได้ด้วย ความมั่นใจและนำไปสู่การตัดสินใจที่มีประสิทธิภาพสูง ความเที่ยงตรงตามพยากรณ์จะสูงหรือต่ำขึ้นอยู่กับความสอดคล้องกันระหว่างค่าจากเครื่องมือวัดกับคะแนนของตัวเกณฑ์ ฉะนั้นถ้าตัวใดไม่มีความเที่ยงตรงและความเชื่อมั่นต่ำก็จะทำให้ค่าการแสดงผลหลักฐานความเที่ยงตรงตามพยากรณ์ต่ำไปด้วย

ส่วนการแสดงผลหลักฐานความเที่ยงตรงตามสภาพ ก็มักมีปัญหาในแง่การแปลความหมาย เพราะดัชนีที่ใช้วัดระดับความสัมพันธ์ เป็นค่าสหสัมพันธ์ระหว่างตัวแปรที่ใช้วัดกับตัวเกณฑ์ ถ้าตัวเกณฑ์ที่วัดได้นั้นเป็น “ตัวเกณฑ์ที่แท้จริง” ค่าสหสัมพันธ์ที่ได้จะแปลความหมายได้มั่นใจว่าตรงกับเกณฑ์นั้นเพียงไร แต่ถ้าค่าวัดที่แทนตัวเกณฑ์นั้นมีความเที่ยงตรงต่ำกล่าวคือ ค่าวัดตัวเกณฑ์ที่ได้มาไม่สามารถเป็นตัวแทนที่ดีของตัวเกณฑ์แล้ว ก็ยากที่จะแปลความหมายของความเที่ยงตรงนี้ได้ ค่าความเที่ยงตรงตามสภาพนี้จะมีความหมายดีก็ต่อเมื่อ ค่าวัดทั้งคู่เป็นตัวแทนที่ดีของตัวแปรนั้น การแสดงผลหลักฐานความเที่ยงตรงตามสภาพจึงจะมีความหมายดี การที่ได้ค่าความเที่ยงตรงระหว่างตัวแปรคู่นี้ต่ำอาจมีความหมายได้หลายอย่าง เช่น ค่าวัดจากตัวเครื่องมือวัดไม่ตรง หรือตัวเกณฑ์มีค่าวัดไม่ตรง หรือไม่ตรงทั้งคู่ ดังนั้น การแสดงผลหลักฐานความเที่ยงตรงตามสภาพจึงต้องคำนึงถึงลักษณะของเกณฑ์ที่ใช้ว่ามีความเกี่ยวข้องกันหรือไม่เพียงไร (บุญเชิด ภิญโญนนตพงษ์. 2545: 175-176)

ความยากลำบากอีกประการที่เกิดขึ้นในการใช้ความเที่ยงตรงของหลักฐานตามเกณฑ์ทั้ง 2 ชนิดก็คือ การขาดตัวแปรที่เหมาะสมกับสังกัดต่าง ๆ ที่นักวิจัยสนใจจะศึกษาสังกัดใด ๆ ที่นำมาศึกษานั้นถ้ายังมีความเป็นนามธรรมมากขึ้นก็จะมีควมลำบากมากขึ้นที่จะสร้างเกณฑ์ที่เหมาะสม ดังนั้น ในขณะที่หลักฐานความเที่ยงตรงตามเกณฑ์นี้จะปรากฏให้ดูเหมือนว่าเป็นวิธีการที่ดึงดูดใจในการหาระดับความเป็นตัวแทนของสิ่งที่ต้องการจะวัด แต่คุณค่านี้ก็เพียงภาพลวงตาในเบื้องต้นเท่านั้น (Zeller. 1988: 326) เนื่องจากการตอบสนองทางทฤษฎีความเที่ยงตรงโดยแท้จริงเป็นสิ่งที่ทำได้ลำบากยิ่ง

วิธีการแสดงหลักฐานความเที่ยงตรงตามเกณฑ์ ขึ้นอยู่กับข้อมูลเชิงประจักษ์ที่ได้จากแนวทางใหญ่ ๆ 2 แนวทาง คือ การพิจารณาระดับความสัมพันธ์ระหว่างคะแนนที่ได้จากการวัดคุณลักษณะที่สนใจกับคะแนนเกณฑ์ ส่วนอีกแนวทางหนึ่งนั้นพิจารณาจากความแตกต่างระหว่างกลุ่มข้อมูล ข้อมูลเชิงประจักษ์ที่ได้มาเช่นนี้นั้น ความสนใจไม่ได้อยู่ที่ความสัมพันธ์ระหว่างคะแนนสอบกับผลการวัดต่าง ๆ หลาย ๆ ตัว ซึ่งเป็นโครงสร้างภายนอกที่เกี่ยวข้องกับคุณลักษณะที่วัด แต่ค่อนข้างจะสนใจแค่ว่าความสัมพันธ์ระหว่างคะแนนสอบกับเกณฑ์บางอย่างที่มีจุดมุ่งหมายในการนำไปใช้เฉพาะด้าน การหาหลักฐานข้อมูลเชิงประจักษ์ทั้ง 2 แนวทางดังกล่าวมีดังนี้ (จินตนา ธนวิบูลย์ชัย, 2535: 77)

การหาความสัมพันธ์

การแสดงผลหลักฐานความเที่ยงตรงตามเกณฑ์ เป็นการหาค่าความสัมพันธ์ของค่าที่วัดได้จากเครื่องมือวัดที่ต้องการกับค่าที่วัดได้จากเกณฑ์นิยมใช้สหสัมพันธ์ของเพียร์สัน (Pearson Product Moment Correlation) วิเคราะห์ข้อมูล ในกรณีที่ระหว่างตัวแปรสองตัวหาความสัมพันธ์นั้นมีข้อตกลงเบื้องต้นว่าสัมพันธ์กันเป็นเส้นตรง แต่ถ้าเมื่อใดมีความสัมพันธ์เป็นเส้นโค้งก็ควรใช้สัมประสิทธิ์ เอต้า (Eta Coefficient) เป็นตัวบ่งชี้ความสัมพันธ์ เช่น ระดับความวิตกกังวล หากสูงมาก ๆ หรือต่ำมาก ๆ ก็จะส่งผลกระทบต่อผลการสอบโดยมีความสัมพันธ์เป็นเส้นโค้ง ดังนั้น ผู้ใช้จะต้องระมัดระวังในเรื่องนี้ด้วยมิใช่จะใช้สหสัมพันธ์ของเพียร์สันทุกครั้ง ค่าสหสัมพันธ์ที่ได้นี้ยังไม่สามารถนำไปใช้พยากรณ์ตัวแปรเกณฑ์ได้ในทันทีเพราะเป็นค่าที่ใช้ความเกี่ยวข้องระหว่างสองตัวแปร แต่ไม่ได้บอกว่าตัวแปรหนึ่งเป็นสาเหตุให้เกิดอีกตัวแปรหนึ่ง จึงยังไม่ควรกล่าวถึงความสัมพันธ์ในลักษณะเป็นเหตุเป็นผลต่อกัน นอกจากนี้ ความมีนัยสำคัญทางสถิติของสัมประสิทธิ์สหสัมพันธ์ก็เป็นสิ่งที่มีความสำคัญต่อการนำค่าสหสัมพันธ์ไปใช้

การพิจารณาความแตกต่างระหว่างกลุ่มข้อมูล

การแสดงผลหลักฐานความเที่ยงตรงตามเกณฑ์ในลักษณะนี้จะพิจารณาว่าคะแนนสอบที่ได้นั้นสามารถแยกแยะกลุ่มออกได้ตามเกณฑ์ที่กำหนด กล่าวคือ ถ้าจะใช้คะแนนสอบในการพยากรณ์ความสำเร็จหรือความล้มเหลวในการศึกษาของบุคคล แบบวัดที่มีความเที่ยงตรงตามเกณฑ์จะต้องแยกกลุ่มคนที่สามารถสำเร็จการศึกษาและคนที่ล้มเหลวได้ตามที่พยากรณ์ไว้ โดยทั่วไปในการคำนวณหาค่าเฉลี่ยของคะแนนแบบวัดภายในกลุ่มต่าง ๆ แล้วนำค่าเฉลี่ยมาเปรียบเทียบกับกันนั้น จะใช้ t-test หรือ ANOVA ในการสรุปการเปรียบเทียบความแตกต่างอย่างมีนัยสำคัญ ข้อมูลที่ได้จากการเปรียบเทียบความแตกต่างนี้สามารถนำมาใช้หาความเที่ยงตรงเชิงสภาพและเชิงทำนายได้เช่นเดียวกับการหาความสัมพันธ์ สำหรับในการทำนายนี้นั้นข้อมูลความแตกต่างของกลุ่มไม่เพียงแต่จะทำให้เห็นระดับความแตกต่างระหว่างค่าเฉลี่ยของกลุ่มเท่านั้น แต่ยังทำให้เห็นการซ้อนทับกัน (overlap) ระหว่างคะแนนของทั้ง 2 กลุ่มได้ด้วย ในการทำนาย

จึงสนใจพิจารณาส่วนที่ซ้อนทับกันมากกว่าการพิจารณาช่วงห่างการกระจายของคะแนนทั้งหมดจากการศึกษาของ ดิลตัน (จินตนา ธนวิบูลย์ชัย. 2535: 77; อ้างอิงจาก Tilton. 1937) ถ้าซ้อนทับกันถึงร้อยละ 75 จะทำให้ทำนายได้น้อย แต่ถ้าต่ำกว่าร้อยละ 50 ก็จะมีประโยชน์ในการทำนายได้ (Walsh; & Betz. 1990: 66) การหา ข้อมูลเชิงประจักษ์ของความเที่ยงตรงตามเกณฑ์ นอกจาก 2 วิธีหลักที่กล่าวมาแล้ว ยังสามารถหาได้จากวิธีการอื่น ๆ เช่น ไบซีเรียล (biserial) หรือ พอยท์ไบซีเรียล (point biserial) การใช้เทคนิคัลทวาริเอท (Multivariate) และการใช้การออกแบบการทดลองโดยใช้การวิเคราะห์ความแปรปรวน เป็นต้น

2.2.3 หลักฐานแสดงความเที่ยงตรงตามโครงสร้าง (Construct – Related

Evidence of Validity)

หลักฐานแสดงความเที่ยงตรงตามโครงสร้าง เป็นการแสดงหลักฐานความเที่ยงตรงว่า เครื่องมือวัดนั้นสามารถวัดภายใต้ขอบเขต ความหมาย หรือคุณลักษณะประจำตามโครงสร้างทางทฤษฎีที่สมมติขึ้นนั้นได้เพียงใด คำว่า “โครงสร้าง” (Construct) ในที่นี้หมายถึง ตัวประกอบ หรือองค์ประกอบ (Factor) จึงเรียกความเที่ยงตรงตามโครงสร้างนี้อีกอย่างหนึ่งว่า “หลักฐานแสดงความเที่ยงตรงตามองค์ประกอบ” แต่ละโครงสร้างต้องเกี่ยวกับทฤษฎี ซึ่งอธิบายและพยากรณ์พฤติกรรมของมนุษย์ การแสดงหลักฐานความเที่ยงตรงตามโครงสร้าง จึงเป็นการหาความสอดคล้องระหว่างองค์ประกอบที่วัดในเครื่องมือวัด และองค์ประกอบที่ต้องการวัดว่า วัดในองค์ประกอบเดียวกันหรือไม่ นั่นคือ วัดลักษณะทางจิตวิทยาหรือคุณสมบัติตามที่ต้องการหรือไม่ และปริมาณที่วัดแต่ละองค์ประกอบเป็นสัดส่วนสอดคล้องกับที่ต้องการหรือไม่ (บุญเชิด ภิญโญนนตพงษ์. 2545: 177) ความเที่ยงตรงชนิดนี้จะเกี่ยวกับความสัมพันธ์ระหว่างคะแนนสอบกับการทำนายคุณลักษณะตามทฤษฎีบางอย่าง ที่ตอบคำถามว่าแบบทดสอบนั้นๆ สามารถวัดได้ในสิ่งที่ต้องการจะวัดหรือไม่ ความเที่ยงตรงตามโครงสร้างจะสามารถตอบคำถามนี้ได้มากกว่าความเที่ยงตรง 2 ชนิดแรกที่กล่าวมาแล้ว การทำความเข้าใจกับหลักฐานความเที่ยงตรงตามโครงสร้างนี้มีความจำเป็นอย่างยิ่งต้องระลึกไว้เสมอว่า ตัวแบบทดสอบไม่ใช่สิ่งที่เราสนใจ แต่ความสนใจจะอยู่ที่แบบทดสอบได้สะท้อนปรากฏการณ์ในโลกของความเป็นจริงบางประการ หรือคุณสมบัติหรือพฤติกรรมบางอย่างที่เราเห็นว่ามีสำคัญต่อปรากฏการณ์หรือคุณสมบัติหรือพฤติกรรมอื่น ๆ นั้นได้หรือไม่ เพราะปรากฏการณ์หรือคุณสมบัติหรือพฤติกรรมต่าง ๆ มีลักษณะที่เป็นระเบียบแบบแผนและมีความเป็นนามธรรม ไม่สามารถสังเกตได้โดยตรงด้วยประสาทสัมผัสทั้ง 5 แต่ต้องอาศัยการอ้างอิงหรือสันนิษฐานเอาจากปรากฏการณ์อื่น ๆ ที่สังเกตได้ สิ่งนี้เองเราเรียกว่า Construct การวัด Construct ทางจิตวิทยานับเป็นเรื่องที่ยาก เพราะ Construct เกิดขึ้นจากกระบวนการทางจิตซึ่งอยู่ภายในตัวบุคคล การวัด Construct จึงกระทำโดยอ้อมเพื่ออธิบายบุคคลโดยยึดตามทฤษฎีคุณลักษณะนั้น ๆ เช่น การวัดความสามารถทางสติปัญญาซึ่งเป็น Construct หนึ่ง ซึ่งตามทฤษฎีคุณลักษณะของสติปัญญาจะแสดงระดับความ

แตกต่างกันในบุคคลต่าง ๆ จึงมีการพัฒนาแบบทดสอบวัดไอคิว (IQ) ขึ้นมา ซึ่งเป็นการสะท้อนความสามารถทางสติปัญญาของบุคคลโดยทางอ้อม ดังนั้น ในการแสดงหลักฐานความเที่ยงตรงตามโครงสร้างก็จะเป็นการรวบรวมข้อมูลมาสนับสนุนว่าแบบทดสอบวัดไอคิวนั้นสามารถสะท้อน Construct ของสติปัญญาได้อย่างแท้จริง (จินตนา ธนวิบูลย์ชัย. 2535: 80)

จะสร้างเครื่องมือวัดให้มีหลักฐานแสดงความเที่ยงตรงตามโครงสร้าง จะต้องให้คำนิยามปฏิบัติการ (Operational Definition) ของคำนั้นโดยอาศัยทฤษฎีใดทฤษฎีหนึ่งเป็นหลัก แล้วจึงลงมือสร้างเครื่องมือวัดสิ่งนั้นตามคำนิยามปฏิบัติการ ความหมายของคำว่า นิยามปฏิบัติการ ประกอบด้วยสามส่วน คือ คุณลักษณะ (Trait) ที่ต้องการวัด สิ่งเร้าหรือสถานการณ์ที่จะกระตุ้นให้คุณลักษณะนั้นแสดงออกมา และการตอบสนองแสดงออกที่สามารถสังเกตเห็นได้ เมื่อสามารถให้คำนิยามปฏิบัติการในสิ่งที่ต้องการจะทดสอบได้แล้ว จึงสร้างข้อคำถามขึ้นตามนิยามปฏิบัติการที่ให้นั้น และนำผลจากการทดสอบไปคำนวณความเที่ยงตรงเพื่อการแสดงหลักฐานความเที่ยงตรงตามโครงสร้างต่อไป หลักฐานความเที่ยงตรงตามโครงสร้างจึงมีลักษณะคล้ายกับหลักฐานความเที่ยงตรงตามเนื้อหาและหลักฐานความเที่ยงตรงตามเกณฑ์ แต่ความแตกต่างอยู่ที่หลักฐานความเที่ยงตรงตามโครงสร้างต้องสะสมรายละเอียดจากหลาย ๆ แหล่งต่าง ๆ กันมากกว่า เพื่อนำมารวมเป็นคำบรรยายพฤติกรรมได้อย่างกว้างขวางกว่า คงทนกว่า และมีลักษณะความเป็นนามธรรมมากกว่า การกำหนดเกณฑ์เพียงเกณฑ์เดียว จึงไม่เพียงพอที่จะอธิบายคุณลักษณะของโครงสร้างได้ จึงต้องอาศัยรายละเอียดที่มากกว่าหรือต้องอาศัยเกณฑ์หลายเกณฑ์ (บุญเชิด ภิญโญนนตพงษ์. 2545: 178)

วิธีการแสดงหลักฐานความเที่ยงตรงตามโครงสร้างนั้น เป็นการแสดงหลักฐานความเที่ยงตรงว่า เครื่องมือวัดนั้นสามารถวัดขอบเขต ความหมาย หรือคุณลักษณะประจำตามโครงสร้างทางทฤษฎีที่สมมติขึ้นนั้นได้เพียงใด การแสดงหลักฐานความเที่ยงตรงตามโครงสร้างสามารถทำได้หลายวิธีการ ดังนี้

2.2.3.1 วิธีพิจารณาเทียบกับโครงสร้างที่กำหนด

เครื่องมือวัดผลการเรียนที่เขียนข้อสอบวัดตามตารางลักษณะเฉพาะ หรือตารางวิเคราะห์หลักสูตรสามารถแสดงหลักฐานความเที่ยงตรงตามโครงสร้างได้ ให้ผู้เชี่ยวชาญวิเคราะห์ว่า ข้อสอบแต่ละข้อเขียนวัดได้ตรงตามพฤติกรรมในตารางวิเคราะห์หลักสูตรหรือไม่ และจำนวนข้อสอบเหล่านั้นมีสัดส่วนเป็นไปตามตารางวิเคราะห์หลักสูตร การใช้ดุลพินิจดังกล่าวสรุปได้ว่า แบบทดสอบวัดผลการเรียนนี้มีหลักฐานแสดงความเที่ยงตรงตามโครงสร้างของการวัดด้านสติปัญญา

2.2.3.2 วิธีเปรียบเทียบจากกลุ่มที่ต่างกัน

การศึกษาว่าเครื่องมือวัดโครงสร้างของสิ่งที่จะวัดได้นั้น ใช้กลุ่มตัวอย่างที่ต่างกันสองกลุ่มที่รู้แจ้งชัดว่า กลุ่มหนึ่งมีคุณลักษณะในสิ่งที่ต้องการวัด ส่วนอีกกลุ่มหนึ่งไม่มีคุณลักษณะในสิ่งนั้น แล้วเปรียบเทียบคะแนนที่ได้จากทั้งสองกลุ่ม โดยใช้สถิติ t-test ทดสอบนำค่า t-test ที่คำนวณได้ไปเทียบกับ ค่า t ที่เปิดจากตารางการแจกแจงของ t ที่มีชั้นความอิสระเท่ากับ n_1+n_2 ซึ่งใช้เป็นค่าเกณฑ์เทียบ ถ้าค่า t ที่คำนวณได้มีค่ามากกว่าค่าเกณฑ์ที่ใช้เทียบ แสดงว่า คะแนนที่ได้จากทั้งสองกลุ่มแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ก็สามารถสรุปว่าเครื่องมือวัดนั้นมีหลักฐานแสดงความเที่ยงตรงตามโครงสร้างสูง

2.2.3.3 วิธีเทียบกับเครื่องมือมาตรฐานที่วัดคุณลักษณะเดียวกัน

ค่าสหสัมพันธ์ของเครื่องมือวัดกับเครื่องมือมาตรฐานที่วัดคุณลักษณะเดียวกันสามารถบ่งชี้หลักฐานความเที่ยงตรงตามโครงสร้างได้ ดังนั้นการแสดงผลหลักฐานความเที่ยงตรงตามโครงสร้าง สามารถทำได้โดยการนำเครื่องมือวัดที่ต้องการกับเครื่องมือที่วัดในคุณลักษณะเดียวกันที่เป็นมาตรฐานแล้วไปสอบกับกลุ่มตัวอย่างกลุ่มเดียวกัน แล้วจึงหาค่าสัมประสิทธิ์สหสัมพันธ์ของข้อมูลทั้งสองชุด

2.2.3.4 วิธีการหาค่าความสอดคล้องภายในเครื่องมือวัด

การแสดงผลหลักฐานความเที่ยงตรงตามโครงสร้างเป็นวิธีการที่ต้องอาศัยสหสัมพันธ์ระหว่างเครื่องมือวัดกับเกณฑ์ภายนอกที่ยอมรับได้ สำหรับวิธีนี้จะอาศัยความสอดคล้องภายในเครื่องมือวัดโดยไม่ใช้เกณฑ์ภายนอก สามารถพิจารณาจากดัชนีต่าง ๆ ดังนี้

1) พิจารณาจากดัชนีอำนาจจำแนกรายข้อเพราะข้อสอบที่มีค่าอำนาจจำแนกสูง เป็นข้อสอบที่วัดในทิศทางเดียวกันกับส่วนรวม ถือว่ามีหลักฐานความเที่ยงตรงตามโครงสร้างภายในสูง

2) พิจารณาจากระดับความสัมพันธ์ระหว่างคะแนนส่วนย่อยภายในเครื่องมือวัดกับคะแนนรวม

3) พิจารณาจากค่าความเชื่อมั่นของเครื่องมือวัดที่หาด้วยสูตรความสอดคล้องภายใน เช่น สูตร KR-20 หรือสูตร แอลฟา ของครอนบัค (Cronbach's Coefficient-Alpha) ดังนั้น เครื่องมือวัดได้มีค่าความเชื่อมั่นสูง ก็สามารถสรุปว่า มีหลักฐานความเที่ยงตรงตามโครงสร้างภายในสูงได้

2.2.3.5 วิธีการวิเคราะห์องค์ประกอบ (Factor Analysis)

การแสดงผลฐานความเที่ยงตรงตามโครงสร้างที่ตรงประเด็นมากที่สุด คือ การวิเคราะห์องค์ประกอบ เพราะเป็นวิธีการทางสถิติที่สามารถตรวจเช็คลักษณะประจำทางจิตวิทยา เนื่องจากตัวแปรต่าง ๆ เมื่อนำมาหาค่าสัมประสิทธิ์สหสัมพันธ์ จะพบว่า มีตัวแปรบางคู่มีความสัมพันธ์กันสูง หรือบางทีก็พบว่า มีกลุ่มตัวแปรบางกลุ่มมีความสัมพันธ์ซึ่งกันและกันสูง นั่นแสดงว่า ตัวแปรเหล่านั้นวัดบางสิ่งบางอย่างที่เป็นองค์ประกอบร่วมกัน การวิเคราะห์องค์ประกอบเป็นการจัดสมรรถภาพหรือคุณลักษณะต่าง ๆ ทางจิตวิทยาที่วัดได้ให้เป็นหมวดหมู่ตามโครงสร้าง คำนำนัยขององค์ประกอบแรกก่อนหมุนแกน จะเป็นค่าที่แสดงถึงหลักฐานความเที่ยงตรงตามโครงสร้างได้

2.2.3.6 วิธีการวิเคราะห์หลายลักษณะหลายวิธี (Multitrait-Multimethods)

การแสดงผลฐานความเที่ยงตรงตามโครงสร้างแตกต่างจากการแสดงผลฐานความเที่ยงตรงตามเกณฑ์ที่ต้องใช้หลักฐานต่าง ๆ มากกว่า โดยอาศัยสมมติฐานที่ว่า ถ้าเครื่องมือวัดกับเกณฑ์มีลักษณะร่วมกันจะมีค่าสหสัมพันธ์กันสูง และถ้าเครื่องมือวัดกับเกณฑ์มีลักษณะต่างกันจะมีค่าสหสัมพันธ์กันต่ำ นำมาวิเคราะห์พร้อมกัน ซึ่งแคมเบลและฟิสส์ (Cammell; & Fiske.1959) ได้พัฒนาแนวคิดนี้ให้เหมาะสมเรียกว่า การวิเคราะห์หลายลักษณะหลายวิธีการ โดยใช้วิธีการเทียบความเที่ยงตรงร่วม (Convergent Validity) กับความเที่ยงตรงแยก (Divergent Validity) ซึ่งความเที่ยงตรงร่วมควรมีค่าสูงกว่า และความเที่ยงตรงแยกควรมีค่าต่ำกว่า แต่ค่าความเชื่อมั่นของเครื่องมือวัดแต่ละวิธีควรมีค่าสูงสุด

การแสดงผลฐานความเที่ยงตรงที่ดีที่สุดควรต้องแสดงให้เห็นถึงความเที่ยงตรงทั้ง 3 ประเภท อย่างไรก็ตาม ในแง่ของการปฏิบัติแล้วไม่มีความจำเป็นที่จะต้องแสดงผลฐานให้ครบทั้ง 3 ประเภท ขึ้นกับคุณลักษณะของแบบวัดที่ต้องการหลักฐานมายืนยันความเที่ยงตรงโดยทั่วไปแล้ว ความเที่ยงตรงเชิงเนื้อหาจะยืนพื้น สำหรับการแสดงผลฐานความเที่ยงตรงตามโครงสร้างจะแสดงเพื่อตอบคำถามเกี่ยวกับความสามารถของแบบทดสอบในการแปลความหมายของคุณลักษณะที่วัดได้ดีเพียงใด ความเที่ยงตรงตามเกณฑ์จะใช้เมื่อแบบวัดนั้นต้องการหลักฐานแสดงว่า แบบวัดสามารถพยากรณ์ความสามารถในการทำงานในอนาคตได้ดีเพียงใดหรือประมาณค่าความสามารถในปัจจุบันจากการวัดแบบอื่น ๆ ที่เป็นเกณฑ์ได้ดีเพียงใด (Linn; & Gronlund. 1995: 50-51)

เนื่องจากธรรมชาติของความเที่ยงตรงมีความเฉพาะเจาะจงในการแปลความหมายผลการวัด ไม่มีแบบทดสอบใดมีความเที่ยงตรงในทุกวัตถุประสงค์ (Gronlund; & Linn. 1990) ในการศึกษาครั้งนี้เป็นการแสดงผลฐานความเที่ยงตรงตามโครงสร้างโดยการยืนยันองค์ประกอบของทักษะกระบวนการทางวิทยาศาสตร์ที่ประยุกต์ใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน

และการศึกษาความไม่แปรเปลี่ยนของโครงสร้างองค์ประกอบ การวิเคราะห์หลายคุณลักษณะหลายวิธีตามแนวทางของทฤษฎีมาตรฐานเดิม (Classical) นอกจากนี้ผู้วิจัยยังได้ศึกษาเพิ่มเติมโดยใช้การวิเคราะห์โมเดลของราสช์แบบหลายองค์ประกอบ (Many-Facets Rasch Model) ตามแนวทางของทฤษฎีการตอบข้อคำถาม ดังมีรายละเอียดต่อไปนี้

3. การวิเคราะห์องค์ประกอบ

การวิเคราะห์ข้อมูลทางสถิติที่มีหลักการเชิงวิชาการ มีวิธีการวิเคราะห์ซับซ้อน เป็นวิธีการที่มีอำนาจสูง และเป็นประโยชน์อย่างยิ่งต่อการวิจัยทางสังคมศาสตร์ และพฤติกรรมศาสตร์ จนได้รับการยกย่องว่าเป็นวิธีการที่เยี่ยมยอดทางการวิเคราะห์ข้อมูลทางสถิติทั้งปวง คือ การวิเคราะห์องค์ประกอบ (เพ็ญนภา กุลนภาดล. 2547: 62; อ้างอิงจาก Kerlinger. 1973: 659) ซึ่งการวิเคราะห์องค์ประกอบ (Factor Analysis) เป็นชื่อทั่วไปที่ใช้เรียกวิธีการวิเคราะห์ข้อมูลที่มีวิธีการ และเป้าหมายการวิเคราะห์ต่างกัน คือ การวิเคราะห์ส่วนประกอบ การวิเคราะห์องค์ประกอบร่วม การวิเคราะห์องค์ประกอบเชิงสำรวจ และการวิเคราะห์องค์ประกอบเชิงยืนยัน วิธี การวิเคราะห์องค์ประกอบเหล่านี้ไม่ว่าวิธีใดวิธีหนึ่งต่างก็เป็นวิธีการที่มีประโยชน์ต่อนักวิจัยทั้งสิ้น

งลักษณ์ วิรัชชัย (2542: 122) กล่าวถึง การวิเคราะห์องค์ประกอบเป็นวิธีการอธิบายข้อมูลให้ง่ายขึ้นด้วยการลดจำนวนตัวแปร (variable reduction) โดยการพยายามหาโครงสร้างตัวประกอบจำนวนน้อย ๆ ที่จะแทนตัวแปรจำนวนมาก ๆ เนื่องจาก การวิจัยทางสังคมศาสตร์ และการวิจัยทางพฤติกรรมศาสตร์นั้น มุ่งเก็บรวบรวมข้อมูลซึ่งเป็นลักษณะภายในที่ไม่สามารถสังเกตได้โดยตรงหรืออาจเรียกว่าตัวแปรแฝงและต้องศึกษาคุณลักษณะดังกล่าวนี้จากพฤติกรรมการแสดงออกของบุคคล โดยการวัดหรือสังเกตพฤติกรรมเหล่านั้น แทนคุณลักษณะที่ต้องการศึกษา ในทางปฏิบัตินักวิจัยจะเก็บรวบรวมข้อมูลได้เป็นตัวแปรสังเกตได้หลายตัว และใช้การวิเคราะห์องค์ประกอบมาวิเคราะห์ข้อมูลเพื่อให้ได้องค์ประกอบอันเป็นคุณลักษณะที่ผู้วิจัยต้องการศึกษา กล่าวได้ว่า วิธีการวิเคราะห์องค์ประกอบเป็นการวิเคราะห์ข้อมูลทางสถิติ ที่ช่วยให้แก่นักวิจัยสร้างองค์ประกอบจากตัวแปรหลาย ๆ ตัวแปรโดยรวมกลุ่มตัวแปรที่เกี่ยวข้องสัมพันธ์กันเป็นองค์ประกอบเดียวกัน และแต่ละองค์ประกอบ คือ ตัวแปรแฝงอันเป็นคุณลักษณะที่ผู้วิจัยต้องการศึกษา

นงลักษณ์ วิรัชชัย (2542: 122) กล่าวว่า การวิเคราะห์องค์ประกอบมีวัตถุประสงค์ 2 ประการ คือ

1. เพื่อสำรวจระบุ หรือศึกษาว่า องค์ประกอบรวมที่สามารถอธิบายความสัมพันธ์ร่วมกันระหว่างตัวแปรต่าง ๆ โดยที่จำนวนองค์ประกอบรวมที่หาได้ จะมีจำนวนน้อยกว่าจำนวนตัวแปรซึ่งเป็นการลดจำนวนตัวแปรลง และได้องค์ประกอบหลัก ๆ เรียกว่า การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis Model) การวิเคราะห์ในลักษณะนี้ทำให้ง่ายต่อการทำความเข้าใจลักษณะของข้อมูล และสะดวกในการแปลความหมาย

2. เพื่อต้องการทดสอบสมมติฐานเกี่ยวกับแบบแผน โครงสร้างขององค์ประกอบว่าองค์ประกอบแต่ละตัว ประกอบด้วยตัวแปรอะไรบ้าง และตัวแปรแต่ละตัวควรมีน้ำหนักหรืออัตราความสัมพันธ์กับองค์ประกอบมากน้อยเพียงใด ตรงกับที่คาดคะเนไว้หรือไม่ หรือสรุปได้ว่าเพื่อต้องการทดสอบว่าองค์ประกอบนี้ ตรงกับโมเดลหรือตรงกับทฤษฎีที่มีอยู่หรือไม่ โมเดลนี้เรียกว่า โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis Model)

แซทฟีวส์ และโคลลิน (เพ็ญนภา กุลนภาดล. 2547: 63; อ้างอิงจาก Chatfield. & Collins. 1980: 89) กล่าวว่า ในปัจจุบันนักวิจัยเริ่มใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis Model: CFA) แทนการวิเคราะห์องค์ประกอบเชิงสำรวจมากขึ้น สาเหตุที่เป็นเช่นนี้เนื่องจาก องค์ประกอบเชิงสำรวจมีรูปแบบวิเคราะห์หลากหลาย และได้ผลการวิเคราะห์ที่ไม่สอดคล้องกัน นอกจากนี้การวิเคราะห์องค์ประกอบเชิงสำรวจยังมีข้อดลงเบื้องต้นที่เข้มงวดและไม่ตรงตามความเป็นจริง เช่น ข้อดลงเบื้องต้นที่ว่า ตัวแปรสังเกตได้ทุกตัวเป็นผลมาจากองค์ประกอบร่วมทุกตัว ส่วนที่เป็นความคลาดเคลื่อนของตัวแปรไม่สัมพันธ์กัน รวมทั้งสเกลองค์ประกอบที่สร้างขึ้นแปลความหมายได้ยาก เพราะในบางครั้งสเกลองค์ประกอบเกิดจากการสุ่มตัวแปรที่ไม่น่าจะมีย้องค์ประกอบร่วมกัน จุดอ่อนของการวิเคราะห์องค์ประกอบเชิงสำรวจนี้ ทำให้ ลอง (Long. 1983: 12) กล่าวว่า สำหรับนักวิจัยหลายคนใช้เทคนิคการวิเคราะห์องค์ประกอบเชิงสำรวจเป็น GIGO model (Garbage In and Garbage Out model) และเสนอว่า ถ้าทำได้นักวิจัยไม่ควรใช้การวิเคราะห์องค์ประกอบเชิงสำรวจเลย

การวิเคราะห์องค์ประกอบจะมีอยู่หลายวิธีและหลักสำคัญของการวิเคราะห์ก็มีลักษณะเป็นแบบเดียวกัน เมื่อเข้าใจหลักการที่สำคัญแล้ว ย่อมสามารถทำความเข้าใจวิธีการต่าง ๆ ที่หลากหลายได้โดยง่าย การศึกษาในครั้งนี้ผู้วิจัยเลือกที่จะนำเสนอเฉพาะกรณีการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งและอันดับสอง ที่ใช้ในการวิเคราะห์ข้อมูลเพื่อหาความเที่ยงตรงตามโครงสร้างเท่านั้น

3.1 การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง

ตัวแปรทางจิตวิทยาหรือทางการศึกษา เช่น เจตคติ แรงจูงใจ ฯลฯ ที่ใช้ในการศึกษาคุณลักษณะหรือพฤติกรรมต่าง ๆ ของมนุษย์ เป็นคุณลักษณะที่ไม่สามารถวัดได้โดยตรง ต้องมีการนิยามคุณลักษณะตามทฤษฎีสำหรับอธิบายคุณลักษณะนั้น ในการสร้างเครื่องมือเพื่อวัดคุณลักษณะดังกล่าว จึงจำเป็นต้องสร้างเครื่องมือตามนิยามที่กำหนดและตรวจสอบคุณภาพของเครื่องมือที่สร้างขึ้นว่า สามารถวัดคุณลักษณะที่ต้องการวัดได้หรือไม่และวัดได้ดีเพียงใด ซึ่งเรียกการตรวจสอบคุณภาพของเครื่องมือในความหมายดังกล่าวว่า เป็นการตรวจสอบความเที่ยงตรงตามโครงสร้าง (Construct Validity) ประเภทหนึ่ง

คุณลักษณะที่ไม่สามารถวัดได้โดยตรง มีชื่อเรียกอื่น ๆ อีก เช่น ตัวแปรแฝงหรือองค์ประกอบ ในการวัดตัวแปรดังกล่าวเราต้องวัดจากตัวแปรสังเกตหนึ่งตัวขึ้นไปตามนิยามที่กำหนดไว้เพื่อเป็นตัวแทนในภาพรวมของตัวแปรแฝงตามทฤษฎี การวิเคราะห์องค์ประกอบเป็นวิธีทางสถิติวิธีหนึ่งที่ใช้ตรวจสอบความสัมพันธ์ระหว่างตัวแปรแฝงกับตัวแปรสังเกต ที่ให้คำตอบว่า ตัวแปรสังเกตที่กำหนดให้เป็นตัวบ่งชี้ของตัวแปรแฝงหนึ่ง ๆ จะมีความสัมพันธ์กับตัวแปรแฝงนั้น และจะไม่มีความสัมพันธ์กับตัวแปรแฝงอื่น ๆ ซึ่งเป็นแนวคิดหลักในการวิเคราะห์องค์ประกอบเชิงยืนยัน ดังนั้น การวิเคราะห์องค์ประกอบเชิงยืนยันจึงเป็นการศึกษาเพื่อยืนยันองค์ประกอบที่กำหนดไว้แล้วล่วงหน้าจากการตรวจสอบเอกสารและ /หรืองานวิจัยที่เกี่ยวข้องมาเป็นอย่างดี

โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันเป็นกรณีเฉพาะของโมเดลสมการเชิงโครงสร้างซึ่งใช้กันอย่างหลากหลายในการศึกษาทางการวัดผล (MacCallum; & Austin. 2000: 208) ขั้นตอนของการวิเคราะห์องค์ประกอบเชิงยืนยัน จึงมีลักษณะคล้ายคลึงกับขั้นตอนของการวิเคราะห์โมเดลสมการเชิงโครงสร้าง ขั้นตอนการวิเคราะห์องค์ประกอบเชิงยืนยันประกอบด้วยขั้นตอนต่าง ๆ ดังนี้ (Bollen. 1989: 226-304; Long. 1994: 228-232; Maruyama. 1998: 140-147; พรทิพย์ ไชยโส. 2536: 5-7; และ นงลักษณ์ วิรัชชัย. 2542)

3.1.1 การระบุรูปแบบโมเดล (Model Specification)

การกำหนดรูปแบบการวิเคราะห์องค์ประกอบเชิงยืนยันเป็นการระบุรูปแบบของตัวแปรทางทฤษฎีที่ผู้วิจัยต้องการยืนยันข้อมูลเชิงประจักษ์ ซึ่งได้มาจากการกำหนดจำนวนตัวแปรแฝง (ξ) จำนวนตัวแปรสังเกต (X) ความแปรปรวนร่วมระหว่างตัวแปรแฝง ความสัมพันธ์ระหว่างตัวแปรสังเกตและตัวแปรแฝง ความสัมพันธ์ระหว่างตัวแปรสังเกตกับความคลาดเคลื่อน และความแปรปรวนและความแปรปรวนร่วมระหว่างความคลาดเคลื่อน (δ) ซึ่งข้อได้เปรียบของการวิเคราะห์องค์ประกอบเชิงยืนยัน คือ สามารถกำหนดส่วนประกอบต่าง ๆ ที่กล่าวมาตามความต้องการของนักวิจัย ความสัมพันธ์เชิงเส้นตรงในองค์ประกอบเชิงยืนยันมีดังนี้

$$\mathbf{X} = \Lambda_X \xi + \delta \quad (1)$$

เมื่อ (กำหนดให้ตัวแปรสังเกต 6 ตัวแปร ตัวแปรแฝง 2 ตัวแปร)

X คือ เวกเตอร์ของตัวแปรสังเกตขนาด $q \times 1$

$$\begin{array}{c} |X_1| \\ |X_2| \\ |X_3| \\ |X_4| \\ |X_5| \\ |X_6| \end{array}$$

ξ คือ เวกเตอร์ของตัวแปรแฝงขนาด $s \times 1$

$$\begin{array}{c} |\xi_1| \\ |\xi_2| \end{array}$$

Λ_X คือ เมตริกของ factor loading ซึ่งแสดงความสัมพันธ์ระหว่างตัวแปรสังเกต (X) กับตัวแปรแฝง (ξ) ขนาด $q \times s$ ซึ่งกำหนดให้ $q > s$

$$\begin{array}{c} |\Lambda_1 0| \\ |\Lambda_2 0| \\ |\Lambda_3 0| \\ |0 \Lambda_4| \\ |0 \Lambda_5| \\ |0 \Lambda_6| \end{array}$$

δ คือ เวกเตอร์ของความคลาดเคลื่อนขนาด $q \times 1$

$$\begin{array}{c} |\delta_1| \\ |\delta_2| \\ |\delta_3| \\ |\delta_4| \\ |\delta_5| \\ |\delta_6| \end{array}$$

โดย δ มี 2 องค์ประกอบ ได้แก่ s คือ ความแปรปรวนเฉพาะ (unique factor) และ e คือ ความคลาดเคลื่อนส่วนที่เหลือเชิงสุ่ม ($\delta = s + e$)

เนื่องจากตัวแปร X ในสมการที่ 1 จะไม่สามารถประมาณค่าตัวแปรที่ไม่สามารถสังเกตได้โดยตรง จำเป็นต้องมีการตรวจสอบจากเมตริกซ์ความแปรปรวนร่วมระหว่างตัวแปรสังเกตซึ่งได้จากสมการความแปรปรวนร่วม ดังนี้

$$\Sigma(\theta) = \Lambda_X \Phi \Lambda_X' + \Theta_\delta \quad (2)$$

เมื่อ $\Sigma(\theta)$ คือ เมตริกซ์ความแปรปรวนร่วมระหว่างตัวแปรสังเกต

$\Sigma(\theta) = E(xx')$ มีลักษณะเป็นเมตริกซ์สมมาตร มีขนาด $q \times q$ สมาชิกตัวที่ i และ j ของ $\Sigma(\theta)$ เขียนได้ว่าเป็น σ_{ij} ถ้า x เป็นคะแนนมาตรฐานแล้ว $E(x_i x_j)$ จะแสดงค่าความสัมพันธ์ระหว่าง x_i และ x_j ในขณะที่ $\Sigma(\theta)$ จะเป็นเมตริกซ์ของสหสัมพันธ์

Λ_X คือ เมตริกของ factor loading และ

Λ_X' คือ transpose ของเมตริก factor loading

Φ คือ เมตริกซ์ความแปรปรวนร่วมระหว่างตัวแปรแฝง (ξ) ซึ่งมีลักษณะเป็นเมตริกซ์สมมาตร ขนาด $s \times s$ สมาชิกในเมตริกซ์ Φ คือ Φ_{ij} แสดงถึงค่าความแปรปรวนร่วมระหว่างตัวแปรแฝง ξ_i และ ξ_j ในกรณีที่ความแปรปรวนระหว่างตัวแปรสังเกตไม่มีความสัมพันธ์กัน สมาชิกนอกเส้นทแยงมุมของ Φ จะเป็นศูนย์ และถ้าตัวแปรสังเกตถูกทำให้เป็นมาตรฐานที่มีความแปรปรวนเป็น 1 Φ จะเป็นเมตริกซ์สหสัมพันธ์ที่มีค่าตามเส้นทแยงเป็น 1 และสหสัมพันธ์ของความแปรปรวนระหว่างตัวแปรสังเกตนอกเส้นทแยงมุมจะเป็นศูนย์

Θ_δ คือ ความแปรปรวนระหว่างส่วนที่เหลือขององค์ประกอบ (residual factor) ซึ่งเป็นเมตริกซ์สมมาตร ขนาด $q \times q$ สมาชิกตัวที่ i และ j ของ Θ_δ คือ $\Theta_{\delta_{ij}}$ เป็นความแปรปรวนร่วมระหว่างองค์ประกอบที่เหลือ ในกรณีที่องค์ประกอบส่วนที่เหลือไม่สัมพันธ์กัน สมาชิกนอกเส้นทแยงมุมของเมตริกซ์ Θ_δ จะเป็นศูนย์

3.1.2 การระบุความเป็นค่าเดียวของโมเดล (Identification)

การระบุความเป็นค่าเดียวของโมเดลเป็นการพิจารณาค่าพารามิเตอร์ที่จะประมาณค่าจากโมเดลที่กำหนดนั้น สามารถประมาณค่าได้หรือไม่ ค่าพารามิเตอร์ที่ประมาณได้นั้นสามารถประมาณค่าที่เป็นค่าเดียวได้หรือไม่ นั่นคือ ค่าพารามิเตอร์ที่ประมาณจะได้ค่าพารามิเตอร์เพียงค่าเดียวที่ให้ผลที่ได้จากค่าประมาณสอดคล้องกับโมเดลเชิงทฤษฎีที่ระบุไว้หรือไม่ ซึ่งความเป็นไปได้ของการระบุความเป็นค่าเดียวมี 3 โอกาสที่เป็นไปได้ดังนี้

Unidentified Model คือ รูปแบบที่มีจำนวนพารามิเตอร์ที่ต้องการประมาณค่ามากกว่าจำนวนสมการที่ใช้ในการหาค่าพารามิเตอร์ ทำให้ไม่สามารถหาค่าจากสมการได้

Just-identified Model คือ รูปแบบที่มีจำนวนพารามิเตอร์ที่จะทำการประมาณค่าเท่ากับจำนวนสมการที่ใช้ในการหาค่าพารามิเตอร์ ทำให้สามารถประมาณค่าพารามิเตอร์ได้เพียงค่าเดียว (unique solution) รูปแบบนี้มีค่าองศาความเป็นอิสระเท่ากับศูนย์

Overidentified Model คือ รูปแบบที่มีจำนวนพารามิเตอร์ที่จะทำการประมาณค่าน้อยกว่าจำนวนสมการที่ใช้ในการหาค่าพารามิเตอร์ รูปแบบนี้มีค่าองศาความเป็นอิสระไม่เท่ากับศูนย์

เนื่องจากสมการที่ใช้ในการหาค่าพารามิเตอร์ได้จากเมตริกซ์ความแปรปรวนร่วมหรือเมตริกซ์สหสัมพันธ์ ดังนั้น การระบุความเป็นค่าเดียวของโมเดลจึงมีลักษณะที่แตกต่างกัน กล่าวคือ ค่าพารามิเตอร์ที่สามารถหาค่าในเมตริกซ์ความแปรปรวนร่วม คือ จำนวนค่าความแปรปรวนและความแปรปรวนร่วมทุกค่าในเมตริกซ์ หรือ $q(q+1)/2$ ตัว ส่วนเมตริกซ์สหสัมพันธ์ คือ จำนวนค่าสหสัมพันธ์ระหว่างตัวแปรสังเกตไม่รวมค่าตามเส้นทแยงมุมที่มีค่าเป็น 1 หรือ $q(q-1)/2$ ตัว ซึ่งทำให้เมตริกซ์ความแปรปรวนร่วมมีจำนวนพารามิเตอร์ที่สามารถประมาณค่าได้มากกว่าเมตริกซ์สหสัมพันธ์

อย่างไรก็ตาม เมตริกซ์ความแปรปรวนร่วมไม่สามารถจะใช้ประโยชน์ในการประมาณค่าพารามิเตอร์ได้ดีกว่าเมตริกซ์สหสัมพันธ์ในทุกกรณี สำหรับการประมาณค่าในโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันในรูปแบบสมการโครงสร้างนั้น สามารถใช้เมตริกซ์ใดก็ได้ในการประมาณค่า แฮร์และคณะ (Hair; & et al. 1995: 648) เสนอแนะให้ใช้เมตริกซ์สหสัมพันธ์ในการวิเคราะห์องค์ประกอบเชิงยืนยันเนื่องจากโมเดลนี้มุ่งศึกษาแบบความสัมพันธ์ (pattern of interrelationship) ซึ่งต้องการมาตราที่มีความเสถียร (scale invariant) ซึ่งหมายถึง ตัวแปรสังเกตของตัวแปรแฝงต่างๆ ต้องมีความเป็นมาตรฐาน จะทำให้ตัวแปรแฝงสามารถเปรียบเทียบกันได้

การเลือกเมตริกซ์สหสัมพันธ์ที่ใช้ในการวิเคราะห์ตัวแปรระดับการวัดในมาตราจัดอันดับ (ordinal scale) เช่น มาตรวัดของไลเคิร์ต (Likert scale) ซึ่งไม่เป็นตัวแปรต่อเนื่องและไม่ควรที่จะอนุมานให้เป็นตัวแปรต่อเนื่องด้วยนั้น ไม่มีหน่วยของการวัด ดังนั้น ค่าเฉลี่ย ความแปรปรวนและความแปรปรวนร่วมของตัวแปรจัดอันดับจึงไม่มีความหมาย การให้ค่าของตัวแปรในมาตราจัดอันดับเป็นตัวแปรต่อเนื่องเพื่อใช้ในการวิเคราะห์ในสมการเชิงโครงสร้าง การกระทำเช่นนี้เป็นสิ่งที่ไม่ดี เนื่องจากส่วนมากการกระจายของข้อมูลไม่เป็นโค้งปกติ และที่สำคัญกว่านั้นคือ ค่าของตัวแปรไม่เป็นค่าต่อเนื่อง ดังนั้น การใช้ตัวแปรจัดอันดับในการวิเคราะห์สมการเชิงโครงสร้างจึงต้องการเทคนิคอื่นที่มากกว่าการใช้เทคนิคในการวิเคราะห์โดยใช้ค่าตัวแปรที่เป็นตัวแปรต่อเนื่อง (Jöreskog. 1994: 383) เบิร์น (Byrne. 1998: 165-166) ได้แนะนำว่า สำหรับตัวแปรที่มีระดับการวัดแบบจัดอันดับ เช่น มาตราส่วนประมาณค่า (rating

scale) นั้น ควรให้เมตริกซ์สหสัมพันธ์แบบ polychoric ในการประมาณค่าเนื่องจากระดับการวัดที่แท้จริงอยู่ในมาตราจัดลำดับเท่านั้น และใช้เมตริกซ์ความแปรปรวนร่วมแบบ asymptotic สำหรับใช้ถ่วงน้ำหนัก การใช้เมตริกซ์ในการประมาณค่าจึงต้องมีเมตริกซ์ทั้งสองเมตริกซ์ในการเขียนคำสั่ง สำหรับวิธีการประมาณค่าควรใช้วิธี WLS (Weighted Least Squares) ในการประมาณค่า ซึ่งการหาค่าสัมประสิทธิ์ทั้งสองเมตริกซ์สามารถกระทำได้โดยการใช้โปรแกรม PRELIS ซึ่งเป็นโปรแกรมย่อยในโปรแกรม LISREL (Jöreskog; & Sörbom. 1996: 16-17)

นอกจากระบุจำนวนพารามิเตอร์สำหรับการวิเคราะห์ข้อมูล โบเลน (Bollen. 1989: 244-246) ได้กำหนดกฎ 3 ตัวบ่งชี้ขึ้น โดยกำหนดว่าโมเดลที่สามารถประมาณค่าได้นั้น จำเป็นต้องมี 3 ตัวบ่งชี้ (ตัวแปรสังเกต) ในแต่ละตัวแปรแฝงหรืออย่างน้อยต้องมีตัวบ่งชี้ 2 ตัวขึ้นไปในกรณีที่ขนาดของกลุ่มตัวอย่างใหญ่พอ จากผลการวิจัยของแอนเดอร์สันและเกอร์บิง (Kelloway. 1998: 63; citing Anderson; & Gerbing. 1984) พบว่า เมื่อกลุ่มตัวอย่างมีขนาดเล็ก (น้อยกว่า 100 คน) การใช้ตัวบ่งชี้เพียง 2 ตัวในแต่ละตัวแปรแฝงจะทำให้โมเดลไม่สามารถวิเคราะห์ข้อมูลได้ (convergence failures) และผลการวิเคราะห์ออกมาไม่เหมาะสม

3.1.3 การประมาณค่า (Estimation)

ขั้นตอนนี้เป็น การประมาณค่าโดยใช้ข้อมูลจากกลุ่มตัวอย่างมาประมาณค่าใน รูปแบบของเมตริกซ์ (S) เพื่อมาหาค่าประมาณของ Λ_X , Φ และ Θ_δ จากโมเดลที่ระบุค่าเริ่มต้น ($\Sigma(\theta)$) การประมาณค่าจะทำซ้ำ (iteration) จนกว่าค่าที่ได้จะใกล้เคียงกับค่าของข้อมูลเชิงประจักษ์มากที่สุด (converges) จึงสิ้นสุดการประมาณค่า เทคนิคที่ใช้ในการประมาณค่าใน โมเดลลิสเรลมีจำนวน 7 วิธีการ (Jöreskog; & Sörbom. 1993: 116) ได้แก่ Instrumental Variables (IV), Two-Stage Least Squares (TSLS), Unweighted Least Squares (ULS), Generalized Least Squares (GLS), Maximum likelihood (ML), Generally Weighted Least Squares (WLS) และ Diagonally Weighted Least Squares (DWLS) ในการเลือกใช้วิธีการ ประมาณค่า เมื่อข้อมูลมีระดับการวัดแบบจัดอันดับควรใช้วิธีการประมาณค่าโดยวิธี WLS หรือ DWLS แต่วิธีการประมาณค่าทั้งสองจะสามารถประมาณค่าได้อย่างเหมาะสมเมื่อมีขนาดกลุ่ม ตัวอย่างที่ใหญ่มาก เช่น มากกว่า 1,000 คน ขึ้นไป ส่วนในกรณีที่ข้อมูลมีระดับการวัดแบบ ต่อเนื่อง วิธีที่นิยมใช้กันคือ ML วิธีการแบบ ML จะสามารถประมาณค่าได้เหมาะสมขนาดกลุ่ม ตัวอย่างควรมีขนาดใหญ่ (เช่น มากกว่า 200 คน ขึ้นไป) และการกระจายของข้อมูลเป็นโค้ง ปกติ

3.1.4 การประเมินโมเดล (Model Evaluation)

เป็นการทดสอบความเหมาะสมระหว่างโมเดลที่กำหนดกับข้อมูลเชิงประจักษ์ในภาพรวม มีดัชนีที่แสดงความเหมาะสมอยู่ 3 กลุ่ม ดังนี้

3.1.4.1 ดัชนีความเหมาะสมสัมบูรณ์ (absolute fit index) ดัชนีกลุ่มนี้มีความสมบูรณ์ในตัวเองไม่ต้องนำไปเปรียบเทียบกับโมเดลอื่น ๆ ได้แก่

สถิติไคสแควร์ ในโปรแกรม LISREL จะให้สถิติไคสแควร์ที่คำนวณจากฟังก์ชันการประมาณค่าคูณกับ $(N-1)$ เมื่อ N คือขนาดของกลุ่มตัวอย่าง แล้วนำค่าไคสแควร์ที่ได้เปรียบเทียบกับค่าวิกฤตที่ df เท่ากับ จำนวนความแปรปรวนร่วม – จำนวนพารามิเตอร์ที่ต้องการประมาณค่า ถ้าค่าที่คำนวณน้อยกว่าค่าวิกฤตแสดงว่า โมเดลมีความเหมาะสมหรือแสดงว่า ความแปรปรวนร่วมระหว่างตัวแปรสังเกตเกิดขึ้นเพราะความสัมพันธ์ตามที่กำหนดไว้ในโมเดลหรืออธิบายข้อมูลได้ดี ข้อจำกัดของไคสแควร์ คือ เป็นค่าสถิติที่มีความไวต่อความเป็นโค้งปกติของข้อมูล สำหรับโมเดลที่มีความซับซ้อนสูง (มีตัวแปรมากและค่าองค์ประกอบมาก) จะทำให้ค่าไคสแควร์มีนัยสำคัญทางสถิติได้ง่ายถึงแม้ข้อมูลจะเหมาะสมกับรูปแบบที่กำหนดไว้ และถ้ากลุ่มตัวอย่างมีขนาดใหญ่มากผลการทดสอบมักจะออกมามีนัยสำคัญหรือถ้ากลุ่มตัวอย่างมีขนาดเล็กเกินไปการประมาณค่าไคสแควร์อาจไม่ถูกต้อง (Marsh; & Hocever. 1985: 567)

ดัชนีอำนาจการทำนาย (Indices of proportion of variance Accounted for) ได้แก่ Goodness of Fit Index (GFI) ซึ่งค่าดัชนีตัวนี้จะไม่แน่นอนเมื่อกลุ่มตัวอย่างมีขนาดเล็ก ส่วนค่า Adjusted GFI (AGFI) เป็นค่าปรับแก้เกี่ยวกับจำนวนของค่าพารามิเตอร์ที่ประมาณค่า (Spector. 2001: 22) ค่าความเป็นไปได้ของดัชนีทั้งสองตัวคือ 0-1 โดยเกณฑ์การพิจารณาความเหมาะสมของโมเดลของดัชนีทั้งสองตัวคือ ควรมีค่าสูงกว่า 0.9 นอกจากนี้ยังมีค่า Critical N (CN) สำหรับใช้ประมาณค่ากลุ่มตัวอย่างที่ยอมรับได้ที่จะทำให้โมเดลมีความเหมาะสม ซึ่งเกณฑ์ในการพิจารณาค่า CN คือ ควรมีค่าสูงกว่า 200 (Marsh, Balla; & McDonald. 1988: 393; citing Hoelter. 1983)

ดัชนีในกลุ่มเศษที่เหลือ (Residual Based Fit Indices) ได้แก่ RMR (Root Mean Square Residual) ใช้ในกรณีที่มีการประมาณค่าโดยใช้เมตริกซ์สหสัมพันธ์ และ SRMR (Standardized Root Mean Square Residual) ใช้ในกรณีที่ประมาณค่าโดยเมตริกซ์ความแปรปรวนและความแปรปรวนร่วม (Spector. 2001: 22) เป็นค่าเฉลี่ยเศษที่เหลือจากการเปรียบเทียบค่าความแปรปรวนและความแปรปรวนร่วมที่ได้จากกลุ่มตัวอย่างกับค่าที่ประมาณจากพารามิเตอร์ ถ้าโมเดลมีความเหมาะสมสูงค่าทั้งสองควรมีค่าที่เล็กมาก โดยทั่วไปไม่ควรเกิน 0.05 ถ้าต้องการสรุปว่าโมเดลเหมาะสม นอกจากนี้ยังมีค่า RMSEA (Root Mean Square of Approximation) ถ้าโมเดลมีความเหมาะสมสูง ค่าดัชนีของกลุ่มนี้ควรมีค่าเล็กมาก หากค่า RMR และ SRMR มีค่าต่ำกว่า 0.08 และค่า RMSEA มีค่าต่ำกว่า 0.06 แสดงว่าโมเดลมีความเหมาะสมระดับดีมาก (Hu; & Bentler. 1999)

นอกจากนี้ นักวิจัยสามารถตรวจสอบความเหมาะสมของโมเดลได้ด้วยวิธีการวิเคราะห์ส่วนที่เหลือ (Analysis of Residual) สามารถพิจารณาจากแหล่งข้อมูล 2 แหล่ง คือ (นงลักษณ์ วิรัชชัย. 2542: 55-57)

เมตริกซ์ส่วนที่เหลือในการเทียบความเหมาะสม (Fitted Residuals Matrix) หมายถึง เมตริกซ์ที่เป็นผลต่างของเมตริกซ์ S และ $\Sigma(\hat{\theta})$ ซึ่งในโปรแกรมลิสเรลจะให้ค่าส่วนที่เหลือทั้งในรูปคะแนนดิบและคะแนนมาตรฐาน ค่าส่วนที่เหลือในรูปคะแนนมาตรฐานนั้น ถ้าโมเดลมีความเหมาะสมกับข้อมูลเชิงประจักษ์ ค่าส่วนที่เหลือในรูปคะแนนมาตรฐานไม่ควรมีค่าเกิน 2.00 ถ้าเกินก็ต้องปรับโมเดล

คิวพล็อต (Q-Plot) เป็นกราฟแสดงความสัมพันธ์ระหว่างค่าส่วนที่เหลือกับค่าควอนไทล์ปกติ (normal quantiles) ถ้าได้เส้นกราฟที่มีความชันมากกว่าเส้นทแยงมุมอันเป็นเกณฑ์ในการเปรียบเทียบ แสดงว่า โมเดลมีความเหมาะสมกับข้อมูลเชิงประจักษ์

3.1.4.2 ดัชนีเหมาะสมเชิงเปรียบเทียบ (comparative fit indices) ใช้ในการเปรียบเทียบค่าไคสแควร์ของโมเดลที่เป็น subset ของกันและกัน เช่น ดัชนี Normed fit Index (NFI) ซึ่งมีแนวโน้มจะให้ค่าต่ำเมื่อจำนวนกลุ่มตัวอย่างน้อย และ NonNormed fit Index (NNFI) (Bentler; & Bonett. 1980; citing Bentler. 1990) เพื่อปรับแก้ความลำเอียงจากขนาดของกลุ่มตัวอย่างของ NFI แต่จะมีการแปรเปลี่ยนของค่าดัชนีค่อนข้างสูง (Spector. 2001: 22) ปี 1990 เบนท์เลอร์ (Bentler. 1990) ได้เสนอดัชนี Comparative Fit Index (CFI) ซึ่งจะใช้ได้ดีเมื่อกลุ่มตัวอย่างมีขนาดเล็ก ดัชนี Incremental Fit Index (IFI) (Bollen. 1989) ซึ่งความผันแปรของผลการประเมินโมเดลจะมีความแปรเปลี่ยนต่ำกว่าดัชนี NNFI (Spector. 2001: 22) ค่าความเป็นไปได้ของดัชนีทั้ง 4 ตัว คือ 0 ถึง 1 โดยค่าสูงกว่า 0.90 แสดงว่า โมเดลมีความเหมาะสมกับข้อมูลเชิงประจักษ์ระดับดี (Kelloway. 1998: 30-31) และสูงกว่า 0.95 แสดงว่า โมเดลมีความเหมาะสมกับข้อมูลเชิงประจักษ์ระดับดีมาก (Hu; & Bentler. 1999)

3.1.4.3 ดัชนีโมเดลเชิงประหยัด (parsimony fit indices) เป็นดัชนีที่ใช้ในการตรวจสอบความเหมาะสมเมื่อมีการลดพารามิเตอร์ในโมเดลลง เช่น ดัชนี PGFI, AIC, CAIC เป็นต้น ผู้วิจัยจะใช้ดัชนีในกลุ่มนี้ เมื่อมีวัตถุประสงค์การวิจัยเพื่อเปรียบเทียบโมเดลคู่แข่งว่าโมเดลใดที่มีความประหยัดมากกว่ากัน กล่าวคือ เมื่อมีโมเดลมากกว่าหนึ่งโมเดลที่มีลักษณะความสัมพันธ์ที่แตกต่างกัน แต่ใช้ข้อมูลในการวิเคราะห์ชุดเดียวกัน โมเดลที่มีความประหยัดมากกว่าคือโมเดลที่มีค่าพารามิเตอร์ในการประมาณค่าน้อยกว่า (Marsh; & Hau. 1996)

3.1.5 การปรับโมเดล (Model Modification)

ในกรณีที่นักวิจัยพบว่า โมเดลเริ่มต้นไม่เหมาะสมกับข้อมูลเชิงประจักษ์ หรือ อาจต้องการปรับโมเดลเพื่อทดสอบสมมติฐานบางสมมติฐาน วิธีการที่ใช้ทดสอบนัยสำคัญของ โมเดลที่ปรับแล้วมีดังนี้ (คุชฎี โยเหลา. ม.ป.ป.: 9)

3.1.5.1 การเปรียบเทียบการเพิ่มขึ้นหรือลดลงของค่าไคสแควร์

กรณีนี้เหมาะสมเมื่อผู้วิจัยมีโมเดล 2 โมเดล ที่โมเดลหนึ่งเป็น subset ของโมเดลที่ใหญ่กว่า โดยคำนวณเป็นค่าสถิติ ดังนี้

$$\chi^2_{diff} = \chi^2_{large} - \chi^2_{small} \quad (3)$$

ค่าที่ได้มีการแจกแจงเป็น χ^2 ที่ $df = df_{large} - df_{small}$

ถ้าค่าที่ได้ลดลงอย่างมีนัยสำคัญทางสถิติ แสดงว่า โมเดลเล็กมีความเหมาะสมกับข้อมูลมากกว่า การทดสอบแบบนี้มีข้อด้อยสำหรับกลุ่มตัวอย่างที่มีขนาดเล็ก เพราะ อาจไม่พบผลที่แตกต่างอย่างมีนัยสำคัญ และผู้วิจัยต้องคำนวณค่าสถิติของทั้งสองโมเดลทำให้ เสียเวลาและใช้หน่วยความจำของคอมพิวเตอร์สูง

3.1.5.2 Lagrangian Multiplier Test (LMT) วิธีการนี้ใช้เปรียบเทียบโมเดลที่เป็น subset กัน เช่นเดียวกับกรณีแรก แต่มีข้อได้เปรียบคือ สามารถประมาณค่าพารามิเตอร์ต่าง ๆ จากโมเดลเพียงโมเดลเดียว จากนั้นทดลองประมาณค่า (free) พารามิเตอร์บางตัว แล้ว ดูว่าผลการทำดังกล่าว จะทำให้โมเดลโดยรวมเหมาะสมสูงขึ้นหรือไม่ ในโปรแกรมลิสเรลจะมี ดัชนีการปรับโมเดลที่เป็นข้อเสนอแนะจากโปรแกรม ในการกำหนดพารามิเตอร์ที่ยังไม่ประมาณค่าให้เป็นพารามิเตอร์ที่ประมาณค่า (free parameter) และจะมีดัชนีบ่งบอกถึงค่าไคสแควร์ที่ ลดลงเมื่อประมาณค่าพารามิเตอร์ตัวนั้น เพื่อเป็นข้อมูลให้นักวิจัยได้ใช้พิจารณาปรับโมเดลให้ เหมาะสมกับข้อมูลเชิงประจักษ์มากขึ้น

3.1.5.3 Wald test เป็นการทดสอบที่ตอบคำถามว่า มีพารามิเตอร์ใดบ้างที่ ควรตัดทิ้งจากโมเดล ดังนั้น ในการคำนวณจึงทำการกำหนดให้พารามิเตอร์ตัวนั้นเป็น 0 (fixed parameter) แล้วดูว่าโมเดลมีความเหมาะสมหรือไม่

3.1.6 การประเมินพารามิเตอร์ (Parameter Evaluation)

การปรับโมเดลเป็นวิธีการเพื่อศึกษาความเหมาะสมระหว่างโมเดลเริ่มต้นกับ เมตริกซ์ความแปรปรวนร่วม/เมตริกซ์สหสัมพันธ์จากข้อมูลเชิงประจักษ์ในภาพรวมทั้งโมเดล อย่างไรก็ตามในแต่ละตัวแปรแฝงยังสามารถทำการประเมินความเหมาะสมแยกจากกันได้อีก ด้วย โดยพิจารณาจากเกณฑ์ดังนี้ (Hair; et al. 1995: 652-654)

การทดสอบความมีนัยสำคัญของค่า factor loadings ของตัวแปรสังเกตแต่ละตัว สถิติที่ใช้ในการทดสอบคือ ค่าสถิติ t โดยการเปรียบเทียบค่าสถิติจากการคำนวณกับค่าวิกฤตที่กำหนด โดยการมีระดับนัยสำคัญทางสถิติที่ .05 จะมีค่าวิกฤตเท่ากับ 1.96 และระดับนัยสำคัญทางสถิติที่ .01 จะมีค่าวิกฤตเท่ากับ 2.576 หากค่า factor loading ในแต่ละตัวแปรสังเกตที่มีต่อตัวแปรแฝง มีค่าสถิติ t ที่มีนัยสำคัญทางสถิติจะเป็นการยืนยันความสัมพันธ์ระหว่างตัวแปรสังเกตกับตัวแปรแฝงว่ามีอยู่จริง

การประเมินความเชื่อมั่นเชิงโครงสร้าง (Construct Reliability) และการประเมินค่า variance extracted ค่าสถิติทั้งสองใช้เพื่อทดสอบความเพียงพอของตัวแปรสังเกตที่จะใช้เป็นตัวแทนของตัวแปรแฝง โดยการประเมินความเพียงพอของความเชื่อมั่นเชิงโครงสร้างใช้เกณฑ์ 0.70 ขึ้นไป ส่วนค่า variance extracted ใช้เกณฑ์ 0.50 ขึ้นไป โดยค่า factor loadings และค่าความคลาดเคลื่อนที่ใช้ในการคำนวณจะใช้คะแนนมาตรฐาน สำหรับสูตรในการคำนวณแต่ละค่ามีดังนี้ (Hair; et al. 1995: 642)

$$\text{Construct Reliability} = \frac{(\sum \Lambda_X)^2}{(\sum \Lambda_X)^2 + \sum (\delta)} \quad (4)$$

$$\text{Variance Extracted} = \frac{\sum \Lambda_X^2}{\sum \Lambda_X^2 + \sum (\delta)} \quad (5)$$

ข้อควรคำนึงอีกประการหนึ่งเกี่ยวกับคุณภาพของตัวแปรสังเกตคือ สัมประสิทธิ์ความแปรปรวนของตัวแปรสังเกตแต่ละตัวที่สามารถอธิบายความแปรปรวนของตัวแปรแฝง (R^2) การอธิบาย R^2 ในโมเดลการวัดของสมการเชิงโครงสร้าง จะมีลักษณะคล้ายกับการอธิบาย R^2 ในสมการถดถอย (regression equation) ซึ่งหมายถึงประสิทธิภาพของตัวแปรอิสระที่สามารถพยากรณ์ตัวแปรตาม (บุญเรียง ขจรศิลป์. 2542: 124) แต่สำหรับ R^2 ในสมการเชิงโครงสร้าง การอธิบายจะหมายถึงขนาดของความแปรปรวนที่สามารถอธิบายได้ในสมการของตัวแปรสังเกตแต่ละตัว

การวิจัยในครั้งนี้ ผู้วิจัยได้ทำการวิเคราะห์องค์ประกอบอันดับหนึ่ง เพื่อยืนยันลักษณะองค์ประกอบของแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ จำนวน 22 ตัวบ่งชี้ว่า มีความสัมพันธ์ร่วมกันในการอธิบายความสามารถของผู้เรียนด้านทักษะกระบวนการทางวิทยาศาสตร์ เป็น 5 องค์ประกอบจริงหรือไม่ เมื่อใช้ข้อมูลเชิงประจักษ์ในขอบเขตที่ทำการเก็บข้อมูล หากผลการประเมินโมเดลแล้วพบว่า ค่าดัชนีความเหมาะสมมีค่าสูงกว่าเกณฑ์ที่กำหนดขึ้น และผลการประมาณค่าอยู่ในช่วงที่เหมาะสม และค่าพารามิเตอร์แต่ละตัวมีคุณภาพแสดงว่า แบบวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ กำหนดเป็น 5 องค์ประกอบ

สามารถนำมาใช้เป็นองค์ประกอบคุณภาพในการประเมินทักษะกระบวนการทางวิทยาศาสตร์ได้อย่างมีความเที่ยงตรงตามโครงสร้าง

3.2 การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง

ในกรณีที่การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งมีจำนวนองค์ประกอบจำนวนมากและองค์ประกอบดังกล่าวอาจสามารถถูกอธิบายจากตัวแปรแฝงอื่น ๆ ที่ไม่มีอิทธิพลทางตรงกับตัวแปรสังเกต (Bollen. 1989: 313-314) ดังนั้นการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองจะเป็นทางเลือกหนึ่งในการลดจำนวนองค์ประกอบที่ใช้ในการอธิบายปรากฏการณ์ลงได้ อย่างไรก็ตาม การกำหนดองค์ประกอบของตัวแปรอันดับสองยังคงยึดหลักการเกี่ยวกับการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งคือต้องมีทฤษฎีหรืองานวิจัยสนับสนุนองค์ประกอบดังกล่าวมาอย่างดี

การศึกษาองค์ประกอบเชิงยืนยันอันดับสองนั้น จะกระทำได้ก็ต่อเมื่อการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งมีความเหมาะสมกับข้อมูลเป็นอย่างดี (Marsh. 1991) และสมมติฐานการทดสอบองค์ประกอบอันดับสองนั้นจะมีลักษณะคล้ายกับการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งแต่องค์ประกอบอันดับสองจะเป็นตัวแปรภายนอกแฝงที่ส่งอิทธิพลไปยังตัวแปรภายในแฝง (องค์ประกอบอันดับหนึ่ง) สมมติฐานแตกต่างจากการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งคือความแปรปรวนร่วมระหว่างองค์ประกอบอันดับหนึ่งสามารถอธิบายได้โดยสัมประสิทธิ์ถดถอย (Factor Loadings) ขององค์ประกอบอันดับสอง โดยไม่รวมความสัมพันธ์ของความคลาดเคลื่อนในการวัดระหว่างองค์ประกอบอันดับหนึ่ง (Byrne. 1998: 170) สมการของโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองมีดังนี้ (Bollen. 1989: 314)

$$\eta = B\eta + \Gamma\xi + \zeta \quad (6)$$

$$y = \Lambda_y\eta + \varepsilon \quad (7)$$

เมื่อ η	คือ	เวกเตอร์ของตัวแปรภายในแฝง E ขนาด (NE x 1)
ξ	คือ	เวกเตอร์ของตัวแปรภายนอกแฝง K ขนาด (NK x 1)
ζ	คือ	เวกเตอร์ความคลาดเคลื่อน z ของตัวแปรภายในแฝง (E) ขนาด (NE x 1)
B	คือ	เมตริกซ์อิทธิพลเชิงสาเหตุระหว่างตัวแปรภายในแฝง (E) ขนาด (NE x NE)
Γ	คือ	เมตริกซ์อิทธิพลเชิงสาเหตุระหว่างตัวแปรภายนอกแฝง (K) ไปตัวแปรภายในแฝง (E) ขนาด (NE x NK) ในที่นี้คือเมตริกซ์ Factor Loading ขององค์ประกอบอันดับสอง

y	คือ	เวกเตอร์ตัวแปรภายในสังเกตได้ y ขนาด $(NY \times 1)$
Λ_y	คือ	เมตริกซ์สัมประสิทธิ์ถดถอยของ y บน E ขนาด $(NY \times NE)$ ในที่นี้คือเมตริกซ์ Factor Loading ขององค์ประกอบอันดับหนึ่ง
ε	คือ	เวกเตอร์ความคลาดเคลื่อน e ในการวัดตัวแปรภายในสังเกต (y) ขนาด $(NY \times 1)$

สมการที่ 6 เป็นการบ่งบอกถึงความสัมพันธ์ระหว่างองค์ประกอบอันดับหนึ่ง (ตัวแปรภายนอกแฝง) กับองค์ประกอบอันดับสอง (ตัวแปรภายในแฝง) คือโมเดลเชิงโครงสร้างนั่นเอง และสมการที่ 7 เป็นการบ่งบอกถึงความสัมพันธ์ระหว่างองค์ประกอบอันดับหนึ่งกับตัวแปรสังเกต

3.3 การวิเคราะห์กลุ่มพหุ

ในการประเมินโมเดลนั้น กลยุทธ์หนึ่งในการทดสอบความเที่ยงตรงของโมเดลคือ การทดสอบความเที่ยงตรงข้ามกลุ่ม (Cross Validation) ที่เป็นอิสระต่อกัน เพื่อใช้เป็นกลุ่มที่ทดสอบความเที่ยงตรงของโมเดล (Jöreskog; & Sörbom. 1993: 128-129) เกี่ยวกับความคงที่ของโมเดลระหว่างประชากรที่มีการวัดซ้ำหลาย ๆ ครั้งได้อย่างดี (นงลักษณ์ วิรัชชัย. 2540: 111) วิธีการวิเคราะห์ข้อมูลเพื่อให้ได้คำตอบดังกล่าวเรียกว่า การวิเคราะห์กลุ่มพหุ (Multigroup Analysis) (นงลักษณ์ วิรัชชัย. 2542: 217)

ในกรณีที่ผู้วิจัยต้องการทราบว่า ค่าพารามิเตอร์ของโมเดลของประชากร กลุ่มหนึ่ง จะเหมือนกับประชากรอีกกลุ่มหนึ่งหรือไม่นั้น จะใช้การวิเคราะห์กลุ่มพหุ โดยการเปรียบเทียบโครงสร้างและค่าพารามิเตอร์ของโมเดล เรียกว่าการศึกษาความไม่แปรเปลี่ยนของโมเดล

เราสนใจศึกษาความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มตัวอย่างที่มีอิสระต่อกัน สำหรับการวิเคราะห์องค์ประกอบเชิงยืนยันนั้นจะมีลักษณะเป็นการทดสอบคือ (1) การทดสอบโมเดลพื้นฐาน (Baseline Models) ของแต่ละกลุ่มตัวอย่างย่อย (2) การทดสอบความเท่าเทียมกันของโครงสร้างองค์ประกอบ (Factor Structure) (Λ_{form}) (3) การทดสอบความเท่าเทียมกันค่าพารามิเตอร์ ซึ่งประกอบด้วย Factor Loading (Λ_x) Factor Variance/Residual (Θ_δ) และ Factor Correlation (Φ) (Marsh; Hau; Chung; & Siu. 1998: 149-150)

3.3.1 การทดสอบโมเดลพื้นฐาน เบิร์น (Byrne. 1998: 262) กล่าวว่าเป็นการประเมินค่าความเหมาะสมระหว่างโมเดลกับข้อมูลเชิงประจักษ์ในแต่ละกลุ่มย่อยทีละครั้งว่ามีโครงสร้างองค์ประกอบที่เหมือนหรือแตกต่างกันอย่างไร เมื่อมีการปรับโมเดลให้เหมาะสมกับข้อมูลเชิงประจักษ์ ความรู้เบื้องต้นนี้จะเป็นประโยชน์ต่อการตั้งสมมติฐานและการทดสอบความไม่แปรเปลี่ยนของโครงสร้างองค์ประกอบและค่าพารามิเตอร์ต่อไป

3.3.2 การทดสอบความเท่าเทียมกันของโครงสร้างองค์ประกอบ เป็นการทดสอบรูปร่างของโมเดลว่าไม่แปรเปลี่ยนจริง ซึ่งจะเกิดขึ้นเมื่อข้อมูลที่เก็บมาในแต่ละกลุ่มมีจำนวนองค์ประกอบและจำนวนข้อความสอดคล้องกันในแต่ละองค์ประกอบ (Cheung; & Rensvold. 2002: 235-237; citing Meredith. 1993) กล่าวคือนัยหนึ่งคือ เป็นการทดสอบว่าโมเดลตามทฤษฎีที่เหมาะสมกับข้อมูลเชิงประจักษ์ของกลุ่มประชากรแต่ละกลุ่มนั้นประกอบด้วยจำนวนตัวแปรในโมเดล ลักษณะโครงสร้างความสัมพันธ์ระหว่างตัวแปรเป็นแบบเดียวกันหมดทุกกลุ่ม เมตริกซ์พารามิเตอร์ในโมเดลการวัดมีขนาดเท่ากัน มีรูปแบบเมตริกซ์ (Matrix Form) และสถานะเมตริกซ์ (Matrix Mode) เป็นแบบเดียวกันทุกกลุ่ม แต่ขนาดพารามิเตอร์ไม่เท่ากัน สมมติฐานสำหรับการทดสอบมีดังนี้ (Cheung; & Rensvold. 2002: 236)

$$H_{form} : \Lambda_{form}^{(1)} = \Lambda_{form}^{(2)} \quad (8)$$

3.3.3 การทดสอบความไม่แปรเปลี่ยนของพารามิเตอร์ในโมเดล เป็นการทดสอบความไม่แปรเปลี่ยนของเมตริกซ์พารามิเตอร์ 3 เมตริกซ์ ซึ่งจะกระทำหลังการทดสอบความไม่แปรเปลี่ยนของรูปแบบของโมเดล โดยหลักการแล้วจะทดสอบความไม่แปรเปลี่ยนของเมตริกซ์พารามิเตอร์ที่มีความเข้มงวดน้อยที่สุด (Least Restriction) ไปจนถึงการทดสอบความไม่แปรเปลี่ยนของเมตริกซ์พารามิเตอร์ที่มีความเข้มงวดมากที่สุด (Most Restriction) (นงลักษณ์ วัชรชัย. 2540: 52) แนวคิดการทดสอบทั้ง 3 พารามิเตอร์มีดังนี้ (Cheung; & Rensvold. 2002: 236) (1) การทดสอบความไม่แปรเปลี่ยนของเมตริกซ์ Factor Loading (Λ_x) เป็นการทดสอบว่าในภาพรวมแล้วลักษณะความสัมพันธ์ระหว่างตัวแปรสังเกตกับตัวแปรแฝงของตัวแปรสังเกตนั้นเหมือนกันระหว่างกลุ่ม (2) การทดสอบความไม่แปรเปลี่ยนของเมตริกซ์ Factor Variance/Residual (Θ_δ) เป็นการทดสอบตัวแปรสังเกตว่ามีความสอดคล้องภายในเหมือนกันระหว่างกลุ่ม หรือกล่าวได้ว่า ในทุกกลุ่ม ตัวแปรสังเกตมีคุณภาพที่จะใช้เป็นตัวบ่งชี้ในตัวแปรแฝงนั้นได้เหมือนกัน (3) การทดสอบความไม่แปรเปลี่ยนของเมตริกซ์ Factor Correlation (Φ) เป็นการทดสอบความสัมพันธ์ระหว่างตัวแปรแฝงว่าเหมือนกันระหว่างกลุ่ม

สำหรับโมเดลการวัดนั้น โบลเลน (Bollen. 1989: 360) ได้นำเสนอสมมติฐานทางสถิติที่มีลักษณะซ้อนกันเป็นระดับลดหลั่นดังนี้

H_{form} : โครงสร้างไม่แปรเปลี่ยน (ขนาดของเมตริกซ์ และสถานะของพารามิเตอร์ เป็นแบบกำหนด, อิสระ และบังคับในเมตริกซ์ Λ , Θ_δ และ Φ เหมือนกัน) (ความหมายเดียวกับสมการที่ 6)

$$\begin{aligned} H_{\Lambda_x} : \quad & \Lambda_x^{(1)} = \Lambda_x^{(2)} \\ H_{\Lambda_x \Theta_\delta} : \quad & \Lambda_x^{(1)} = \Lambda_x^{(2)} \quad \Theta_\delta^{(1)} = \Theta_\delta^{(2)} \\ H_{\Lambda_x \Theta_\delta \Phi} : \quad & \Lambda_x^{(1)} = \Lambda_x^{(2)} \quad \Theta_\delta^{(1)} = \Theta_\delta^{(2)} \quad \Phi^{(1)} = \Phi^{(2)} \end{aligned} \quad (9)$$

สำหรับสมมติฐานการทดสอบความไม่แปรเปลี่ยนของการวิเคราะห์องค์ประกอบ อันดับสองมีพารามิเตอร์ที่ต้องทดสอบความไม่แปรเปลี่ยนเพิ่มอีก 2 พารามิเตอร์คือ (1) การทดสอบความไม่แปรเปลี่ยนของค่าพารามิเตอร์ในเมตริกซ์อิทธิพลเชิงสาเหตุจากตัวแปรภายนอกไปยังตัวแปรภายในแฝง (Γ) และ (2) การทดสอบความไม่แปรเปลี่ยนของค่าพารามิเตอร์ความแปรปรวนและความแปรปรวนร่วมของความคลาดเคลื่อนในการวัดตัวแปรภายในแฝง (Ψ) โดยลักษณะการทดสอบยังคงมีลักษณะซ้อนกันเป็นระดับลดหลั่นดังนี้

H_{form} : รูปแบบไม่แปรเปลี่ยน (ขนาดของเมตริกซ์และสถานะของพารามิเตอร์ เป็นแบบกำหนด, อิสระ และบังคับในเมตริกซ์ Λ , Γ , Θ_δ , Φ และ Ψ เหมือนกัน)

$$\begin{aligned} H_{\Lambda_x} : \quad & \Lambda_x^{(1)} = \Lambda_x^{(2)} \\ H_{\Lambda_x \Gamma} : \quad & \Lambda_x^{(1)} = \Lambda_x^{(2)} \quad \Gamma^{(1)} = \Gamma^{(2)} \\ H_{\Lambda_x \Gamma \Theta_\delta} : \quad & \Lambda_x^{(1)} = \Lambda_x^{(2)} \quad \Gamma^{(1)} = \Gamma^{(2)} \quad \Theta_\delta^{(1)} = \Theta_\delta^{(2)} \\ H_{\Lambda_x \Gamma \Theta_\delta \Phi} : \quad & \Lambda_x^{(1)} = \Lambda_x^{(2)} \quad \Gamma^{(1)} = \Gamma^{(2)} \quad \Theta_\delta^{(1)} = \Theta_\delta^{(2)} \quad \Phi^{(1)} = \Phi^{(2)} \\ H_{\Lambda_x \Gamma \Theta_\delta \Phi \Psi} : \quad & \Lambda_x^{(1)} = \Lambda_x^{(2)} \quad \Gamma^{(1)} = \Gamma^{(2)} \quad \Theta_\delta^{(1)} = \Theta_\delta^{(2)} \quad \Phi^{(1)} = \Phi^{(2)} \\ & \Psi^{(1)} = \Psi^{(2)} \end{aligned} \quad (10)$$

อย่างไรก็ตามในการทดสอบความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มตัวอย่าง ไม่มีเกณฑ์ตายตัวว่าต้องทำการทดสอบความไม่แปรเปลี่ยนของค่าพารามิเตอร์ใดบ้าง โดยมีเงื่อนไขขั้นต่ำในการทดสอบคือ การทดสอบความเท่ากันของโครงสร้างองค์ประกอบและความเท่ากันของค่า Factor Loading หรืออาจจะกล่าวได้ว่า สารสำคัญที่ควรสนใจคือการทดสอบความไม่แปรเปลี่ยนของความสัมพันธ์ระหว่างตัวแปร (Marsh; Hau; Chung; & Siu. 1998: 149)

การประเมินโมเดลในการทดสอบความไม่แปรเปลี่ยนในโมเดลการวัดนั้น ใช้ Likelihood - Ratio Test (LRT) หรือที่เรียกว่า การทดสอบความแตกต่างของค่าไคสแควร์ ($\Delta\chi^2$) ที่ละสมมติฐาน (Bollen. 1989: 368) โดยการประเมินความเหมาะสมพิจารณาว่า นัยสำคัญของค่าความแตกต่างขององศาอิสระ (Δdf) (Cheung; & Rensvold. 2002: 239) หากไม่มีนัยสำคัญทางสถิติหมายความว่าโมเดลมีความเหมาะสมระหว่างกลุ่มตัวอย่าง นอกจากนี้ยังพิจารณาถึงค่า GFI's และค่า Δ 's หากมีค่าสูง หมายความว่าโมเดลมีความเหมาะสมระหว่างกลุ่มตัวอย่างเช่นกัน

4. เมตริกซ์หลายคุณลักษณะหลายวิธี

เมตริกซ์หลายคุณลักษณะหลายวิธี (Multitrait-Multimethod Matrix) เป็นกระบวนการประเมินความเที่ยงตรงตามโครงสร้างของเครื่องมือวัดคุณลักษณะ โดยอาศัยสมมติฐานที่ว่า ถ้าเครื่องมือวัดกับเกณฑ์มีลักษณะร่วมกัน จะมีค่าสหสัมพันธ์กันสูง ถ้าเครื่องมือวัดกับเกณฑ์มีลักษณะต่างกัน จะมีค่าสหสัมพันธ์กันต่ำ แคมป์เบล และ ฟิสต์ (Campbell; & Fiske. 1959) ได้พยายามคิดค้นวิธีดำเนินการในทางปฏิบัติเพื่อให้นักวิจัยสามารถนำไปใช้ได้ ในเมตริกซ์หลายคุณลักษณะหลายวิธีนั้น แคมป์เบล และ ฟิสต์ ได้เสนอแนะความเที่ยงตรงแบบใหม่ 2 แบบ คือ ความเที่ยงตรงร่วม (Convergent Validity) และความเที่ยงตรงแยก (Discriminant Validity) โทรชิม (Trochim. 1999) ให้ความหมายความเที่ยงตรงร่วมว่า เป็นระดับของความคล้ายคลึงกันของแบบวัดที่ควรจะคล้ายคลึงกันตามทฤษฎี ตัวอย่างเช่น ความเที่ยงตรงร่วมของแบบทดสอบวัดความสามารถทางการคำนวณ เราอาจจะนำไปสัมพันธ์คะแนนของแบบวัดความสามารถทางการคำนวณฉบับอื่น ๆ ที่มีจุดมุ่งหมายเพื่อวัดความสามารถทางคณิตศาสตร์เหมือนกับฉบับที่เราสร้าง เมื่อเราหาสหสัมพันธ์ระหว่างคะแนนของแบบวัดทั้งสองฉบับแล้วมีความสัมพันธ์กันสูง ก็เป็นหลักฐานบ่งชี้ว่ามีความเที่ยงตรงร่วม ส่วนความเที่ยงตรงแยก โทรชิมให้ความหมายว่าเป็นระดับของความแตกต่างกันของแบบวัดที่ควรจะแตกต่างกันตามทฤษฎี ตัวอย่างเช่น ความเที่ยงตรงแยกของแบบทดสอบวัดความสามารถทางการคำนวณ ที่นำไปหาความสัมพันธ์กับคะแนนของแบบวัดทางภาษา เมื่อแบบวัดทั้งสองฉบับนี้มีความสัมพันธ์ต่ำก็เป็นหลักฐานบ่งชี้ว่ามีความเที่ยงตรงแยก ซึ่งความเที่ยงตรงร่วมควรมีค่าสูงกว่า ความเที่ยงตรงแยกควรมีค่าต่ำกว่า และความเชื่อมั่นของเครื่องมือวัดแต่ละวิธีควรมีค่าสูงสุด (บุญเชิด ภิญญอนันตพงษ์. 2545: 189)

เมตริกซ์หลายคุณลักษณะหลายวิธี เมื่ออยู่ในรูปของเมตริกซ์หรือตารางแสดงความสัมพันธ์แล้ว การประเมินความเที่ยงตรงตามโครงสร้างก็จะง่ายขึ้น ภายในเมตริกซ์หลายคุณลักษณะหลายวิธี จะแสดงชุดของข้อมูลที่วัดในแต่ละคุณลักษณะ (traits) ที่หลากหลาย โดยใช้วิธีการวัดหลายวิธี เช่น การสอบแบบเขียนตอบ (paper and pencil test) การสังเกตโดยตรง (direct observation) การวัดการปฏิบัติ (performance measure) แต่ในเมตริกซ์หลายคุณลักษณะหลายวิธี ก็เป็นวิธีที่มีข้อจำกัดคือจะต้องวัดแต่ละคุณลักษณะในทุกวิธีการที่กำหนด

		Method ₁			Method ₂			Method ₃		
	Traits	A ₁	B ₁	C ₁	A ₂	B ₂	C ₂	A ₃	B ₃	C ₃
Method ₁	A ₁	(.89)								
	B ₁	.51	(.89)							
	C ₁	.38	.37	(.76)						
Method ₂	A ₂	.57	.22	.09	(.93)					
	B ₂	.22	.57	.10	.68	(.94)				
	C ₂	.11	.11	.46	.59	.58	(.84)			
Method ₃	A ₃	.56	.22	.11	.67	.42	.33	(.94)		
	B ₃	.23	.58	.12	.43	.66	.34	.67	(.92)	
	C ₃	.11	.11	.45	.34	.32	.58	.58	.60	(.85)

ภาพประกอบ 3 ลักษณะของเมตริกซ์หลายคุณลักษณะหลายวิธี (Trochim. 1999)

โครงสร้างของเมตริกซ์หลายคุณลักษณะหลายวิธี มีลักษณะเป็นเมตริกซ์สหสัมพันธ์ ดังภาพประกอบ 3 แสดงเมตริกซ์หลายคุณลักษณะหลายวิธี สำหรับวัด 3 คุณลักษณะ (A, B และ C) แต่ละคุณลักษณะจะวัดด้วยวิธีการที่แตกต่างกัน 3 วิธี (วิธีที่ 1, 2 และ 3) ค่าที่แสดงในกรอบแต่ละกรอบ คือ วิธีการแต่ละวิธี เมตริกซ์หลายคุณลักษณะหลายวิธีจะแสดงเมตริกซ์สหสัมพันธ์ระหว่างชุดข้อมูล ยกเว้นในแนวไดอะกอนอลของเมตริกซ์ซึ่งโดยปกติควรจะมีค่าเป็น 1 เพราะสหสัมพันธ์ระหว่างตัวมันเองจะมีค่าเป็น 1.00 เสมอ เราจะแทนที่ค่านั้นด้วยค่าความเชื่อมั่นของเครื่องมือแต่ละชุดของข้อมูล ก่อนที่จะแปลความหมายของเมตริกซ์หลายคุณลักษณะหลายวิธี จะต้องทำความเข้าใจในส่วนต่าง ๆ ของเมตริกซ์เสียก่อน อันดับแรกจะสังเกตว่าค่าในเมตริกซ์จะแสดงค่าสหสัมพันธ์มีลักษณะเป็นเมตริกซ์สมมาตร ดังนั้น เราจะมองเพียงครึ่งหนึ่งของเมตริกซ์ อันดับที่สอง สหสัมพันธ์แยกออกเป็น 3 กลุ่มใหม่ แสดงด้วยรูปแบบ 3 รูปคือ ไดอะกอนอล, สามเหลี่ยม และสี่เหลี่ยม ดังนี้

ไดอะกอนอลที่แสดงสหสัมพันธ์ระหว่างการวัดของคุณลักษณะที่เหมือนกันโดยใช้วิธีการเดียวกัน (Monotrait - Monomethod Diagonal) เรียกว่า ไดอะกอนอลของความเชื่อมั่น (The Reliability Diagonal) การประมาณค่าความเชื่อมั่นสำหรับเครื่องมือวัดแต่ละชุด สามารถเลือกวิธีประมาณค่าได้หลายวิธี (เช่น การทดสอบซ้ำ, การหาความสอดคล้องภายใน ฯลฯ)

เนื่องจากเครื่องมือมี 9 ชุด ดังนั้นจึงมีค่าความเชื่อมั่นอยู่ 9 ค่า ความเชื่อมั่นค่าแรกในตัวอย่าง (.89) เป็นค่าความเชื่อมั่นระหว่างคุณลักษณะ A วิธีการที่ 1 หรือค่าความเชื่อมั่นของเครื่องมือวัดคุณลักษณะ A ที่วัดด้วยวิธีการที่ 1 นั่นเอง ค่าความเชื่อมั่นค่าที่ 2 (.89) เป็นค่าความเชื่อมั่นระหว่างคุณลักษณะ B วิธีการที่ 1 หรือค่าความเชื่อมั่นของเครื่องมือวัดคุณลักษณะ B ที่วัดด้วยวิธีการที่ 1 เป็นต้น

ไดอะกอนอลที่แสดงสหสัมพันธ์ระหว่างการวัดของคุณลักษณะที่เหมือนกันโดยใช้วิธีการที่แตกต่างกัน (Monotrait - Heteromethod Diagonal) เรียกว่า ไดอะกอนอลความเที่ยงตรง (The Validity Diagonal) เมตริกซ์หลายคุณลักษณะหลายวิธีจะแสดงค่าสหสัมพันธ์ภายในกล่องสี่เหลี่ยม นั่นคือ ความเที่ยงตรงจะแสดงในเมตริกซ์ไดอะกอนอลที่อยู่ภายในกล่องสี่เหลี่ยมแต่ละกล่อง ในรูปภาพ 1 สหสัมพันธ์ระหว่าง A1-A2 มีค่าเป็น 0.57 นั่นคือ สหสัมพันธ์ระหว่างวิธีการวัด 2 วิธี (1 และ 2) ที่วัดคุณลักษณะเดียวกัน (A) ค่าเหล่านี้ควรจะมีความสัมพันธ์กันสูง นั่นหมายความว่า ยังมีค่าสหสัมพันธ์กันสูงยิ่งจะทำให้มีความเที่ยงตรงสูง ความเที่ยงตรงในที่นี้ก็คือ “ความเที่ยงตรงร่วม” (Convergent Validity)

ปกติแล้วชุดของเครื่องมือที่ดีควรมีค่าในแนวไดอะกอนอลดังกล่าวสูงเพื่อแสดงว่าตัวเครื่องมือสามารถวัดสิ่งที่มุ่งทำการวัดได้เป็นอย่างดี แต่อย่างไรก็ตามค่าเหล่านี้มักจะต่ำกว่าค่าในไดอะกอนอลของความเชื่อมั่น

สหสัมพันธ์ระหว่างการวัดคุณลักษณะต่างกันแต่ใช้วิธีการเดียวกัน (The Heterotrait-Monomethod Triangles) ค่าเหล่านี้เรียกว่า ค่าความเที่ยงตรงแยก (Discriminant Validity) ตัวอย่างเช่น สหสัมพันธ์ระหว่าง A1-B1 มีค่าเป็น 0.51 อยู่ในส่วนบนทางซ้ายของสามเหลี่ยมคุณลักษณะต่างกันและวิธีการเดียวกัน สังเกตว่าเป็นสหสัมพันธ์ของคุณลักษณะที่ต่างกัน (A และ B) แต่วัดในวิธีการเดียวกัน (วิธีการที่ 1) ค่าเหล่านี้ควรจะมีค่าความสัมพันธ์กันต่ำ ซึ่งหมายถึงมี “ความเที่ยงตรงแยก” สูง ถ้าสหสัมพันธ์มีค่าสูง นั่นคงเป็นเพราะประสิทธิภาพของวิธีการที่ใช้วัด แสดงว่า เครื่องมือแต่ละอย่างใช้วัดความสามารถได้หลายอย่างหรือแสดงว่าเครื่องมือ “แกร่ง”

สหสัมพันธ์ระหว่างการวัดคุณลักษณะและใช้วิธีการที่แตกต่างกัน (The Heterotrait - Heteromethod Triangles) หรือเป็นค่าความเที่ยงตรงแยก (Discriminant Validity) อีกชุดหนึ่งที่ทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างผลการวัดความสามารถหลายอย่างด้วยเครื่องมือหลายชนิด เช่น สหสัมพันธ์ระหว่าง A1-B2 มีค่า 0.22 โดยทั่วไปแล้วเครื่องมือวัดแต่ละชนิดมักที่จะมุ่งวัดความเที่ยงตรงของความสามารถอย่างใดอย่างหนึ่งเท่านั้น ดังนั้นสัมประสิทธิ์สหสัมพันธ์ของการวัดคุณลักษณะและวิธีการดังกล่าวนี้ย่อมมีค่าต่ำที่สุดในเมตริกซ์และไม่มีนัยสำคัญ

ค่าสหสัมพันธ์ของเครื่องมือวัดชนิดเดียวกับความสามารถเดียวเท่านั้น (Monomethod Blocks) ซึ่งจะมีจำนวนบล็อกทั้งหมดเท่ากับจำนวนเครื่องมือ ในตัวอย่างได้แก่ บล็อกในแนวไดอะกอนอลของความเชื่อมั่นนั่นเอง

ค่าสหสัมพันธ์ทั้งหมดที่เกิดจากวิธีการวัดต่างกัน (Heteromethod Blocks) นั่นคือจะมี $(K(K-1)/2)$ ค่าที่แสดงค่าสหสัมพันธ์ของวิธีการวัดต่างกัน เมื่อ K = จำนวนวิธีการวัด ในตัวอย่างนี้มีวิธีการวัด 3 วิธี ดังนั้น จะได้กล่องสี่เหลี่ยมที่แสดงวิธีการวัดต่างกันจำนวน $(3(3-1)/2 = 3)$ กล่อง ก็คือ กล่อง 3 กล่องที่อยู่ครึ่งล่างของเมตริกซ์

หลักการแปลความหมาย

ในการแปลความหมาย ผู้วิจัยจะต้องตัดสินใจแปลความหมายของ เมตริกซ์หลายคุณลักษณะหลายวิธี ด้วยตัวเอง โดยยึดหลักการหรือกฎในการแปลความหมาย แต่ก็มีบางครั้ง ที่กฎการแปลความหมายบางอย่างจะถูกฝ่าฝืน ผู้วิจัยอาจจะแปลผลได้อย่างราบรื่นว่าเครื่องมือวัดมีความเที่ยงตรงตามโครงสร้างอย่างดี โดยที่ผู้วิจัยไม่จำเป็นจะต้องยึดมั่นมากนักในการนำหลักการไปใช้ในการวิจัย

ตัวอย่างในการศึกษากับเด็กเกรด 6 เราต้องการวัดคุณลักษณะ 3 คุณลักษณะ คือ คุณลักษณะความภูมิใจในตนเอง (SE) คุณลักษณะการเปิดเผยตน (SD) และการควบคุมตนเอง (LC) ยิ่งกว่านั้น เราจะวัดคุณลักษณะทั้ง 3 ด้วยวิธีการวัด 3 วิธี คือ การประเมินตนเองของเด็ก โดยการใช้การเขียนตอบ (P&P) การประเมินโดยครู (Teacher) และการประเมินโดยครอบครัว (Parent) ผลที่ได้จะแสดงใน เมตริกซ์หลายคุณลักษณะหลายวิธี ดังนี้

	Traits	P&P			Teacher			Parent		
		SE ₁	SD ₁	LC ₁	SE ₂	SD ₂	LC ₂	SE ₃	SD ₃	LC ₃
P&P	SE ₁	(.89)								
	SD ₁	.51	(.89)							
	LC ₁	.38	.37	(.76)						
Teacher	SE ₂	.57	.22	.09	(.93)					
	SD ₂	.22	.57	.10	.68	(.94)				
	LC ₂	.11	.11	.46	.59	.58	(.84)			
Parent	SE ₃	.56	.22	.11	.67	.42	.33	(.94)		
	SD ₃	.23	.58	.12	.43	.66	.34	.67	(.92)	
	LC ₃	.11	.11	.45	.34	.32	.58	.58	.60	(.85)

ภาพประกอบ 4 เมตริกซ์สหสัมพันธ์ของ 3 คุณลักษณะและ 3 วิธีการวัด

การแปลความหมายอาศัยหลักการพื้นฐานของเมตริกซ์หลายคุณลักษณะหลายวิธี คือ

- 1) สัมประสิทธิ์ในไดอะกอนอลที่แสดงค่าความเชื่อมั่นจะต้องมีค่าสูงในเมตริกซ์
- 2) สัมประสิทธิ์ในไดอะกอนอลที่แสดงค่าความเที่ยงตรงควรจะมีนัยสำคัญแตกต่าง

จากศูนย์และมีค่าสูงเพียงพอ นั่นคือ เป็นหลักฐานที่แสดงความเที่ยงตรงร่วม สหสัมพันธ์ทั้งหมดในรูปภาพ 2 เป็นไปตามเกณฑ์นี้

- 3) สัมประสิทธิ์ในไดอะกอนอลที่แสดงค่าความเที่ยงตรงควรจะมีค่าสูงกว่า

สัมประสิทธิ์ที่แสดงคุณลักษณะต่างกันและวิธีการต่างกัน ในตัวอย่างนี้ (SE P&P)-(SE Teacher) ควรจะสูงกว่า (SE P&P)-(SD Teacher), (SE P&P)-(LC Teacher), (SD P&P)-(SE Teacher) และ (LC P&P)-(SE Teacher) นั่นคือ เป็นจริงในทุก ๆ กรณีสำหรับตัวอย่างนี้

- 4) สัมประสิทธิ์ความเที่ยงตรงควรจะสูงกว่าทุก ๆ สัมประสิทธิ์ในสามเหลี่ยมของ

คุณลักษณะต่างกันแต่วิธีการเดียวกัน นั่นคือเน้นว่าองค์ประกอบคุณลักษณะจะต้องมีความแข็งแกร่งมากกว่าองค์ประกอบวิธีการ สังเกตว่าไม่เป็นจริงในทุก ๆ กรณี ตัวอย่างเช่น (LC P&P)-(LC Teacher) มีสหสัมพันธ์เป็น 0.46 น้อยกว่า (SE Teacher)-(SD Teacher), (SE Teacher)-(LC Teacher) และ (SD Teacher)-(LC Teacher) จะเห็นว่า ไม่เป็นไปตามเกณฑ์ เป็นเพราะองค์ประกอบวิธีการโดยเฉพาะการประเมินของครู อาจเป็นไปได้การประเมินของครูเป็นวิธีการวัดที่มีประสิทธิภาพมาก

- 5) ความสัมพันธ์ระหว่างคุณลักษณะจะมีรูปแบบคล้าย ๆ กัน ดูในสามเหลี่ยมทั้งหมด

ในตัวอย่างของเราจะเห็นความชัดเจนในเกณฑ์ที่กล่าวมานี้ สังเกตว่าในสามเหลี่ยมทั้งหมด SE-SD มีความสัมพันธ์กันสูงมากกว่าความสัมพันธ์กับ LC

ข้อเด่นและข้อด้อยของเมตริกซ์หลายคุณลักษณะหลายวิธี

แนวคิดของ เมตริกซ์หลายคุณลักษณะหลายวิธี เป็นการประเมินความเที่ยงตรงตามโครงสร้าง ในเมตริกซ์หนึ่ง ๆ มันเป็นไปได้ที่เราจะตรวจสอบทั้งความเที่ยงตรงร่วมและความเที่ยงตรงแยกพร้อมกัน แม้จะมีประโยชน์ เมตริกซ์หลายคุณลักษณะหลายวิธี ก็ยังมีผู้นำไปใช้น้อย มีเหตุผลด้วยกันหลายประการ ประการแรกคือรูปแบบที่สมบูรณ์ (purest form) เมตริกซ์หลายคุณลักษณะหลายวิธี จะต้องมีการวัดทั้งภายในกลุ่มและข้ามกลุ่ม คือ มีการวัดคุณลักษณะที่หลากหลายโดยใช้วิธีการวัดหลายแบบ ขณะที่ Campbell และ Fiske กล่าวว่าไว้อย่างชัดเจนว่า สามารถใช้รูปแบบที่ไม่สมบูรณ์ได้ โดยจะเน้นหนักความสำคัญของรูปแบบจำลองของคุณลักษณะเดียวกันที่ใช้วิธีการแตกต่างกันในการประยุกต์ใช้กับเนื้อหาในการวิจัย มันเป็นไปได้ที่การวัดคุณลักษณะทั้งหมดจะใช้วิธีการที่แตกต่างกันทั้งหมด ในการนำไปใช้ในการวิจัยทางสังคมศาสตร์โดยมาก เป็นไปไม่ได้ที่จะใช้วิธีการที่กำหนดแน่นอนในทุก ๆ ส่วนของรูปแบบการวิจัย ประการที่สองการตัดสินใจ โดยธรรมชาติของ เมตริกซ์หลายคุณลักษณะหลายวิธี อาจจะถูกตัดสินทวนซ้ำใหม่จนได้ผลสรุปที่ถูกต้อง ผู้วิจัยบางคนต้องการ

ทดสอบสำหรับความเที่ยงตรงตามโครงสร้างโดยให้ผลเป็นสัมประสิทธิ์ทางสถิติเพียงตัวเดียวในการทดสอบ ซึ่งจะเหมือนกับสัมประสิทธิ์สหสัมพันธ์ มันเป็นไปได้ในการใช้ เมตริกซ์หลายคุณลักษณะหลายวิธี ที่จะแสดงปริมาณของความเที่ยงตรงตามโครงสร้างด้วยค่าเพียงค่าเดียว ในที่สุดการตัดสินใจโดยธรรมชาติของ เมตริกซ์หลายคุณลักษณะหลายวิธี ก็คือ ผู้วิจัยที่แตกต่างกันจะมีการสรุปผลความเที่ยงตรงตามโครงสร้างที่แตกต่างกัน

5. โมเดลของราส์ชแบบหลายองค์ประกอบ

การวิเคราะห์คุณภาพของเครื่องมือวัดด้วยวิธีการตามแนวทางของทฤษฎีมาตรฐานเดิม (Classical) ถึงแม้จะได้รับความนิยมและใช้กันอย่างแพร่หลายมาเป็นเวลาช้านานแล้วก็ตาม แต่ก็ยังมีปัญหาบางประการเกิดขึ้นอยู่บ้าง เช่น การวิเคราะห์หาค่าความเชื่อมั่น (Reliability) และค่าความคลาดเคลื่อนมาตรฐานของการวัด (Standard Error of Measurement) แฮมเบิลตันและสวามินาธาน (Hambleton; & Swaminathan. 1985: 2-3) กล่าวว่า ความเชื่อมั่นที่นิยามในลักษณะของแบบทดสอบคู่ขนานนั้นยากจะสัมฤทธิ์ผลในทางปฏิบัติ เพราะผลการตอบของผู้ตอบอาจจะไม่เหมือนเดิมในการตอบครั้งที่สอง ส่วนค่าความคลาดเคลื่อนมาตรฐานของการวัดซึ่งกำหนดให้มีค่าเท่ากันทุกคนนั้น ลอร์ด และโนวิก (Lord; & Novick. 1968) ทำการวิจัยแล้วพบว่า ค่าความคลาดเคลื่อนมาตรฐานของการวัดสำหรับคะแนนที่มีค่าปานกลางจะมีขนาดเล็กกว่าค่าความคลาดเคลื่อนมาตรฐานของการวัดสำหรับคะแนนที่มีค่าสูงหรือคะแนนที่มีค่าต่ำ นอกจากนี้การแปลความหมายผลการสอบด้วยคะแนนดิบเป็นปัญหาสำคัญอีกประการหนึ่งที่ไรท์และมาสเตอร์ (Wright; & Masters. 1982: 28-34) ชี้ให้เห็นว่า คะแนนดิบที่ได้จากการตอบแบบทดสอบไม่ใช่คะแนนในมาตราอันตรภาคชั้น (Interval Scale) ไม่สามารถแสดงความหมายของคะแนนลงบนมาตราเชิงเส้นตรง (Linear Scale) ได้ การได้คะแนนรวมเท่ากันของผู้ตอบอาจมาจากการตอบแบบทดสอบต่างข้อความกัน เป็นผลให้การแปลความหมายของคะแนนรวม ไม่สามารถอธิบายคุณลักษณะที่เราต้องการวัดได้อย่างชัดเจน รวมทั้งคะแนนรวมของข้อความและของผู้ตอบจะเปลี่ยนแปลงไปตามการเพิ่มหรือลดของจำนวนข้อความหรือจำนวนของผู้ตอบ โดยพบว่าเมื่อมีการลดจำนวนข้อความจะทำให้คะแนนรวมของผู้ตอบต่ำลง

จากปัญหาที่กล่าวมาแล้ว นักวัดผลจึงได้เสนอวิธีการตอบข้อคำถาม (Item Response Theory; IRT) ขึ้น เพื่อนำไปใช้แก้ปัญหาทางการวัดผลที่สำคัญ ๆ โดยเฉพาะอย่างยิ่งปัญหาที่ไม่สามารถแก้ไขได้ด้วยทฤษฎีแบบมาตรฐานเดิม ทฤษฎีนี้มีจุดเด่นคือ สามารถประมาณค่าความสามารถหรือคุณลักษณะของผู้ตอบเป็นอิสระจากชุดของข้อคำถามหรือข้อความที่เลือกใช้ ไม่ว่าจะเป็นเลือกกลุ่มตัวอย่างย่อยกลุ่มใดมาใช้ในการประมาณค่าพารามิเตอร์ของข้อความ ค่าประมาณเหล่านั้นจะมีค่าคงที่ ซึ่งคุณสมบัติข้อนี้เป็นประโยชน์อย่างมากสำหรับนำมาใช้ในการสร้างและพัฒนาแบบทดสอบในปัจจุบัน (ผจงจิต อินทสุวรรณ. 2525: 66-67) และสืบเนื่องมาจากความแตกต่างทางด้านแนวคิดของบรรดานักวัดผลที่เกี่ยวกับการใช้โมเดลทาง

คณิตศาสตร์ เพื่ออธิบายความสัมพันธ์ระหว่างระดับความสามารถ (Ability) หรือคุณลักษณะภายในตัวบุคคล (Trait) กับพฤติกรรมการตอบสนองต่อข้อคำถามหรือข้อความ (Performance) ซึ่งเป็นหลักการสำคัญของทฤษฎีนี้แตกต่างกัน เป็นผลให้มีโมเดลที่ใช้อธิบายความสัมพันธ์ดังกล่าวหลายโมเดล แต่โมเดลที่มีผู้นิยมนำมาใช้กันมากในปัจจุบัน ได้แก่ ราสช์ โมเดล (Rasch Model; Rasch. 1960) เพราะโมเดลนี้มีจำนวนพารามิเตอร์ที่ต้องประมาณค่าน้อยกว่าโมเดลอื่น ๆ ทำให้การวิเคราะห์ง่าย ไม่ยุ่งยากซับซ้อน (Hambleton. 1989: 157) โดยเฉพาะอย่างยิ่ง ถ้ากลุ่มตัวอย่างที่ใช้ในการวิเคราะห์มีจำนวนไม่มากนัก สามารถใช้เครื่องคิดเลขคำนวณด้วยมือได้เลย (Wright; & Stone. 1979: 28-45) นอกจากนี้ ราสช์ โมเดล ยังสามารถประมาณค่าพารามิเตอร์ให้ค่า ถูกต้องกว่าโมเดลแบบอื่น ๆ เมื่อนำไปใช้กับกลุ่มตัวอย่างขนาดเล็ก (van de Vijver. 1986: 48; & Lord. 1980: 190) ด้วยเหตุผลดังกล่าว จึงทำให้มีการนำเอาโมเดลนี้มาใช้ในการประมาณค่าพารามิเตอร์ได้อย่างมีประสิทธิภาพทั้งประหยัดเวลาและค่าใช้จ่ายในการวิเคราะห์ข้อมูล ในระยะที่ผ่านมาได้มีการนำ ราสช์ โมเดล มาประยุกต์ใช้กับการวัดพฤติกรรมทางด้านพุทธิพิสัยใน แบบวัดที่มีการให้คะแนนเป็นสองค่า (Dichotomous) ทั้งที่การวัดพฤติกรรมศาสตร์มิได้จำกัดอยู่เฉพาะการวัดความสามารถทางสมองประการเดียวเท่านั้น แต่ยังรวมไปถึงการวัดเจตคติ การวัดบุคลิกภาพ หรือการวัดภาคปฏิบัติที่ให้คะแนนเป็นแบบหลายค่า (Polytomous) ด้วยเหตุนี้จึงทำให้ในระยะต่อมา นักวัดผลได้ปรับขยายโมเดลต่าง ๆ ที่พัฒนาขึ้นในระยะแรก เพื่อนำไปใช้วัดพฤติกรรมทางด้านคุณลักษณะและทักษะการปฏิบัติ (Performance) เพิ่มขึ้นด้วย อาทิ โมเดลที่พัฒนาขึ้นจากราสช์ โมเดล เช่น เรตติ้ง สเกล โมเดล (Rating Scale Model) ของแอนดริช (Andrich. 1978) พาเชียล เครดิต โมเดล (Partial Credit Model) ของมาสเตอร์ (Masters. 1980) และนอกจากนี้ยังมี เกรด เรสพอน โมเดล (Grade Response Model) ของ ซามิโงมา (Samejima. 1969) ซึ่งพัฒนามาจากโลจิสติก โมเดล ที่ใช้พารามิเตอร์สองตัว (Two – Logistic Model; 2PL) ของเบิร์นบอม (Birnbbaum. 1968) เป็นต้น

ต่อมา ลินาเคอร์ (Linacre. 1989) ได้พัฒนาโมเดลที่มีความแตกต่างจากโมเดลของทฤษฎีการตอบข้อคำถามแบบปกติทั่วไป คือมีการขยายแนวคิดโมเดลของทฤษฎีการตอบข้อคำถามจากที่ให้ค่าพารามิเตอร์แค่ 2 ตัว คือ พารามิเตอร์ของข้อคำถาม (Item Parameter) ที่ประกอบด้วย ค่าความยาก ค่าอำนาจจำแนก และค่าการเดา กับ พารามิเตอร์ของผู้เข้าสอบ (Examinee Parameter) ซึ่งได้แก่ ค่าความสามารถของผู้ตอบ โมเดลที่ลินาเคอร์พัฒนาขึ้นมาคือ โมเดลของราสช์แบบหลายองค์ประกอบ (Many-Facets Rasch Model: FACETS) ที่ใช้ในการตรวจสอบคุณภาพของการให้คะแนนการตอบที่เป็นลำดับขั้น (Order Response Categories) ลินาเคอร์ ได้พัฒนาขึ้นมาจากแนวคิดของราสช์โมเดล (Engelhard. 1994: 94; citing Andrich. 1988; Rasch. 1980; Wright; & Masters. 1982) โดยการเพิ่มค่าพารามิเตอร์ที่สนใจให้ศึกษาได้มากกว่า 2 พารามิเตอร์ เช่น พารามิเตอร์ของผู้ตรวจให้คะแนนในกรณีที่มีการใช้ผู้ตรวจให้คะแนนมากกว่า 2 คนขึ้นไป เป็นต้น โมเดลของราสช์แบบหลายองค์ประกอบใช้การตรวจสอบ

ความเหมาะสมของข้อมูลกับโมเดลในการแสดงความเที่ยงตรงตามโครงสร้าง ที่ต้องการตรวจสอบคุณภาพในการให้คะแนน นอกจากนั้น โมเดลยังสามารถบ่งชี้พฤติกรรมการตรวจให้คะแนนของผู้ตรวจ (Rater) ว่ามีความเคร่งครัดหรือตรวจให้คะแนนอย่างมั่งก่ายเกินไป ฟาเซตโมเดลมีลักษณะที่สำคัญ คือ เป็นโมเดลเส้นตรงเชิงบวก (Additive linear model) ที่อาศัยการแปลงคะแนนสังเกตของผู้ตอบแต่ละคน แต่ละรายการของกลุ่มตัวอย่างให้อยู่ในรูปของคะแนนโลจิต (Logit Score) ใช้วิธีการวิเคราะห์ข้อมูลตามแนวทางของ Regression ที่มีอัตราส่วนของชั้นความน่าจะเป็นที่ต่อเนื่องกัน (Successive Category Probabilities; log odds) เป็นตัวแปรตาม และฟาเซต (Facets) เป็นตัวแปรอิสระ ถ้าวัดความสามารถทางการเขียนด้วยข้อสอบเลือกตอบ (multiple choice items) โมเดลของราส์ซที่ เหมาะสมกับข้อมูลเชิงประจักษ์ที่มีการให้คะแนนแบบสองค่า (dichotomous) สามารถเขียนเป็นสมการได้ (Engelhard. 1994: 95) ดังนี้

$$\ln [P_{ni1} / P_{ni0}] = \beta_n - \delta_i \quad (11)$$

เมื่อ P_{ni1}	แทน	ความน่าจะเป็นที่ผู้ตอบคนที่ n จะตอบข้อสอบข้อที่ i ได้ถูกต้อง (x=1)
P_{ni0}	แทน	ความน่าจะเป็นที่นักเรียน n จะตอบข้อสอบข้อที่ i ไม่ถูกต้องหรือตอบผิด (x=0)
β_n	แทน	ระดับความสามารถของผู้ตอบคนที่ n
δ_i	แทน	ค่าความยากของข้อสอบข้อที่ i

โมเดลนี้มี 2 องค์ประกอบ คือ องค์ประกอบความสามารถของนักเรียน (Student Competence) และองค์ประกอบความยากของข้อสอบ (Item Difficulty) โมเดลนี้สามารถใช้กับข้อมูลมาตราประเมิน (rating scale) ที่มีหลายองค์ประกอบถ้าเป็นโมเดลที่มี 3 องค์ประกอบ (ดังเช่น ความสามารถของนักเรียน, ความยากของข้อสอบ และความเคร่งครัดของผู้ตรวจ) โมเดลของราส์ซที่เหมาะสมกับข้อมูล สามารถเขียนเป็นสมการได้ดังนี้ (Engelhard. 1994: 95)

$$\ln [P_{nijk} / P_{nijk-1}] = \beta_n - \delta_i - \lambda_j - \tau_k \quad (12)$$

เมื่อ P_{nijk}	แทน	ความน่าจะเป็นที่ผู้ตรวจ j จะประเมินผู้ตอบคนที่ n ให้ระดับคะแนน k จากการตอบข้อความรายการที่ i
P_{nijk-1}	แทน	ความน่าจะเป็นที่ผู้ตรวจ j จะประเมินผู้ตอบคนที่ n ให้ระดับคะแนน k-1 จากการตอบข้อความรายการที่ i
β_n	แทน	ระดับความสามารถของผู้ตอบคนที่ n

δ_i	แทน	ค่าความยากของข้อความรายการที่ i
λ_j	แทน	ค่าความเคร่งครัดของผู้ตรวจ j
τ_k	แทน	ค่าความยากประจำชั้นที่ k สัมพันธ์กับชั้นที่ k-1

สำหรับ τ_k ถือเป็นค่าพารามิเตอร์ที่ไม่นับเป็นองค์ประกอบของโมเดล

ฟาเซตโมเดลเป็นโมเดลเอกมิติ (Unidimensional) ที่สามารถรวมค่าพารามิเตอร์ความสามารถของนักเรียนเข้ากับองค์ประกอบอื่น ๆ ได้ เช่น รวมเข้ากับองค์ประกอบความยากของข้อคำถามและองค์ประกอบของผู้ประเมิน เป็นต้น ฟาเซตโมเดลมีข้อตกลงเบื้องต้นว่า “ชุดของประชากรที่ได้รับการวัดและชุดของข้อคำถามที่ใช้วัดกลุ่มประชากรนั้น สามารถจัดเป็นลำดับเฉพาะ (unique ordering) ในแต่ละทอมตามลำดับของความสามารถและความยากของข้อคำถาม” (Engelhard. 1994: 95; citing Choppin. 1987: 111) ถ้าข้อมูลมีความเหมาะสมกับโมเดลและลำดับเฉพาะนี้เป็นจริง ก็จะประสบผลสำเร็จตามลักษณะความต้องการของการวัดลักษณะดังกล่าวประกอบด้วย (1) ความสามารถในการแยกค่าพารามิเตอร์กับสถิติที่พอเพียงออกจากกันในการประมาณค่าพารามิเตอร์ (2) ความคงที่ในการประมาณค่าความสามารถ (ability or competence), ความเคร่งครัดของผู้ประเมิน (rater severity) และความยากของข้อคำถาม (item difficulty) (ซึ่งเป็นผลที่เกิดจากคุณสมบัติที่เรียกว่า specific objectivity ของราสส์โมเดล) และ (3) ช่วงสเกลของการวัดที่เท่าเทียมกัน (equal-interval scale) ฟาเซตโมเดลมีส่วนที่คล้ายกับ equating model โดยเฉพาะผู้ประเมิน (raters) เทียบได้กับแบบทดสอบที่มีหลากหลายรูปแบบ เนื่องจากนักเรียนมีความแตกต่างกัน ถูกประเมินด้วยผู้ประเมินที่แตกต่างกัน ดังนั้น จึงจำเป็นต้อง equate หรือปรับสถิติเพื่อให้เหมาะสมกับผู้ประเมินในกลุ่มนั้น

ตัวอย่างของการประเมินโดยใช้โมเดลของราสส์แบบหลายองค์ประกอบ ดังเช่น การประเมินงานเขียนที่รวมเอาหลายฟาเซต (facets) เข้าด้วยกัน โดยทั่วไปผู้ประเมินหลายคนจะใช้มาตรประเมิน (rating scale) ประเมินความสามารถในการเขียนของนักเรียนจากงานหลายชิ้นในหลาย ๆ พฤติกรรม กลุ่มตัวอย่างที่มีขนาดใหญ่สามารถประเมินได้ในหลาย ๆ แนวทาง เช่น การประเมินพฤติกรรมของนักเรียนโดยใช้เกณฑ์การประเมินแบบองค์รวม (Holistic scoring) (Engelhard. 1994: 96; citing Cooper. 1977) นักเรียนอาจเขียนตอบในแบบทดสอบเพียงอย่างเดียว หรืองานที่นักเรียนเขียนนั้นอาจถูกประเมินโดยผู้ประเมินเพียงคนเดียว เป็นต้น ดังเช่น การประเมินงานเขียนของนักเรียนเกรด 8 ที่ประเมินจากพฤติกรรม 5 ด้าน มีผู้ประเมิน 2 คน โดยใช้มาตรประเมิน 4 ระดับ จากฟาเซตโมเดลในสมการที่ 12 ความน่าจะเป็นของนักเรียนคนที่ n ที่มีความสามารถ β_n ได้รับการให้คะแนน x ($x = 0, 1, \dots, m$) จากรายการประเมินที่ δ_i โดยผู้ประเมิน λ_j ที่มีความยากประจำชั้น τ_k เขียนเป็นสมการได้ดังนี้

$$\pi_{ijk} = \frac{1}{1 + \sum_{k=1}^m \exp \left[k (\beta_n - \delta_i - \lambda_j) - \sum_{s=1}^k \tau_s \right]} \quad (13)$$

เมื่อ $x = 0$ และ

$$\pi_{ijk} = \frac{\exp \left[x (\beta_n - \delta_i - \lambda_j) - \sum_{s=1}^x \tau_s \right]}{1 + \sum_{k=1}^m \exp \left[k (\beta_n - \delta_i - \lambda_j) - \sum_{s=1}^k \tau_s \right]} \quad (14)$$

เมื่อ $x = 1, \dots, m$.

การวิเคราะห์ค่าสถิติความเหมาะสมกับโมเดล

การตรวจสอบความเหมาะสมของข้อมูลเชิงประจักษ์กับโมเดลสามารถตรวจสอบได้หลายแนวทาง แนวทางหนึ่งที่นิยมก็คือ การใช้ค่ามาตรฐานของส่วนที่เหลือ (Standardized Residuals; Z_{nij}) เป็นดัชนีบ่งชี้คุณภาพของการให้คะแนน นอกจากนี้ การตรวจสอบความเหมาะสมของข้อมูลกับโมเดลยังมีประโยชน์เกี่ยวกับการวิเคราะห์ค่าความเที่ยงตรง (Validity) และค่าความเชื่อมั่น (Reliability) ของแบบทดสอบอีกด้วย การตรวจสอบความเหมาะสมของข้อมูลกับโมเดลสามารถคำนวณได้จากสมการ (Engelhard. 1994: 96)

$$Z_{nij} = \frac{X_{nij} - E_{nij}}{\left[\sum_{k=0}^m (k - E_{nij})^2 \pi_{ijk} \right]^{\frac{1}{2}}} \quad (15)$$

เมื่อ

$$E_{nij} = \sum_{k=0}^m k \pi_{ijk}$$

X_{nij} คือ คะแนนจากผู้ประเมินคนที่ j สังเกตนักเรียนคนที่ n ในรายการข้อที่ i

E_{nij} คือ ค่าคาดหวังของการประเมินบนโมเดลการวัด

ค่ามาตรฐานของส่วนที่เหลือ (Standardized Residuals) นี้ ให้รายละเอียดที่เกี่ยวกับการตรวจสอบความเหมาะสมระหว่างคะแนนสังเกตกับรูปแบบการประเมินที่คาดหวัง (Expected rating patterns) ของโมเดล ค่ามาตรฐานของส่วนที่เหลือสามารถสรุปความแตกต่างขององค์ประกอบ (Facets) และความแตกต่างของสมาชิก (element) ในองค์ประกอบได้ ค่า

มาตรฐานของส่วนที่เหลือจึงเป็นดัชนีที่แสดงความเหมาะสมของข้อมูลเชิงประจักษ์และโมเดลทางทฤษฎี

สถิติ OUTFIT เป็นดัชนีของค่าเฉลี่ยยกกำลังสองของส่วนที่เหลือแบบไม่ถ่วงน้ำหนัก (unweighted mean square residual) ที่แสดงความแตกต่างระหว่างคะแนนสังเกตกับค่าความคาดหวัง สถิติ OUTFIT, u_j สำหรับองค์ประกอบของผู้ประเมิน สามารถคำนวณได้จากสมการดังนี้

$$u_j = \sum_{n=1}^N \sum_{i=1}^I Z_{nij}^2 / (N + I) \quad (16)$$

เนื่องจากค่าสถิติ OUTFIT เป็นสถิติที่มีค่าค่อนข้างไว (sensitivity) ต่อการระบุข้อความใดที่ไม่เหมาะสมกับโมเดล (Misfit) ดังนั้น ค่าสถิตินี้จึงมีจุดอ่อนในบางลักษณะของการตอบจากกลุ่มผู้ตอบ เช่น ในกรณีที่ผู้ตอบส่วนใหญ่ตอบข้อความใดที่เหมาะสมกับค่าคาดหวังตามโมเดล แต่ปรากฏว่า มีผู้ตอบบางคนที่มีรูปแบบการตอบที่แตกต่างไปจากลักษณะการตอบคำถามของกลุ่มผู้ตอบโดยรวม เช่น คนที่มีคุณลักษณะ (Trait) ในระดับสูงตอบข้อความในบางรายการในระดับต่ำ ๆ หรือในทางตรงกันข้าม ผู้ตอบที่มีคุณลักษณะระดับต่ำตอบกลับข้อความบางรายการในระดับสูง ๆ ซึ่งลักษณะการตอบดังกล่าวนี้ ถือว่าไม่เหมาะสมกับค่าความน่าจะเป็น (Probability) ที่เขาควรจะตอบ จึงทำให้ค่าสถิติปฏิเสธข้อความในบางรายการทั้งไปทั้งที่ความจริงแล้วข้อความดังกล่าวยังใช้ได้ เพียงแต่ผู้ตอบบางคนตอบข้อความรายการนั้นด้วยความเสแสร้งเท่านั้น (Wright; & Masters. 1982: 99) ดังนั้น จึงมีสถิติอีกค่าหนึ่งซึ่งละเอียดขึ้นโดยมี W_{nij} เป็นตัวถ่วงน้ำหนักดังนี้

สถิติ INFIT แทน ค่าเฉลี่ยยกกำลังสองแบบถ่วงน้ำหนัก (weighted mean squares) กับสัดส่วนความแปรปรวนส่วนที่เหลือยกกำลังสองแบบถ่วงน้ำหนัก (weighted proportional to the variance of the square residuals) ที่สามารถคำนวณแต่ละสมาชิกของฟาเซ็ทได้ สถิติ INFIT มีความไว (sensitivity) ต่ำต่อผลการประเมินที่ไม่คาดหวัง (outlying unexpected) สถิติ INFIT, v_j สำหรับองค์ประกอบของผู้ประเมิน สามารถคำนวณได้ดังสมการ

$$v_j = \sum_{n=1}^N \sum_{i=1}^I W_{nij} Z_{nij}^2 / \sum_{n=1}^N \sum_{i=1}^I W_{nij} \quad (17)$$

เมื่อ

$$W_{nij} = \sum_{k=0}^m (k - E_{nij})^2 \pi_{nijk}$$

สถิติที่มีความเหมาะสมกับโมเดลสำหรับองค์ประกอบอื่นๆและสมาชิกภายในองค์ประกอบนั้น นิยมคำนวณแล้วสรุปบนค่าที่เหมาะสม เมื่อโมเดลมีความเหมาะสมกับข้อมูลเชิงประจักษ์ ค่าคาดหวังของค่าเฉลี่ยยกกำลังสอง (Mean square) จะมีค่าเท่ากับ 1 ลันซ์, ไรท์ และลินาเคอร์ (Engelhard. 1994: 97; citing Lunz; Wright; & Linacre. 1990) แนะนำว่าขอบเขตของการยอมรับได้ของโมเดลที่เหมาะสมกับข้อมูลจะมีค่าอยู่ระหว่าง .6 ถึง 1.5 ส่วนไรท์และมาสเตอร์ (Wright; & Masters. 1982: 102) ได้เสนอขอบเขตที่ยอมรับได้ของโมเดลที่เหมาะสมกับข้อมูลเมื่อคำนวณจากค่าสถิติ INFIT และ OUTFIT ควรมีค่าอยู่ระหว่าง -2 ถึง +2 เกณฑ์ต่าง ๆ ที่เป็นเป้าหมายในการศึกษาเพื่อการตรวจสอบคุณภาพของการประเมินอยู่ภายใต้ข้อสรุปของค่ามาตรฐานของส่วนที่เหลือที่แสดงในรูปของค่าเฉลี่ยยกกำลังสองแบบไม่ถ่วงน้ำหนัก (OUTFIT) และค่าเฉลี่ยยกกำลังสองแบบถ่วงน้ำหนัก (INFIT) ในโมเดล 3 องค์ประกอบ (ผู้ตรวจ, ขอบเขตเนื้อหา และตัวนักเรียน)

หนึ่งในทางปฏิบัติค่าสถิติ INFIT ที่คำนวณได้มักใช้เป็นดัชนี (Index) ตัวแรกในการบ่งชี้ข้อความหรือผู้ตอบที่ไม่เหมาะสม (Misfit) กับโมเดล อย่างไรก็ตามถ้าพิจารณาร่วมกับค่าสถิติ OUTFIT ก็จะทำให้ทราบข้อมูลเกี่ยวกับความเหมาะสมกับโมเดลเพิ่มขึ้น จะนำไปสู่การตัดสินใจในการคัดเลือกข้อความและผู้ตอบที่มีรูปแบบการตอบเหมาะสมกับโมเดลได้อย่างถูกต้องมากขึ้น (Wright; & Linacre. 1984: 4-21, 4-29)

สำหรับการศึกษาคครั้งนี้ผู้วิจัยเลือกทดสอบความเที่ยงตรงตามโครงสร้าง โดยใช้เกณฑ์ของลันซ์, ไรท์และลินาเคอร์ (Engelhard. 1994: 97; citing Lunz; Wright; & Linacre. 1990) ที่กำหนดขอบเขตของการยอมรับได้ของโมเดลที่เหมาะสมกับข้อมูลอยู่ระหว่าง .6 ถึง 1.5 ในกรณีที่ผลการวิจัยไม่เป็นไปตามที่คาดหวังผู้วิจัยอาจจะเลือกพิจารณาใช้เกณฑ์ของไรท์และมาสเตอร์ (Wright; & Masters. 1982: 102) ที่กำหนดขอบเขตของการยอมรับได้ของโมเดลที่เหมาะสมกับข้อมูลที่มีค่าอยู่ระหว่าง -2 ถึง +2 เนื่องจากเป็นช่วงที่กว้างกว่าเกณฑ์ของ ลันซ์, ไรท์และลินาเคอร์

สถิติอื่น ๆ ที่มีประโยชน์ในการตรวจสอบตัวแปรแฝง เรียกว่า ดัชนีแยกแยะความเชื่อมั่น (Reliability of Separation Index) ดัชนีนี้ให้ข้อมูลเกี่ยวกับสมาชิกภายใต้ฟ้าเซตที่แยกลำดับของความเชื่อมั่น ที่มีความคล้ายคลึงกับดัชนีความเชื่อมั่นแบบมาตรฐานเดิมที่แสดงอัตราส่วนของความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนการสังเกต ดังสมการ

$$R = \frac{SD^2 - MSE}{SD^2} \quad (18)$$

เมื่อ SD^2 คือ ความแปรปรวนของคะแนนสังเกตของค่าความยากสำหรับองค์ประกอบที่อยู่บนสเกลของตัวแปรแฝง และ MSE คือ ค่าประมาณความคลาดเคลื่อนของค่าเฉลี่ยยกกำลังสองที่คล้ายกับค่าเฉลี่ยของการวัดความแปรปรวนของความคลาดเคลื่อนสำหรับแต่ละสมาชิกภายในองค์ประกอบ

6. งานวิจัยที่เกี่ยวข้อง

งานวิจัยในต่างประเทศ

แทรนท์ (Trant. 1986: 104) ได้ศึกษาความเที่ยงตรงรวมและความเที่ยงตรงแยกของแบบวัดลักษณะความเป็นผู้นำ โดยกลุ่มตัวอย่างเป็นนักเรียนเกรด 11 และ 12 จำนวน 94 คน ใช้วิธีการวัด 2 วิธีคือ ให้นักเรียนทำแบบสำรวจแบบรายงานตนเองและครูให้ทำแบบประเมินนักเรียน พบว่า ด้านลำดับขั้นทักษะการสื่อสารมีความเที่ยงตรงรวมและความเที่ยงตรงแยก

วีสสัน, คลาค และเทลเกน (Watson; Clark; & Tellegen. 1988: 1063-1070) ได้ศึกษาความเที่ยงตรงของแบบวัดอารมณ์ทางด้านบวกและลบ (PANAS Scale) โดยวัดลักษณะอารมณ์ 2 ด้าน คือ ด้านบวกและด้านลบ ใช้วิธีวิเคราะห์หลายลักษณะหลายวิธี พบว่า ความเที่ยงตรงรวมมีค่า 0.89 ถึง 0.95 และความเที่ยงตรงแยกมีค่าต่ำกว่าความเที่ยงตรงรวม

ไบเดอร์แมน (Biederman. 1993) ทำการวิเคราะห์หาความเที่ยงตรงตามโครงสร้างของเครื่องมือวิจัยชุดหนึ่ง ซึ่งประกอบด้วยแบบสอบถามเพื่อรายงานตนเอง (Self-report Inventories) เพื่อวัดความต้องการในการทำงานจำนวน 4 ชุด และแบบสอบถามเพื่อรายงานความสำคัญของงาน (Job-rating Inventories) เพื่อวัดความต้องการในการทำงาน (work reinforces) อีก 3 ชุด ปรากฏว่า การวิเคราะห์ดังกล่าวทำให้สามารถหาความเที่ยงตรงรวมและความเที่ยงตรงแยกได้เป็นอย่างดี

เอ็ดเดนแมน และ ไรเดน (Adelman; & Reidel. 1993) ได้วิเคราะห์ความเที่ยงตรงตามโครงสร้างของระบบฐานความรู้ (Knowledge-Based System) ของ 3 วิชาชีพที่ทำการวัดหลายระดับ โดยการวิเคราะห์แบบหลายลักษณะหลายวิธี (Multitrait-Multimethod Analysis) ปรากฏว่า ความเที่ยงตรงรวมของเครื่องมือชุดดังกล่าวมีค่าสูงกว่าความเที่ยงตรงแยกอีก 2 ชนิดมาก ทำให้ได้เครื่องมือที่ดีเพื่อนำไปใช้ต่อไป

เลเวน (Lievens. 1997) ได้ศึกษาความเที่ยงตรงตามโครงสร้างของวิธีการประเมินผลแบบหนึ่งๆ เรียกว่า “ศูนย์การประเมินค่า” (Assessment Center) โดยการนำวิธีวิเคราะห์แบบหลายลักษณะหลายวิธี (Multitrait-Multimethod Analysis) มาใช้ ศูนย์การประเมินดังกล่าวประกอบด้วยนักประเมินค่ากลุ่มหนึ่งๆ ที่ทำการสังเกตและประเมินค่าเพื่อให้คะแนนผู้มาสมัครสอบในหลายด้าน เช่น การวางแผนและการแก้ปัญหาในสถานการณ์ต่าง ๆ การเล่นบทบาทสมมติ และการอภิปราย เป็นต้น กระบวนการเก็บข้อมูลทั้งหมดใช้สถานการณ์จำลอง โดยมีนักประเมิน

264 คน ทำการสังเกตและประเมินค่าให้คะแนนความสามารถของผู้มาสมัครสอบ 3 ด้าน ในสถานการณ์ 3 อย่าง จากแบบบันทึกวิธีทัศน์ ผลการศึกษาปรากฏว่า การวิเคราะห์แบบหลายลักษณะหลายวิธี (Multitrait-Multimethod Analysis) สามารถหาความเที่ยงตรงตามโครงสร้างของการประเมินค่าแบบ “ศูนย์การประเมินค่า” ได้เป็นอย่างดี

บาคาร์เนลลี และ เคปาลา (Bargaranelli; & Caprara. 2000) ได้้นำการวิเคราะห์แบบหลายลักษณะหลายวิธี (Multitrait-Multimethod Analysis) มาใช้ในการหาความเที่ยงตรงตามโครงสร้างของเครื่องมือทดสอบทางจิตวิทยาชุดหนึ่งประกอบด้วยแบบสอบถาม 2 ชนิด และแบบประเมินค่าอีก 2 ชนิด เพื่อประเมินผลตนเองในหลาย ๆ ด้าน โดยใช้กลุ่มตัวอย่างจำนวน 200 คน ซึ่งแบ่งเป็น 2 กลุ่มคือ ชายและหญิงในจำนวนที่เท่ากันคือกลุ่มละ 100 คน กลุ่มตัวอย่างแต่ละคนจะได้รับการประเมินผลจากเพื่อนสนิทจำนวน 6 คน ผู้วิจัยนำข้อมูลจากการวิเคราะห์แบบหลายลักษณะหลายวิธี ไปทำการวิเคราะห์อีกครั้งด้วยวิธี Structural Equation Models อีก 4 วิธี ผลปรากฏว่า การวิเคราะห์ด้วยวิธีวิเคราะห์แบบหลายลักษณะหลายวิธีได้ผลดีกว่าการวิเคราะห์ด้วยวิธี Structural Equation Models อีก 4 วิธี

งานวิจัยในประเทศ

จิรพร ปัทมะสุคนธ์ (2543: 48-49) ศึกษาความเที่ยงตรงตามโครงสร้างของแบบวัดบุคลิกภาพความเป็นทหาร โดยศึกษาแก่นักเรียนนายร้อยพระจุลจอมเกล้า นักเรียนนายเรือ นักเรียนนายเรืออากาศ ชั้นปีที่ 4 และ 5 สังกัดกระทรวงกลาโหม จำนวน 505 คน ผลการศึกษาพบว่า ความเที่ยงตรงตามโครงสร้างที่วิเคราะห์ด้วยวิธีวิเคราะห์องค์ประกอบเชิงสำรวจ แบบวัดบุคลิกภาพความเป็นทหารชนิดสถานการณ์มีจำนวน 14 องค์ประกอบ มีค่าน้ำหนักองค์ประกอบอยู่ระหว่าง 0.318 ถึง 0.834 ชนิดข้อความมีจำนวน 9 องค์ประกอบ มีค่าน้ำหนักองค์ประกอบอยู่ระหว่าง 0.308 ถึง 0.858 เมื่อวิเคราะห์ด้วยวิธีหลายลักษณะหลายวิธี พบว่า มีค่าสัมประสิทธิ์รวมอยู่ระหว่าง 0.532 ถึง 0.639 ซึ่งมีค่ามากกว่า 0.50 แสดงว่า แบบวัดบุคลิกภาพความเป็นทหารมีความเที่ยงตรงรวม แต่ไม่มีความเที่ยงตรงแยก เพราะพบว่า ค่าสหสัมพันธ์แยกมีค่าอยู่ระหว่าง -0.091 ถึง 0.673 ซึ่งมีบางลักษณะเท่านั้นที่มีค่าสหสัมพันธ์แยกต่ำกว่า 0.50

ดวงใจ บุญยะภาส (2543: 45) ศึกษาความเที่ยงตรงตามโครงสร้างของแบบทดสอบวัดบุคลิกภาพตามทฤษฎีบุคลิกภาพของไอแซกส์ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 4 ในจังหวัดราชบุรี ซึ่งมีจำนวนกลุ่มตัวอย่าง 2,214 คน ผลการศึกษาพบว่า จากการวิเคราะห์องค์ประกอบแบบทดสอบวัดบุคลิกภาพมีจำนวน 3 องค์ประกอบ คือ ด้านการแสดงตัว ด้านอารมณ์แปรปรวน และด้านพฤติกรรมทางจิต สอดคล้องกับแนวคิดของไอแซกส์และทุกด้านวัดในแนวคิดร่วมกัน ส่วนความเที่ยงตรงตามโครงสร้างที่วิเคราะห์ด้วยวิธีวิเคราะห์หลายลักษณะหลายวิธี พบว่า ค่าความเที่ยงตรงรวมมีสหสัมพันธ์กันสูง มีค่าอยู่ระหว่าง 0.564 ถึง 0.632 ส่วน

ค่าความเที่ยงตรงแยกมีสหสัมพันธ์กันต่ำ มีค่าอยู่ระหว่าง -0.016 ถึง 0.381 ซึ่งมีค่าต่ำกว่าความเที่ยงตรงรวม แสดงว่า แบบทดสอบที่สร้างขึ้นมีความเที่ยงตรงตามโครงสร้าง

นพวรรณ ศรีเกตุ. (2550: บทคัดย่อ). ศึกษาคุณภาพแบบวัดด้านทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการ โดยการแสดงหลักฐานด้านความเที่ยงตรงและความเชื่อมั่น กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 5 โรงเรียนมัธยมศึกษาที่อยู่ในเขตกรุงเทพมหานคร ภาคเรียนที่ 2 ปีการศึกษา 2549 จำนวน 185 คน ซึ่งเลือกมาโดยการสุ่มแบบสองขั้นตอน ผลการวิจัยพบว่า แบบวัดทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการที่สร้างขึ้นมีคุณภาพใช้ได้ตามเกณฑ์ โดยมีหลักฐานการแสดงความเที่ยงตรงตามโครงสร้างคำนวณด้วยสูตรของไฮเซนและบอร์นสเทร็ด (Heise and Bohrnstedt) มีค่า .70 มีหลักฐานความเที่ยงตรงเชิงโครงสร้างภายในฉบับด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งและอันดับสอง โมเดลกับข้อมูลเชิงประจักษ์มีความเหมาะสมกันพอดี ดัชนีความสอดคล้องของผู้ประเมิน 3 คน มีค่า .93 ความเชื่อมั่นที่ประมาณค่าด้วยสัมประสิทธิ์การสุปร้องอิง มีค่า .68 ซึ่งมีค่าความเชื่อมั่นอยู่ในเกณฑ์สูง

อัญชลี ศรีกลชาญ (2546: บทคัดย่อ) ศึกษาการพัฒนาแบบวัดพหุปัญญาสำหรับนักเรียนชั้นมัธยมศึกษาตอนต้นตามทฤษฎีพหุปัญญาของการ์ดเนอร์ ประกอบด้วย 8 องค์ประกอบ ได้แก่ ภาษา ตรรกะและคณิตศาสตร์ มิติ ร่างกายและการเคลื่อนไหว ดนตรี มนุษยสัมพันธ์ การเข้าใจตนเอง และธรรมชาติ โดยการตรวจสอบความเที่ยงตรงเชิงโครงสร้างของแบบวัดและสร้างปกติวิสัย กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาตอนต้น สังกัดกรมสามัญศึกษา จังหวัดชลบุรี ปีการศึกษา 2545 จำนวน 800 คน จากผลการศึกษาพบว่า แบบวัดพหุปัญญามีความเที่ยงตรงเชิงโครงสร้างอยู่ในเกณฑ์ดี ค่าไคสแควร์ เท่ากับ 1204.7 ($df = 1461$, $p = 1.00$) $GFI = 0.95$, $AGFI = 0.94$, $CFI = 1.00$, $RMR = 0.03$ และ $RMSEA = 0.00$ องค์ประกอบทั้ง 8 องค์ประกอบมีนัยสำคัญกับข้อสอบทั้ง 60 ข้อ ที่ระดับ .05 ปกติวิสัยของแบบวัดพหุปัญญาจำแนกเป็น 3 ระดับ ได้แก่ ผู้ที่มีพหุปัญญาระดับสูง มีตำแหน่งเปอร์เซ็นต์ไทล์ตั้งแต่ 78 ขึ้นไป ผู้ที่มีพหุปัญญาระดับปานกลาง มีตำแหน่งเปอร์เซ็นต์ไทล์อยู่ระหว่าง 24.0 ถึง 77.9 และผู้ที่มีพหุปัญญาระดับต่ำ มีตำแหน่งเปอร์เซ็นต์ไทล์ต่ำกว่า 24.0

อรอุมา สงวนญาติ (2544: 71-73) ศึกษาความเที่ยงตรงตามโครงสร้างของแบบวัดความภาคภูมิใจในตนเองที่มีรูปแบบการตอบและวิธีวิเคราะห์ต่างกัน กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 กลุ่มที่ 3 โรงเรียนมัธยมศึกษาสังกัดกรมสามัญศึกษา ในกรุงเทพมหานคร จำนวน 1,064 คน เครื่องมือที่ใช้ในการศึกษาเป็นแบบวัดความภาคภูมิใจในตนเองที่ผู้วิจัยสร้างขึ้น 2 ฉบับ เป็นฉบับแบบข้อความและแบบสถานการณ์ จากผลการศึกษาพบว่า ความเที่ยงตรงตามโครงสร้างของแบบวัดความภาคภูมิใจในตนเองที่วิเคราะห์ด้วยวิธีหลายลักษณะหลายวิธี มีค่าสหสัมพันธ์รวม ตั้งแต่ 0.639 ถึง 0.758 ซึ่งมีค่าสูงกว่า 0.50 และสหสัมพันธ์แยกมีค่าตั้งแต่ -0.003 ถึง 0.449 ซึ่งมีค่าต่ำกว่าความเที่ยงตรงรวม ความเที่ยงตรงตามโครงสร้างของแบบวัด

ชนิดข้อความและสถานการณ์ที่วิเคราะห์ค่าความสอดคล้องภายในด้วยค่าสัมประสิทธิ์สหสัมพันธ์ของแบบวัดทั้งฉบับมีค่าเท่ากับ 0.939 และ 0.886 ตามลำดับ ความเที่ยงตรงตามโครงสร้างที่วิเคราะห์ด้วยวิธีวิเคราะห์องค์ประกอบแบบสำรวจ พบว่า แบบวัดทั้งสองชนิดมีเพียง 3 องค์ประกอบที่สอดคล้องกัน ได้แก่ ด้านการยอมรับนับถือตนเอง ด้านการได้รับการยอมรับจากครอบครัว และด้านการได้รับการยอมรับจากบุคคลในโรงเรียน

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการตามขั้นตอนดังนี้

1. การกำหนดประชากรและการเลือกกลุ่มตัวอย่าง
2. การสร้างเครื่องมือที่ใช้ในการวิจัย
3. การเก็บรวบรวมข้อมูล
4. การจัดกระทำและการวิเคราะห์ข้อมูล

การกำหนดประชากรและการเลือกกลุ่มตัวอย่าง

ประชากร

ประชากรที่ใช้ในการวิจัยเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 โรงเรียนมัธยมศึกษาสังกัดเขตพื้นที่การศึกษานครราชสีมา เขต 3 ปีการศึกษา 2550 จำนวน 30,560 คน

การเลือกกลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการวิจัย แบ่งออกเป็น 2 กลุ่ม ได้แก่

(1) กลุ่มตัวอย่างที่ใช้ในการเก็บข้อมูลเพื่อแสดงหลักฐานความเที่ยงตรงตามโครงสร้างและความเชื่อมั่นของการประเมินทักษะกระบวนการวิทยาศาสตร์ โดยใช้ทฤษฎีมาตรฐานเดิม (Classical) และทฤษฎีการตอบข้อคำถามตามโมเดลของราสช์แบบหลายองค์ประกอบ (Many-Facet Rasch Model) โดยมีตัวแปรที่เกี่ยวข้อง 3 ตัวแปร คือ ทักษะ (Task), นักเรียนหรือผู้เข้าสอบ (Examinee) และผู้ตรวจให้คะแนน (Rater) กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 โรงเรียนมัธยมศึกษา สังกัดเขตพื้นที่การศึกษานครราชสีมา เขต 3 ปีการศึกษา 2550 จำนวน 330 คน ซึ่งเลือกมาโดยการสุ่มแบบสองขั้นตอน (Two - stage random sampling) โดยมีขั้นตอนในการสุ่มดังนี้

ขั้นที่ 1 ประเมินการโรงเรียนที่เป็นกลุ่มตัวอย่างที่ใช้ในการวิจัยด้วยการเลือกสุ่มแบบเจาะจง (Purposive sampling) เลือกโรงเรียนโศคชัยสามัคคี เป็นกลุ่มตัวอย่างที่ใช้ในการวิจัย เนื่องจากผลการทดสอบของนักเรียนเป็นเพียงข้อมูลประกอบในการวิจัยเท่านั้น ข้อมูลหลักที่ใช้ตอบคำถามในการวิจัยก็คือ ผลคะแนนของผู้ตรวจที่จะนำมาพิจารณาความเที่ยงตรงและความเชื่อมั่นจากทั้งสองทฤษฎี ผู้วิจัยจึงตัดสินใจเลือกกลุ่มตัวอย่างเพียงโรงเรียนเดียวเพื่อใช้ในการวิจัย

ขั้นที่ 2 ใช้วิธีการสุ่มอย่างง่าย (Simple random sampling) แบ่งกลุ่มตัวอย่างออกเป็น 2 กลุ่ม โดยใช้วิธีการจับฉลาก คือ

1. กลุ่มตัวอย่างที่ใช้ในการหาคุณภาพของเครื่องมือ จำนวน 50 คน
2. กลุ่มตัวอย่างที่ใช้ในการวิจัย จำนวน 280 คน

(2) กลุ่มตัวอย่างที่เป็นผู้ตรวจให้คะแนน (Rater) แบบวัดทักษะกระบวนการทางวิทยาศาสตร์ โดยการเลือกจากครูผู้สอนวิทยาศาสตร์ที่มีประสบการณ์ทางด้านวิทยาศาสตร์ โดยคัดเลือกมาจำนวน 14 คน เพื่อตรวจคะแนนแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ให้กับนักเรียนจำนวน 280 คน โดยการเลือกแบบเจาะจง (Purposive sampling)

การสร้างเครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการศึกษาครั้งนี้ เป็นแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการที่ผู้วิจัยสร้างขึ้น มีลักษณะเป็นคำถามปลายเปิดและกำหนดสถานการณ์ที่เกี่ยวข้องกับวิทยาศาสตร์ให้นักเรียนเขียนตอบ ประกอบด้วยทักษะกระบวนการจำนวน 5 ทักษะ แบบวัดทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการฉบับนี้ใช้เวลาในการทดสอบ 120 นาที โดยมีการดำเนินการตามขั้นตอนดังนี้

ขั้นตอนในการสร้างเครื่องมือ

ในการสร้างแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการ ผู้วิจัยได้ดำเนินการตามขั้นตอนดังนี้

1. กำหนดจุดมุ่งหมายในการสร้างแบบวัด

1.1 เพื่อสร้างแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการจำนวน 5 ทักษะ ลักษณะเป็นข้อคำถามปลายเปิดและกำหนดสถานการณ์ให้ตอบ สำหรับนักเรียนชั้นมัธยมศึกษาตอนปลาย

1.2 เพื่อศึกษาค่าความเที่ยงตรงตามโครงสร้าง และค่าความเชื่อมั่นของการตรวจให้คะแนนตามทฤษฎีมาตรฐานเดิม (Classical) และทฤษฎีการตอบสนองข้อคำถามตามโมเดลของราส์ชแบบหลายองค์ประกอบ (Many-Facet Rasch Model)

2. ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง

2.1 ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับความหมาย ประเภทของทักษะกระบวนการทางวิทยาศาสตร์ การวัดผลและประเมินผล และการตรวจให้คะแนนโดยใช้กฎเกณฑ์การให้คะแนน (Scoring Rubric)

2.2 ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับความเที่ยงตรงตามโครงสร้าง ความเชื่อมั่นตามทฤษฎีมาตรฐานเดิมและทฤษฎีการตอบสนองข้อคำถามตามโมเดลของราล์สแบบหลายองค์ประกอบ เพื่อใช้เป็นแนวทางในการดำเนินการ

3. วิเคราะห์ทักษะกระบวนการทางวิทยาศาสตร์ และเขียนพฤติกรรมที่บ่งชี้ทักษะกระบวนการทางวิทยาศาสตร์ย่อยทั้ง 5 ทักษะ กำหนดการวัดในแต่ละทักษะโดยเชิญครูผู้สอนวิทยาศาสตร์ซึ่งถือว่าเป็นผู้เชี่ยวชาญด้านเนื้อหามาร่วมกันวิเคราะห์ ผลการวิเคราะห์ดังแสดงในตาราง 2

ตาราง 2 ผลการวิเคราะห์พฤติกรรมที่บ่งชี้ทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ

ทักษะ	พฤติกรรมที่บ่งชี้
- การตั้งสมมติฐาน	คาดคะเนคำตอบก่อนการทดลองได้สอดคล้องกับตัวแปรต้นและตัวแปรตามโดยอาศัยความรู้และประสบการณ์เดิม
- การกำหนดนิยามเชิงปฏิบัติการ	การกำหนดความหมายและขอบเขตของคำหรือตัวแปรต่าง ๆ ในปฏิบัติการทดลองให้สามารถทดสอบหรือวัดได้
- การกำหนดและควบคุมตัวแปร	การบ่งชี้และกำหนดตัวแปรต้น ตัวแปรตาม และควบคุมตัวแปรที่เป็นสาเหตุอื่น ๆ ในปรากฏการณ์หนึ่ง ๆ ที่ต้องการศึกษา
- การทดลอง	การออกแบบการทดลอง ระบุวัสดุอุปกรณ์ที่ใช้ในการทดลอง อธิบายขั้นตอนการทดลอง และเสนอแนะการบันทึกผลการทดลอง
- การตีความหมายข้อมูลและลงข้อสรุป	การแปลความหมายหรือบรรยายลักษณะและคุณสมบัติของข้อมูลพร้อมทั้งสรุปความสัมพันธ์ของตัวแปรที่อยู่ในข้อมูลที่ต้องการศึกษา

4. เขียนแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ที่เป็นแบบเขียนตอบตามพฤติกรรมที่บ่งชี้ในตาราง 2 กำหนดน้ำหนักคะแนนในแต่ละทักษะกระบวนการ พร้อมทั้งจัดทำคู่มือการตรวจให้คะแนน (เฉลย) และสร้างกฎเกณฑ์การให้คะแนนแบบเฉพาะ (Specific Rubric) แยกเป็นรายทักษะ เพื่อให้ผู้ตรวจได้ใช้ในการให้คะแนนแต่ละทักษะ

ตัวอย่างเครื่องมือที่ใช้ในการวิจัย

แบบวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ

ข้อ 0 ทักษะการตั้งสมมติฐาน

1. จากการทดลองละลายสาร A ในน้ำเย็นปรากฏว่า สาร A 21 กรัม ละลายได้พอดีในน้ำ 250 ลูกบาศก์เซนติเมตร จากการทดลองนี้ ถ้าผู้ทดลองมีความเชื่อว่า การอุ่นสารละลายให้ร้อนขึ้นจะช่วยให้สาร A ละลายได้เพิ่มขึ้นอีก ผู้ทดลองควรตั้งสมมติฐานว่าอย่างไร

.....

2. นำลูกกลมตันที่ทำด้วยวัตถุต่างชนิดกัน ได้แก่ วัตถุ A , B และวัตถุ C ซึ่งมีเส้นผ่าศูนย์กลาง เท่ากันมา 3 ลูก ปล่อยให้กลิ้งบนพื้นเอียงเดียวกัน แล้วจับเวลาที่ลูกกลมแต่ละชนิดใช้ในการกลิ้งจากปลายบนสุดถึงปลายล่างสุดของพื้นเอียง ทำซ้ำแบบเดิม 3 ครั้ง หาค่าเฉลี่ยและนำมาเปรียบเทียบกัน กรณีดังกล่าว ถ้านักเรียนเป็นผู้ทดลองจะตั้งสมมติฐานอย่างไร

.....

3. จงศึกษาข้อมูลจากตารางต่อไปนี้

ชนิดของ สาร	ความสามารถในการละลายของสารที่อุณหภูมิห้อง (กรัม)	
	ในน้ำ 100 ลบ.ซม.	ในเอทิลแอลกอฮอล์ 100 ลบ.ซม.
A	140	5
B	39	1
C	200	0.6
D	36	0
E	6	0

ข้อมูลในการทดลองจากตารางนี้ ควรตั้งสมมติฐานอย่างไร

.....

แนวทางในการตอบ

ข้อ 0 ทักษะการตั้งสมมติฐาน

1. การเพิ่มอุณหภูมิให้แก่น้ำจะมีผลต่อปริมาณการละลายของสาร A
2. ชนิดของวัตถุที่ใช้ทำลูกกลมตันมีผลต่อเวลาในการกลิ้งของลูกกลมตันบนพื้นเอียง
3. ความสามารถในการละลายของสารแต่ละชนิดขึ้นอยู่กับชนิดของตัวทำละลาย หรือ ชนิดของตัวทำละลายมีผลต่อความสามารถในการละลายของสารแต่ละชนิด

กฎเกณฑ์การให้คะแนนทักษะการตั้งสมมติฐาน

ระดับคุณภาพ	คำอธิบาย
4 ดีมาก	หมายถึง สามารถคาดคะเนคำตอบก่อนการทดลองได้สอดคล้องกับตัวแปรต้น และตัวแปรตาม การแสดงเหตุผลและแนวคิดชัดเจนถูกต้องทั้งหมด
3 ดี	หมายถึง สามารถคาดคะเนคำตอบก่อนการทดลองได้สอดคล้องกับตัวแปรต้น และตัวแปรตาม การแสดงเหตุผลและแนวคิดชัดเจนถูกต้อง 2 ข้อใน 3 ข้อ
2 ใช้ได้	หมายถึง สามารถคาดคะเนคำตอบก่อนการทดลองได้สอดคล้องกับตัวแปรต้น และตัวแปรตาม การแสดงเหตุผลและแนวคิดชัดเจนถูกต้อง 1 ข้อใน 3 ข้อ
1 ใช้ไม่ได้	หมายถึง สามารถคาดคะเนคำตอบก่อนการทดลองไม่สอดคล้องกับตัวแปรต้น และตัวแปรตาม การแสดงเหตุผลและแนวคิดมีแนวโน้มที่จะได้คำตอบ
0	ไม่ตอบหรือตอบไม่ถูกเลย

วิธีการหาคุณภาพของเครื่องมือ

แบบวัดทักษะกระบวนการทางวิทยาศาสตร์ที่ผู้วิจัยสร้างขึ้น มีการดำเนินการหาคุณภาพของเครื่องมือตามขั้นตอนดังนี้

1. นำข้อคำถามที่ผู้วิจัยสร้างขึ้นไปให้ผู้เชี่ยวชาญ ประกอบด้วยผู้เชี่ยวชาญ ด้านการวัดผล จำนวน 3 ท่าน และผู้เชี่ยวชาญด้านการสอนวิทยาศาสตร์ จำนวน 2 ท่าน ที่มีวุฒิ ตั้งแต่ปริญญาโทขึ้นไป ตรวจสอบความเที่ยงตรงเชิงเนื้อหา (Content Validity) และความ เป็นปรนัย (Objectivity) ของข้อคำถามและคู่มือการตรวจให้คะแนน นำผลการวิเคราะห์ความ เที่ยงตรงเชิงเนื้อหาโดยใช้ดัชนีความสอดคล้อง (IOC) รวมความเห็นในช่อง เห็นด้วย (+1) ไม่

แน่ใจ (0) และไม่เห็นด้วย (-1) ตามวิธีของโรวินลลีและแฮมเบิลตัน คัดเลือกข้อที่มีค่า IOC ตั้งแต่ 0.5 ขึ้นไป ปรากฏผลการคำนวณค่า IOC มีค่าระหว่าง 0.6 – 1.0 ดังแสดงในภาคผนวก

2. นำแบบทดสอบที่ผ่านการตรวจสอบคุณภาพในขั้นต้น ไปทดสอบกับกลุ่มตัวอย่างที่ใช้หาคุณภาพของเครื่องมือ คือ นักเรียนชั้นมัธยมศึกษาปีที่ 4 โรงเรียนโชคชัยสามัคคี จำนวน 50 คน โดยการให้นักเรียนทำข้อสอบทีละข้อเพื่อตรวจสอบเวลาที่เหมาะสมในการทำข้อสอบ โดยพิจารณาจากจำนวนนักเรียนที่ทำเสร็จเรียบร้อยประมาณร้อยละ 80 ของนักเรียนที่เข้าสอบ

3. นำผลการทดสอบมาตรวจให้คะแนนโดยผู้วิจัยเป็นผู้ดำเนินการเอง แล้วจึงวิเคราะห์หาคุณภาพของข้อสอบ เพื่อคัดเลือกและปรับปรุง ได้ค่าความยากอยู่ระหว่าง .246 - .789 และค่าอำนาจจำแนกเป็นรายข้อ มีค่าอยู่ระหว่าง .332 – .845 โดยใช้ Item total Correlation ได้ข้อสอบจำนวน 22 ข้อ คือ ข้อ 1.1, ข้อ 1.2, ข้อ 1.3, ข้อ 2.1, ข้อ 2.2, ข้อ 2.4, ข้อ 3.1, ข้อ 3.2, ข้อ 3.3, ข้อ 3.4, ข้อ 4.3, ข้อ 5.1.1, ข้อ 5.1.2 ข้อ 5.2.1, ข้อ 5.2.2, ข้อ 5.3.1, ข้อ 5.3.2, ข้อ 5.4.1 และข้อ 5.4.2 ซึ่งข้อที่ผ่านการคัดเลือกมีค่าความยากอยู่ระหว่าง .316 - .789 และค่าอำนาจจำแนกมีค่าอยู่ระหว่าง .529 – .845 ดังแสดงในตารางภาคผนวก

4. จัดทำคู่มือดำเนินการสอบและจัดพิมพ์

การเก็บรวบรวมข้อมูล

ผู้วิจัยได้ดำเนินการเก็บรวบรวมข้อมูลในการวิจัย ดังนี้

1. ติดต่อประสานงาน ขออนุญาตผู้บริหารและครูผู้สอนของโรงเรียนกลุ่มตัวอย่าง เพื่อกำหนดวัน เวลา สถานที่ และวิธีดำเนินการสอบ

2. คัดเลือกผู้ตรวจให้คะแนน (Rater) จำนวน 14 คน (ที่มีคุณสมบัติดังที่กล่าวไว้ในการคัดเลือกกลุ่มตัวอย่างในการวิจัยกลุ่มที่ 2) จัดประชุมชี้แจงผู้ตรวจเพื่อซักซ้อมและทำความเข้าใจการตรวจให้คะแนน

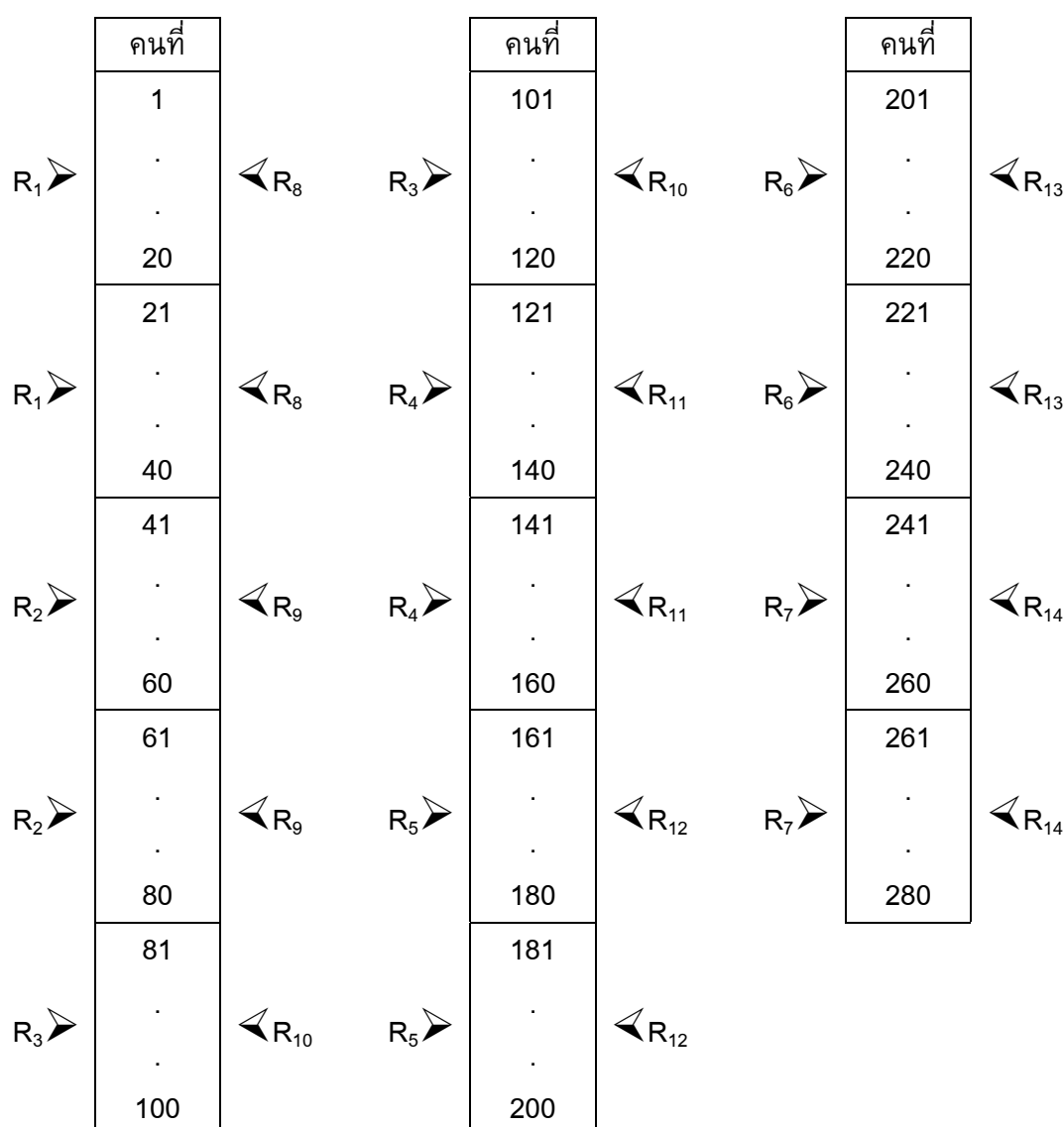
3. นำแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ที่ผ่านการวิเคราะห์หาคุณภาพรายข้อแล้ว ไปทดสอบกับกลุ่มตัวอย่าง จำนวน 280 คน ผู้ดำเนินการสอบ 2 คน ต่อนักเรียน 50 คน ซึ่งผู้ดำเนินการสอบต้องชี้แจงการปฏิบัติในการดำเนินการสอบ ดังนี้

3.1 อธิบายให้ผู้เข้าสอบเข้าใจวัตถุประสงค์ และผลประโยชน์ที่จะได้รับการทำแบบทดสอบ

3.2 อธิบายให้ผู้เข้าสอบทุกคนเข้าใจถึงวิธีดำเนินการสอบ วิธีการปฏิบัติและวิธีตอบ ก่อนที่จะให้ผู้เข้าสอบลงมือทำ

การจัดกระทำข้อมูลและการวิเคราะห์ข้อมูล

1. นำกระดาษคำตอบของผู้เข้าสอบทั้งหมดมาถ่ายเอกสารจำนวน 2 ชุด จัดให้ผู้ตรวจจำนวน 14 คน ตรวจให้คะแนนตามแนวทางการตอบและกฎเกณฑ์การให้คะแนนในคู่มือการตรวจที่ผู้วิจัยได้สร้างขึ้น โดยมีเงื่อนไข นักเรียน 1 คน จะต้องถูกตรวจให้คะแนนจากผู้ตรวจ 2 คน โดยไม่ซ้ำกัน เพราะฉะนั้น ผู้ตรวจ 1 คน จะต้องตรวจผลงานนักเรียนจำนวน 40 ชุด จึงจะครบตามจำนวนนักเรียน 280 คน พอดี แล้วบันทึกผลการตรวจให้คะแนน ดังภาพประกอบ 5



ภาพประกอบ 5 แสดงการจัดจับคู่ในการตรวจให้คะแนนระหว่างนักเรียนและผู้ตรวจ

หมายเหตุ R_{1-14} แทน ผู้ตรวจคนที่ 1 - 14

คนที่ หมายถึง นักเรียนคนที่

จากภาพประกอบ 5 นักเรียนคนที่ 1 จะถูกตรวจให้คะแนนโดยผู้ตรวจคนที่ 1 และผู้ตรวจคนที่ 8 นักเรียนคนที่ 41 จะถูกตรวจให้คะแนนโดยผู้ตรวจคนที่ 2 และผู้ตรวจคนที่ 9 ซึ่งเป็นไปตามเงื่อนไข นักเรียน 1 คน จะต้องถูกตรวจให้คะแนนจากผู้ตรวจ 2 คน โดยไม่ซ้ำกันทุกประการ

2. นำผลที่ได้จากการตรวจให้คะแนนมาวิเคราะห์ข้อมูล ดังนี้

2.1 วิเคราะห์ค่าสถิติพื้นฐานของแบบวัด

2.2. วิเคราะห์ค่าความเที่ยงตรงเชิงโครงสร้าง (Construct Validity) จากผลการตรวจให้คะแนนแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ โดยใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirm Factor Analysis)

2.3 วิเคราะห์ค่าความเที่ยงตรงรวม (Convergent Validity) และความเที่ยงตรงแยก (Discriminant Validity) จากผลของคะแนนแบบวัดทักษะกระบวนการทางวิทยาศาสตร์

2.4 วิเคราะห์ค่าความเที่ยงตรงเชิงโครงสร้าง (Construct Validity) จากผลการตรวจให้คะแนนแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ โดยใช้ทฤษฎีการตอบข้อคำถามตามโมเดลของราสช์แบบหลายองค์ประกอบ (Many-Facet Rasch Model) ที่มีองค์ประกอบจำนวน 3 องค์ประกอบ ได้แก่ ค่าความยากของทักษะกระบวนการ, ค่าความสามารถของผู้ตอบคำถาม และค่าความเคร่งครัดของผู้ตรวจให้คะแนน

2.5 วิเคราะห์ค่าความเชื่อมั่นของผลการตรวจให้คะแนน (Reliability of rating) และค่าความเชื่อมั่นระหว่างผู้ตรวจให้คะแนน (Reliability of raters) แบบวัดทักษะกระบวนการทางวิทยาศาสตร์

2.6 วิเคราะห์ค่าความเชื่อมั่นเชิงโครงสร้าง (Construct Reliability) และค่า Variance Extracted ของแบบวัดทักษะกระบวนการทางวิทยาศาสตร์

3. สถิติที่ใช้ในการวิเคราะห์ข้อมูล

3.1 ค่าสถิติพื้นฐานของการทดสอบ ได้แก่ ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้ และค่าความโด่ง

3.2 สถิติที่บ่งชี้คุณภาพของแบบวัด ได้แก่ ค่าความยาก ค่าอำนาจจำแนกของข้อสอบรายข้อ และค่าความเชื่อมั่นของแบบวัดทั้งฉบับ

3.3 ความเที่ยงตรงเชิงโครงสร้าง (Construct Validity) ของผลการตรวจให้คะแนน โดยใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งและอันดับสอง ในการทดสอบโมเดลองค์ประกอบของแบบวัดทักษะกระบวนการทางวิทยาศาสตร์และเปรียบเทียบความเหมาะสมของโมเดลกับข้อมูลเชิงประจักษ์ ดังนี้

3.3.1 ดัชนีความเหมาะสมสัมบูรณ์ (absolute fit index) พิจารณาจากค่าสถิติไคสแควร์ โดยพิจารณาจากค่าองศาอิสระที่ไม่มีนัยสำคัญที่ระดับ .05 ซึ่งแสดงว่าโมเดลที่กำหนดขึ้นมีความเหมาะสมกับข้อมูลเชิงประจักษ์ อย่างไรก็ตามค่าไคสแควร์จะมีแนวโน้มที่จะมีนัยสำคัญเมื่อนำขนาดกลุ่มตัวอย่างที่ใช้ในการศึกษามีขนาดใหญ่ ดังนั้น จึงพัฒนาดัชนีอื่น ๆ เพื่อใช้แทนค่าไคสแควร์ โดยเรียกดัชนีพวกนี้ว่าดัชนีบังคับซึ่งใกล้เคียงความเหมาะสม (Closed Fit) ดัชนีเหล่านี้จะเปิดโอกาสให้มีความคลาดเคลื่อนในการประมาณค่าได้บ้างในเกณฑ์ที่เหมาะสม การวิจัยในครั้งนี้ใช้ค่า GFI และ AGFI เกณฑ์การพิจารณาว่าโมเดลมีความเหมาะสม กับข้อมูลเชิงประจักษ์ในระดับที่ยอมรับได้คือ มีค่ามากกว่า .90 และเกณฑ์ในระดับดีคือมากกว่า .95 ส่วนดัชนีในกลุ่มเศษเหลือใช้ค่า RMSEA เกณฑ์พิจารณาคือค่าดัชนีขนาดไม่เกิน .10 และ .06 ตามลำดับ

3.3.2 ดัชนีเหมาะสมเชิงเปรียบเทียบเพื่อเปรียบเทียบค่าไคสแควร์ของโมเดลที่เป็น Subset ของกันและกัน เพื่อตรวจสอบการกำหนดคุณสมบัติ (Specification) ของโมเดลย่อยว่ามีความคลาดเคลื่อนไปจากโมเดลที่ใหญ่กว่าหรือไม่ โมเดลย่อยที่มีความเหมาะสม กับโมเดลที่ใหญ่กว่าจะมีค่าของดัชนีเชิงเปรียบเทียบใกล้ 1 การวิจัยในครั้งนี้ใช้ค่า CFI รายงานผลการประเมินโมเดลในกลุ่มเชิงเปรียบเทียบ เนื่องจากดัชนี CFI มีความคงที่มากกว่าดัชนีอื่น ๆ และไม่มีผลเกี่ยวข้องเกี่ยวกับขนาดของกลุ่มตัวอย่าง (Bentler. 1990; Hu; & Bentler. 1999)

สำหรับดัชนีในกลุ่มประหยัดผู้วิจัยไม่ได้นำเสนอในการวิจัยครั้งนี้ เนื่องจากวัตถุประสงค์ในการใช้ดัชนีในกลุ่มประหยัดนั้นใช้เพื่อเปรียบเทียบโมเดลคู่แข่งว่าโมเดลใดมีความประหยัดมากกว่ากัน กล่าวคือ เมื่อมีโมเดลมากกว่าหนึ่งโมเดลที่มีลักษณะความสัมพันธ์ที่แตกต่างกัน แต่ใช้ข้อมูลในการวิเคราะห์ชุดเดียวกัน โมเดลที่มีความประหยัดมากกว่าคือโมเดลที่มีค่าพารามิเตอร์ในการประมาณค่าน้อยกว่า (Marsh; & Hau. 1996) แต่วัตถุประสงค์การวิจัยในครั้งนี้ไม่มีการเปรียบเทียบความประหยัดระหว่างโมเดลที่กำหนดขึ้น จึงไม่มีความจำเป็นในการแสดงค่าดัชนีดังกล่าว

3.4 วิเคราะห์ความเที่ยงตรงร่วม (Convergent Validity) และความเที่ยงตรงแยก (Discriminant Validity) ของผลการตรวจให้คะแนน ใช้สูตรสหสัมพันธ์แบบเพียร์สัน (บุญเชิด ภิญญอนันตพงษ์. 2545 : 84)

$$r_{xy} = \frac{N\sum XY - \sum X \sum Y}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

เมื่อ	r_{xy}	แทน	สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน
	N	แทน	จำนวนกลุ่มตัวอย่าง
	$\sum X$	แทน	ผลรวมของคะแนนรายข้อ (item)
	$\sum Y$	แทน	ผลรวมของคะแนนรวม (total)

ΣXY	แทน	ผลรวมของผลคูณของคะแนน X และคะแนน Y
ΣX^2	แทน	ผลรวมกำลังสองของคะแนน X
ΣY^2	แทน	ผลรวมกำลังสองของคะแนน Y

3.5 ความเที่ยงตรงเชิงโครงสร้าง (Construct Validity) จากผลการตรวจให้คะแนนแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ โดยใช้ทฤษฎีการตอบข้อคำถามตามโมเดลของราส์ชแบบหลายองค์ประกอบ (Many-Facet Rasch Model) ที่มีจำนวน 3 องค์ประกอบ ได้แก่ ความยากของทักษะกระบวนการ, ความสามารถของผู้ตอบคำถาม และความเคร่งครัดของผู้ตรวจให้คะแนน จากสมการ

$$\ln[P_{nijk} / P_{nijk-1}] = \beta_n - \delta_i - \lambda_j - \tau_k$$

เมื่อ P_{nijk}	แทน	ความน่าจะเป็นที่ผู้ตรวจ (j) จะประเมินผู้ตอบคนที่ (n) ให้ระดับคะแนน (k) จากการตอบข้อความรายการที่ (i)
P_{nijk-1}	แทน	ความน่าจะเป็นที่ผู้ตรวจ (j) จะประเมินผู้ตอบคนที่ (n) ให้ระดับคะแนน (k-1) จากการตอบข้อความรายการที่ (i)
β_n	แทน	ระดับความสามารถของผู้ตอบคนที่ (n)
δ_i	แทน	ค่าความยากของข้อความรายการที่ (i)
λ_j	แทน	ค่าความเคร่งครัดของผู้ตรวจ (j)
τ_k	แทน	ค่าความยากประจำขั้นที่ (k) ของการประเมินสัมพัทธ์กับค่าความยากประจำขั้นที่ (k-1)

3.6 ความเชื่อมั่นของผลการตรวจให้คะแนน (Reliability of rating) แบบความสอดคล้องภายใน (Internal consistency) โดยใช้สูตร r_{F-R} ของ Feldt – Raju (บุญเชิด ภิญโญนนตพงษ์. 2545: 139) จากสมการ

$$r_{F-R} = \frac{1}{[1 - \Sigma \lambda_i^2]} \cdot \frac{[S_x^2 - \Sigma S_i^2]}{S_x^2}$$

เมื่อ r_{F-R}	แทน	ความเชื่อมั่นของผลการตรวจให้คะแนน
λ_i^2	แทน	สัดส่วนระหว่างผลรวมของเมตริกในแถวที่ i กับความแปรปรวนของคะแนนรวม
S_i^2	แทน	ความแปรปรวนของคะแนนในแต่ละด้าน
S_x^2	แทน	ความแปรปรวนของคะแนนของรวม

3.7 ดัชนีความสอดคล้องของผู้ตรวจให้คะแนน โดยใช้สัมประสิทธิ์ Kappa จาก

สูตร

$$K = \frac{P_o - P_e}{1 - P_e}$$

เมื่อ K แทน ดัชนีความสอดคล้องของผู้ตรวจให้คะแนน

P_o แทน ค่าความสอดคล้องของคะแนนสังเกต

P_e แทน ค่าคาดหวังของความสอดคล้อง

3.8 ความเชื่อมั่นเชิงโครงสร้าง (Construct Reliability) และค่า Variance

Extracted ขององค์ประกอบตัวแปรแฝงของแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ (Hair; et al. 1995: 642) จากสูตร

$$\text{Construct Reliability} = \frac{(\sum \Lambda_x)^2}{(\sum \Lambda_x)^2 + \sum (\delta)}$$

$$\text{Variance Extracted} = \frac{\sum \Lambda_x^2}{\sum \Lambda_x^2 + \sum (\delta)}$$

เมื่อ Λ_x แทน น้ำหนักองค์ประกอบ (Factor Loading)

δ แทน ความคลาดเคลื่อน (Error)

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การศึกษาวิจัยครั้งนี้มีความมุ่งหมายที่สำคัญ เพื่อแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของผลการประเมินด้านทักษะกระบวนการทางวิทยาศาสตร์ โดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน การวิเคราะห์ความเที่ยงตรงรวมและความเที่ยงตรงแยกตามแนวทางทฤษฎีมาตรฐานเดิมและการวิเคราะห์โมเดลของราล์ชแบบหลายองค์ประกอบตามแนวทางของทฤษฎีการตอบข้อคำถาม พร้อมทั้งแสดงหลักฐานความเชื่อมั่นของผลการประเมินจากการให้คะแนนของผู้ประเมินมากกว่าหนึ่งคนในทั้งสองทฤษฎี ผู้วิจัยแบ่งการนำเสนอผลการวิจัยเป็น 5 ตอน คือ ตอนที่ 1 เป็นการวิเคราะห์ข้อมูลพื้นฐานของตัวแปรที่ใช้ในการประเมินด้านทักษะกระบวนการทางวิทยาศาสตร์ ตอนที่ 2 การแสดงหลักฐานความเที่ยงตรงตามโครงสร้าง โดยการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งและอันดับสอง ตอนที่ 3 เป็นผลการวิเคราะห์ความเที่ยงตรงรวม (Convergence Validity) และความเที่ยงตรงแยก (Discriminant Validity) ตามแนวทางทฤษฎีมาตรฐานเดิม ตอนที่ 4 เป็นการแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของการประเมินทักษะกระบวนการทางวิทยาศาสตร์โดยการวิเคราะห์โมเดลของราล์ชแบบสามองค์ประกอบตามแนวทางของทฤษฎีการตอบข้อคำถาม และตอนที่ 5 เป็นผลการวิเคราะห์ความเชื่อมั่นของผลการตรวจให้คะแนน ความสอดคล้องของผู้ตรวจให้คะแนน และความเชื่อมั่นเชิงโครงสร้าง

สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล

การนำเสนอผลการวิเคราะห์ข้อมูลในการวิจัยในครั้งนี้ เพื่อให้เกิดความเข้าใจเกี่ยวกับความหมายในการนำเสนอผลการวิเคราะห์ที่ตรงกัน ผู้วิจัยได้กำหนดสัญลักษณ์ และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล ดังต่อไปนี้

Mean	แทน	ค่าเฉลี่ย
SD	แทน	ส่วนเบี่ยงเบนมาตรฐาน
Kurt	แทน	ค่าความโด่ง
Skew	แทน	ค่าความเบ้
SE	แทน	ค่าความคลาดเคลื่อนมาตรฐาน
t	แทน	ค่า สถิติ t
INFIT	แทน	ค่าเฉลี่ยยกกำลังสองแบบถ่วงน้ำหนัก
OUTFIT	แทน	ค่าเฉลี่ยยกกำลังสองแบบไม่ถ่วงน้ำหนัก
df	แทน	ระดับองศาอิสระ

p	แทน	ค่าความน่าจะเป็น
GFI	แทน	ดัชนีวัดระดับความสอดคล้อง
AGFI	แทน	ดัชนีวัดระดับความสอดคล้องที่ปรับแก้แล้ว
CFI	แทน	ดัชนีวัดระดับความสอดคล้องเปรียบเทียบ
RMSEA	แทน	ดัชนีความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์
$\Sigma \lambda_i^2$	แทน	สัดส่วนระหว่างผลรวมของเมตริกในแถวที่ i กับความแปรปรวนของคะแนนรวม
ΣS_i^2	แทน	ความแปรปรวนของคะแนนในแต่ละด้าน
S_X^2	แทน	ความแปรปรวนของคะแนนของรวม
r_{F-R}	แทน	ความเชื่อมั่นของผลการตรวจให้คะแนน
Λ_x	แทน	น้ำหนักองค์ประกอบ
δ	แทน	ความคลาดเคลื่อน
Σ	แทน	ผลรวม

1. การวิเคราะห์ข้อมูลพื้นฐานของตัวแปรที่ใช้ในการประเมินทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการ

การวิเคราะห์ตอนนี้ ผู้วิจัยนำข้อมูลจากการวัดมาวิเคราะห์ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้และค่าความโด่งของข้อมูลตามตัวแปรในแต่ละทักษะเพื่อพิจารณาถึงข้อตกลงเบื้องต้นเกี่ยวกับการกระจายเป็นโค้งปกติของข้อมูล ปรากฏผลดังตาราง 3

ตาราง 3 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้และค่าความโด่งของข้อมูลจำแนกตามตัวแปร

ตัวแปร	Mean	SD	Kurt	Skew
H01	0.662	0.343	-0.799	-0.522
H02	0.780	0.298	0.012	-1.005
H03	0.454	0.359	-1.044	0.139
D01	0.871	0.235	1.290	-1.529
D02	0.357	0.335	-0.785	0.406
D03	0.498	0.363	-1.093	0.005

ตาราง 3 (ต่อ)

ตัวแปร	Mean	SD	Kurt	Skew
CT01	0.523	0.347	-0.913	-0.062
CT02	0.388	0.348	-0.909	0.335
CT03	0.830	0.262	0.413	-1.199
CT04	0.771	0.299	-0.129	-0.929
E01	0.716	0.323	-0.523	-0.706
E02	0.750	0.322	-0.226	-0.929
E03	0.600	0.348	-0.916	-0.292
E04	0.662	0.330	-0.731	-0.465
CC01	0.507	0.354	-0.996	-0.020
CC02	0.432	0.337	-0.806	0.169
CC03	0.695	0.337	-0.654	-0.654
CC04	0.498	0.358	-1.039	0.005
CC05	0.487	0.360	-1.064	0.037
CC06	0.314	0.319	-0.652	0.511
CC07	0.439	0.357	-1.018	0.181
CC08	0.541	0.370	-1.164	-0.132
รวม	0.581	0.332	-0.625	-0.303

ผลการวิเคราะห์ตามตาราง 3 สรุปได้ดังนี้

ค่าสถิติพื้นฐาน เมื่อพิจารณาถึงลักษณะการกระจายของตัวแปรที่ใช้ในการประเมินทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการแต่ละตัว พบว่า ค่าความเบ้และค่าความโด่งของข้อมูล มีแนวโน้มที่จะกระจายเป็นโค้งปกติ เนื่องจากว่าทั้งค่าความเบ้และความโด่งส่วนใหญ่มีค่าใกล้เคียง 1 ถึง -1 ยกเว้นบางตัวแปรที่มีค่าเกิน -1 ไปบ้างเล็กน้อย ซึ่งก็ไม่ได้ถือว่ามากมายอยู่ในเกณฑ์ที่สามารถยอมรับได้ ข้อมูลที่เป็นโค้งปกติค่าความเบ้เท่ากับ 0 และค่าความโด่งไม่เกิน 3 (กนกทิพย์ พัฒนาพัวพันธ์. 2529: 101)

2. การแสดงหลักฐานความเที่ยงตรงตามโครงสร้าง โดยวิธีการวิเคราะห์องค์ประกอบ เชิงยืนยันอันดับหนึ่งและอันดับสอง

ผู้วิจัยได้นำข้อมูลที่ผ่านการตรวจจากผู้ตรวจ จำนวน 22 ตัวแปร มาวิเคราะห์เพื่อตรวจสอบความเที่ยงตรงตามโครงสร้างของตัวแปรที่มีองค์ประกอบจำนวน 5 องค์ประกอบ โดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งเพื่อตรวจสอบความสัมพันธ์ระหว่างตัวแปรแฝงกับตัวแปรสังเกต และการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง เพื่อเป็นทางเลือกในการลดจำนวนองค์ประกอบที่ใช้ในการอธิบายปรากฏการณ์

2.1 การเปรียบเทียบความเหมาะสมกับข้อมูลเชิงประจักษ์ของโมเดลองค์ประกอบเชิงยืนยันอันดับหนึ่ง เพื่อพิจารณาความสัมพันธ์ของข้อมูลในแต่ละตัวแปรว่ามีความเหมาะสมกับโมเดลการวิจัย โดยพิจารณาจากค่าดัชนีความเหมาะสมของค่าพารามิเตอร์ในแต่ละทักษะ ปรากฏผลดังตาราง 4

ตาราง 4 ผลการประมาณค่าการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง

ตัวแปร	องค์ประกอบ	สัมประสิทธิ์	SE	t	p
H01	การตั้งสมมติฐาน	0.06	0.03	2.23	>.05
H02		0.07	0.03	2.50	>.05
H03		0.08	0.03	2.46	>.05
D01	การกำหนดนิยามเชิงปฏิบัติการ	0.06	0.02	3.60	>.05
D02		0.15	0.02	6.21	>.05
D03		0.19	0.03	7.13	>.05
CT01	การกำหนดและควบคุมตัวแปร	0.15	0.03	5.30	>.05
CT02		0.12	0.03	4.50	>.05
CT03		0.08	0.02	4.01	>.05
CT04		0.02	0.02	0.83	<.05
E01	การทดลอง	0.06	0.02	2.59	>.05
E02		0.05	0.02	2.42	>.05
E03		0.13	0.03	3.73	>.05
E04		0.06	0.02	2.77	>.05

ตาราง 4 (ต่อ)

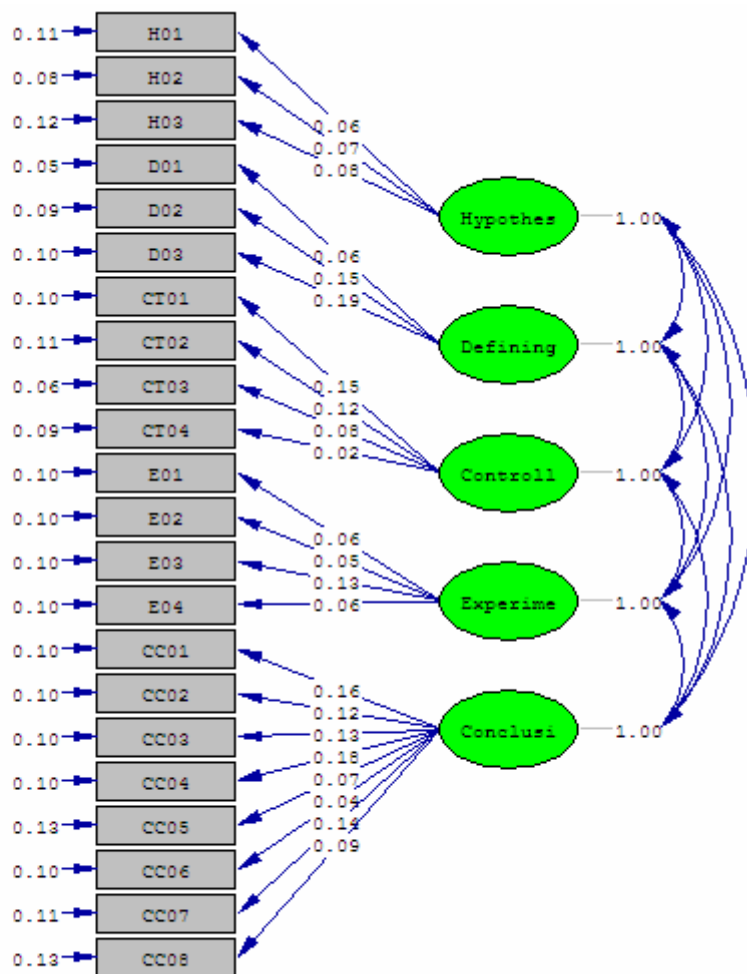
ตัวแปร	องค์ประกอบ	สัมประสิทธิ์	SE	t	p
CC01	การตีความหมายข้อมูลและลงข้อสรุป	0.16	0.02	7.08	>.05
CC02		0.12	0.02	5.66	>.05
CC03		0.13	0.02	5.74	>.05
CC04		0.18	0.02	7.68	>.05
CC05		0.07	0.02	2.78	>.05
CC06		0.04	0.02	2.11	>.05
CC07		0.14	0.02	5.85	>.05
CC08		0.09	0.02	3.59	>.05

Chi-square = 205.65 df = 199 P = 0.36

GFI = 0.94 , AGFI = 0.92 , RMSEA = 0.01 , CFI = 0.98

ผลการวิเคราะห์ตามตาราง 4 สรุปได้ดังนี้

การวิเคราะห์ดัชนีความเหมาะสมของโมเดลองค์ประกอบเชิงยืนยันอันดับหนึ่งจำนวน 5 องค์ประกอบ พบว่า ค่าไคสแควร์ = 205.65 (df = 199 P = 0.36) ค่า GFI = 0.94 ค่า AGFI = 0.92 ค่า RMSEA = 0.01 และ CFI = 0.98 ค่าดัชนีความเหมาะสมของทุกตัวแปร บ่งชี้ว่า โมเดลการวิจัยมีความเหมาะสมกับข้อมูลเชิงประจักษ์ เนื่องจากเป็นไปตามเกณฑ์ที่กำหนดไว้ที่ว่า ไคสแควร์ต้องไม่มีนัยสำคัญทางสถิติ ค่า GFI และ AGFI ควรสูงกว่า .90 และ เกณฑ์ในระดับดีคือมากกว่า .95 ส่วนค่า RMSEA ควรต่ำกว่า .06 และค่า CFI ควรสูงกว่า .95 ตามลำดับ แสดงว่า โมเดลองค์ประกอบเชิงยืนยันอันดับหนึ่ง 5 องค์ประกอบ มีความเหมาะสมกับข้อมูลเชิงประจักษ์ และเป็นไปตามสมมติฐานการวิจัยข้อที่ 1 ทุกประการ สำหรับโมเดล องค์ประกอบเชิงยืนยันอันดับหนึ่ง 5 องค์ประกอบ ดังภาพประกอบ 6



Chi-Square=205.65, df=199, P-value=0.35827, RMSEA=0.011

ภาพประกอบ 6 ผลการประมาณค่าสัมประสิทธิ์องค์ประกอบโมเดลการวิเคราะห์องค์ประกอบ
เชิงยืนยันอันดับหนึ่ง

2.2 การเปรียบเทียบความเหมาะสมกับข้อมูลเชิงประจักษ์ของโมเดลองค์ประกอบเชิงยืนยันอันดับสอง เพื่อค้นหาโมเดลทางเลือกในการอธิบายผลการประเมินทักษะกระบวนการทางวิทยาศาสตร์ในมิติอื่น ๆ ที่ใช้ฐานองค์ประกอบอันดับหนึ่ง 5 องค์ประกอบ โดยการพิจารณาจากค่าดัชนีความเหมาะสมของค่าพารามิเตอร์ในแต่ละทักษะ ปรากฏผลดังตาราง 5

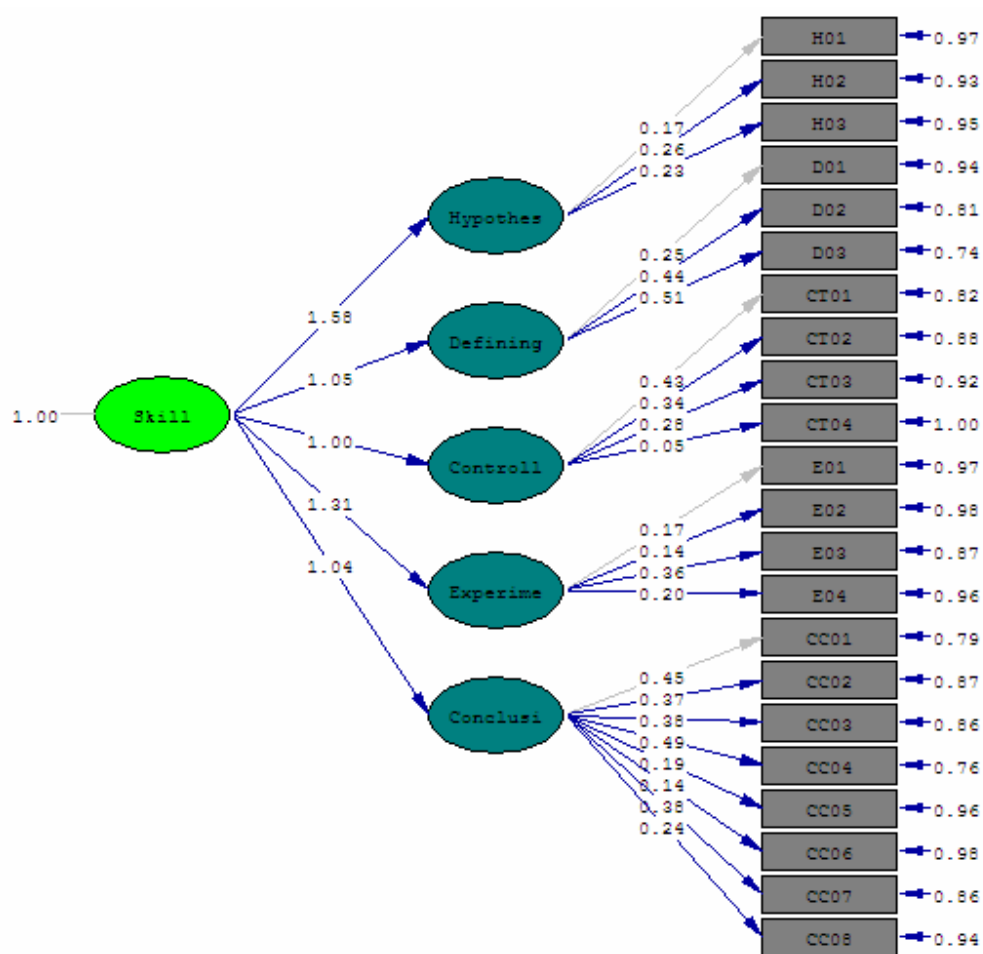
ตาราง 5 ผลการประมาณค่าการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง

ตัวแปร	องค์ประกอบ	สัมประสิทธิ์	SE	t	p
Hypoth	การตั้งสมมติฐาน	1.58	1.51	3.95	>.05
Definin	การกำหนดนิยามเชิงปฏิบัติการ	1.05	0.11	3.89	>.05
Control	การกำหนดและควบคุมตัวแปร	1.00	0.01	6.59	>.05
Experim	การทดลอง	1.31	0.72	3.40	>.05
Conclus	การตีความหมายข้อมูลและลงข้อสรุป	1.04	0.09	7.14	>.05

Chi-square = 209.90 df = 204 P = 0.37
 GFI = 0.94 , AGFI = 0.92 , RMSEA = 0.01 , CFI = 0.98

ผลการวิเคราะห์ตามตาราง 5 สรุปได้ดังนี้

การวิเคราะห์ดัชนีความเหมาะสมของโมเดลองค์ประกอบเชิงยืนยันอันดับสอง พบว่า ค่าไคสแควร์ = 209.90 (df = 204 P = 0.37) ค่า GFI = 0.94 ค่า AGFI = 0.92 ค่า RMSEA = 0.01 และ CFI = 0.98 ค่าดัชนีความเหมาะสมของทุกตัวแปรบ่งชี้ว่า โมเดลการวิจัยมีความเหมาะสมกับข้อมูลเชิงประจักษ์ เนื่องจากเป็นไปตามเกณฑ์ที่กำหนดไว้ที่ว่า ไคสแควร์ต้องไม่มีนัยสำคัญทางสถิติ ค่า GFI และ AGFI ควรสูงกว่า .90 และเกณฑ์ในระดับดีคือมากกว่า .95 ส่วนค่า RMSEA ควรต่ำกว่า .06 และค่า CFI ควรสูงกว่า .95 ตามลำดับ แสดงว่า โมเดลองค์ประกอบเชิงยืนยันอันดับสอง 1 องค์ประกอบ มีความเหมาะสมกับข้อมูลเชิงประจักษ์ และเป็นไปตามสมมติฐานการวิจัยข้อที่ 2 ทุกประการ สำหรับโมเดลองค์ประกอบเชิงยืนยันอันดับสอง 1 องค์ประกอบ ดังภาพประกอบ 7



Chi-Square=209.90, df=204, P-value=0.37361, RMSEA=0.010

ภาพประกอบ 7 ผลการประมาณค่าสัมประสิทธิ์องค์ประกอบโมเดลการวิเคราะห์องค์ประกอบ
เชิงยืนยันอันดับสอง

3. การวิเคราะห์ความเที่ยงตรงรวมและความเที่ยงตรงแยก

ผู้วิจัยได้นำตัวแปรสังเกตทั้ง 22 ตัวแปร ซึ่งจัดอยู่ในกลุ่ม 5 องค์ประกอบย่อย ของทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการมาจัดกระทำโดยการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ทั้งหมดของตัวแปรย่อยแต่ละตัว โดยใช้สูตรสหสัมพันธ์แบบเพียร์สัน (บุญเชิด ภิญโญนนตพงษ์. 2545 : 84) นำผลที่ได้มาใส่ในเมทริกซ์ที่แสดงความสัมพันธ์ระหว่างข้อมูล แล้วจึงรายงานผลการวิเคราะห์ข้อมูลตามวิธีการวิเคราะห์หลายลักษณะหลายวิธี ปรากฏผลดังตาราง 6

ผลการวิเคราะห์ตามตาราง 6 สรุปได้ดังนี้

สัมประสิทธิ์สหสัมพันธ์ภายในทักษะการตั้งสมมติฐาน (Hypothesis) อยู่ระหว่าง .34 - .55 สัมประสิทธิ์สหสัมพันธ์ภายในทักษะการกำหนดนิยามเชิงปฏิบัติการ (Defining) อยู่ระหว่าง .50 - .53 สัมประสิทธิ์สหสัมพันธ์ภายในทักษะการกำหนดและควบคุมตัวแปร (Controlling) อยู่ระหว่าง .47 - .54 สัมประสิทธิ์สหสัมพันธ์ภายในทักษะการทดลอง (Experimenting) อยู่ระหว่าง .39 - .53 และสัมประสิทธิ์สหสัมพันธ์ภายในทักษะการตีความหมายข้อมูลและลงข้อสรุป (Conclusion) อยู่ระหว่าง .29 - .59 ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างทักษะต่าง ๆ นั้นต่ำกว่าค่าสัมประสิทธิ์สหสัมพันธ์ภายในทักษะ แสดงให้เห็นว่า ความเที่ยงตรงรวมมีค่าสูงกว่าความเที่ยงตรงแยก

4. การแสดงหลักฐานความเที่ยงตรงตามโครงสร้าง โดยวิธีการวิเคราะห์โมเดลของราล์ช

แบบสามองค์ประกอบตามแนวทางของทฤษฎีการตอบข้อคำถาม

ในขั้นตอนนี้ ผู้วิจัยได้ดำเนินการนำผลการตอบของผู้เข้าสอบทั้งหมดมาถ่ายเอกสาร 2 ชุด จัดให้ผู้ตรวจจำนวน 14 คน ตรวจให้คะแนนตามแนวทางการตอบและกฎเกณฑ์การให้คะแนนใน คู่มือการตรวจที่ผู้วิจัยได้สร้างขึ้น โดยมีเงื่อนไข นักเรียน 1 คน จะต้องถูกตรวจให้คะแนนจากผู้ตรวจ 2 คน โดยไม่ซ้ำกัน แล้วจึงนำผลการตรวจให้คะแนนทั้งหมดมาวิเคราะห์ด้วยโปรแกรม FACETS ของ Linacre ภายใต้เงื่อนไขการวิเคราะห์แบบ 3 องค์ประกอบคือ องค์ประกอบของทักษะกระบวนการ องค์ประกอบของผู้ตรวจให้คะแนน และองค์ประกอบของนักเรียน ปรากฏผลดังตาราง 7 - 9

ตาราง 7 ผลการวิเคราะห์ค่าความยากของทักษะกระบวนการทางวิทยาศาสตร์ ตามโมเดลของราล์ช แบบ 3 องค์ประกอบ (จำแนกตามทักษะ)

ที่	ความยาก (Logit)	SE	INFIT (MnSq)	OUTFIT (MnSq)	ทักษะ
1	-.39	.04	0.8	1.0	การตั้งสมมติฐาน
2	-.15	.04	0.6	0.7	การกำหนดนิยามเชิงปฏิบัติการ
3	.14	.04	0.7	0.8	การกำหนดและควบคุมตัวแปร
4	-.15	.04	0.8	0.9	การทดลอง
5	.56	.04	0.7	0.9	การตีความหมายข้อมูลและลงข้อสรุป
Mean	.00	.04	0.7	0.9	
SD	.33	.00	0.1	0.1	

Task Separation Reliability = .98

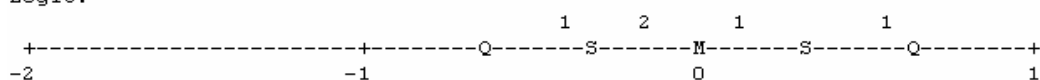
ผลการวิเคราะห์ตามตาราง 7 สรุปได้ดังนี้

ค่าความยากของทักษะกระบวนการ (Task Difficulty) ในแต่ละทักษะอยู่ในช่วงระหว่าง $-.39$ ถึง $+.56$ ($SD = 0.33$, $Mean SE = 0.4$) ทักษะการตั้งสมมติฐาน ทักษะการกำหนดนิยามเชิงปฏิบัติการ และทักษะการทดลอง มีค่าความยาก เท่ากับ $-.39$, $-.15$ และ $-.15$ ตามลำดับ ส่วนทักษะการตีความหมายข้อมูลและลงข้อสรุปและทักษะการกำหนดและควบคุมตัวแปร มีค่าความยาก เท่ากับ $.56$ และ $.14$ ตามลำดับ ทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการทั้ง 5 ทักษะ มีค่าความยากอยู่ในเกณฑ์ที่ยอมรับได้ คืออยู่ในช่วงระหว่าง -2.5 ถึง $+2.5$ (ศิริชัย กาญจนวาสี. 2545: 48)

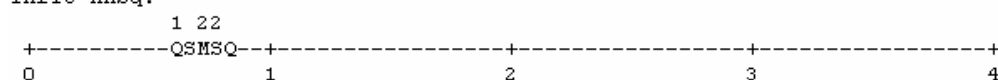
การเปรียบเทียบความเหมาะสมของข้อมูลเชิงประจักษ์กับโมเดลของราส์ชแบบสามองค์ประกอบ โดยใช้ Weighted mean square residual : INFIT และ Unweighted mean square residual : OUTFIT ในการทดสอบ ผลการวิเคราะห์ INFIT ในแต่ละทักษะอยู่ในแต่ละทักษะอยู่ในช่วงระหว่าง 0.6 ถึง 0.8 ($Mean = 0.7$, $SD = 0.1$) และ OUTFIT ในแต่ละทักษะอยู่ในช่วงระหว่าง 0.7 ถึง 1.0 ($Mean = 0.9$, $SD = 0.1$) ทั้งสองค่าอยู่ในเกณฑ์ที่ยอมรับได้ตามเกณฑ์ของลันซ์, ไรท์และลินาเคอร์ (Engelhard. 1994: 97; citing Lunz; Wright; & Linacre. 1990) ที่กำหนดขอบเขตของการยอมรับได้ของโมเดลที่เหมาะสมกับข้อมูลอยู่ระหว่าง $.6$ ถึง 1.5 ความเชื่อมั่นองค์ประกอบของทักษะกระบวนการ (Task Separation Reliability) เท่ากับ $.98$ สำหรับค่าความยากของทักษะกระบวนการ และการเปรียบเทียบความเหมาะสมของข้อมูลกับโมเดล รายละเอียดดังภาพประกอบ 8

Ratings of Science Process skill 08-18-2007 23:24:10
Table 6.3 Process Skill Facet Summary.

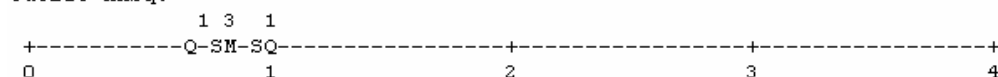
Logit:



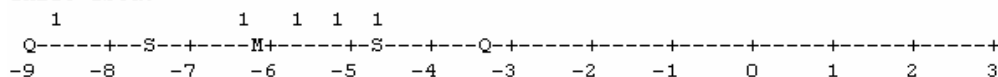
Infitt MnSq:



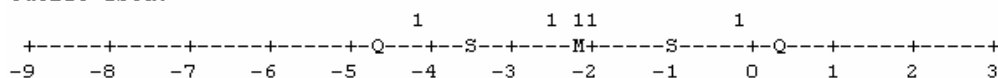
Outfit MnSq:



Infitt ZStd:



Outfit ZStd:



ภาพประกอบ 8 ค่าความยากของทักษะกระบวนการ และการเปรียบเทียบความเหมาะสมของข้อมูลกับโมเดล (เฉพาะองค์ประกอบของทักษะกระบวนการ)

ตาราง 8 ผลการวิเคราะห์ความสามารถของผู้ตรวจให้คะแนนทักษะกระบวนการทาง
วิทยาศาสตร์ตามโมเดลของราล์ชแบบ 3 องค์ประกอบ (จำแนกตามผู้ตรวจ)

ผู้ตรวจ	ความสามารถ (Logit)	SE	INFIT (MnSq)	OUTFIT (MnSq)
1	-.34	.07	0.7	0.9
2	-.05	.07	0.7	0.8
3	-.13	.07	0.8	0.8
4	-.16	.07	0.7	0.8
5	.22	.07	0.7	0.9
6	-.13	.07	0.7	0.8
7	-.09	.07	0.7	0.7
8	-.19	.07	0.8	0.9
9	-.26	.07	0.8	1.0
10	-.09	.07	0.7	0.8
11	-.54	.08	0.8	1.1
12	.01	.07	0.6	0.7
13	-.13	.07	0.6	0.7
14	-.24	.07	0.8	0.9
Mean	-.15	.07	0.7	0.9
SD	.17	.00	0.1	0.1

Rater Separation Reliability = .83

ผลการวิเคราะห์ตามตาราง 8 สรุปได้ดังนี้

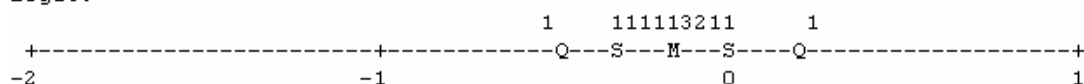
การวิเคราะห์ความสามารถของผู้ตรวจให้คะแนนทักษะกระบวนการทางวิทยาศาสตร์
ขั้นบูรณาการ พบว่า ค่าความสามารถของผู้ตรวจ (Severity/Leniency) อยู่ในช่วงระหว่าง -.54
ถึง +.22 (SD = 0.17 , Mean SE = 0.07) ผู้ตรวจที่อยู่ในกลุ่มเคร่งครัด ได้แก่ ผู้ตรวจคนที่ 5
และ 12 ซึ่งมีค่าความสามารถของผู้ตรวจ เท่ากับ .22 และ .01 ตามลำดับ ที่เหลือนอกนั้นเป็น
ผู้ตรวจในกลุ่มใจดี ได้แก่ ผู้ตรวจคนที่ 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 13 และ 14 โดยมีค่า
ความสามารถของผู้ตรวจ เท่ากับ -.34, -.05, -.13, -.16, -.13, -.09, -.19, -.26, -.09, -.54, -.13
และ -.24 ตามลำดับ ก็ยังถือว่าผู้ตรวจแต่ละคนมีความสามารถใกล้เคียงกัน และไม่ได้แตกต่างกัน
มากนัก

การเปรียบเทียบความเหมาะสมของข้อมูลเชิงประจักษ์กับโมเดลของราส์ชแบบสามองค์ประกอบ โดยใช้ Weighted mean square residual : INFIT และ Unweighted mean square residual : OUTFIT ในการทดสอบ ผลการวิเคราะห์ INFIT ของผู้ตรวจอยู่ในช่วงระหว่าง 0.6 ถึง 0.8 (Mean = 0.7 , SD = 0.1) และ OUTFIT ของผู้ตรวจอยู่ในช่วงระหว่าง 0.7 ถึง 1.1 (Mean = 0.9 , SD = 0.1) ทั้งสองค่าอยู่ในเกณฑ์ที่ยอมรับได้ตามเกณฑ์ของลินซ์, ไรท์ และลินาเคอร์ (Engelhard. 1994: 97; citing Lunz; Wright; & Linacre. 1990) ที่กำหนดขอบเขตการยอมรับได้ของโมเดลที่เหมาะสมกับข้อมูลอยู่ระหว่าง .6 ถึง 1.5 ความเชื่อมั่นองค์ประกอบของผู้ตรวจ (Rater Separation Reliability) เท่ากับ .83 สำหรับค่าความสามารถของผู้ตรวจและการเปรียบเทียบความเหมาะสมของข้อมูลกับโมเดลรายละเอียดดังภาพประกอบ 9

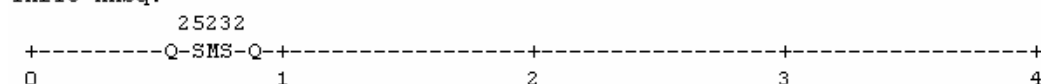
Ratings of Science Process skill 08-18-2007 23:24:10

Table 6.1 Teacher Facet Summary.

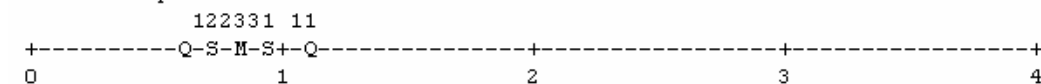
Logit:



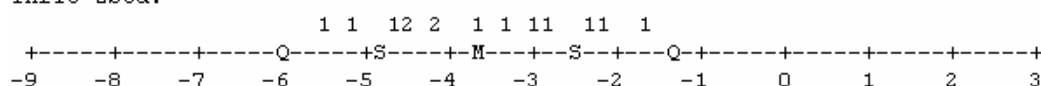
Infit MnSq:



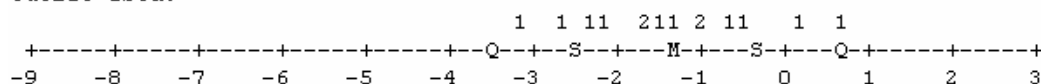
Outfit MnSq:



Infit ZStd:



Outfit ZStd:



ภาพประกอบ 9 ค่าความสามารถของผู้ตรวจ และการเปรียบเทียบความเหมาะสมของข้อมูลกับโมเดล (เฉพาะองค์ประกอบของผู้ตรวจ)

ตาราง 9 ผลการวิเคราะห์ความสามารถของนักเรียนตามโมเดลของราส์ช แบบ 3 องค์ประกอบ
(จำแนกตามนักเรียน)

นักเรียน คนที่	ความสามารถ (Logit)	INFIT (MnSq)	OUTFIT (MnSq)	Mean Rating From 2 Rater	Expected Rater from Rasch Model
203	.86	0.9	0.9	3.6	3.64
217	.86	0.9	0.9	3.6	3.64
37	.83	1.2	2.1	3.3	3.63
4	.82	0.8	0.9	3.5	3.62
7	.78	0.7	0.8	3.4	3.59
41	.78	1.1	1.5	3.2	3.59
160	.78	1.0	0.9	3.4	3.59
76	.78	1.3	2.2	3.2	3.59
147	.72	0.9	1.4	3.1	3.55
139	.56	1.5	2.6	3.0	3.43
146	.42	2.6	2.7	2.9	3.29
22	.27	1.0	1.1	2.8	3.12
188	-.82	1.2	2.3	2.0	1.61
233	-1.04	1.6	3.4	1.8	1.43
23	-1.09	0.6	0.8	1.7	1.40
71	-1.09	0.4	0.7	1.7	1.40
170	-1.09	1.2	1.5	1.7	1.40
211	-1.09	0.9	1.1	1.7	1.40
245	-1.09	0.4	0.6	1.7	1.40
ฯลฯ					
Mean	0.00	0.8	0.9	2.6	2.69
SD	0.56	0.4	0.5	0.5	0.71

Examinee Separation Reliability = .67

ผลการวิเคราะห์ตามตาราง 9 สรุปได้ดังนี้

จากการวิเคราะห์ความสามารถด้านทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ
ของนักเรียน (ซึ่งนำมาเสนอบางส่วน รายละเอียดในภาคผนวก) พบว่า ระดับความสามารถของ
นักเรียน (Ability) แต่ละคนอยู่ในช่วงระหว่าง -1.09 ถึง +.86 (SD = 0.56 , Mean SE = 0.32)

ระดับความสามารถของนักเรียนแต่ละคนมีการกระจายแตกต่างกันไป ความสามารถด้านทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการสูงสุด ได้แก่ นักเรียนคนที่ 203 และ 217 ระดับความสามารถมีค่า .86 ทั้งสองคน (SE = 0.4) ต่ำที่สุด ได้แก่ นักเรียนคนที่ 23 , 71 , 170 , 211 และ 245 ระดับความสามารถมีค่า -1.09 ทุกคน (SE. = 0.35 , 0.38 , 0.47 , 0.39 , และ 0.38 ตามลำดับ) ระดับความสามารถของนักเรียนทุกคนอยู่ในเกณฑ์ที่ยอมรับได้ คือ -3 ถึง +3 (ศิริชัย กาญจนวาสี. 2545: 47)

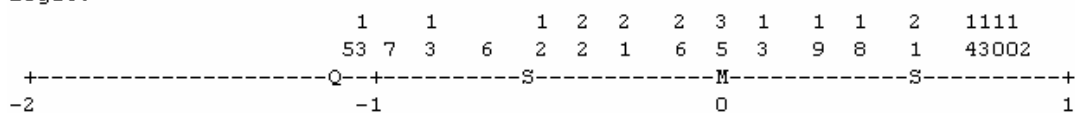
การเปรียบเทียบความเหมาะสมของข้อมูลเชิงประจักษ์กับโมเดลของราส์ชแบบสามองค์ประกอบ โดยใช้ Weighted mean square residual : INFIT และ Unweighted mean square residual : OUTFIT ในการทดสอบ ผลการวิเคราะห์ INFIT ของนักเรียนอยู่ในช่วงระหว่าง 0.2 ถึง 2.6 (Mean = 0.8 , S.D. = 0.4) และ OUTFIT ของนักเรียนอยู่ในช่วงระหว่าง 0.2 ถึง 3.4 (Mean = 0.9 , S.D. = 0.5) ถ้าพิจารณาในภาพรวม พบว่า ทั้งสองค่าอยู่ในเกณฑ์ที่ยอมรับได้ตามเกณฑ์ของลินซ์, ไรท์และลินาเคอร์ (Engelhard. 1994: 97; citing Lunz; Wright; & Linacre. 1990) ที่กำหนดขอบเขตของการ ยอมรับได้ของโมเดลที่เหมาะสมกับข้อมูลอยู่ระหว่าง .6 ถึง 1.5 แต่ถ้าพิจารณาเป็นรายบุคคลจะพบว่ามี มีบางคนที่ INFIT และ OUTFIT ไม่เหมาะสมกับโมเดล แต่ถ้านำไปเปรียบเทียบกับเกณฑ์ของไรท์และมาสเตอร์ (Wright; & Masters. 1982: 102) ที่กำหนดขอบเขตการยอมรับได้ของโมเดลมีค่าอยู่ระหว่าง -2 ถึง +2 จะพบว่า ข้อมูลมีความเหมาะสมกับโมเดลตามเกณฑ์ของไรท์และมาสเตอร์เป็นอย่างดี ยกเว้น OUTFIT ของนักเรียนคนที่ 233 ซึ่งมีค่าเท่ากับ 3.4 และ นักเรียนคนที่ 188 , 76 , 139 , 146 และ 37 ซึ่งมีค่าเท่ากับ 2.3 , 2.2 , 2.6 , 2.7 และ 2.1 ตามลำดับ นอกนั้นถือว่าเป็นไปตามเกณฑ์ของไรท์และมาสเตอร์ทุกประการ ความเชื่อมั่นองค์ประกอบของนักเรียน (Examinee Separation Reliability) เท่ากับ .67

การวิเคราะห์เปรียบเทียบค่าเฉลี่ยของการตรวจให้คะแนนจากผู้ตรวจ 2 คน (Mean Rating From 2 Rater) ที่พิจารณาจากผลของคะแนนดิบ (Raw Scores) กับค่าคาดหวังของผู้ตรวจจากโมเดลของราส์ช (Expected Rater from Rasch Model) พบว่า มีค่าใกล้เคียงกัน เช่น นักเรียนคนที่ 203 ค่าเฉลี่ยของผู้ตรวจที่พิจารณาจากคะแนนดิบเท่ากับ 3.6 ค่าคาดหวังของผู้ตรวจจากโมเดลของราส์ช เท่ากับ 3.64 หรือ นักเรียนคนที่ 160 ค่าเฉลี่ยของผู้ตรวจที่พิจารณาจากคะแนนดิบเท่ากับ 3.4 ค่าคาดหวังของผู้ตรวจจากโมเดลของราส์ช เท่ากับ 3.59 เป็นต้น แสดงว่า การวิเคราะห์ความสามารถของผู้เรียนโดยพิจารณาจากค่าเฉลี่ยของการตรวจให้คะแนนจากผู้ตรวจที่พิจารณาจากผลของคะแนนดิบตามทฤษฎีมาตรฐานเดิม และค่าคาดหวังของผู้ตรวจจากโมเดลของราส์ชตามทฤษฎีการตอบข้อคำถาม ให้ผลใกล้เคียงกัน สำหรับค่าความสามารถของนักเรียนและการเปรียบเทียบความเหมาะสมของข้อมูลกับโมเดล รายละเอียดดังภาพประกอบ 10

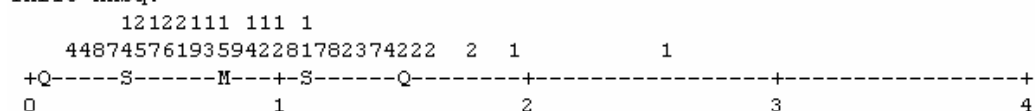
Ratings of Science Process skill 08-18-2007 23:24:10

Table 6.2 Student Facet Summary.

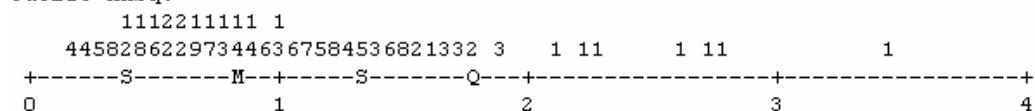
Logit:



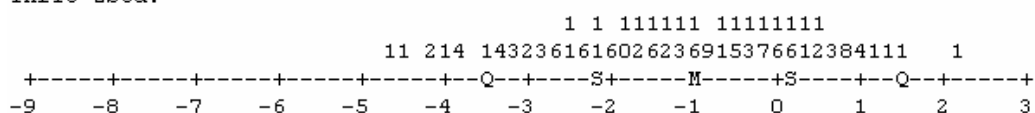
Infit MnSq:



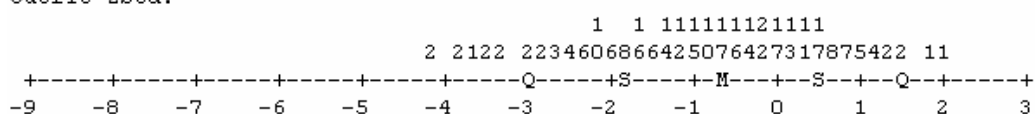
Outfit MnSq:



Infit 2Std:

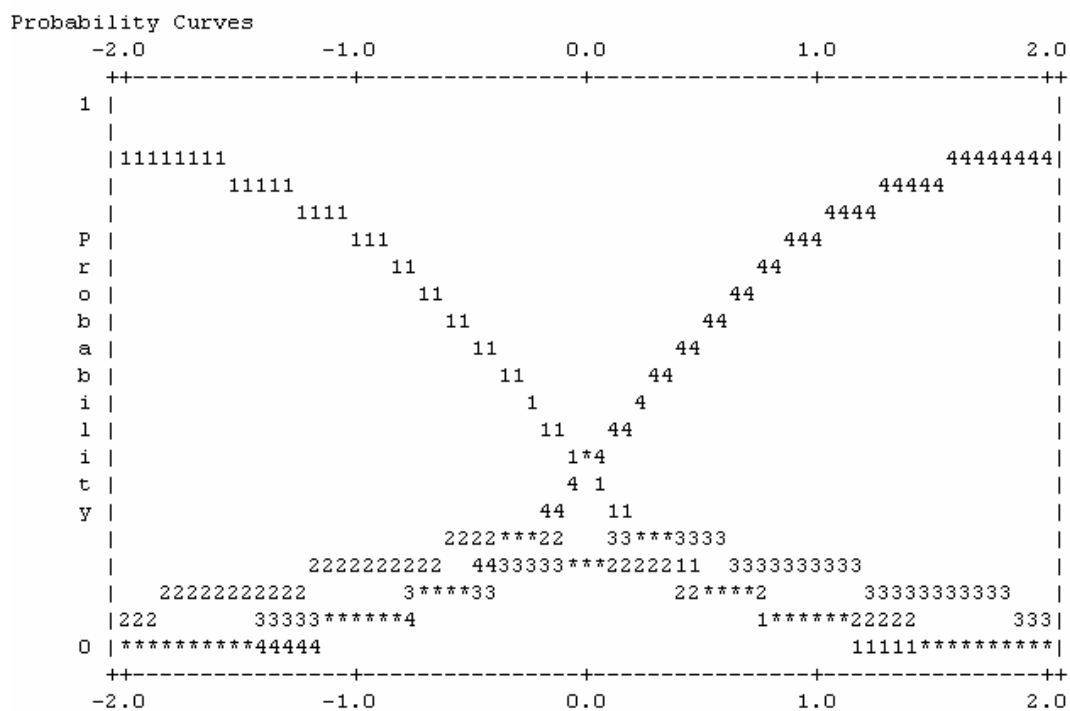


Outfit 2Std:

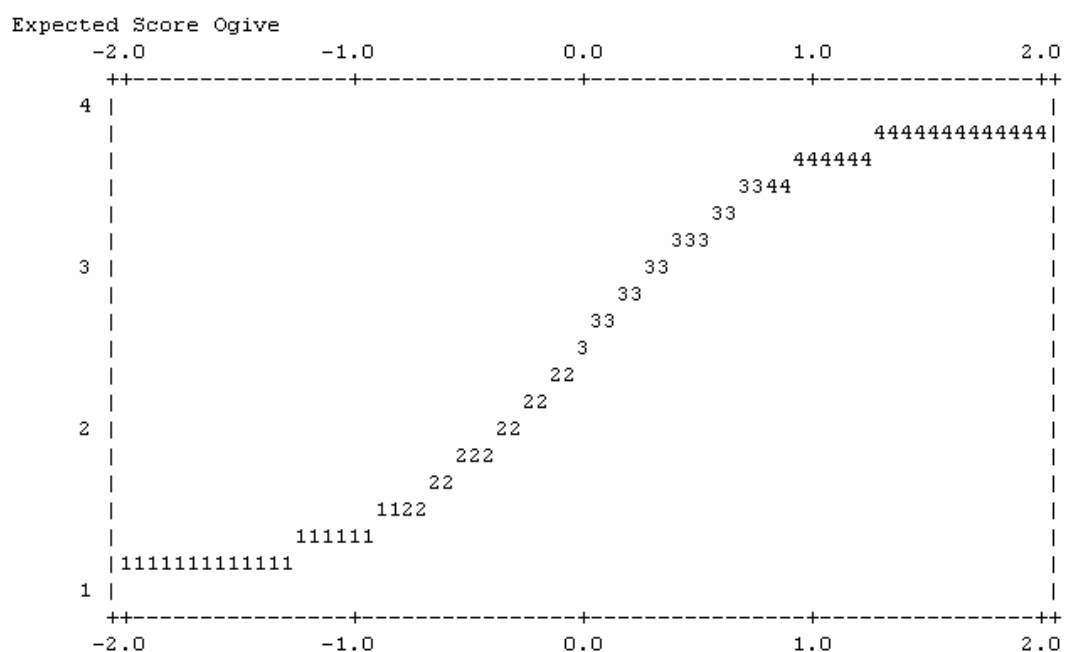


ภาพประกอบ 10 ค่าความสามารถของนักเรียนและการเปรียบเทียบความเหมาะสมของข้อมูล
กับโมเดล (เฉพาะองค์ประกอบของนักเรียน)

สำหรับความยากของทักษะกระบวนการ ความสามารถของผู้ตรวจให้คะแนน
ความสามารถของนักเรียน และระดับของทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการ
สามารถนำเสนอบนมาตราอันเดียวกัน (Common Scale) เพื่อแสดงความสัมพันธ์กันทั้ง 3
องค์ประกอบดังภาพประกอบ 11



ภาพประกอบ 12 โค้งลักษณะลำดับขั้นของทักษะในการประเมินทักษะกระบวนการทาง
วิทยาศาสตร์ที่มี 4 ระดับ ของฟาเซ็ทโมเดล



ภาพประกอบ 13 โค้งความน่าจะเป็นในการประเมินทักษะกระบวนการทางวิทยาศาสตร์ ของ
ฟาเซ็ทโมเดล

5. ผลการวิเคราะห์ความเชื่อมั่นของผลการตรวจให้คะแนน ความสอดคล้องของผู้ตรวจให้คะแนน และความเชื่อมั่นเชิงโครงสร้าง

5.1 ผลการวิเคราะห์ความเชื่อมั่นของผลการตรวจให้คะแนน จากสูตร r_{F-R} ของ Feldt - Raju (บุญเชิด ภิญโญนันตพงษ์. 2545 : 150) ผู้วิจัยนำผลการตรวจของผู้ตรวจแต่ละคน มาคำนวณหาค่าความเชื่อมั่นของผลการตรวจให้คะแนนแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการทั้งฉบับ ใช้สูตร r_{F-R} ของ Feldt - Raju แยกคำนวณตามผลการตรวจของผู้ตรวจแต่ละคน จึงนำผลการตรวจให้คะแนนของผู้ตรวจทุกคนมาหาค่าความเชื่อมั่นรวมโดยใช้สูตร r_{F-R} ของ Feldt - Raju ตามขั้นตอนเดิม ปรากฏผลดังตาราง 10

ตาราง 10 ค่าความเชื่อมั่นของแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการ

ผู้ตรวจคนที่	$\sum \lambda_i^2$	$\sum S_i^2$	S_x^2	r_{F-R}
1	0.270	6.875	15.744	0.772
2	0.232	7.726	19.533	0.787
3	0.273	8.266	18.962	0.776
4	0.278	5.275	10.756	0.706
5	0.222	4.953	10.100	0.655
6	0.255	7.252	15.512	0.715
7	0.235	6.894	16.656	0.766
8	0.277	7.634	16.435	0.740
9	0.270	7.065	15.669	0.752
10	0.294	8.221	19.795	0.722
11	0.270	7.355	15.003	0.772
12	0.226	5.688	11.410	0.648
13	0.240	6.750	16.789	0.787
14	0.284	7.783	14.969	0.670
ค่าความเชื่อมั่นรวมทั้งฉบับเท่ากับ .8921				

ผลการวิเคราะห์ข้อมูลตามตาราง 11 มีดังนี้

ค่าความเชื่อมั่นของการประเมินทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการ คำนวณโดยใช้สูตร r_{F-R} ของ Feldt-Raju จำแนกตามผู้ตรวจคนที่ 1-14 มีค่าอยู่ระหว่าง 0.648 ถึง 0.787 เมื่อนำผลการตรวจให้คะแนนของผู้ตรวจทุกคนมาหาค่าความเชื่อมั่นรวมโดยใช้สูตร r_{F-R} ของ Feldt – Raju ตามขั้นตอนเดิม พบว่า มีค่าความเชื่อมั่น เท่ากับ 0.8921

5.2 ผลการวิเคราะห์ความสอดคล้องของผู้ตรวจให้คะแนน โดยใช้สัมประสิทธิ์แคปปา ในขั้นตอนนี้ ผู้วิจัยได้นำเอาผลการตรวจให้คะแนนของผู้ตรวจทุกคนมาพิจารณาความเห็นที่ สอดคล้องกันเป็นรายคู่ แล้วคำนวณค่าสัมประสิทธิ์แคปปาของแต่ละคู่ ปรากฏผลดังตาราง 11

ตาราง 11 ความสอดคล้องระหว่างคู่ของผู้ตรวจให้คะแนนคำนวณโดยสัมประสิทธิ์แคปปา

	R01	R02	R03	R04	R05	R06	R07	R08	R09	R10	R11	R12	R13	R14
R01														
R02	0.76													
R03	0.81	0.75												
R04	0.75	0.74	0.76											
R05	0.79	0.73	0.84	0.77										
R06	0.77	0.76	0.82	0.78	0.80									
R07	0.05	0.76	0.81	0.81	0.85	0.82								
R08	0.82	0.75	0.87	0.74	0.83	0.80	0.78							
R09	0.82	0.74	0.86	0.76	0.79	0.83	0.72	0.88						
R10	0.69	0.73	0.80	0.75	0.85	0.76	0.81	0.78	0.72					
R11	0.81	0.79	0.82	0.72	0.79	0.80	0.77	0.81	0.85	0.78				
R12	0.81	0.76	0.85	0.67	0.84	0.83	0.78	0.82	0.85	0.81	0.81			
R13	0.81	0.74	0.74	0.80	0.74	0.76	0.67	0.78	0.80	0.63	0.79	0.80		
R14	0.80	0.78	0.79	0.77	0.84	0.81	0.82	0.81	0.81	0.81	0.78	0.80	0.77	
เฉลี่ยค่าสัมประสิทธิ์แคปปาเท่ากับ .7785														

ผลการวิเคราะห์ข้อมูลตามตาราง 11 มีดังนี้

สัมประสิทธิ์แคปปาให้ค่าสูงสุด คือ คู่ของผู้ตรวจคนที่ 9 และคนที่ 8 เท่ากับ .88 รองลงมา คือ คู่ของผู้ตรวจคนที่ 8 และคนที่ 4 กับคู่ของผู้ตรวจคนที่ 9 และคนที่ 3 มีค่า เท่ากับ .87 และ .86 ตามลำดับ ส่วนสัมประสิทธิ์แคปปาที่ให้ค่าต่ำสุด คือ คู่ของผู้ตรวจคนที่ 13 และคนที่ 10 มีค่าเท่ากับ .63 ค่าสัมประสิทธิ์แคปปาเฉลี่ยทุกคู่เท่ากับ .7785 หมายถึง ความเห็นของ ผู้ตรวจทั้งสองคนมีความสอดคล้องกันมาก (รายละเอียดการแปลความหมาย ของสัมประสิทธิ์แคปปาอยู่ในภาคผนวก)

5.3 ผลการวิเคราะห์ความเชื่อมั่นเชิงโครงสร้าง ในขั้นตอนนี้ ผู้วิจัยได้นำค่าน้ำหนักองค์ประกอบ (Factor Loading) และ ค่าความคลาดเคลื่อน (Error) ของตัวแปรสังเกตที่เป็นผลจากการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งและอันดับสองมาคำนวณหาค่า Construct Reliability และ Variance Extracted เพื่อทดสอบความเพียงพอของตัวแปรสังเกตที่จะใช้เป็นตัวแทนของตัวแปรแฝง การประเมินความเพียงพอของค่าความเชื่อมั่นเชิงโครงสร้างใช้เกณฑ์ตั้งแต่ 0.70 ขึ้นไป ส่วนค่า Variance Extracted ใช้เกณฑ์ตั้งแต่ 0.50 ขึ้นไป ปรากฏผลดังตาราง 12 - 13

ตาราง 12 ค่าน้ำหนักองค์ประกอบและความคลาดเคลื่อนของโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง

ตัวแปร	Λ_x	Λ_x^2	δ
H01	0.0600	0.0036	0.0300
H02	0.0700	0.0049	0.0300
H03	0.0800	0.0064	0.0300
D01	0.0600	0.0036	0.0200
D02	0.1500	0.0225	0.0200
D03	0.1900	0.0361	0.0300
CT01	0.1500	0.0225	0.0300
CT02	0.1200	0.0144	0.0300
CT03	0.0800	0.0064	0.0200
CT04	0.0200	0.0004	0.0200
E01	0.0600	0.0036	0.0200
E02	0.0500	0.0025	0.0200
E03	0.1300	0.0169	0.0300
E04	0.0600	0.0036	0.0200
CC01	0.1600	0.0256	0.0200
CC02	0.1200	0.0144	0.0200
CC03	0.1300	0.0169	0.0200
CC04	0.1800	0.0324	0.0200

ตาราง 12 (ต่อ)

ตัวแปร	Λ_x	Λ_x^2	δ
CC05	0.0700	0.0049	0.0200
CC06	0.0400	0.0016	0.0200
CC07	0.1400	0.0196	0.0200
CC08	0.0900	0.0081	0.0200
Σ	2.2100	0.2709	0.5100
Construct Reliability = .9055, Variance Extracted = .3469			

ตาราง 13 ค่าน้ำหนักองค์ประกอบและความคลาดเคลื่อนของโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง

ตัวแปร	Λ_x	Λ_x^2	δ
Hypoth	1.58	2.50	1.51
Definin	1.05	1.10	0.11
Control	1.00	1.00	0.01
Experim	1.31	1.72	0.72
Conclus	1.04	1.08	0.09
Σ	5.98	7.40	2.44
Construct Reliability = .9361, Variance Extracted = .7519			

ผลการวิเคราะห์ข้อมูลตามตาราง 12 - 13 มีดังนี้

ในการทดสอบความเพียงพอของตัวแปรสังเกตที่จะใช้เป็นตัวแทนของตัวแปรแฝง โดยการประเมินค่าความเชื่อมั่นเชิงโครงสร้าง (Construct Reliability) และการประเมินค่า Variance Extracted พบว่า โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง Construct Reliability เท่ากับ .9055 และ Variance Extracted เท่ากับ .3469 ส่วนโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง Construct Reliability เท่ากับ .9361 และ Variance Extracted เท่ากับ .7519 ค่า Construct Reliability ทั้งสองโมเดลเกิน 0.70 ขึ้นไป ส่วนค่า Variance Extracted โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองเท่านั้นที่มีค่าเกิน 0.50 อันดับหนึ่งถือว่า ไม่ผ่านเกณฑ์ (Hair; et al. 1995: 642)

บทที่ 5

สรุปผล อภิปราย และข้อเสนอแนะ

การวิจัยนี้ มีความมุ่งหมายสำคัญในการแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของคะแนนการประเมินทักษะกระบวนการทางวิทยาศาสตร์ โดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน การวิเคราะห์ความเที่ยงตรงรวมและความเที่ยงตรงแยกตามแนวทางทฤษฎีมาตรฐานเดิม และการวิเคราะห์โมเดลของราล์ชแบบหลายองค์ประกอบตามแนวทางของทฤษฎีการตอบข้อคำถาม พร้อมทั้งแสดงหลักฐานความเชื่อมั่นในการประเมินจากการให้คะแนนของผู้ประเมินมากกว่าหนึ่งคนในทั้งสองทฤษฎี ก่อนที่จะสรุปผลการวิจัย อภิปรายผล และให้ข้อเสนอแนะเพื่อให้เห็นกรอบของการวิจัย จะขอกล่าวถึงความมุ่งหมาย สมมติฐาน และวิธีการวิจัยโดยสังเขป ดังนี้

ความมุ่งหมายของการวิจัย

ความมุ่งหมายของการวิจัยได้แยกเป็นความมุ่งหมายเฉพาะดังนี้

1. เพื่อแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของคะแนนการประเมินทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการโดยใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยันตามแนวทางของทฤษฎีมาตรฐานเดิม ดังนี้

- 1.1 ศึกษาโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งของแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการที่มี 5 องค์ประกอบ ได้แก่ องค์ประกอบของทักษะการตั้งสมมติฐาน ทักษะการกำหนดและควบคุมตัวแปร ทักษะการกำหนดนิยามเชิงปฏิบัติการ ทักษะการทดลอง และทักษะการตีความหมายข้อมูลและลงข้อสรุป

- 1.2 ศึกษาโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองของแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ที่มี 1 องค์ประกอบ คือ องค์ประกอบของทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการ

2. เพื่อแสดงหลักฐานความเที่ยงตรงรวมและความเที่ยงตรงแยกของคะแนนการประเมินทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการตามแนวทางของทฤษฎีมาตรฐานเดิม

3. เพื่อแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของคะแนนการประเมินทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการโดยการวิเคราะห์โมเดลของราล์ชแบบหลายองค์ประกอบตามแนวทางของทฤษฎีการตอบข้อคำถาม

4. เพื่อแสดงหลักฐานความเชื่อมั่นในการให้คะแนนการวัดทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการที่ใช้ผู้ตรวจให้คะแนนมากกว่าหนึ่งคน โดยใช้สถิติดังนี้

4.1 ความเชื่อมั่นของผลการตรวจให้คะแนน (Reliability of rating) แบบความสอดคล้องภายใน (Internal consistency) โดยใช้สูตร r_{F-R} ของ Feldt – Raju

4.2 ดัชนีความสอดคล้องของผู้ตรวจให้คะแนน โดยใช้สัมประสิทธิ์ Kappa

4.3 ความเชื่อมั่นเชิงโครงสร้าง (Construct Reliability) และค่า Variance

Extracted

สมมติฐานของการวิจัย

การวิจัยครั้งนี้ตั้งสมมติฐานไว้ว่า ผลการตรวจให้คะแนนแบบวัดทักษะกระบวนการทางวิทยาศาสตร์มีความเที่ยงตรงตามโครงสร้าง โดยมีสมมติฐานย่อยดังนี้

1. องค์ประกอบตัวแปรแฝง 5 องค์ประกอบของโมเดล ได้แก่ องค์ประกอบของการตั้งสมมติฐาน การกำหนดนิยามเชิงปฏิบัติการ การกำหนดและควบคุมตัวแปร การทดลอง และการตีความหมายข้อมูลและลงข้อสรุป มีความเหมาะสมกับข้อมูลเชิงประจักษ์

2. องค์ประกอบเชิงยืนยันอันดับหนึ่ง สามารถอธิบายองค์ประกอบใหม่เป็นองค์ประกอบอันดับสองของตัวแปรแฝง 1 องค์ประกอบ คือ องค์ประกอบทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ

วิธีดำเนินการวิจัย

ผู้วิจัยได้ดำเนินการวิจัยตามขั้นตอน ดังนี้

1. ติดต่อประสานงานกับโรงเรียนที่เป็นกลุ่มตัวอย่าง แล้วจึงคัดเลือกผู้ตรวจให้คะแนน (Rater) จำนวน 14 คน จัดประชุมชี้แจงเพื่อซักซ้อมและทำความเข้าใจขั้นตอนการตรวจให้คะแนน เมื่อเตรียมการทุกอย่างเสร็จเรียบร้อยแล้ว นำแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ที่ผ่านการวิเคราะห์หาคุณภาพรายข้อแล้ว ไปทดสอบกับกลุ่มตัวอย่าง จำนวน 280 คน

2. นำกระดาษคำตอบของผู้เข้าสอบทั้งหมดมาถ่ายเอกสารจำนวน 2 ชุด จัดให้ผู้ตรวจจำนวน 14 คน ตรวจให้คะแนนตามแนวทางการตอบและกฎเกณฑ์การให้คะแนนในคู่มือการตรวจที่ผู้วิจัยได้สร้างขึ้น โดยมีเงื่อนไข นักเรียน 1 คน จะต้องถูกตรวจให้คะแนนจากผู้ตรวจ 2 คน โดยไม่ซ้ำกัน เพราะฉะนั้น ผู้ตรวจ 1 คน จะต้องตรวจผลงานนักเรียนจำนวน 40 ชุด จึงจะครบตามจำนวนนักเรียน 280 คน พอดี แล้วบันทึกผลการตรวจให้คะแนน

3. การวิเคราะห์ข้อมูล ได้ดำเนินการตามขั้นตอน ดังนี้

3.1 วิเคราะห์ข้อมูลพื้นฐานของตัวแปรที่ใช้ในการวิจัย โดยนำเสนอค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้และค่าความโด่งของข้อมูลตามตัวแปรในแต่ละทักษะ เพื่อพิจารณาถึงข้อตกลงเบื้องต้นเกี่ยวกับการกระจายเป็นโค้งปกติของข้อมูล

3.2 วิเคราะห์ข้อมูลเพื่อตอบสมมติฐานการวิจัย ใช้โปรแกรม LISREL 8.54 ของ Jöreskog & Sörbom โดยมีลำดับการวิเคราะห์ข้อมูลคือ เริ่มจากการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง เพื่อทดสอบโมเดลองค์ประกอบอันดับหนึ่งที่มี 5 องค์ประกอบว่า มีความเหมาะสมกับข้อมูลเชิงประจักษ์หรือไม่ แล้วจึงวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองที่มี 1 องค์ประกอบ เพื่อทดสอบแบบจำลององค์ประกอบอันดับสองว่ามีความเหมาะสมกับข้อมูลเชิงประจักษ์หรือไม่เช่นเดียวกัน ค่าสถิติที่ใช้ในการตรวจสอบความเที่ยงตรงตามโครงสร้างของโมเดลการวิจัยตามสมมติฐานกับข้อมูลเชิงประจักษ์ ได้แก่ ค่าดัชนีความเหมาะสมสัมบูรณ์ที่พิจารณาค่าไคสแควร์ โดยพิจารณาจากค่าองศาอิสระที่ไม่มีนัยสำคัญที่ระดับ .05 ค่า GFI และ AGFI มีค่ามากกว่า .90 ในกลุ่มเศษเหลือใช้ค่า RMSEA เกณฑ์พิจารณาคือขนาดดัชนีไม่เกิน .06 ดัชนีเหมาะสมเชิงเปรียบเทียบ พิจารณาค่า CFI ที่สูงกว่า .90 การประเมินในรายละเอียด พิจารณาจากความมีนัยสำคัญของค่าสถิติ t ของค่า Factor Loading ของแต่ละองค์ประกอบ เป็นดัชนีที่บ่งบอกถึงคุณภาพของตัวแปรสังเกตในการอธิบายตัวแปรแฝง

3.3 วิเคราะห์ค่าความเที่ยงตรงรวมและค่าความเที่ยงตรงแยก โดยคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ทั้งหมดของตัวแปรย่อยทั้ง 22 ตัวแปร ใช้สูตรสหสัมพันธ์แบบเพียร์สัน นำผลที่ได้มาใส่ในเมทริกซ์ที่แสดงความสัมพันธ์ระหว่างข้อมูล การแปลผลเปรียบเทียบระหว่างค่าสัมประสิทธิ์สหสัมพันธ์ภายในทักษะกับค่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างทักษะ

3.4 วิเคราะห์โมเดลของราล์ชแบบหลายองค์ประกอบ โดยการนำผลการตรวจให้คะแนนจากผู้ตรวจจำนวน 14 คน ตรวจให้คะแนนตามแนวทางการตอบและกฎเกณฑ์การให้คะแนนในคู่มือการตรวจที่ผู้วิจัยได้สร้างขึ้น ภายใต้เงื่อนไข นักเรียน 1 คน จะต้องถูกตรวจให้คะแนนจากผู้ตรวจ 2 คน โดยไม่ซ้ำกัน มาวิเคราะห์ด้วยโปรแกรม FACETS ของ Linacre โดยการวิเคราะห์แบบ 3 องค์ประกอบคือ องค์ประกอบของทักษะกระบวนการ องค์ประกอบของผู้ตรวจให้คะแนน และองค์ประกอบของนักเรียน

3.5 วิเคราะห์ค่าความเชื่อมั่นของผลการตรวจให้คะแนน ความสอดคล้องของผู้ตรวจให้คะแนนและความเชื่อมั่นเชิงโครงสร้าง การวิเคราะห์ความเชื่อมั่นของผลการตรวจให้คะแนนใช้สูตร r_{F-R} ของ Feldt-Raju คำนวณค่าความเชื่อมั่นจำแนกตามผู้ตรวจแต่ละคน แล้วจึงนำผลการตรวจให้คะแนนของผู้ตรวจทุกคนมาหาค่าความเชื่อมั่นรวมโดยใช้สูตรเดิม ได้ค่าความเชื่อมั่นรวมของแบบวัดทั้งฉบับ การวิเคราะห์ความสอดคล้องของผู้ตรวจให้คะแนน โดยการนำผลการตรวจให้คะแนนของผู้ตรวจทุกคนมาพิจารณาความเห็นที่สอดคล้องกันเป็นรายคู่ แล้วคำนวณค่าสัมประสิทธิ์แคปปาของแต่ละคู่ จึงหาค่าเฉลี่ยรวม และวิเคราะห์ความเชื่อมั่นเชิงโครงสร้าง โดยคำนวณจากน้ำหนักองค์ประกอบและค่าความคลาดเคลื่อนของตัวแปรสังเกตที่เป็นผลจากการวิเคราะห์องค์ประกอบเชิงยืนยัน ค่า Construct Reliability ใช้เกณฑ์ .70 ขึ้นไป ส่วนค่า Variance Extracted ใช้เกณฑ์ .50 ขึ้นไป

สรุปผลการวิจัย

การสรุปผลการวิจัย แบ่งออกเป็น 4 ตอน คือ (1) สรุปผลการแสดงความเที่ยงตรงตามโครงสร้างโดยใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยันตามแนวทางของทฤษฎีมาตรฐานเดิม (2) สรุปผลการวิเคราะห์ความเที่ยงตรงรวมและความเที่ยงตรงแยก ตามแนวทางของทฤษฎีมาตรฐานเดิม (3) สรุปผลการแสดงความเที่ยงตรงตามโครงสร้าง โดยใช้วิธีการวิเคราะห์โมเดลของราล์ชแบบหลายองค์ประกอบตามแนวทางของทฤษฎีการตอบข้อคำถาม (4) สรุปผลการแสดงหลักฐานความเชื่อมั่นของผลการตรวจให้คะแนน ความสอดคล้องของผู้ตรวจให้คะแนน และความเชื่อมั่นเชิงโครงสร้าง ดังมีรายละเอียดต่อไปนี้

1. การแสดงหลักฐานความเที่ยงตรงตามโครงสร้างโดยใช้วิธีการวิเคราะห์

องค์ประกอบเชิงยืนยันตามแนวทางของทฤษฎีมาตรฐานเดิม

การสรุปผลในตอนนี้ ผู้วิจัยนำเอาข้อมูลการตรวจให้คะแนนจากผู้ตรวจ 2 ชุด มาจับคู่แล้วหาค่าเฉลี่ยของการตรวจในแต่ละคู่ กลายเป็นข้อมูลชุดเดียวกันที่มีจำนวนนักเรียน 280 คน เป็นกลุ่มตัวอย่างในการวิจัย มีตัวแปรสังเกตจำนวน 22 ตัวแปร แล้วจึงนำข้อมูลทั้งหมดมาวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งและอันดับสอง ด้วยโปรแกรม LISREL 8.54 แยกผลการวิเคราะห์ออกเป็น 2 ตอน ดังนี้

1.1 ผลการเปรียบเทียบความเหมาะสมของข้อมูลเชิงประจักษ์ของโมเดลองค์ประกอบเชิงยืนยันอันดับหนึ่ง พบว่า ดัชนีความเหมาะสมสัมบูรณ์ทุกตัวให้ผลการวิเคราะห์เป็นไปตามเกณฑ์ทุกประการ ไคสแควร์ เท่ากับ 205.65 ($df = 199$, $p=0.36$) ไม่มีนัยสำคัญทางสถิติ อัตราส่วนระหว่างไคสแควร์กับระดับองศาอิสระ มีค่าไม่เกิน 2 ($\chi^2 / df = 1.0334$) GFI และ AGFI มีค่าเท่ากับ 0.94 และ 0.92 ตามลำดับ ทั้งสองค่าเป็นไปตามเกณฑ์ที่กำหนดคือ .90 ขึ้นไป ดัชนีกลุ่มเศษเหลือ RMSEA เท่ากับ 0.01 ซึ่งมีค่าไม่เกิน .06 และดัชนีเหมาะสมเชิงเปรียบเทียบ CFI เท่ากับ 0.98 สูงกว่า .90 ซึ่งถือว่าเป็นค่าที่ดีมาก ๆ จึงสรุปได้ว่า โมเดลองค์ประกอบเชิงยืนยันอันดับหนึ่ง 5 องค์ประกอบ มีความเหมาะสมกับข้อมูลเชิงประจักษ์ ตัวแปรสังเกตทั้ง 22 ตัวแปร มีความสัมพันธ์กับตัวแปรแฝงที่ประกอบด้วยทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการจำนวน 5 องค์ประกอบ และเป็นไปตามสมมติฐานการวิจัยข้อที่ 1 ทุกประการ

1.2 ผลการเปรียบเทียบความเหมาะสมของข้อมูลเชิงประจักษ์ของโมเดลองค์ประกอบเชิงยืนยันอันดับสอง พบว่า ดัชนีความเหมาะสมสัมบูรณ์ทุกตัวให้ผลการวิเคราะห์เป็นไปตามเกณฑ์ทุกประการ ไคสแควร์ เท่ากับ 209.90 ($df = 204$, $p=0.37$) ไม่มีนัยสำคัญทางสถิติ อัตราส่วนระหว่างไคสแควร์กับระดับองศาอิสระ มีค่าไม่เกิน 2 ($\chi^2 / df = 1.0289$) GFI และ AGFI มีค่าเท่ากับ 0.94 และ 0.92 ตามลำดับ ทั้งสองค่าเป็นไปตามเกณฑ์ที่กำหนดคือ .90 ขึ้น

ไป ดัชนีกลุ่มเศษเหลือ RMSEA เท่ากับ 0.01 ซึ่งมีค่าไม่เกิน .06 และดัชนีเหมาะสมเชิงเปรียบเทียบ CFI เท่ากับ 0.98 สูงกว่า .90 ซึ่งถือว่าเป็นค่าที่ดีมาก ๆ จึงสรุปได้ว่า โมเดลองค์ประกอบเชิงยืนยันอันดับสอง 1 องค์ประกอบ มีความเหมาะสมกับข้อมูลเชิงประจักษ์ ตัวแปรแฝงภายนอก 5 องค์ประกอบ สามารถอธิบายได้ด้วยตัวแปรแฝงภายใน 1 องค์ประกอบ คือองค์ประกอบของทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการและเป็นไปตามสมมติฐานการวิจัยข้อที่ 2 ทุกประการ

2. การวิเคราะห์ความเที่ยงตรงรวมและความเที่ยงตรงแยกตามแนวทางของทฤษฎีมาตรฐานเดิม

สรุปผลการวิจัยในตอนนี้ผู้วิจัยได้นำตัวแปรสังเกตทั้ง 22 ตัวแปร ซึ่งจัดอยู่ในกลุ่ม 5 องค์ประกอบ ของทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ มาจัดกระทำโดยคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ทั้งหมดของตัวแปรย่อยแต่ละตัว โดยใช้สูตรสหสัมพันธ์แบบเพียร์สัน (บุญเชิด ภิญโญนนตพงษ์. 2545: 84) นำผลที่ได้มาใส่ในเมทริกซ์แสดงความสัมพันธ์ระหว่างข้อมูล แล้วจึงรายงานผลการวิเคราะห์ข้อมูลตามวิธีการวิเคราะห์หลายลักษณะหลายวิธี ดังนี้

จากการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ภายใน พบว่า ทักษะการตั้งสมมติฐานมีค่าอยู่ระหว่าง .34 - .55 ทักษะการกำหนดนิยามเชิงปฏิบัติการมีค่าอยู่ระหว่าง .50 - .53 ทักษะการกำหนดและควบคุมตัวแปรมีค่าอยู่ระหว่าง .47 - .54 ทักษะการทดลองมีค่าอยู่ระหว่าง .39 - .53 และทักษะการตีความหมายข้อมูลและลงข้อสรุปมีค่าอยู่ระหว่าง .29 - .59 โดยภาพรวมพบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างทักษะมีค่าต่ำกว่าค่าสัมประสิทธิ์สหสัมพันธ์ภายในทักษะ แสดงให้เห็นว่า ความเที่ยงตรงรวมมีค่าสูงกว่าความเที่ยงตรงแยก เป็นไปตามเงื่อนไขการแสดงหลักฐานความเที่ยงตรงตามโครงสร้างโดยวิธีการวิเคราะห์หลายลักษณะหลายวิธีทุกประการ

3. การแสดงหลักฐานความเที่ยงตรงตามโครงสร้างโดยใช้วิธีการวิเคราะห์โมเดลของราล์ชแบบสามองค์ประกอบตามแนวทางของทฤษฎีการตอบข้อคำถาม

สรุปผลการวิจัยในขั้นตอนนี้ ผู้วิจัยได้ดำเนินการนำผลการตอบของผู้เข้าสอบทั้งหมดมาถ่ายเอกสาร 2 ชุด จัดให้ผู้ตรวจจำนวน 14 คน ตรวจให้คะแนนตามแนวทางของการตอบและกฎเกณฑ์การให้คะแนนในคู่มือการตรวจที่ผู้วิจัยได้สร้างขึ้น โดยมีเงื่อนไข นักเรียน 1 คน จะต้องถูกตรวจให้คะแนนจากผู้ตรวจ 2 คน โดยไม่ซ้ำกัน แล้วจึงนำผลการตรวจให้คะแนนทั้งหมดมาวิเคราะห์ด้วยโปรแกรม FACETS ของ Linacre ภายใต้เงื่อนไขการวิเคราะห์แบบ 3 องค์ประกอบคือ องค์ประกอบของทักษะกระบวนการ องค์ประกอบของผู้ตรวจให้คะแนน และองค์ประกอบของนักเรียน การสรุปผลจำแนกตามองค์ประกอบดังนี้

3.1 ผลการวิเคราะห์ความยากของทักษะกระบวนการและการเปรียบเทียบความเหมาะสมของข้อมูลกับโมเดล (เฉพาะองค์ประกอบของทักษะ)

ค่าความยากของทักษะ (Task Difficulty) ในแต่ละทักษะอยู่ในช่วงระหว่าง $-.39$ ถึง $+.56$ ($SD = 0.33$, $Mean SE = 0.4$) ทักษะการตั้งสมมติฐาน ทักษะการกำหนดนิยามเชิงปฏิบัติการ และทักษะการทดลอง มีค่าความยาก เท่ากับ $-.39$, $-.15$ และ $-.15$ ตามลำดับ ส่วนทักษะการตีความหมายข้อมูลและลงข้อสรุป และทักษะการกำหนดและควบคุมตัวแปร มีค่าความยาก เท่ากับ $.56$ และ $.14$ ตามลำดับ ทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ ทั้ง 5 ทักษะ มีค่าความยากอยู่ในเกณฑ์ที่ยอมรับได้ คืออยู่ในช่วงระหว่าง -2.5 ถึง $+2.5$ (ศิริชัย กาญจนวาสี. 2545: 48)

การเปรียบเทียบความเหมาะสมของข้อมูลเชิงประจักษ์กับโมเดลของราล์ชแบบสามองค์ประกอบ โดยใช้ Weighted mean square residual : INFIT และ Unweighted mean square residual : OUTFIT ในการทดสอบ ผลการวิเคราะห์ INFIT ในแต่ละทักษะอยู่ในช่วงระหว่าง 0.6 ถึง 0.8 ($Mean = 0.7$, $SD = 0.1$) และ OUTFIT ในแต่ละทักษะอยู่ในช่วงระหว่าง 0.7 ถึง 1.0 ($Mean = 0.9$, $SD = 0.1$) ทั้งสองค่าอยู่ในเกณฑ์ที่ยอมรับได้ตามเกณฑ์ของลินซ์, ไรท์และลินาเคอร์ (Engelhard. 1994: 97; citing Lunz; Wright; & Linacre. 1990) ที่ได้กำหนดขอบเขตของการยอมรับได้ของโมเดลที่เหมาะสมกับข้อมูลอยู่ระหว่าง $.6$ ถึง 1.5 ความเชื่อมั่นองค์ประกอบของทักษะ (Task Separation Reliability) เท่ากับ $.98$

3.2 ผลการวิเคราะห์ความสามารถของผู้ตรวจ และการเปรียบเทียบความเหมาะสมของข้อมูลกับโมเดล (เฉพาะองค์ประกอบของผู้ตรวจ)

ค่าความสามารถของผู้ตรวจ (Severity/Leniency) อยู่ในช่วงระหว่าง $-.54$ ถึง $+.22$ ($SD = 0.17$, $Mean SE = 0.07$) ผู้ตรวจที่อยู่ในกลุ่มเคร่งครัด ได้แก่ ผู้ตรวจคนที่ 5 และ 12 ซึ่งมีค่าความสามารถของผู้ตรวจ เท่ากับ $.22$ และ $.01$ ตามลำดับ ที่เหลือนอกนั้นเป็นผู้ตรวจในกลุ่มใจดี ได้แก่ ผู้ตรวจคนที่ 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 13 และ 14 โดยมีค่าความสามารถของผู้ตรวจ เท่ากับ $-.34$, $-.05$, $-.13$, $-.16$, $-.13$, $-.09$, $-.19$, $-.26$, $-.09$, $-.54$, $-.13$ และ $-.24$ ตามลำดับ ก็ยังถือว่าผู้ตรวจแต่ละคนมีความสามารถใกล้เคียงกัน และไม่ได้แตกต่างกันมากนัก

การเปรียบเทียบความเหมาะสมของข้อมูลเชิงประจักษ์กับโมเดลของราล์ชแบบสามองค์ประกอบ โดยใช้ Weighted mean square residual : INFIT และ Unweighted mean square residual : OUTFIT ในการทดสอบ ผลการวิเคราะห์ INFIT ของผู้ตรวจอยู่ในช่วงระหว่าง 0.6 ถึง 0.8 ($Mean = 0.7$, $SD = 0.1$) และ OUTFIT ของผู้ตรวจอยู่ในช่วงระหว่าง 0.7 ถึง 1.1 ($Mean = 0.9$, $SD = 0.1$) ทั้งสองค่าอยู่ในเกณฑ์ที่ยอมรับได้ตามเกณฑ์ของลินซ์, ไรท์และลินาเคอร์ (Engelhard. 1994: 97; citing Lunz; Wright; & Linacre. 1990) ที่ได้กำหนดขอบเขตของการยอมรับได้ของโมเดลที่เหมาะสมกับ ข้อมูลอยู่ระหว่าง $.6$ ถึง 1.5 ความเชื่อมั่นองค์ประกอบของผู้ตรวจ (Rater Separation Reliability) เท่ากับ $.83$

3.3 ผลการวิเคราะห์ความสามารถของนักเรียน และการเปรียบเทียบความเหมาะสมของข้อมูลกับโมเดล (เฉพาะองค์ประกอบของนักเรียน)

ระดับความสามารถของนักเรียน (Ability) แต่ละคนอยู่ในช่วงระหว่าง -1.09 ถึง +.86 (SD = 0.56 , Mean SE = 0.32) ความสามารถของนักเรียนแต่ละคนมีการกระจายแตกต่างกันไป ความสามารถด้านทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการสูงสุดมี 2 คน ต่ำที่สุดมี 5 คน ระดับความสามารถของนักเรียนทุกคนอยู่ในเกณฑ์ที่ยอมรับได้ คือ -3 ถึง +3 (ศิริชัย กาญจนวาสี. 2545: 47)

การเปรียบเทียบความเหมาะสมของข้อมูลเชิงประจักษ์กับโมเดลของราส์ชแบบสามองค์ประกอบ โดยใช้ Weighted mean square residual : INFIT และ Unweighted mean square residual : OUTFIT ในการทดสอบ ผลการวิเคราะห์ INFIT ของนักเรียนอยู่ในช่วงระหว่าง 0.2 ถึง 2.6 (Mean = 0.8 , S.D. = 0.4) และ OUTFIT ของนักเรียนอยู่ในช่วงระหว่าง 0.2 ถึง 3.4 (Mean = 0.9 , S.D. = 0.5) ถ้าพิจารณาในภาพรวม พบว่า ทั้งสองค่าอยู่ในเกณฑ์ที่ยอมรับได้ตามเกณฑ์ของลินซ์, ไรท์และลินาเคอร์ (Engelhard. 1994: 97; citing Lunz; Wright; & Linacre. 1990) ที่ได้กำหนดขอบเขตของการยอมรับได้ของโมเดลที่เหมาะสมกับข้อมูลอยู่ระหว่าง .6 ถึง 1.5 ถ้าพิจารณาเป็นรายบุคคลจะพบว่า มีบางคนที่มี INFIT และ OUTFIT ไม่เหมาะสมกับโมเดล แต่ถ้าเปรียบเทียบกับเกณฑ์ของไรท์และมาสเตอร์ (Wright; & Masters. 1982: 102) ที่กำหนดขอบเขตของการยอมรับได้ของโมเดลที่เหมาะสมกับข้อมูลที่มีค่าอยู่ระหว่าง -2 ถึง +2 จะพบว่า ข้อมูลมีความเหมาะสมกับโมเดลตามเกณฑ์ของไรท์และมาสเตอร์เป็นอย่างดี ยกเว้น OUTFIT ของนักเรียนบางคนซึ่งมีค่าเกิน +2 นอกนั้นก็ถือว่า เป็นไปตามเกณฑ์ของไรท์และมาสเตอร์ทุกประการ ส่วนความเชื่อมั่นองค์ประกอบของนักเรียน (Examinee Separation Reliability) เท่ากับ .67

เปรียบเทียบค่าเฉลี่ยของการตรวจให้คะแนนจากผู้ตรวจ 2 คน (Mean Rating From 2 Rater) ที่พิจารณาจากผลของคะแนนดิบ (Raw Scores) กับค่าคาดหวังของผู้ตรวจจากโมเดลของราส์ช (Expected Rater from Rasch Model) พบว่า ส่วนมากมีค่าใกล้เคียงกัน แสดงว่าการประเมินความสามารถของผู้เรียนโดยพิจารณาจากค่าเฉลี่ยของผู้ตรวจให้คะแนนที่เป็นผลของคะแนนดิบตามทฤษฎีมาตรฐานเดิม และค่าคาดหวังของผู้ตรวจจากโมเดลของราส์ชตามทฤษฎีการตอบข้อคำถาม ให้ผลที่ใกล้เคียงกัน

4. การแสดงหลักฐานความเชื่อมั่นของผลการตรวจให้คะแนน ความสอดคล้องของผู้ตรวจให้คะแนน และความเชื่อมั่นเชิงโครงสร้าง

ผู้วิจัยได้แบ่งการดำเนินการออกเป็น 3 ตอน ดังนี้

4.1 ผลการวิเคราะห์ความเชื่อมั่นของผลการตรวจให้คะแนน จากสูตร r_{F-R} ของ Feldt - Raju (บุญเชิด ภิญโญนนิตพงษ์. 2545 : 139) ผู้วิจัยนำผลการตรวจของผู้ตรวจแต่ละคน มาคำนวณหาความเชื่อมั่นของผลการตรวจให้คะแนนแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการทั้งฉบับ โดยจำแนกตามผลการตรวจของผู้ตรวจแต่ละคน แล้วจึงนำค่าความเชื่อมั่นของผู้ตรวจทุกคนมาหาความเชื่อมั่นรวมทั้งฉบับโดยใช้สูตรและวิธีการเดิม ผลการวิเคราะห์พบว่า ค่าความเชื่อมั่นของการประเมินทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการ คำนวณโดยใช้สูตร r_{F-R} ของ Feldt-Raju มีค่าอยู่ระหว่าง 0.6478 ถึง 0.7874 เมื่อนำผลการตรวจให้คะแนนของผู้ตรวจทุกคนมาหาความเชื่อมั่นรวมโดยใช้สูตร r_{F-R} ของ Feldt - Raju ตามขั้นตอนเดิม พบว่า มีค่าความเชื่อมั่นของผู้ตรวจทุกคน เท่ากับ 0.8921

4.2 ผลการวิเคราะห์ความสอดคล้องของผู้ตรวจให้คะแนน โดยใช้สัมประสิทธิ์แคปปา ผู้วิจัยได้นำเอาผลการตรวจให้คะแนนของผู้ตรวจทุกคนมาพิจารณาความเห็นที่สอดคล้องกันเป็นรายคู่ แล้วคำนวณค่าสัมประสิทธิ์ Kappa ของแต่ละคู่ พบว่า สัมประสิทธิ์แคปปาให้ค่าสูงสุดคือ คู่ของผู้ตรวจคนที่ 9 และคนที่ 8 เท่ากับ .88 รองลงมา คือ คู่ของผู้ตรวจคนที่ 8 และคนที่ 4 กับคู่ของผู้ตรวจคนที่ 9 และคนที่ 3 มีค่าเท่ากับ .87 และ .86 ตามลำดับ ส่วนสัมประสิทธิ์แคปปาที่ให้ค่าต่ำสุด คือ คู่ของผู้ตรวจคนที่ 13 และคนที่ 10 มีค่าเท่ากับ .63 ค่าสัมประสิทธิ์แคปปาเฉลี่ยทุกคู่เท่ากับ .7785 หมายถึง ความเห็นของผู้ตรวจทั้งสองคนมีความสอดคล้องกันมาก

4.3 ผลการวิเคราะห์ความเชื่อมั่นเชิงโครงสร้าง โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง Construct Reliability เท่ากับ .9055 และ Variance Extracted เท่ากับ .3469 ส่วนโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง Construct Reliability เท่ากับ .9631 และ Variance Extracted เท่ากับ .7519 ค่า Construct Reliability ทั้งสองโมเดลเกิน 0.70 ขึ้นไป ส่วนค่า Variance Extracted โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองเท่านั้นที่มีค่าเกิน 0.50 ส่วนอันดับหนึ่งถือว่า ไม่ผ่านเกณฑ์ (Hair; et al. 1995: 642)

อภิปรายผล

การแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของคะแนนการประเมินทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ เป็นการตรวจสอบหาความถูกต้อง แม่นยำ ของค่าวัดที่ใช้สำหรับพยากรณ์หรือลงความเห็นจากแบบวัด เป็นการตรวจสอบความสมเหตุสมผลของการแปลความหมายของคะแนน การตรวจสอบดังกล่าว เป็นการเปรียบเทียบค่าที่วัดได้จากแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการกับผู้วิจัยสร้างขึ้นกับตัวแปรที่ถือว่าเป็นตัวแปรเกณฑ์ เป้าหมายในการแสดงหลักฐานความเที่ยงตรงตามโครงสร้างเพื่อแปลความหมายของข้อมูลที่ได้จากการวัดว่า ข้อมูลที่ได้มานั้นเป็นตัวแทนที่ดีในการบรรยายคุณลักษณะทางด้านทักษะกระบวนการทางวิทยาศาสตร์ และพยากรณ์คุณลักษณะนั้น ๆ ได้ดีเพียงใด (บุญเชิด ภิญโญนนตพงษ์. 2545: 171-172) กรอนลันและลินน์ (Gronlund; & Linn. 1990: 48-49) กล่าวว่า เมื่อใช้คำว่าความเที่ยงตรงของในแง่ของการทดสอบและการประเมินจะต้องคำนึงถึงความเหมาะสมของการแปลความหมายจากผลการทดสอบหรือการประเมินเครื่องมือสำหรับกลุ่มใดกลุ่มหนึ่ง ไม่ได้ประเมินที่ตัวเครื่องมือ แต่เป็นความเที่ยงตรงของการแปลความหมายที่เกิดขึ้นจากผลการประเมิน ความเที่ยงตรงจึงมีลักษณะเฉพาะเจาะจงในการแปลความหมาย ซึ่งจะต้องพยายามแปลความหมายให้เกิดความเป็นเหตุเป็นผลให้มากที่สุด ความเป็นเหตุเป็นผลจะเชื่อถือได้หรือไม่ ขึ้นอยู่กับว่าได้พยายามใช้วิธีการทางวิทยาศาสตร์เข้ามาเป็นตัวช่วยในการแปลความหมายของคะแนนและสามารถอ้างอิงไปสู่จุดมุ่งหมายที่ต้องการได้มากน้อยเพียงใด (จินตนา ธนวิบูลย์ชัย. 2535: 64) การแสดงหลักฐานความเที่ยงตรงตามโครงสร้างเป็นการแสดงหลักฐานเพื่อตอบคำถามเกี่ยวกับความสามารถในการแปลความหมายของคุณลักษณะที่ต้องการวัด (Linn; & Gronlund. 1995: 50-51) จากแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ การทำความเข้าใจเกี่ยวกับหลักฐานความเที่ยงตรงตามโครงสร้างในครั้งนี้ มีความจำเป็นต้องระลึกไว้เสมอว่า ตัวแบบวัดไม่ใช่สิ่งที่ผู้วิจัยสนใจ แต่ความสนใจจะอยู่ที่แบบวัดได้สะท้อนปรากฏการณ์ในโลกของความเป็นจริงมากน้อยเพียงใด หลักฐานความเที่ยงตรงตามโครงสร้างต้องสะสมรายละเอียดจากหลายๆ แหล่งที่แตกต่างกัน เพื่อนำมารวมเป็นคำบรรยายพฤติกรรมได้อย่างกว้างขวาง คงทน และมีลักษณะความเป็นนามธรรม การกำหนดเกณฑ์เพียงเกณฑ์เดียวไม่เพียงพอที่จะอธิบายคุณลักษณะของโครงสร้างได้ ต้องอาศัยรายละเอียดที่มากกว่าหรือต้องอาศัยเกณฑ์หลายเกณฑ์ (บุญเชิด ภิญโญนนตพงษ์. 2545: 178)

จากประเด็นที่กล่าวมาแล้วข้างต้น ผู้วิจัยจึงดำเนินการสะสมรายละเอียดจากหลักฐานหลาย ๆ แหล่ง ทั้งจากทฤษฎีมาตรฐานเดิม (Classical Theory) และทฤษฎีการตอบข้อคำถาม (Item Response Theory) ทฤษฎีมาตรฐานเดิมใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน การวิเคราะห์ความเที่ยงตรงรวมและความเที่ยงตรงแยก สำหรับการวิเคราะห์องค์ประกอบเชิงยืนยันวิเคราะห์เปรียบเทียบความเหมาะสมของข้อมูลเชิงประจักษ์กับโมเดลองค์ประกอบเชิงยืนยัน

อันดับหนึ่ง พิจารณาถึงความสัมพันธ์ของข้อมูลในแต่ละตัวแปรว่า มีความเหมาะสมกับโมเดล การวิจัย ดัชนีความเหมาะสมสัมบูรณ์ (Absolute fit index) ดัชนีเหมาะสมเชิงเปรียบเทียบ (Comparative fit indices) ทุกตัวแปรอยู่ในเกณฑ์ที่ใช้พิจารณาความเหมาะสมของโมเดล และเมื่อเปรียบเทียบความเหมาะสมของข้อมูลเชิงประจักษ์กับโมเดลองค์ประกอบเชิงยืนยันอันดับสอง เพื่อค้นหาโมเดลทางเลือกในการอธิบายผลการประเมินทักษะกระบวนการทางวิทยาศาสตร์ ในมิติอื่น ๆ ที่ใช้เป็นฐานองค์ประกอบอันดับหนึ่ง 5 องค์ประกอบ โดยการพิจารณาจากค่าดัชนีความเหมาะสมของค่าพารามิเตอร์ในแต่ละทักษะ ค่าดัชนีความเหมาะสมของทุกตัวแปรบ่งชี้ว่า โมเดลการวิจัยมีความเหมาะสมกับข้อมูลเชิงประจักษ์ตามเกณฑ์ที่ใช้ในการพิจารณาความเหมาะสมของข้อมูล ทั้งโมเดลองค์ประกอบเชิงยืนยันอันดับหนึ่งและอันดับสอง สอดคล้องกับ นพวรรณ ศรีเกตุ (2550: บทคัดย่อ) หาคุณภาพของแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ ขั้นบูรณาการโดยการแสดงหลักฐานความเที่ยงตรงตามโครงสร้างภายในฉบับด้วยการวิเคราะห์ องค์ประกอบเชิงยืนยันอันดับหนึ่งและอันดับสอง โมเดลการวิจัยมีความเหมาะสมกับข้อมูลเชิงประจักษ์เช่นเดียวกัน ทั้งนี้เพราะโดยธรรมชาติของทักษะกระบวนการทางวิทยาศาสตร์จะมี ลักษณะเฉพาะตัวที่เป็นรูปธรรมมีความชัดเจนในตัวเอง มีลำดับขั้นของทักษะโดยเริ่มจากทักษะ ขั้นพื้นฐาน คือ การสังเกต การจำแนกประเภท การวัด การคำนวณ การหาความสัมพันธ์ ระหว่างสเปสและเวลา การลงความเห็นจากข้อมูล การจัดกระทำและสื่อความหมายข้อมูล การ พยากรณ์ ตามลำดับ ผู้มีทักษะทางวิทยาศาสตร์ ขั้นสูง เช่น การตั้งสมมติฐาน การกำหนดนิยาม เชิงปฏิบัติการ การกำหนดและควบคุมตัวแปร การทดลอง และการตีความหมายข้อมูลและ ลงข้อสรุป จะเกิดจากการสะสมทักษะขั้นต้น ๆ มาก่อน จากความชัดเจนและความเป็นโมเดลที่ มาตรฐานของกระบวนการทางวิทยาศาสตร์จึงทำให้มีความชัดเจนในการวัดแต่ละทักษะ ผู้ตรวจ ให้คะแนนมีความเป็นปรนัยในการตรวจให้คะแนน และเป็นไปในทิศทางเดียวกัน

การวิเคราะห์ความเที่ยงตรงรวมและความเที่ยงตรงแยก โดยอาศัยการพิจารณาจาก สัมประสิทธิ์สหสัมพันธ์ภายในทักษะและสัมประสิทธิ์สหสัมพันธ์ระหว่างทักษะ แล้วเปรียบเทียบ สัมประสิทธิ์สหสัมพันธ์ทั้งสองค่า ผลการวิเคราะห์พบว่า แนวโน้มของค่าสัมประสิทธิ์สหสัมพันธ์ ภายในทักษะมีค่าสูงกว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างทักษะอย่างเห็นได้ชัด สัมประสิทธิ์ สหสัมพันธ์ภายในทักษะที่มีค่าสูงจะเป็นหลักฐานแสดงถึงความเที่ยงตรงรวม (Convergent Validity) สัมประสิทธิ์สหสัมพันธ์ระหว่างทักษะที่มีค่าต่ำกว่าเป็นหลักฐานแสดงถึงความเที่ยงตรง แยก (Discriminant Validity) แนวคิดนี้สอดคล้องกับหลักการวัดตามทฤษฎีที่โครงสร้างคล้ายกัน จะมีความสัมพันธ์กันสูง ส่วนโครงสร้างที่แตกต่างกันควรจะมีค่าความสัมพันธ์กันต่ำ รูปแบบการ วิเคราะห์หลายลักษณะหลายวิธีที่ใช้ในการวิจัยครั้งนี้ ถือว่า เป็นรูปแบบการวิเคราะห์หลาย ลักษณะหลายวิธีที่ไม่สมบูรณ์ การพิจารณาหลักฐานในการแสดงความเที่ยงตรงตามโครงสร้าง จะเปรียบเทียบเฉพาะค่าสัมประสิทธิ์สหสัมพันธ์ภายในทักษะและค่าสัมประสิทธิ์สหสัมพันธ์ ระหว่างทักษะเท่านั้น ดังเช่นที่ Campbell และ Fiske กล่าวไว้อย่าง ชัดเจนว่าสามารถใช้

รูปแบบที่ไม่สมบูรณ์ได้ โดยจะเน้นหนักความสำคัญของรูปแบบจำลองของคุณลักษณะเดียวกันที่ใช้วิธีการแตกต่างกันในการประยุกต์ใช้กับเนื้อหาในการวิจัย

การวิเคราะห์ข้อมูลทั้งสองวิธีตามแนวทางของทฤษฎีมาตรฐานเดิม ให้ผลสอดคล้องกับวัตถุประสงค์และสมมติฐานของการวิจัยทุกประการ จึงเป็นหลักฐานพยานช่วยในการยืนยันผลการประเมินทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการที่มีคุณภาพด้านความเที่ยงตรงตามโครงสร้าง แสดงถึงประสิทธิภาพของการให้คะแนนที่มีความถูกต้อง แม่นยำ ในการแปลความหมายของคุณลักษณะที่ต้องการวัดและสามารถบอกถึงระดับของทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการของนักเรียนได้

ส่วนการวิเคราะห์โมเดลของราสช์แบบ 3 องค์ประกอบ (Three – Facets Rasch Model) ตามแนวทางของทฤษฎีการตอบข้อคำถาม ในแต่ละองค์ประกอบ (Facet) มีรายงานผลอยู่ 2 ส่วน คือ การรายงานค่าพารามิเตอร์ของแต่ละองค์ประกอบและการเปรียบเทียบความเหมาะสมของข้อมูลเชิงประจักษ์กับโมเดลของราสช์แบบ 3 องค์ประกอบ ค่าพารามิเตอร์ของแต่ละองค์ประกอบ เช่น ค่าความยากของทักษะกระบวนการทางวิทยาศาสตร์ (Task Difficulty) ในแต่ละทักษะอยู่ในช่วงระหว่าง -0.39 ถึง $+0.56$ ($SD = 0.33$, $Mean SE = 0.4$) ทักษะการตีความหมายข้อมูลและลงข้อสรุป และทักษะการกำหนดและควบคุมตัวแปร มีค่าความยากสูงสุดและรองลงมาตามลำดับ ทักษะการตั้งสมมติฐานมีค่าความยากต่ำสุด ถ้าพิจารณาจากขอบเขตของการยอมรับได้ของค่าความยากตามแนวทางของทฤษฎีการตอบข้อคำถามที่อยู่ในช่วงระหว่าง -2.5 ถึง $+2.5$ (ศิริชัย กาญจนวาสี. 2545: 48) จะพบว่า ทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการทั้ง 5 ทักษะ มีค่าความยากอยู่ในเกณฑ์ที่ยอมรับได้ตามแนวทางของทฤษฎีการตอบข้อคำถาม สามารถที่จะนำแบบวัดนี้ไปใช้ในการประเมินทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการของนักเรียนในระดับที่ใกล้เคียงกันได้เป็นอย่างดี ค่าความสามารถของผู้ตรวจ (Severity/Leniency) แต่ละคนอยู่ในช่วงระหว่าง -0.54 ถึง $+0.22$ ($SD = 0.17$, $Mean SE = 0.07$) ผู้ตรวจที่เคร่งครัด (Severity) และตั้งใจตรวจตามกฎเกณฑ์การให้คะแนนมีอยู่ 2 คน คือ ผู้ตรวจคนที่ 5 และ 12 นอกนั้นจัดอยู่ในกลุ่มที่ไม่ค่อยเคร่งครัด (Leniency) เท่าที่ควร แต่ก็ยังถือว่าผู้ตรวจแต่ละคนมีความสามารถใกล้เคียงกัน และไม่ได้แตกต่างกันมากนัก ความสามารถของผู้ตรวจให้คะแนนยังไม่ได้มีการกำหนดขอบเขตของการยอมรับได้เหมือนค่าพารามิเตอร์ขององค์ประกอบอื่น ๆ ผู้วิจัยจึงขอใช้เกณฑ์ร่วมกันกับเกณฑ์ความสามารถของผู้เข้าสอบคืออยู่ในช่วงระหว่าง -3 ถึง $+3$ (ศิริชัย กาญจนวาสี. 2545: 47) เป็นเกณฑ์ในการพิจารณา ดังนั้น จึงถือว่าผู้ตรวจให้คะแนนด้านทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการในการวิจัยครั้งนี้ มีความระดับสามารถอยู่ในเกณฑ์ที่ยอมรับได้ตามแนวทางของทฤษฎีการตอบข้อคำถาม สำหรับองค์ประกอบของนักเรียน ระดับความสามารถของนักเรียน (Ability) แต่ละคนอยู่ในช่วงระหว่าง -1.09 ถึง $+0.86$ ($SD = 0.56$, $Mean SE = 0.32$) ความสามารถของนักเรียน แต่ละคนมีการกระจายแตกต่างกันไป ความสามารถด้าน

ทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการสูงสุดมี 2 คน ต่ำที่สุดมี 5 คน ระดับความสามารถของนักเรียนทุกคนอยู่ในเกณฑ์ที่ยอมรับได้ คืออยู่ในช่วงระหว่าง -3 ถึง +3 (ศิริชัย กาญจนวาสี. 2545: 47)

การเปรียบเทียบความเหมาะสมของข้อมูลเชิงประจักษ์กับโมเดลของราส์ชแบบสามองค์ประกอบ โดยการใช้ Weighted mean square residual; INFIT และ Unweighted mean square residual; OUTFIT ในการทดสอบองค์ประกอบของทักษะกระบวนการทางวิทยาศาสตร์ ค่า INFIT ในแต่ละทักษะอยู่ในช่วงระหว่าง 0.6 ถึง 0.8 (Mean = 0.7 , SD = 0.1) และค่า OUTFIT ในแต่ละทักษะอยู่ในช่วงระหว่าง 0.7 ถึง 1.0 (Mean = 0.9 , SD = 0.1) องค์ประกอบของผู้ตรวจให้คะแนน ค่า INFIT ของผู้ตรวจอยู่ในช่วงระหว่าง 0.6 ถึง 0.8 (Mean = 0.7 , SD = 0.1) และค่า OUTFIT ของผู้ตรวจอยู่ในช่วงระหว่าง 0.7 ถึง 1.1 (Mean = 0.9 , SD = 0.1) ทั้งสององค์ประกอบมีค่าตามเกณฑ์ที่ยอมรับได้ของลินซ์, ไรท์และลินาเคอร์ (Engelhard. 1994: 97; citing Lunz; Wright; & Linacre. 1990) ที่ได้กำหนดขอบเขตของการยอมรับได้ของโมเดลที่เหมาะสมกับข้อมูลอยู่ระหว่าง .6 ถึง 1.5 ส่วนองค์ประกอบของนักเรียน ค่า INFIT อยู่ในช่วงระหว่าง 0.2 ถึง 2.6 (Mean = 0.8 , S.D. = 0.4) และค่า OUTFIT อยู่ในช่วงระหว่าง 0.2 ถึง 3.4 (Mean = 0.9 , S.D. = 0.5) มีนักเรียนบางคนที่มีค่า INFIT และ OUTFIT ไม่สอดคล้องกับเกณฑ์ของลินซ์, ไรท์และลินาเคอร์ แต่ถ้าพิจารณาในภาพรวม ค่าเฉลี่ย INFIT จะอยู่ที่ 0.8 และค่าเฉลี่ย OUTFIT จะอยู่ที่ 0.9 เป็นไปตามเกณฑ์ของลินซ์, ไรท์และลินาเคอร์ทุกประการ ด้วยเหตุผลดังกล่าว เฉพาะในส่วนขององค์ประกอบของนักเรียน ผู้วิจัยจึงขอใช้เกณฑ์ของไรท์และมาสเตอร์ (Wright; & Masters. 1982: 102) กำหนดขอบเขตของการยอมรับได้ของโมเดลที่เหมาะสมกับข้อมูลมีค่าอยู่ระหว่าง -2 ถึง +2 จะพบว่า ข้อมูลมีความสอดคล้องกับโมเดลตามเกณฑ์ของไรท์และมาสเตอร์เป็นอย่างดี ยกเว้น OUTFIT ของนักเรียนบางคนซึ่งมีค่าเกิน +2 นอกนั้นถือว่า เป็นไปตามเกณฑ์ของไรท์และมาสเตอร์ทุกประการ

ในการตรวจสอบองค์ประกอบของทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ ซึ่งจัดว่าเป็นตัวแปรแฝง ในโมเดลของราส์ชแบบสามองค์ประกอบยังมีสถิติอีกค่าหนึ่งที่เรียกว่า ดัชนีแยกแยะความเชื่อมั่น (Reliability of Separation Index) ซึ่งเป็นดัชนีที่ให้ค่าความเชื่อมั่นแยกกันไปในแต่ละองค์ประกอบ การวิจัยในครั้งนี้จึงมีค่าดัชนีแยกแยะความเชื่อมั่นอยู่สามค่าคือ ดัชนีความเชื่อมั่นของทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ มีค่าเท่ากับ .98 ดัชนีความเชื่อมั่นของผู้ตรวจให้คะแนน มีค่าเท่ากับ .83 และดัชนีความเชื่อมั่นของนักเรียน มีค่าเท่ากับ .67 ดัชนีความเชื่อมั่นของทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการมีค่าสูงสุด และดัชนีความเชื่อมั่นของนักเรียนมีค่าต่ำสุด

เปรียบเทียบค่าเฉลี่ยของการตรวจให้คะแนนจากผู้ตรวจ 2 คน (Mean Rating From 2 Rater) ที่พิจารณาจากผลของคะแนนดิบ (Raw Scores) กับค่าคาดหวังของผู้ตรวจจากโมเดลของราส์ช (Expected Rater from Rasch Model) พบว่า ส่วนมากมีค่าใกล้เคียงกัน แสดงว่า

การประเมินความสามารถของผู้เรียนโดยพิจารณาจากค่าเฉลี่ยของผู้ตรวจให้คะแนนที่เป็นผลของคะแนนดิบตามทฤษฎีมาตรฐานเดิม และค่าคาดหวังของผู้ตรวจจากโมเดลของราล์ชตามทฤษฎีการตอบข้อคำถาม ให้ผลที่ใกล้เคียงกัน

การวิเคราะห์โมเดลของราล์ชแบบสามองค์ประกอบตามทฤษฎีการตอบข้อคำถาม ให้ผลสอดคล้องกับวัตถุประสงค์ของการวิจัยทั้งสามองค์ประกอบ การวิเคราะห์ค่าพารามิเตอร์ของแต่ละองค์ประกอบ อยู่ในขอบเขตที่ยอมรับได้ตามแนวทางของทฤษฎีการตอบข้อคำถาม ส่วนผลการเปรียบเทียบความเหมาะสมกับข้อมูลเชิงประจักษ์ของโมเดลของราล์ชแบบสามองค์ประกอบ เฉพาะองค์ประกอบของทักษะกระบวนการและองค์ประกอบของผู้ตรวจเท่านั้นที่พบว่า โมเดลการวิจัยมีความเหมาะสมกับข้อมูลเชิงประจักษ์ ส่วนองค์ประกอบของนักเรียนพิจารณาที่ OUTFIT จะพบว่า มีนักเรียนจำนวนหนึ่งที่มีค่า OUTFIT ไม่เหมาะสมกับข้อมูลเชิงประจักษ์ แม้จะใช้เกณฑ์ของไรท์และมาสเตอร์ (Wright; & Masters. 1982: 102) แล้วก็ตาม ทั้งนี้เนื่องมาจากว่า ค่า OUTFIT เป็นสถิติที่มีความไว (Sensitivity) ต่อการระบุข้อคำถามที่ไม่สอดคล้องกับความน่าจะเป็น (Probability) ที่นักเรียนควรจะตอบ ถ้าเปรียบเทียบกับ INFIT ซึ่งเป็นค่าสถิติที่มีการถ่วงน้ำหนักจะพบว่า โมเดลการวิจัยมีความเหมาะสมกับข้อมูลเชิงประจักษ์มากขึ้น ในทางปฏิบัติ INFIT มักใช้เป็นดัชนี (Index) ตัวแรกในการบ่งชี้ข้อคำถามหรือผู้ตอบที่ไม่สอดคล้อง (Misfit) กับโมเดล ถ้าพิจารณาร่วมกับ OUTFIT ก็จะทำให้ทราบข้อมูลเกี่ยวกับความเหมาะสมกับโมเดลเพิ่มมากขึ้น จะนำไปสู่การตัดสินใจในการคัดเลือกข้อคำถามและผู้ตอบที่มีรูปแบบการตอบเหมาะสมกับโมเดลได้อย่างถูกต้องมากขึ้น (Wright; & Linacre. 1984: 4-21, 4-29) สรุปได้ว่า คะแนนการประเมินทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการมีความเที่ยงตรงตามโครงสร้างตามแนวทางของทฤษฎีการตอบข้อคำถามและให้ผลที่สอดคล้องกับทฤษฎีมาตรฐานเดิมทุกประการ

สำหรับค่าความเชื่อมั่นของการประเมินทักษะกระบวนการทางวิทยาศาสตร์ ผู้วิจัยได้ดำเนินการรายงานใน 3 ลักษณะ คือ ความเชื่อมั่นของผลการตรวจให้คะแนน ความสอดคล้องของผู้ตรวจให้คะแนน และความเชื่อมั่นเชิงโครงสร้าง ความเชื่อมั่นของผลการตรวจให้คะแนนคำนวณจากสูตร r_{F-R} ของ Feldt - Raju (บุญเชิด ภิญโญนนตพงษ์. 2545 : 139) จำแนกตามผู้ตรวจแต่ละคน แล้วจึงนำผลการตรวจให้คะแนนของผู้ตรวจทุกคนมาหาค่าความเชื่อมั่นรวมโดยใช้สูตร r_{F-R} ของ Feldt - Raju ตามขั้นตอนเดิม ได้ค่าความเชื่อมั่นทั้งฉบับเท่ากับ 0.8921 ค่าความเชื่อมั่นของการให้คะแนนทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการที่คำนวณโดยใช้สูตร r_{F-R} ของ Feldt-Raju จำแนกตามผู้ตรวจที่ตรวจให้คะแนนคนละกลุ่มกัน ทำให้ผลของการประมาณค่าความเชื่อมั่นมีความแตกต่างกันออกไป ซึ่งมีค่าอยู่ระหว่าง 0.6478 ถึง 0.7874 ทั้งนี้เนื่องมาจากว่า กลุ่มผู้เข้าสอบที่แยกให้ผู้ตรวจแต่ละคนตรวจให้คะแนน มีความสามารถที่ต่างกัน สูตร r_{F-R} ของ Feldt-Raju เป็นสัมประสิทธิ์ความเชื่อมั่นที่คำนวณผลจากความแปรปรวนของคะแนนส่วนย่อยและความแปรปรวนของคะแนนรวม ค่า

ความเชื่อมั่นจะมากหรือน้อยจึงขึ้นอยู่กับความแปรปรวนของกลุ่มผู้เข้าสอบ การกระจายคะแนนของแต่ละคนก็ส่งผลให้ความเชื่อมั่นเพิ่มขึ้นมากขึ้นหรือน้อยลงได้เช่นกัน ถ้าการกระจายของคะแนนมีค่าน้อย ความแปรปรวนของคะแนนน้อยลง ความเชื่อมั่นจึงต่ำลงด้วย กรอนลันด์ (Gronlund. 1976 : 117) กล่าวว่า ถ้าความสามารถของผู้เข้าสอบในกลุ่มแตกต่างกันมาก การกระจายของคะแนนจะมากทำให้ความเชื่อมั่นสูง ถ้าความสามารถของกลุ่มผู้เข้าสอบใกล้เคียงกันมาก การกระจายของคะแนนมีค่าน้อย ทำให้ความเชื่อมั่นต่ำ สอดคล้องกับ คันทิงแฮม (Cunningham. 1986 : 112-118) ที่พบว่า ความสามารถผู้เข้าสอบที่หลากหลายจะส่งผลให้ค่าความเชื่อมั่นสูงขึ้นเมื่อความแปรปรวนของคะแนนเพิ่มขึ้น การที่ให้ผู้ตรวจแยกกันตรวจให้คะแนนแต่ละกลุ่มที่มีการกระจายของคะแนนแตกต่างกัน จึงส่งผลให้ความแปรปรวนของแต่ละกลุ่มแตกต่างกัน ความเชื่อมั่นที่คำนวณได้ในแต่ละกลุ่มจึงมีค่าแตกต่างกันด้วย ทราูป (Traub. 1994: 38-39) กล่าวว่า ขนาดของสัมประสิทธิ์ความเชื่อมั่นขึ้นอยู่กับความแปรปรวนของแต่ละบุคคลบนคะแนนจริงและคะแนนความคลาดเคลื่อน ถ้าผู้เข้าสอบได้คะแนนต่ำ ความแปรปรวนของคะแนนจริงต่ำและความเชื่อมั่นจึงต่ำตามไปด้วย ความเชื่อมั่นเปลี่ยนไปเมื่อความแปรปรวนของกลุ่มตัวอย่างเปลี่ยนไป แต่เมื่อนำผลการตรวจให้คะแนนของผู้ตรวจทุกคนที่มีความแตกต่างกันมาคำนวณค่าความเชื่อมั่นรวมโดยใช้สูตร r_{F-R} ของ Feldt – Raju ปฏิบัติตามขั้นตอนเดิม ได้ความเชื่อมั่นรวมเท่ากับ 0.8921 จะพบว่า ความเชื่อมั่นรวมมีค่าสูงขึ้น เนื่องจาก คะแนนของแต่ละคนมีความแตกต่างกัน การกระจายของคะแนนจะมาก และความแปรปรวนของคะแนนเพิ่มขึ้น จึงทำให้ค่าความเชื่อมั่นรวมมีค่าสูงขึ้น จึงเป็นการยืนยันผลการให้คะแนนและยืนยันคุณภาพของ แบบวัดทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการทั้งฉบับ

ความสอดคล้องของผู้ตรวจให้คะแนน พิจารณาจากความเห็นที่สอดคล้องกันเป็นรายคู่ แล้วคำนวณความเชื่อมั่นแต่ละคู่ด้วยสัมประสิทธิ์แคปปา ได้ค่าความเชื่อมั่นอยู่ในช่วงระหว่าง .63 ถึง .88 เมื่อนำค่าความเชื่อมั่นทุกคู่มาเฉลี่ย ได้ค่าความเชื่อมั่นของทุกคู่เท่ากับ .7785 ความสอดคล้องของผู้ตรวจให้คะแนนที่คำนวณจากสูตรแคปปา ถือเป็นค่าสัมประสิทธิ์ความสอดคล้องที่ใช้กับมาตราวัดแบบนามบัญญัติ (Nominal Categories) การคำนวณค่าสัมประสิทธิ์แคปปาใช้สัดส่วนของผู้ตัดสินที่เห็นสอดคล้องและสัดส่วนที่คาดหวัง โคเฮน (Cohen. 1960) แนะนำว่า ค่าสัมประสิทธิ์แคปปาเป็นดัชนีที่เหมาะสมในการประเมินความสอดคล้องของผู้สังเกต คำตอบของการคำนวณค่าสัมประสิทธิ์แคปปาอยู่ที่ ผู้วิจัยต้องการวัดความสอดคล้องในการสังเกตว่ามีความแตกต่างอย่างมีนัยสำคัญจากความสอดคล้องที่คาดหวังหรือไม่ (Kvalseth; & Tarald. 1991: 95-101) ถ้ามีความสอดคล้องกันอย่างสมบูรณ์ระหว่างผู้สังเกต 2 คน แล้วสัมประสิทธิ์แคปปาจะมีค่าเท่ากับ 1 ถ้าสัมประสิทธิ์แคปปามีค่าเป็น 0 แสดงว่าความสอดคล้องของผู้สังเกตเท่ากับความสอดคล้องที่คาดหวัง และถ้าสัมประสิทธิ์แคปปามีค่าติดลบ แสดงว่าความสอดคล้องของผู้สังเกตมีค่าน้อยกว่าความสอดคล้องที่คาดหวัง

วัตถุประสงค์การวิเคราะห์สัมประสิทธิ์แคปปาในครั้งนี้ เพื่อต้องการดูความเชื่อมั่นของผู้ตรวจให้

คะแนนเป็นรายคู่ ในภาพรวมพบว่า แนวโน้มความเห็นของผู้ตรวจแต่ละคนมีความสอดคล้องกันในระดับมากถึงมากที่สุด ผลการคำนวณค่าเฉลี่ยของผู้ตรวจทุกคู่ มีความเห็นความสอดคล้องกันในระดับมาก ความคลาดเคลื่อนในการประเมินมีน้อย เลนและคณะ (Lane, S.; & others. 1969: 71-79) แนะนำว่า ความคลาดเคลื่อนเนื่องมาจากผู้ตรวจจะมีค่าน้อยมาก เมื่อมีการตรวจให้คะแนนแบบประเมินรวม (Holistic) และมีเกณฑ์การให้คะแนน ผู้วิจัยได้จัดประชุมชี้แจงผู้ตรวจให้คะแนนเพื่อซักซ้อมและทำความเข้าใจในการตรวจ ผู้ตรวจให้คะแนนทุกคนได้ศึกษาคู่มือการตรวจให้คะแนน กฎเกณฑ์การให้คะแนน (Scoring Rubric) ทักษะกระบวนการทางวิทยาศาสตร์ ขั้นบูรณาการที่ผู้วิจัยสร้างขึ้น จึงส่งผลให้ผลของการวิจัยออกมาในลักษณะน่าพึงพอใจ

ความเชื่อมั่นเชิงโครงสร้าง พิจารณาค่ามาตรฐานของน้ำหนักองค์ประกอบ (Factor Loading) และ ค่าความคลาดเคลื่อน (Error) ของตัวแปรสังเกตที่เป็นผลจากการวิเคราะห์ องค์ประกอบเชิงยืนยันอันดับหนึ่งและอันดับสองมาคำนวณค่า Construct Reliability และค่า Variance Extracted เพื่อทดสอบความเพียงพอของตัวแปรสังเกตที่จะใช้เป็นตัวแทนของตัวแปรแฝง โดยการประเมินความเพียงพอของค่าความเชื่อมั่นเชิงโครงสร้างใช้เกณฑ์ 0.70 ขึ้นไป ส่วนค่า variance extracted ใช้เกณฑ์ 0.50 ขึ้นไป ผลการวิเคราะห์พบว่า โมเดลการวิเคราะห์ องค์ประกอบเชิงยืนยันอันดับหนึ่ง Construct Reliability เท่ากับ .9055 และ Variance Extracted เท่ากับ .3469 ส่วนโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง Construct Reliability เท่ากับ .9361 และ Variance Extracted เท่ากับ .7519 ค่า Construct Reliability ทั้งสองโมเดลเกิน 0.70 ขึ้นไป ส่วนค่า Variance Extracted โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองเท่านั้นที่มีค่าเกิน 0.50 อันดับหนึ่งถือว่า ไม่ผ่านเกณฑ์ (Hair; et al. 1995: 642)

ผลของการเปรียบเทียบความเหมาะสมกับข้อมูลเชิงประจักษ์ของโมเดลการวิเคราะห์ องค์ประกอบเชิงยืนยันอันดับหนึ่งและอันดับสอง ผลของการวิเคราะห์ความเที่ยงตรงรวมและความเที่ยงตรงแยกตามแนวทางของทฤษฎีมาตรฐานเดิม และผลของการประมาณประมาณค่าพารามิเตอร์และเปรียบเทียบความเหมาะสมกับข้อมูลเชิงประจักษ์ของโมเดลของรัสชแบบสามองค์ประกอบตามแนวทางของทฤษฎีการตอบข้อคำถาม อีกทั้งผลจากการวิเคราะห์ความเชื่อมั่นของผลการตรวจให้คะแนน ความสอดคล้องระหว่างผู้ตรวจ และค่าความเชื่อมั่นเชิงโครงสร้าง ที่เป็นไปตามวัตถุประสงค์และสมมติฐานของการวิจัยทุกประการ ช่วยยืนยันการลงความเห็น (Inference) จากคะแนนที่วัดได้อย่างเหมาะสม และมีความหมาย ชนิดของหลักฐานส่วนใหญ่ก็ได้มาจากการเก็บรวบรวมข้อมูลหลายวิธี มีปริมาณเพียงพอที่จะยืนยันได้ว่า คะแนนของการประเมินจากแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ มีความเที่ยงตรงตามโครงสร้างและความเชื่อมั่นตามเกณฑ์คุณภาพที่กำหนด สามารถนำไปใช้ในการประเมินทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการของนักเรียนในกลุ่มที่ใกล้เคียงกันได้ อย่างถูกต้อง แม่นยำ และมีคุณภาพ

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 การแสดงหลักฐานความเที่ยงตรงตามโครงสร้างโดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันทั้งอันดับหนึ่งและอันดับสอง ในกรณีถ้าต้องการศึกษาสารสนเทศด้านคุณภาพของแบบวัด ก็ควรศึกษาเพิ่มเติมด้วยค่าสหสัมพันธ์และ R^2 ที่จะทำให้คุณภาพของแบบวัดเป็นรายข้อโดยเฉพาะถ้าต้องการดูคุณภาพของเครื่องมือ การศึกษาในครั้งนี้ไม่เน้นคุณภาพของแบบวัดแต่สนใจที่จะแปลความหมายของผลการตรวจให้คะแนนเท่านั้น

1.2 กรณีนำรูปแบบของการวิจัยไปใช้ในการประเมินผลภายในชั้นเรียน โมเดลของราล์ชแบบหลายองค์ประกอบให้สารสนเทศของนักเรียนเป็นรายบุคคล สามารถพิมพ์ชื่อนักเรียนในขั้นตอนของการวิเคราะห์ข้อมูลได้เลย จึงมีประโยชน์ในการนำไปใช้พัฒนา/ปรับปรุงการเรียนการสอนของนักเรียน ถือว่าเป็นการแก้ปัญหาได้ตรงจุด เมื่อครูผู้สอนได้ดำเนินการวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการของผู้เรียนและทราบผลของคำตอบเรียบร้อยแล้ว ควรนำผลไปดำเนินการปรับปรุง / แก้ไข และพัฒนาให้เกิดพฤติกรรมทักษะกระบวนการทางวิทยาศาสตร์ในตัวผู้เรียนต่อไป

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

ควรศึกษาโมเดลของราล์ชแบบหลายองค์ประกอบตามทฤษฎีการตอบข้อคำถามในด้านการวิเคราะห์ความลำเอียงของข้อสอบ การปรับเทียบคะแนนระหว่างข้อสอบ หรือการเพิ่มจำนวนฟาเซทที่ใช้ในการศึกษาให้มากขึ้น เป็นต้น

บรรณาธิการ

บรรณานุกรม

- กนกทิพย์ พัฒนาพัฑฒ์. (2529). การวิเคราะห์ข้อมูลเชิงปริมาณเพื่อการวิจัยการศึกษา. ภาควิชาประเมินผลและวิจัยการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยเชียงใหม่. กรมวิชาการ กระทรวงศึกษาธิการ. (2545). สารและมาตรฐานการเรียนรู้กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ในหลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2544. กรุงเทพฯ: กระทรวงฯ.
- จินตนา ธนวิบูลย์ชัย. (2535). ความเที่ยงตรง : ปัญหาและแนวคิดใหม่. วารสารวัดผล การศึกษา. 13(39): 63-90.
- จิรพร ปัทมะสุคนธ์. (2543). การศึกษาความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบวัด บุคลิกภาพความเป็นทหารของนักเรียนทหาร. ปรินิพนธ์ กศ.ม. (การวัดผล การศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- จำนง พรายแถมแซ. (2529). เทคนิคการวัดและประเมินผลการเรียนรู้กับการสอนซ่อมเสริม (ตามกระบวนการทางวิทยาศาสตร์). กรุงเทพฯ: ไทยวัฒนาพานิช.
- ดวงใจ บุญยะภาส. (2543). การศึกษาความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบวัด บุคลิกภาพตามแนวทฤษฎีบุคลิกภาพของไอแซกส์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 4 ในจังหวัดราชบุรี.. ปรินิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- ดุขฎิ โยเหลา. (มปป). เอกสารประกอบการอบรม การแปลผลการวิเคราะห์ข้อมูลด้วย คอมพิวเตอร์ เรื่อง: การวิเคราะห์โมเดลเชิงสาเหตุด้วยโปรแกรม LISREL. สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- นงลักษณ์ วิรัชชัย. (2540). ความไม่แปรเปลี่ยนของแบบจำลองการเป็นสมาชิกด้วยใจรักของ ครูระหว่างบุคลากรครู 2 กลุ่ม: การประยุกต์ใช้การสร้างแบบจำลองสมการโครงสร้าง ชนิดกลุ่มพหุ. กรุงเทพฯ: ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- _____. (2542). โมเดลลิสเรล สถิติวิเคราะห์สำหรับการวิจัย. กรุงเทพฯ: ศูนย์หนังสือแห่ง จุฬาลงกรณ์มหาวิทยาลัย.
- นิคม ทาแดง และสุจินต์ วิศวรธีรานนท์. (2529). เอกสารการสอนชุดวิชาวิทยาศาสตร์ 3 หน่วยที่ 1. พิมพ์ครั้งที่ 2. กรุงเทพฯ : นกัการพิมพ์.
- นพวรรณ ศรีเกตุ. (2550). การแสดงหลักฐานความเที่ยงตรงและความเชื่อมั่นของแบบวัดทักษะ กระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ. ปรินิพนธ์ กศ.ม. (การวัดผล การศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร

- บุญเชิด ภิญโญนนตพงษ์. (2544). การประเมินการเรียนรู้ที่เน้นผู้เรียนเป็นสำคัญ : แนวคิดและวิธีการ. กรุงเทพฯ : สำนักงานคณะกรรมการการศึกษาแห่งชาติ.
- _____. (2545). "คุณภาพของเครื่องมือวัด," ใน ประมวลสาระชุดวิชา การพัฒนาเครื่องมือสำหรับการประเมินการศึกษา หน่วยที่ 3. กรุงเทพฯ : มหาวิทยาลัยสุโขทัยธรรมาธิราช.
- บุญเรียง ขจรศิลป์. (2542). สถิติวิจัย 1. (พิมพ์ครั้งที่ 7). กรุงเทพฯ: ศูนย์หนังสือแห่งจุฬาลงกรณ์มหาวิทยาลัย.
- พรทิพย์ ไชยโส. (2536). CONSTRUCT VALIDITY: หลักการและแนวทางการวิเคราะห์ด้วย CONFIRMATORY FACTOR ANALYSIS. วารสารพัฒนาบริหารศาสตร์. 33(1): 1-16.
- เพ็ญนภา กลุณภาดล. (2547). การศึกษาและการพัฒนา การปฏิบัติหน้าที่ของสมาชิกครอบครัวในครอบครัวของผู้ป่วยโรคซึมเศร้า. ปรินญาณพนธ์ กศ.ด. (จิตวิทยาการให้คำปรึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- ผจงจิต อินทสุวรรณ. (2525). "Latent Trait Theory." วารสารวัดผลการศึกษา. 3(3): 51-69. มหาวิทยาลัย, ทบวง. คณะอนุกรรมการพัฒนาการสอนและผลดีวัสดุอุปกรณ์วิทยาศาสตร์. (2525). ชุดการเรียนการสอนสำหรับครูวิทยาศาสตร์ เล่ม 1-3. กรุงเทพฯ : ทบวงมหาวิทยาลัย.
- ยุพา ตันติเจริญ. (2531). "การเรียนการสอนวิทยาศาสตร์ในระดับประถมศึกษาและมัธยมศึกษาในชนบท." วารสารคณะกรรมการการศึกษาแห่งชาติว่าด้วยการศึกษาสหประชาชาติ. 20: 41-42.
- ส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี, สถาบัน. (2518). รายงานการสร้างแบบทดสอบทักษะกระบวนการทางวิทยาศาสตร์. กรุงเทพฯ : หน่วยทดสอบและประเมินผล, สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี.
- _____. (2519). เอกสารประกอบการสอนวิชาวิทยาศาสตร์กายภาพ. กรุงเทพฯ : สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี.
- _____. (2524). ทักษะกระบวนการทางวิทยาศาสตร์. กรุงเทพฯ : สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี.
- ศิริชัย กาญจนวาสี. (2545). ทฤษฎีการทดสอบแนวใหม่. กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- วัฒนชัย จันทร์วิบูล. (2538). การสร้างแบบทดสอบภาคปฏิบัติวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นพื้นฐาน. ปรินญาณพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

- วรรณทิพา รอดแรงคำ และพิมพ์พันธ์ เตชะคุปต์. (2532). *ทักษะกระบวนการทางวิทยาศาสตร์ สำหรับครู*. กรุงเทพฯ : เดอะมาสเตอร์กรุ๊ปแมเนจเม้นท์.
- อนันต์ จันทร์गी. (2525). *ผลการใช้คำถามของครูที่มีต่อทักษะกระบวนการทางวิทยาศาสตร์ ผลสัมฤทธิ์และทัศนคติของนักเรียน ม.ศ. 2 และ ม.2. ปรินญานิพนธ์ กศ.ด. (การวิจัย และพัฒนาหลักสูตร)* กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- อัญชลี ศรีกลชาญ. (2546). *การพัฒนาแบบวัดพหุปัญญาสำหรับนักเรียนชั้นมัธยมศึกษา ตอนต้น*. วิทยานิพนธ์ วท.ม. (เทคโนโลยีการวัดผลทางการศึกษา) ชลบุรี : บัณฑิตวิทยาลัย มหาวิทยาลัยบูรพา. ถ่ายเอกสาร.
- อรอุมา สงวนญาติ. (2544). *การศึกษาความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบวัด ความภาคภูมิใจในตนเองที่มีรูปแบบการตอบและวิธีการวิเคราะห์ต่างกัน*. ปรินญานิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- Adelman, L.; & Riedel, S.L. (1993). *Using the multitrait-multimethod matrix to evaluate knowledge-based systems*. Retrieved on April 10, 2002 from <http://www.bacon.gmu.edu/c3i/PubAbstract/CommdSupport.html>.
- American Association for The Advancement of Science. (AAAS.) (1970). *Science A Process Approach*. New York : Comantary for Teacher. AAAS. Xerox.
- Anderson, C.R. (1978, August). The Effectiveness of A Simulation Learning Game in Teaching Consumer Gredit to Senior Approach to Instruction. *Dissertation Abstract International*. 31(2) : 127-132.
- Andrich, David. (1978b, December). A Rating Formulation for Ordered Response Categories. *Psychometrika*. 43(4): 561-574.
- Bargaraneli, C.; & Caprara, G.V. (2000). *Measuring the big five in self-report and other ratings: a multitrait-mutimethod study*. Retrieved on April 10, 2002 from <http://www.ocs.mq.edu.au/~abaillie/node35.html>
- Bentler, P.M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin*. 107(2): 238-246.
- Biederman, J.L. (1998). *The dimensionality of work-related needs and work reinforcers*. Retrieved on April 10, 2002 from http://www.lib.ncsu.edu/etd/public/etd-5718123109823060/ed_title.html.

- Bollen, Kenneth A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley&Sons.
- Byrne, Barbara M. (1998). *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming*. New Jersey: Lawrence Erlbaum.
- Campbell, D.T.; & Fiske, D.W. (1959). Convergent and Discriminant Validation by the Multitrait-multimethod matrix. *Psychological Bulletin*. 56. 81-10555.
- Cheung, G. W.; & Rensvold R. B. (2002). Evaluation Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling*. 9(2): 233-255.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 20: 36-37.
- Cunningham, George K. (1986). *Education and Psychological Measurement*. New York: Macmillian.
- Duren, Redeney L. (1978, January - March). Measuring the Process of Science Objectives. *Science Education*. 62(1) : 24-25.
- Engelhard, G. (1994). Examining rater errors in the assessment of written composition with a many-faceted Rasch model. *Journal of Education Measurement*. 31(2): 93-112.
- Garland, Nell A.C. and Others. (1973). *Elementary Science Learning by Investigation*. 2nd ed. New York Rand McNally.
- Gay, L.R. (1992). *Educational Evaluation and Measurement Competencies for Analysis and Application*. 2nd ed. New York : Merrill.
- Gronlund, Norman E. (1976). *Measurement and Evaluation in Testing*. 3rd ed. New York: Macmillian.
- Groulund, Norman E.; & Linn, Robert L. (1990). *Measurement and Evaluation in Teaching*. 6th edition. New York: Macmillan.
- Hair, JR Joseph F.; Anderson, Rolph E.; Tatham, Ronald L.; & Black, William C. (1995). *Multivariate Data Analysis with Readings*. 4th edition. New Jersey: Prentice Hall.
- Hambleton. Ronald K. (1989). Principles and Selected Applications of Item Response Theory. In *Educational Measurement*. Edited by Linn, Robert L. p. 147-200. London: Callier Macmillian Publishers.

- Hambleton, Ronald K.; & H. Swaminathan. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer-Nijhoff Publishing.
- Hopkins, C. D.; & Antes, R. L. (1990) *Classroom Measurement and Evaluation*. 3rd ed. Itasca, IL: F. E. Peacock.
- Hu, L. T.; & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure Analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*. 6: 1-55.
- Jacobson, Willard J.; & Bergman Abby Barry. (1980). *Science for Children*. Englewood Cliffs, New Jersey : Prentice-Hall.
- Jöreskog, Karl G. (1994). On The Estimation of Polychoric Correlations and Their Asymptotic Covariance Matrix. *Psychometrika*. 59(3): 381-389.
- Jöreskog, Karl G.; & Sörbom, Dag. (1993). *LISREL 8: Structural Equation Modeling with The SIMPLIS Command Language*. Chicago: Scientific Software International.
- _____. (1996). *LISREL 8: User's Reference Guide*. Chicago: Scientific Software International.
- Kelloway, E. Kenvin. (1998). *Using LISREL for Structural Equation Modeling A Researcher's Guide*. Thousand Oaks, CA: Sage.
- Klopfer, L.E. (1971, Summer). Evaluation of Learning in Science. *An Handbook on Formative and Summative Evaluation of Student Learning*. New York : McGraw-Hill Book Co.
- Kvalseth.; & Tarald O. (1991). A Coefficient of Agreement for Nominal Scales: An Asymmetric Version of Kappa. *Education and Psychological Measurement*. New York: Macmillian.
- Lane, S.; & others. (1996, Spring). Generalizability and Validity of a Mathematics Performance Assessment. *Journal of Educational Measurment*. 33(1): 71-91.
- Lievens, F. (1997). The construct validity of assessment center ratings: results from an assessment center simulation. Retrieved on April 10, 2002 from http://www.ulb.ac.be/bps/docs/activities/bps97/porg/BPS97_P56.html.
- Linacre, J.M. (1989). *Many-faceted Rasch Measurement*. Chicago: MESA.
- Linn, Robert L.; & Groulund, Norman E. (1995). *Measurement and Assessment in Teaching*. (7th edition). New Jersey: Prentice Hall.

- Long, J.Scott. (1983). *Confirmatory Factor Analysis*. Beverly Hills, California : Sage Publications, Inc.
- _____. (1994). Confirmatory Factor Analysis: A Preface to LISREL. In: Lewis-Beck, Michael S., editors. *Factor Analysis and Related Techniques*. Singapore: Sage.
- Lord, Federic M. (1980). *Application of Item Response Theory to Practical Testing Problems*. New Jersey: Lawrence Erlbaum Associates Publishers.
- Lord, Federic M.; & Mavin R. Novick. (1968). *Statistical Theory of Mental Test Scores*. Massachusetts: Addison – Wesley Publishing Company Inc.
- Lunstrom, Donald ; & Lawrence Lowery. (1978). Process pattern and Structural Themes in Science. *In Inquiry Technique for Teaching Science*. William D. Romey. Englewood Cliffs, N.J. : Prentice-Hall.
- MacCallum, R. C.; & Austin, J. T. (2000). Applications of Structural Equation Modeling In Psychological Research. *Annual Review of Psychology*. 51: 201-226.
- Masters, Geoffery N. (1980, July). A Rasch Model for Rating Scales. *Dissertation Abstracts International*. 41(1): 215-A-215-A.
- Marsh, H. W.; Balla, J. R.; & McDonald, R. P. (1988). Goodness-of-Fit Indexes in Confirmatory Factor Analysis: The Effect of Sample Size. *Psychological Bulletin*. 103(3): 391-410.
- Marsh, H. W.; Hau. Kit-Tai. (1996). Assessing Goodness of Fit: Is Patrimony Always Desirable?. *Journal of Experimental Education*. 64(4): 364-391.
- Marsh, H. W.; Hau. Kit-Tai.; Chung, Choi-Man.; & Sui, T.L.P. (1998). Confirmatory Factor Analyses of Chinese Students' Evaluations of University Teaching. *Structural Equation Modeling*. 5(2): 143-164.
- Marsh, H. W.; & Hocevar, D. (1985). Applications of Confirmatory Factor Analysis to the Study of Self-Concept: First- and Higher Order Factor Models and Their Invariance Across Groups. *Psychological Bulletin*. 97(3): 562-582.
- Marsh, H. W. (1991). Multidimensional Student's Evaluations of Testing Effectiveness: A Test of Alternative Higher-Order Structures. *Journal of Educational Psychology*. 83(2): 285-296.
- Maruyama, Geoffrey M. (1998). *Basics of Structural Equation Modeling*. Thousand Oaks, CA: Sage.

- McDonald, Roderick P. (1999). *Test Theory: A Unified Treatment*. New Jersey: Lawrence Erlbaum.
- Messick, Samuel. (1989). Validity. In: Linn, Robert L.; editors. *Educational Measurement*. New York: The American Council on Education and Macmillan Publishing Company.
- Muraki, E., Hormbo. C.M.; & Lee, Yong-Won. (2000, December). Equating and Linking of Performance Assessments. In *Applied Psychological Measurement*, 24(4): 325-337.
- Okey, James R.; & Ronald L. Field. (1973). *Basic Process Skills Program*. Bloomington : Indiana University.
- Peterson, K.D. (1978, March). Scientific Inquiry Training for high School Student. *Journal of Research_In Science Teaching*. 15(3) : 153.
- Popham, W. James. (1990). *Modern Educational Measurement; A Practitioner's Perspective*. Prentice Hall Ince. Englewood Cliffs, New Jersey, U.S.A.
- Samejima, Fumiko. (1969). Amount of Information.; Estimation of Mental Ability Using a Response Pattern of Graded Scores. *Psychometika Monograph Supplement*. 34(4). Part 2: 37-46.
- Showalter, V.M. (1974, March). What Is Unified Science Education 7 (Part 6) : Programs Objectives And Scientific Literacy. *Prism II*, 2(4) : 1-8.
- Spector, Paul E. (2001). Research Methods in Industrial and Organizational Psychology: Data Collection and Data Analysis with Special Consideration to International Issues. In: Anderson, Neil.; Ones, Deniz S.;Sinangil, Handan K.; & Viswesvaran, Chockalingam., editors. *Handbook of Industial, Work & Organizational Psychology Volume 1 Personel Psychology*. London: Sage.
- Sund, Robert B.; & A.J. Picard. (1972). *Behavioral Objectives and Evaluational Measure : Science And Mathematics*. Ohio, Charlee E. Merrill.
- Traub, Ross E. (1994). *Reliability for the Social Sciences : Theory and Applications*. California: SAGE Publications, Inc.
- Trochim, W.M.K. (1999). *Idea of Construct Validity*. Retrieved on April 10. 2002 from <http://trochim.human.cornell.edu/kb/considea.htm>.
- Van de Vijver, J.R. Fons. (1986, March). The Robustness of Rasch Estimated. *Applied Psychological Measurement*. 10(1): 45-57.

- Walsh, Bruce.; & Betz, Nancy. (1990). *Test and Assessment*. 2nd ed. Prentice Hall Ince. Englewood Cliffs, New Jersey, U.S.A.
- Ward, Annie W. (1999). *Assessment in Classroom*. Wadsworth Publishing Company.
- William Wiersma.; & Stehpen G. Jurs. (1985). *Education Measurement and Testing*. Allyn and Bacon, Inc.
- Wright, Benjamin. D.; & John M. Linacre. (1984). MICROSCALE Manual. *Rasch Model Computer Program*: MediAx Interactive Technologies.
- Wright, Benjamin. D.; & Mark H. Stone. (1979). *Best Test Design*. Chicago, Illinois: MESA Press.
- Wright, B.D.; & Master, G.N. (1982). Rating scale anlysis. Chicago, IL: Mesa Press.
- Zeller, R.A. (1988). Validity. *Educational Research, Methodology and Measurement and International Handbook*. John P. Keeves, Editor. Pergamon Press. Australia.

คำอธิบาย:

ตารางนี้มีไว้สำหรับการพิมพ์รายงาน
ภาษาอังกฤษ ห้ามลบแถว!!! ของตารางนี้
มีเซนนั้นจะเกิด Error ในการเลือกคำสั่งจาก
เมนู Biblio แต่หลังจากที่ได้เลือกการ
จากเมนู Biblio ไปแล้วอย่างน้อย 1 ครั้ง จะลบ
แถวนี้ทิ้งก็ได้

ภาคผนวก

คู่มือการตรวจให้คะแนน แบบวัดทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการ

คำชี้แจง

1. คู่มือการตรวจให้คะแนนแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการ ประกอบด้วย 2 องค์ประกอบด้วยกัน คือ
 - 1.1 แบบวัดทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการ
 - 1.2 แนวทางในการตอบและกฎเกณฑ์การให้คะแนน (Scoring Rubric)
2. การตรวจให้คะแนนแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการ กำหนดให้ตรวจสอบความถูกต้องของคำตอบในแต่ละข้อโดยใช้แนวทางในการตอบประกอบการพิจารณาให้คะแนนและนำผลการตรวจมาจัดระดับคุณภาพของผู้เข้าสอบแยกเป็นรายทักษะโดยใช้กฎเกณฑ์การให้คะแนน (Scoring Rubric) ประกอบการพิจารณา กำหนดระดับคุณภาพเป็น 4 ระดับ ทุกทักษะกระบวนการย่อย

1.1 แบบวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการ

คำชี้แจง

1. แบบวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการฉบับนี้ มีคำถามทั้งหมด 5 ข้อใหญ่ ใช้เวลาในการทำแบบทดสอบ 120 นาที ห้ามขีดเขียนข้อความใด ๆ ลงในแบบวัดโดยเด็ดขาด
2. แบบวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการฉบับนี้ ประกอบด้วยทักษะกระบวนการย่อยจำนวน 5 ทักษะ คือ ทักษะการตั้งสมมติฐาน ทักษะการกำหนดนิยามเชิงปฏิบัติการ ทักษะการกำหนดและควบคุมตัวแปร ทักษะการทดลอง และทักษะการตีความหมายข้อมูลและลงข้อสรุป
3. แบบวัดทุกข้อเป็นแบบอัตนัย โดยแต่ละข้อกำหนดโจทย์หรือสถานการณ์ให้นักเรียนได้ศึกษาก่อน แล้วจึงเขียนตอบลงในกระดาษคำตอบที่ได้จัดเตรียมไว้ในแต่ละข้อ ดังตัวอย่างข้อ (0)

ตัวอย่างแบบทดสอบ

ข้อ (0) สมมติว่านายวิจิตรเคยรับประทานปูนามาแล้ว 6 ครั้ง ผลปรากฏว่า หลังจากรับประทานปูนาแล้วทุกครั้งเขาจะมีอาการท้องเดิน ถ้าต้องการทดลองว่านายวิจิตรรับประทานปูแล้วจะเกิดผลอย่างไร ควรตั้งสมมติฐานของการทดลองในครั้งนี้อย่างไร

.....

ฯลฯ

4. ควรศึกษาข้อความในแต่ละสถานการณ์ให้เข้าใจเสียก่อนแล้วจึงลงมือตอบ
5. ถ้าพบข้อยากอย่าเพิ่งท้อใจ เว้นข้ามไปทำข้ออื่น ๆ ต่อไปก่อนเมื่อมีเวลาเหลือจึงย้อนกลับมาทำใหม่
6. อย่าเปิดแบบทดสอบจนกว่าจะได้รับสัญญาณให้ลงมือทำ
7. เมื่อได้รับสัญญาณเตือนหมดเวลาในการทำข้อสอบ ให้นักเรียนหยุดการทำข้อสอบในทันที

ข้อ 1 ทักษะการตั้งสมมติฐาน

1.1) จากการทดลองละลายสาร A ในน้ำเย็นปรากฏว่า สาร A 21 กรัม ละลายได้พอดีในน้ำ 250 ลูกบาศก์เซนติเมตร จากการทดลองนี้ ถ้าผู้ทดลองมีความเชื่อว่า การอุ่นสารละลายให้ร้อนขึ้นจะช่วยให้สาร A ละลายได้เพิ่มขึ้นอีก ผู้ทดลองควรตั้งสมมติฐานว่าอย่างไร

.....

.....

.....

1.2) นำลูกกลมตันที่ทำด้วยวัตถุต่างชนิดกัน ได้แก่ วัตถุ A , B และวัตถุ C ซึ่งมีเส้นผ่าศูนย์กลาง เท่ากันมา 3 ลูก ปล่อยให้กลิ้งบนพื้นเอียงเดียวกัน แล้วจับเวลาที่ลูกกลิ้งแต่ละชนิดใช้ในการกลิ้งจากปลายบนสุดถึงปลายล่างสุดของพื้นเอียง ทำซ้ำแบบเดิม 3 ครั้ง หาค่าเฉลี่ยและนำมาเปรียบเทียบกัน กรณีดังกล่าว ถ้านักเรียนเป็นผู้ทดลองจะตั้งสมมติฐานอย่างไร

.....

.....

.....

1.3) จงศึกษาข้อมูลจากตารางต่อไปนี้

ชนิดของสาร	ความสามารถในการละลายของสารที่อุณหภูมิห้อง (กรัม)	
	ในน้ำ 100 ลบ.ซม.	ในเอซิลแอลกอฮอล์ 100 ลบ. ซม.
A	140	5
B	39	1
C	200	0.6
D	36	0
E	6	0

ข้อมูลในการทดลองจากตารางนี้ ควรตั้งสมมติฐานอย่างไร

.....

.....

.....

ข้อ 2 ทักษะการกำหนดนิยามเชิงปฏิบัติการ

2.1) คำว่า “น้ำใส” จะกำหนดนิยามอย่างไร ให้สามารถใช้เป็นแนวในการสังเกตและวัดได้ตรงกัน เมื่อต้องการจำแนกน้ำใสออกจากน้ำที่มีลักษณะอื่น ๆ

.....

.....

.....

2.2) ในการศึกษาเรื่อง นก จากประโยคที่ว่า “นกเป็นสัตว์ปีก ส่วนมากบินได้” คำว่า “บินได้” ของนกควรนิยามอย่างไร จึงจะสังเกตและวัดได้ตรงกัน

.....

.....

.....

2.3) คำว่า “ก๊าซคาร์บอนไดออกไซด์” ควรนิยามอย่างไร จึงจะสังเกตและวัดได้ตรงกัน เมื่อต้องการศึกษาปฏิกิริยาของก๊าซคาร์บอนไดออกไซด์

.....

.....

.....

ข้อ 3 ทักษะการกำหนดและควบคุมตัวแปร

3.1) ชนจะได้ทำการทดลองโดยใส่คัลเชื่อมคลอไรด์ 1 ช้อนเบอร์ 1 เท่ากัน ลงในปีกเกอร์ใบเล็กขนาดเดียวกัน 4 ใบ ซึ่งแต่ละใบบรรจุน้ำที่มีปริมาณแตกต่างกันและคนให้ทั่ว หลังจากใส่คัลเชื่อมคลอไรด์ละลายหมด เขาก็วัดอุณหภูมิของน้ำภายในปีกเกอร์แต่ละใบ จากข้อความดังกล่าวจงระบุ ตัวแปรต้น ตัวแปรตาม และตัวแปรที่ต้องควบคุม

ตัวแปรต้น

ตัวแปรตาม

ตัวแปรที่ต้องควบคุม

.....

.....

.....

3.2) เพื่อศึกษาว่าปริมาณของวิตามินเอที่หนูได้รับมีผลต่อน้ำหนักของหนูหรือไม่ นักวิทยาศาสตร์ได้ทดลองเลี้ยงหนู โดยแบ่งหนูออกเป็น 5 กลุ่ม แต่ละกลุ่มจะได้รับอาหารที่เหมือนกัน แต่ได้รับปริมาณของวิตามินที่แตกต่างกัน หลังจากเลี้ยงหนูไว้ 3 สัปดาห์ เขาก็ชั่งน้ำหนักของหนูแต่ละตัว จากข้อความ ดังกล่าวจงระบุ ตัวแปรต้น ตัวแปรตาม และตัวแปรที่ต้องควบคุม

ตัวแปรต้น

ตัวแปรตาม

ตัวแปรที่ต้องควบคุม

.....

.....

.....

.....

3.3) แม่เหล็กไฟฟ้าจะดูดจำนวนตะปูได้มากขึ้นหรือไม่ ถ้าแม่เหล็กไฟฟ้านั้นมีจำนวนแบตเตอรี่เพิ่มขึ้น จากข้อความดังกล่าวจงระบุ ตัวแปรต้น ตัวแปรตาม และตัวแปรที่ต้องควบคุม

ตัวแปรต้น

ตัวแปรตาม

ตัวแปรที่ต้องควบคุม

.....

.....

.....

.....

3.4) การงอกของเมล็ดข้าวโพดในเวลาที่ต่างกันขึ้นอยู่กับปริมาณของน้ำที่เมล็ดข้าวโพดได้รับ จากข้อความดังกล่าวจงระบุ ตัวแปรต้น ตัวแปรตาม และตัวแปรที่ต้องควบคุม

ตัวแปรต้น

ตัวแปรตาม

ตัวแปรที่ต้องควบคุม

.....

.....

.....

.....

ข้อ 4 ทักษะการทดลอง

ชาวสวนคนหนึ่งมีที่ดินอยู่หลายแปลง เขาสังเกตว่าที่ดินแต่ละแปลงที่ใช้ในการปลูกพืชต้องรดน้ำในปริมาณที่ต่างกัน เขาเกิดความสงสัยว่า ดินต่างชนิดกันจะอุ้มน้ำไว้ในปริมาณที่ต่างกันหรือไม่ จึงได้ตั้งสมมติฐานว่า “ชนิดของดินมีผลต่อปริมาณของน้ำที่อุ้มไว้”

4.1) ถ้านักเรียนเป็นชาวสวนคนนั้นจะมีวิธีออกแบบการทดลองอย่างไร

.....

.....

.....

4.2) จะใช้วัสดุอุปกรณ์ประกอบการทดลองอะไรบ้าง

.....

.....

.....

4.3) รายละเอียดขั้นตอนการดำเนินการทดลองเป็นอย่างไร

.....

.....

.....

4.4) การบันทึกผลการทดลองควรออกแบบอย่างไร

.....

.....

.....

ข้อ 5 ทักษะการตีความหมายข้อมูลและลงข้อสรุป

5.1) จงศึกษาข้อมูลในตารางที่แสดงปริมาณการจับปลาทะเลของเรือประมงลำหนึ่ง โดยใช้เครื่องมืออวนลากในเวลา 1 ชั่วโมง ระหว่างปี พ.ศ. 2508 – 2520 แล้วตอบคำถามลงใน ข้อ 5.1.1 – 5.1.2

พ.ศ.	จำนวนปลาที่เรือจับได้ (kg)
2508	232
2509	132
2510	115
2511	106
2512	103
2513	97
2514	66
2515	63
2516	52
2517	58
2518	47
2519	57
2520	47

5.1.1) ตั้งแต่ปี พ.ศ. 2508 – 2520 จำนวนปลาที่จับได้ในแต่ละชั่วโมงเปลี่ยนแปลงอย่างไร

.....

.....

.....

5.1.2) จากข้อมูลข้างต้นลงข้อสรุปได้ว่าอย่างไร

.....

.....

.....

5.2) จากการทดลองละลายสาร A ในของเหลว B จำนวน 50 ลูกบาศก์เซนติเมตร ที่อุณหภูมิต่าง ๆ กัน ได้ผลการทดลองดังตาราง ให้ใช้เป็นข้อมูลตอบคำถามข้อ 5.2.1 – 5.2.2

อุณหภูมิของของเหลว B ($^{\circ}\text{C}$)	ปริมาณของสาร A ที่ละลายในของเหลว B (g)
20	5
30	10
40	20
50	40

5.2.1) ที่อุณหภูมิ 40°C และ 50°C สาร A ละลายในของเหลว B ได้แตกต่างกันเท่าใด

.....

.....

.....

5.2.2) จากข้อมูลในตารางเมื่ออุณหภูมิสูงขึ้น การละลายของสาร A เป็นอย่างไร

.....

.....

.....

5.3) จากการวัดความดันบรรยากาศและความหนาแน่นของอากาศที่ระดับความสูงต่าง ๆ ได้ผลการทดลองดังนี้

ความสูงจากระดับน้ำทะเล (km)	ความดันบรรยากาศ (mm ของปรอท)	ความหนาแน่นของอากาศ (g/cm^3)
0	760	0.00130
2	600	0.00100
4	470	0.00082
6	360	0.00066
8	280	0.00052
10	210	0.00041
12	160	0.00031

5.3.1) ความดันบรรยากาศที่ระดับความสูง 6 กิโลเมตร จากระดับน้ำทะเลมีค่าเท่าใด

.....

.....

.....

5.3.2) จากตารางข้อมูลข้างต้น สรุปความสัมพันธ์ระหว่างระดับความสูง ความดันบรรยากาศและความหนาแน่นของอากาศได้อย่างไร

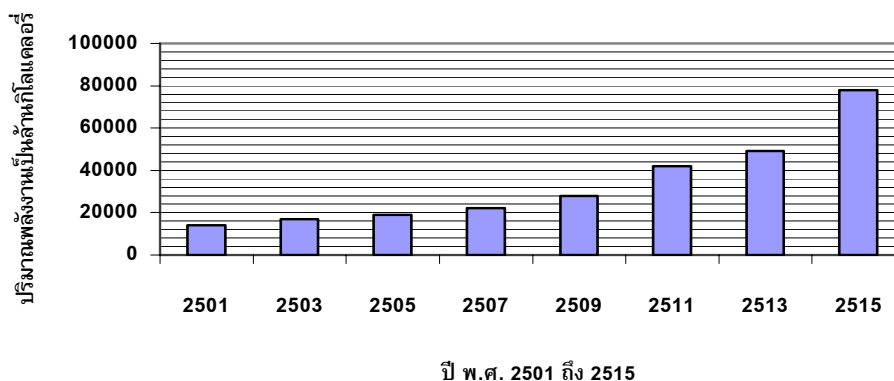
.....

.....

.....

5.4) จากการสำรวจพลังงานที่ใช้ในประเทศไทยจาก ปี พ.ศ. 2501 ถึง พ.ศ. 2515 ได้ผลดังแผนภูมิข้างล่าง ให้ใช้เป็นข้อมูลตอบคำถามข้อ 5.4.1 – 5.4.2

กราฟแสดงปริมาณพลังงานที่ใช้ในประเทศไทยจากปี พ.ศ. 2501 ถึง 2515



5.4.1) พลังงานที่ใช้ในปี พ.ศ. 2505 กับปี พ.ศ. 2515 ต่างกันเท่าใด

.....

.....

.....

5.4.2) จากข้อมูลในแผนภูมิข้างต้น สรุปความสัมพันธ์ระหว่างปริมาณ พลังงานที่ใช้ในประเทศไทยกับระยะเวลาในทุก ๆ 2 ปีได้อย่างไร

.....

.....

.....

1.2 แนวทางในการตอบและกฎเกณฑ์การให้คะแนน

ข้อ 1 ทักษะการตั้งสมมติฐาน

1. ในแต่ละข้อย่อย เช่น ข้อ 1.1, 1.2 และ 1.3 กำหนดน้ำหนักคะแนนข้อละ 1 คะแนน ถ้าตอบถูกต้องตามแนวทางในการตอบ หรือ มีส่วนใกล้เคียง หรือ มีแนวทางที่จะนำไปสู่คำตอบ ให้ข้อย่อยละ 1 คะแนน ถ้าไม่เป็นไปตามเงื่อนไขที่กำหนด ให้ 0 คะแนน

2. นำผลการตรวจให้คะแนนมาจัดระดับคุณภาพ โดยใช้กฎเกณฑ์การให้คะแนนทักษะการตั้งสมมติฐานประกอบการพิจารณาพฤติกรรมที่บ่งชี้ความสามารถของผู้เข้าสอบ กำหนดคุณภาพเป็น 4 ระดับ บันทึกผลการให้คะแนนในภาพรวมของทักษะ

แนวทางในการตอบ

- 1.1. การเพิ่มอุณหภูมิให้แก่น้ำจะมีผลต่อปริมาณการละลายของสาร A
- 1.2. ชนิดของวัตถุที่ใช้ทำลูกกลมตันมีผลต่อเวลาในการกลิ้งของลูกกลมตันบนพื้นเอียง
- 1.3. ความสามารถในการละลายของสารแต่ละชนิดขึ้นอยู่กับชนิดของตัวทำละลาย หรือ ชนิดของตัวทำละลายมีผลต่อความสามารถในการละลายของสารแต่ละชนิด

กฎเกณฑ์การให้คะแนนทักษะการตั้งสมมติฐาน

ระดับคุณภาพ	คำอธิบาย
4 ดีมาก	หมายถึง สามารถคาดคะเนคำตอบก่อนการทดลองได้สอดคล้องกับตัวแปรต้นและตัวแปรตาม การแสดงเหตุผลและแนวคิดชัดเจนถูกต้องทั้งหมด
3 ดี	หมายถึง สามารถคาดคะเนคำตอบก่อนการทดลองได้สอดคล้องกับตัวแปรต้นและตัวแปรตาม การแสดงเหตุผลและแนวคิดชัดเจนถูกต้อง 2 ข้อใน 3 ข้อ
2 ใช้ได้	หมายถึง สามารถคาดคะเนคำตอบก่อนการทดลองได้สอดคล้องกับตัวแปรต้นและ ตัวแปรตาม การแสดงเหตุผลและแนวคิดชัดเจนถูกต้อง 1 ข้อใน 3 ข้อ
1 ใช้ไม่ได้	หมายถึง สามารถคาดคะเนคำตอบก่อนการทดลองไม่สอดคล้องกับตัวแปรต้นและตัวแปรตาม การแสดงเหตุผลและแนวคิดมีแนวโน้มที่จะได้คำตอบ
0	ไม่ตอบหรือตอบไม่ถูกเลย

ข้อ 2 ทักษะการกำหนดนิยามเชิงปฏิบัติการ

1. ในแต่ละข้อย่อย เช่น ข้อ 2.1 , 2.2 และ 2.3 กำหนดน้ำหนักคะแนนข้อละ 1 คะแนน ถ้าตอบถูกต้องตามแนวทางในการตอบ หรือ มีส่วนใกล้เคียง หรือ มีแนวทางที่จะนำไปสู่คำตอบ ให้ข้อย่อยละ 1 คะแนน ถ้าไม่เป็นไปตามเงื่อนไขที่กำหนด ให้ 0 คะแนน

2. นำผลการตรวจให้คะแนนมาจัดระดับคุณภาพ โดยใช้กฎเกณฑ์การให้คะแนนทักษะการกำหนดนิยามเชิงปฏิบัติการประกอบการพิจารณาพฤติกรรมที่บ่งชี้ความสามารถของผู้เข้าสอบ กำหนดคุณภาพเป็น 4 ระดับ บันทึกผลการให้คะแนนในภาพรวมของทักษะ

แนวทางในการตอบ

2.1. น้ำใส หมายถึง น้ำที่สามารถมองเห็นวัตถุที่อยู่อีกด้านหนึ่งได้ชัดเจนเสมือนหนึ่งเห็นวัตถุนั้นโดยตรง

2.2. บินได้ หมายถึง การที่ร่างกายทุก ๆ ส่วนของนกลอยอยู่ในอากาศหรือเคลื่อนที่ไปมาในอากาศได้ด้วยการใช้ปีกช่วยในการเคลื่อนที่

2.3. ก๊าซคาร์บอนไดออกไซด์ หมายถึง ก๊าซชนิดหนึ่งที่มีอยู่ในอากาศ เมื่อผ่านลงในน้ำปูนใสแล้วจะทำให้ น้ำปูนใสขุ่น

กฎเกณฑ์การให้คะแนนทักษะการกำหนดนิยามเชิงปฏิบัติการ

ระดับคุณภาพ	คำอธิบาย
4 ดีมาก	หมายถึง ระบุความหมายและขอบเขตของคำหรือตัวแปรได้ชัดเจน การแสดงเหตุผล ถูกต้องทั้งหมด พร้อมทั้งสามารถเสนอแนะแนวทางในการตรวจสอบและวัดได้
3 ดี	หมายถึง ระบุความหมายและขอบเขตของคำหรือตัวแปรได้ชัดเจน การแสดงเหตุผล ถูกต้อง 2 ข้อใน 3 ข้อ พร้อมทั้งเสนอแนะแนวทางในการตรวจสอบและวัดได้
2 ใช้ได้	หมายถึง ระบุความหมายและขอบเขตของคำหรือตัวแปรได้ชัดเจน การแสดงเหตุผล ถูกต้อง 1 ข้อใน 3 ข้อ พร้อมทั้งเสนอแนะแนวทางในการตรวจสอบและวัดได้
1 ใช้ไม่ได้	หมายถึง ระบุความหมายและขอบเขตของคำหรือตัวแปรได้ชัดเจน การแสดงเหตุผล ถูกต้อง 1 ข้อใน 3 ข้อ แต่ไม่สามารถเสนอแนะแนวทางในการตรวจสอบและวัดได้
0	ไม่ตอบหรือตอบไม่ถูกเลย

ข้อ 3 ทักษะการกำหนดและควบคุมตัวแปร

1. ในแต่ละข้อย่อย เช่น ข้อ 3.1 – 3.4 กำหนดน้ำหนักคะแนนข้อละ 1 คะแนน ถ้าตอบ ถูกตามแนวทางในการตอบ หรือ มีส่วนใกล้เคียง หรือ มีแนวทางที่จะนำไปสู่คำตอบ โดยที่ผู้เข้าสอบจะต้องตอบถูกให้ครบทั้งสามตัวแปรในแต่ละข้อย่อย จึงจะได้ 1 คะแนน ในข้อย่อยนั้น ถ้าไม่เป็นไปตามเงื่อนไขที่กำหนด ให้ 0 คะแนน

2. นำผลการตรวจให้คะแนนมาจัดระดับคุณภาพ โดยใช้กฎเกณฑ์การให้คะแนนทักษะการกำหนดและควบคุมตัวแปรประกอบการพิจารณาพฤติกรรมที่บ่งชี้ความสามารถของผู้เข้าสอบ กำหนดคุณภาพเป็น 4 ระดับ บันทึกผลการให้คะแนนในภาพรวมของทักษะ

แนวทางในการตอบ

3.1. ตัวแปรต้น คือ ความเข้มข้นของสารละลายคลอไรด์

ตัวแปรตาม คือ อุณหภูมิของสารละลายคลอไรด์

ตัวแปรที่ต้องควบคุม คือ ปริมาณของคลอไรด์, ขนาดของช้อนตักสาร , เทอร์มอมิเตอร์ที่ใช้วัดอุณหภูมิ ฯลฯ

3.2. ตัวแปรต้น คือ ปริมาณของวิตามิน เอ

ตัวแปรตาม คือ น้ำหนักของหนู

ตัวแปรที่ต้องควบคุม คือ อาหารที่ใช้เลี้ยงหนู , ชนิดและขนาดของหนู , ระยะเวลาในการเลี้ยง, สถานที่ ฯลฯ

3.3. ตัวแปรต้น คือ จำนวนแบตเตอรี่

ตัวแปรตาม คือ จำนวนตะปูที่ถูกดูด

ตัวแปรที่ต้องควบคุม คือ ชนิดของแม่เหล็กไฟฟ้า , ชนิดของแบตเตอรี่ ฯลฯ

3.4. ตัวแปรต้น คือ ปริมาณของน้ำ

ตัวแปรตาม คือ เวลาในการงอกของเมล็ดข้าวโพด

ตัวแปรที่ต้องควบคุม คือ ชนิดและความสมบูรณ์ของเมล็ดข้าวโพด , ชนิดและปริมาณของดินที่ใช้ปลูกข้าวโพด , แสงสว่าง , อุณหภูมิ ฯลฯ

กฎเกณฑ์การให้คะแนนทักษะการกำหนดและควบคุมตัวแปร

ระดับคุณภาพ	คำอธิบาย
4 ดีมาก	หมายถึง กำหนดตัวแปรต้น ตัวแปรตาม และตัวแปรที่ต้องควบคุมได้ถูกต้องทั้งหมด
3 ดี	หมายถึง กำหนดตัวแปรต้น ตัวแปรตาม และตัวแปรที่ต้องควบคุมได้ถูกต้อง 3 ข้อใน 4 ข้อ
2 ใช้ได้	หมายถึง กำหนดตัวแปรต้น ตัวแปรตาม และตัวแปรที่ต้องควบคุมได้ถูกต้อง 2 ข้อใน 4 ข้อ
1 ใช้ไม่ได้	หมายถึง กำหนดตัวแปรต้น ตัวแปรตาม และตัวแปรที่ต้องควบคุมได้ถูกต้อง 1 ข้อใน 4 ข้อ
0	ไม่ตอบหรือตอบไม่ถูกเลย

ข้อ 4 ทักษะการทดลอง

ในข้อนี้มีคุณลักษณะพิเศษคือ ข้อย่อยแต่ละข้อมีความสัมพันธ์เกี่ยวข้องกับสถานการณ์ในข้อใหญ่ จนไม่สามารถแยกคะแนนให้เป็นรายข้อย่อยเหมือนข้ออื่น ๆ ได้ อีกทั้งการตอบของผู้เข้าสอบอาจไม่ตรงกับแนวทางในการตอบที่ผู้วิจัยเสนอไว้ ผู้ตรวจต้องใช้ประสบการณ์การสอนวิทยาศาสตร์ร่วมกับกฎเกณฑ์การให้คะแนนทักษะการทดลอง ประกอบการพิจารณาพฤติกรรมที่บ่งชี้ความสามารถของผู้เข้าสอบ บันทึกผลการให้คะแนนในภาพรวมของทักษะ (สามารถใช้คะแนนระดับคุณภาพเป็นคะแนนดิบโดยไม่ต้องไปพิจารณาข้อย่อยเหมือนทักษะอื่น)

แนวทางในการตอบ

4.1) การออกแบบการทดลอง โดยการนำเอาตัวอย่างดินต่างชนิดกันมาทดสอบความสามารถในการอุ้มน้ำ การทดลองนี้จะต้องคำนึงถึงปริมาณของดิน ,ปริมาณของน้ำ ,ขนาดของภาชนะที่บรรจุ ,ระยะเวลาในการทดลอง ฯลฯ เพื่อลดข้อผิดพลาดในการทดลอง

4.2) วัสดุอุปกรณ์ประกอบการทดลอง ได้แก่

- 1) ครอบงม หรืออาจเลือกใช้ภาชนะอื่นที่สามารถนำมาทดลองได้ ที่มีรูปร่างและขนาดเดียวกัน 6 ใบ โดยที่ 3 ใบ เจาะรูที่กันครอบพอประมาณ
- 2) ดินเหนียว ดินร่วน และดินทราย
- 3) น้ำ
- 4) ครอบกตวงขนาด 100 ลูกบาศก์เซนติเมตร
- 5) ผ้าขาวบางขนาด 15x15 ตารางเซนติเมตร จำนวน 3 ผืน
- 6) นาฬิกาจับเวลา

4.3) รายละเอียดขั้นตอนการดำเนินการทดลองเป็นดังนี้

- 1) นำดินแต่ละชนิด คือ ดินเหนียว ดินร่วน และดินทราย มาใส่ครอบงมที่เจาะรูแต่ละใบ ในปริมาณ $\frac{2}{3}$ ของครอบงม
- 2) นำครอบงมที่ใส่ดินเหล่านี้วางซ้อนบนครอบงมที่ไม่ได้เจาะรูแต่ละใบโดยมีผ้าขาวบางกั้นไว้ระหว่างครอบงมทั้งสอง
- 3) ตวงน้ำปริมาตร 100 ลูกบาศก์เซนติเมตร แล้วเทลงในครอบงมที่ใส่ดินแต่ละชนิด
- 4) จับเวลาของน้ำที่ไหลจากครอบงมบนสู่ครอบงมล่าง จนครบทั้งหมด
- 5) ปลอ่ยให้น้ำไหลออกจากครอบงมบนสู่ใบล่างประมาณ 2 นาที แล้วนำน้ำในครอบงมใบล่างแต่ละใบไปตวงหาปริมาตร จนครบทั้งหมด แล้วบันทึกผล
- 6) ทำการทดลองซ้ำตามข้อ 1-5 อีก 2 ครั้ง บันทึกผลการทดลอง

4.4) ถ้าตอบตามแนวทางข้อ 4.1-4.3 การออกแบบการบันทึกผลการทดลองควรเป็นดังนี้

ตารางบันทึกผลการทดลอง

ชนิดของดินในครอบงม	ปริมาตรของน้ำก่อนเท (cm ³)	ปริมาตรน้ำที่ไหลผ่านดิน (cm ³)				ปริมาตรน้ำที่เหลือในครอบงมดิน (cm ³)			
		ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3	เฉลี่ย	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3	เฉลี่ย
ดินเหนียว	100								
ดินร่วน	100								
ดินทราย	100								

หมายเหตุ

ถ้าผู้เข้าสอบตอบหรือเสนอวิธีการทดลองที่ไม่ตรงกับแนวทางในการตอบ ให้ผู้ตรวจใช้ประสบการณ์ในการสอนวิทยาศาสตร์พิจารณาให้คะแนนได้ตามความเหมาะสม

กฎเกณฑ์การให้คะแนนทักษะการทดลอง

ระดับคุณภาพ	คำอธิบาย
4 ดีมาก	หมายถึง ออกแบบการทดลองได้ถูกต้องและสอดคล้องกับสมมติฐาน ระบุวัสดุอุปกรณ์ที่ใช้ในการทดลองได้ครอบคลุม อธิบายขั้นตอนการทดลองได้ถูกต้อง และเสนอแนะการบันทึกผลการทดลองได้สอดคล้องกับข้อมูล
3 ดี	หมายถึง ออกแบบการทดลองได้ถูกต้องและสอดคล้องกับสมมติฐาน ระบุวัสดุอุปกรณ์ที่ใช้ในการทดลองไม่ครอบคลุม อธิบายขั้นตอนการทดลองได้ถูกต้อง และเสนอแนะการบันทึกผลการทดลองได้สอดคล้องกับข้อมูล
2 ใช้ได้	หมายถึง ออกแบบการทดลองได้ถูกต้องและสอดคล้องกับสมมติฐาน ระบุวัสดุอุปกรณ์ที่ใช้ในการทดลองไม่ครอบคลุม แต่ไม่อธิบายขั้นตอนการทดลองหรือเสนอแนะการบันทึกผลการทดลองอย่างใดอย่างหนึ่ง
1 ใช้ไม่ได้	หมายถึง ออกแบบการทดลองไม่สอดคล้องกับสมมติฐาน ระบุวัสดุอุปกรณ์ที่ใช้ในการทดลองไม่ครอบคลุม ไม่อธิบายขั้นตอนการทดลอง และไม่เสนอแนะการบันทึกผลการทดลอง
0	ไม่ตอบหรือตอบไม่ถูกเลย

ข้อ 5 ทักษะการตีความหมายข้อมูลและลงข้อสรุป

1. ในแต่ละข้อย่อย เช่น ข้อ 1 – 4 กำหนดน้ำหนักคะแนนข้อละ 1 คะแนน ถ้าตอบถูกต้องตามแนวทางในการตอบ หรือ มีส่วนใกล้เคียง หรือ มีแนวทางที่จะนำไปสู่คำตอบ โดยที่ผู้เข้าสอบจะต้องตอบข้อย่อยถูกต้องทั้งสองข้อ เช่น ตอบข้อ 1.1 และ 1.2 ถูกทั้งสองข้อ จึงจะได้ 1 คะแนน ในข้อ 1 ถ้าไม่เป็นไปตามเงื่อนไขที่กำหนด ให้ 0 คะแนน

2. นำผลการตรวจให้คะแนนมาจัดระดับคุณภาพ โดยใช้กฎเกณฑ์การให้คะแนนทักษะการการตีความหมายข้อมูลและลงข้อสรุปประกอบการพิจารณาพฤติกรรมที่บ่งชี้ความสามารถของผู้เข้าสอบ กำหนดคุณภาพเป็น 4 ระดับ บันทึกผลการให้คะแนนในภาพรวมของทักษะ

แนวทางในการตอบ

5.1.1) เปลี่ยนแปลง กล่าวคือ จำนวนปลาที่จับได้ในแต่ละปีลดน้อยลง ยกเว้น ปี พ.ศ. 2517 และ พ.ศ. 2519 ที่เพิ่มขึ้นเล็กน้อยจากปีก่อนหน้านั้น แต่กลับลดลงในปีถัดไป

5.1.2) ปริมาณปลาในทะเลบริเวณที่เคยจับได้ลดลงทุก ๆ ปี

5.2.1) ที่อุณหภูมิ 40°C และ 50°C สาร A ละลายในของเหลว B ได้แตกต่างกัน 20 กรัม

5.2.2) จากข้อมูลในตารางเมื่ออุณหภูมิสูงขึ้น สรุปการละลายของสาร A ได้ว่า ยิ่งอุณหภูมิสูงขึ้น สาร A ยิ่งละลายในของเหลว B ได้มากขึ้น

5.3.1) ความดันบรรยากาศที่ระดับความสูง 6 กิโลเมตรจากระดับน้ำทะเลมีค่าเท่ากับ 360 มิลลิเมตรปรอท

5.3.2) จากตารางข้อมูลข้างต้น สรุปความสัมพันธ์ระหว่างระดับความสูง ความดันบรรยากาศและความหนาแน่นของอากาศได้ว่า ยิ่งความสูงจากระดับน้ำทะเลเพิ่มมากขึ้น ความดันบรรยากาศและความหนาแน่นของอากาศมีค่าลดลง

5.4.1) พลังงานที่ใช้ในปี พ.ศ. 2505 กับปี พ.ศ. 2515 ต่างกันประมาณ 59,000 ล้าน กิโลแคลอรี

5.4.2) จากข้อมูลในแผนภูมิข้างต้น สรุปความสัมพันธ์ระหว่างปริมาณพลังงานที่ใช้ในประเทศไทยกับระยะเวลาในทุก ๆ 2 ปีได้ว่า ปริมาณพลังงานที่ใช้ในประเทศไทยจากปี พ.ศ. 2501 ถึง พ.ศ. 2515 มีปริมาณเพิ่มมากขึ้นเรื่อย ๆ

กฎเกณฑ์การให้คะแนนทักษะการตีความหมายข้อมูลและลงข้อสรุป

ระดับคุณภาพ	คำอธิบาย
4 ดีมาก	หมายถึง แปลความหมายของข้อมูลที่จัดไว้ในรูปแบบต่าง ๆ ที่ใช้ในการสื่อความหมายข้อมูลในเชิงสถิติ พร้อมทั้งสรุปความสัมพันธ์ของข้อมูลได้ถูกต้องทั้งหมด
3 ดี	หมายถึง แปลความหมายของข้อมูลที่จัดไว้ในรูปแบบต่าง ๆ ที่ใช้ในการสื่อความหมายข้อมูลในเชิงสถิติ พร้อมทั้งสรุปความสัมพันธ์ของข้อมูลได้ถูกต้อง 3 ข้อใน 4 ข้อ
2 ใช้ได้	หมายถึง แปลความหมายของข้อมูลที่จัดไว้ในรูปแบบต่าง ๆ ที่ใช้ในการสื่อความหมายข้อมูลในเชิงสถิติ พร้อมทั้งสรุปความสัมพันธ์ของข้อมูลได้ถูกต้อง 2 ข้อใน 4 ข้อ
1 ใช้ไม่ได้	หมายถึง แปลความหมายของข้อมูลที่จัดไว้ในรูปแบบต่าง ๆ ที่ใช้ในการสื่อความหมายข้อมูลในเชิงสถิติ พร้อมทั้งสรุปความสัมพันธ์ของข้อมูลได้ถูกต้อง 1 ข้อใน 4 ข้อ
0	ไม่ตอบหรือตอบไม่ถูกเลย

ตาราง 14 ดัชนีความสอดคล้องของการประเมินจากผู้เชี่ยวชาญในการตรวจสอบคุณภาพของ
เครื่องมือ (IOC)

ทักษะ	ข้อ	ผู้เชี่ยวชาญคนที่					รวม	IOC	แปลผล
		1	2	3	4	5			
การตั้งสมมติฐาน	1.1	1	1	0	1	0	3	0.6	ใช้ได้
	1.2	0	1	1	1	1	4	0.8	ใช้ได้
	1.3	1	1	1	1	1	5	1.0	ใช้ได้
การกำหนดนิยาม เชิงปฏิบัติการ	2.1	1	1	1	1	1	5	1.0	ใช้ได้
	2.2	1	1	1	1	0	4	0.8	ใช้ได้
	2.3	1	1	1	1	1	5	1.0	ใช้ได้
	2.4	1	1	1	1	1	5	1.0	ใช้ได้
การกำหนดและควบคุม ตัวแปร	3.1	1	1	1	0	1	4	0.8	ใช้ได้
	3.2	0	1	1	1	0	3	0.6	ใช้ได้
	3.3	1	1	1	0	1	4	0.8	ใช้ได้
	3.4	1	1	1	0	1	4	0.8	ใช้ได้
	3.5	1	1	1	1	0	4	0.8	ใช้ได้
การทดลอง	4.1	0	1	0	1	1	3	0.6	ใช้ได้
	4.2	0	1	1	0	1	3	0.6	ใช้ได้
	4.3	1	1	0	1	1	4	0.8	ใช้ได้
การตีความหมายข้อมูล และลงข้อสรุป	5.1.1	0	1	1	1	1	4	0.8	ใช้ได้
	5.1.2	1	1	1	1	0	4	0.8	ใช้ได้
	5.2.1	0	1	1	1	1	4	0.8	ใช้ได้
	5.2.2	1	1	0	1	1	4	0.8	ใช้ได้
	5.3.1	0	1	1	1	1	4	0.8	ใช้ได้
	5.3.2	1	1	1	1	0	4	0.8	ใช้ได้
	5.4.1	0	1	1	1	1	4	0.8	ใช้ได้
	5.4.2	1	1	0	1	1	4	0.8	ใช้ได้

ตาราง 15 ค่าความยาก (p) ค่าอำนาจจำแนก (r_{xy}) ของแบบวัดทักษะกระบวนการทาง
วิทยาศาสตร์ขั้นบูรณาการ

ทักษะ	ข้อ	ค่าความยาก (p)	ค่าอำนาจจำแนก (r_{xy})	ผลการ คัดเลือก
การตั้งสมมติฐาน	1.1	0.316	.668**	ใช้ได้
	1.2	0.368	.636**	ใช้ได้
	1.3	0.421	.684**	ใช้ได้
การกำหนดนิยาม เชิงปฏิบัติการ	2.1	0.395	.620**	ใช้ได้
	2.2	0.474	.645**	ใช้ได้
	2.3	0.368	.332	คัดออก
	2.4	0.553	.750**	ใช้ได้
การกำหนดและควบคุม ตัวแปร	3.1	0.474	.671**	ใช้ได้
	3.2	0.316	.678**	ใช้ได้
	3.3	0.509	.706**	ใช้ได้
	3.4	0.561	.845**	ใช้ได้
	3.5	0.246	.476*	คัดออก
การทดลอง	4.1	0.322	.582**	ใช้ได้
	4.2	0.401	.391	คัดออก
	4.3	0.401	.624**	ใช้ได้
การตีความหมายข้อมูล และลงข้อสรุป	5.1.1	0.447	.649**	ใช้ได้
	5.1.2	0.789	.677**	ใช้ได้
	5.2.1	0.421	.529*	ใช้ได้
	5.2.2	0.474	.781**	ใช้ได้
	5.3.1	0.423	.610**	ใช้ได้
	5.3.2	0.765	.612**	ใช้ได้
	5.4.1	0.456	.549*	ใช้ได้
	5.4.2	0.498	.721**	ใช้ได้

หมายเหตุ ทักษะการทดลองเป็นทักษะที่ต้องใช้เวลามาก หลังจากที่ยืนยันเวลาดูแล้ว ถ้าในแบบทดสอบใช้ทักษะนี้มากกว่า 1 ข้อ จะใช้เวลาเกินกว่าที่กำหนด ดังนั้น ทักษะนี้ผู้วิจัยจึงเลือกเอาเพียงข้อเดียว คือ ข้อ 4.3 ที่สามารถจำแนกกลุ่มสูงและกลุ่มต่ำได้สูงกว่าข้อ 4.1

ตาราง 16 ผลการวิเคราะห์ความสามารถของนักเรียนตามโมเดลของราส์ชแบบ 3 องค์ประกอบ
(จัดเรียงตามความสามารถของนักเรียน)

นักเรียน คนที่	ความสามารถ (Logit)	INFIT (MnSq)	OUTFIT (MnSq)	Mean Rating From 2 Rater	Expected Rater from Rasch Model
203	.86	0.9	0.9	3.6	3.64
217	.86	1.0	0.7	3.6	3.64
13	.83	1.2	1.3	3.3	3.63
18	.83	0.8	1.1	3.3	3.63
33	.83	1.1	1.5	3.3	3.63
37	.83	1.2	2.1	3.3	3.63
39	.83	1.4	1.5	3.3	3.63
122	.83	1.1	1.0	3.3	3.63
123	.83	1.0	1.3	3.3	3.63
145	.83	1.8	2.8	3.3	3.63
226	.83	0.9	1.0	3.3	3.63
254	.83	0.6	0.8	3.3	3.63
4	.82	0.8	0.9	3.5	3.62
10	.82	0.8	0.9	3.5	3.62
77	.82	0.4	0.5	3.7	3.62
83	.82	1.2	0.8	3.5	3.62
103	.82	0.3	0.3	3.7	3.62
126	.82	0.5	0.6	3.5	3.62
153	.82	0.9	1.2	3.5	3.62
209	.82	0.6	0.6	3.5	3.62
232	.82	0.6	0.6	3.5	3.62
249	.82	0.7	0.7	3.5	3.62
7	.78	0.7	0.8	3.4	3.59
41	.78	1.1	1.5	3.2	3.59
50	.78	1.2	1.5	3.2	3.59
63	.78	1.2	1.5	3.2	3.59
76	.78	1.3	2.2	3.2	3.59

ตาราง 16 (ต่อ)

นักเรียน คนที่	ความสามารถ (Logit)	INFIT (MnSq)	OUTFIT (MnSq)	Mean Rating From 2 Rater	Expected Rater from Rasch Model
85	.78	0.6	0.7	3.2	3.59
110	.78	0.7	0.8	3.2	3.59
128	.78	0.6	0.5	3.4	3.59
129	.78	0.7	1.2	3.2	3.59
160	.78	1.0	0.9	3.4	3.59
235	.78	0.7	0.6	3.2	3.59
241	.78	0.8	0.8	3.2	3.59
6	.72	1.2	1.4	3.1	3.55
12	.72	1.8	1.8	3.1	3.55
42	.72	1.0	1.1	3.1	3.55
43	.72	1.4	1.5	3.1	3.55
118	.72	0.6	0.6	3.1	3.55
125	.72	1.1	1.2	3.1	3.55
131	.72	1.4	1.3	3.1	3.55
147	.72	0.9	1.4	3.1	3.55
149	.72	1.4	1.4	3.1	3.55
244	.72	1.6	1.6	3.1	3.55
263	.72	0.6	0.6	3.1	3.55
264	.72	1.1	1.1	3.1	3.55
277	.72	1.0	1.4	3.1	3.55
279	.72	1.1	1.4	3.1	3.55
11	.56	0.7	1.0	3.0	3.43
14	.56	1.9	1.9	3.0	3.43
16	.56	0.9	1.0	3.0	3.43
35	.56	1.1	1.0	3.0	3.43
40	.56	0.7	0.9	3.0	3.43
46	.56	1.0	1.4	3.0	3.43
51	.56	1.5	1.7	3.0	3.43
52	.56	0.9	1.5	3.0	3.43

ตาราง 16 (ต่อ)

นักเรียน คนที่	ความสามารถ (Logit)	INFIT (MnSq)	OUTFIT (MnSq)	Mean Rating From 2 Rater	Expected Rater from Rasch Model
54	.56	1.5	1.6	3.0	3.43
100	.56	1.5	1.4	3.0	3.43
114	.56	1.0	1.0	3.0	3.43
127	.56	1.0	1.3	3.0	3.43
139	.56	1.5	2.6	3.0	3.43
175	.56	0.6	0.6	3.0	3.43
176	.56	0.7	0.7	3.0	3.43
195	.56	0.4	0.4	3.0	3.43
201	.56	0.6	0.5	3.0	3.43
216	.56	0.6	0.7	3.0	3.43
243	.56	0.5	0.5	3.0	3.43
250	.56	1.3	1.2	3.0	3.43
257	.56	1.2	1.2	3.0	3.43
49	.42	0.7	0.9	2.9	3.29
55	.42	0.9	0.9	2.9	3.29
56	.42	0.4	0.5	2.9	3.29
59	.42	0.7	0.8	2.9	3.29
93	.42	0.6	0.6	2.9	3.29
113	.42	0.9	1.0	2.9	3.29
115	.42	1.4	1.4	2.9	3.29
140	.42	1.2	1.0	2.9	3.29
142	.42	0.7	1.1	2.9	3.29
146	.42	2.6	2.7	2.9	3.29
171	.42	0.5	0.6	2.9	3.29
194	.42	0.2	0.2	2.9	3.29
224	.42	0.8	0.8	2.9	3.29
230	.42	0.7	0.6	2.9	3.29
239	.42	0.5	0.5	2.9	3.29
259	.42	1.1	1.2	2.9	3.29

ตาราง 16 (ต่อ)

นักเรียน คนที่	ความสามารถ (Logit)	INFIT (MnSq)	OUTFIT (MnSq)	Mean Rating From 2 Rater	Expected Rater from Rasch Model
272	.42	1.0	0.9	2.9	3.29
22	.27	1.0	1.1	2.8	3.12
26	.27	0.5	0.7	2.8	3.12
29	.27	1.0	1.1	2.8	3.12
30	.27	1.5	1.7	2.8	3.12
53	.27	0.9	0.9	2.8	3.12
67	.27	0.4	0.4	2.8	3.12
86	.27	1.1	1.3	2.8	3.12
87	.27	1.6	1.7	2.8	3.12
101	.27	1.4	1.4	2.8	3.12
105	.27	0.5	0.6	2.8	3.12
106	.27	0.8	1.0	2.8	3.12
132	.27	0.9	1.3	2.8	3.12
148	.27	0.9	0.8	2.8	3.12
150	.27	0.9	0.9	2.8	3.12
204	.27	1.2	1.2	2.8	3.12
214	.27	0.3	0.3	2.8	3.12
260	.27	0.6	0.7	2.8	3.12
261	.27	0.8	.07	2.8	3.12
275	.27	0.9	1.0	2.8	3.12
73	.14	0.4	0.4	2.7	2.94
78	.14	0.6	0.6	2.7	2.94
117	.14	0.5	0.5	2.7	2.94
155	.14	0.5	0.8	2.7	2.94
156	.14	1.2	1.1	2.7	2.94
177	.14	0.8	0.7	2.7	2.94
179	.14	0.9	1.1	2.7	2.94
180	.14	0.6	0.6	2.7	2.94

ตาราง 16 (ต่อ)

นักเรียน คนที่	ความสามารถ (Logit)	INFIT (MnSq)	OUTFIT (MnSq)	Mean Rating From 2 Rater	Expected Rater from Rasch Model
210	.14	0.5	0.6	2.7	2.94
227	.14	0.6	0.6	2.7	2.94
242	.14	1.0	1.1	2.7	2.94
248	.14	0.2	0.2	2.7	2.94
267	.14	1.3	1.6	2.7	2.94
8	.00	0.4	0.4	2.6	2.74
24	.00	0.5	0.5	2.6	2.74
28	.00	0.3	0.3	2.6	2.74
47	.00	0.7	0.8	2.6	2.74
72	.00	1.0	1.0	2.6	2.74
75	.00	1.3	1.2	2.6	2.74
84	.00	0.5	0.5	2.6	2.74
89	.00	0.4	0.4	2.6	2.74
95	.00	1.0	1.0	2.6	2.74
99	.00	1.1	1.2	2.6	2.74
124	.00	0.9	1.2	2.6	2.74
136	.00	0.6	0.5	2.6	2.74
144	.00	0.4	0.4	2.6	2.74
151	.00	0.7	0.7	2.6	2.74
161	.00	0.7	0.7	2.6	2.74
162	.00	0.7	0.7	2.6	2.74
166	.00	0.9	1.0	2.6	2.74
167	.00	0.5	0.5	2.6	2.74
181	.00	0.3	0.3	2.6	2.74
184	.00	0.4	0.4	2.6	2.74
192	.00	0.4	0.4	2.6	2.74
198	.00	0.6	0.6	2.6	2.74
200	.00	0.4	0.5	2.6	2.74
206	.00	0.2	0.2	2.6	2.74

ตาราง 16 (ต่อ)

นักเรียน คนที่	ความสามารถ (Logit)	INFIT (MnSq)	OUTFIT (MnSq)	Mean Rating From 2 Rater	Expected Rater from Rasch Model
207	.00	0.4	0.4	2.6	2.74
212	.00	0.4	.04	2.6	2.74
215	.00	0.7	0.7	2.6	2.74
219	.00	0.6	0.6	2.6	2.74
220	.00	0.8	0.8	2.6	2.74
229	.00	0.4	0.4	2.6	2.74
231	.00	0.5	0.5	2.6	2.74
252	.00	0.5	0.6	2.6	2.74
256	.00	1.0	1.0	2.6	2.74
266	.00	0.5	0.5	2.6	2.74.
269	.00	1.3	1.4	2.6	2.74
2	-.13	0.7	0.7	2.5	2.53
3	-.13	0.6	0.5	2.5	2.53
5	-.13	0.2	0.2	2.5	2.53
21	-.13	1.1	1.3	2.5	2.53
25	-.13	0.4	0.4	2.5	2.53
38	-.13	0.6	0.6	2.5	2.53
65	-.13	0.5	0.6	2.5	2.53
66	-.13	0.5	0.5	2.5	2.53
69	-.13	0.7	0.7	2.5	2.53
91	-.13	0.8	0.8	2.5	2.53
98	-.13	0.6	0.6	2.5	2.53
102	-.13	0.7	0.7	2.5	2.53
107	-.13	0.3	0.4	2.5	2.53
111	-.13	0.6	0.6	2.5	2.53
119	-.13	0.9	0.8	2.5	2.53
121	-.13	0.9	1.0	2.5	2.53
141	-.13	0.7	0.7	2.5	2.53
158	-.13	0.4	0.4	2.5	2.53

ตาราง 16 (ต่อ)

นักเรียน คนที่	ความสามารถ (Logit)	INFIT (MnSq)	OUTFIT (MnSq)	Mean Rating From 2 Rater	Expected Rater from Rasch Model
159	-.13	0.3	0.3	2.5	2.53
164	-.13	0.5	0.5	2.5	2.53
189	-.13	0.7	0.7	2.5	2.53
196	-.13	0.4	0.4	2.5	2.53
218	-.13	0.3	0.3	2.5	2.53
240	-.13	0.5	0.4	2.5	2.53
246	-.13	0.4	0.4	2.5	2.53
20	-.26	0.3	0.3	2.4	2.32
27	-.26	0.4	0.5	2.4	2.32
32	-.26	0.4	0.4	2.4	2.32
74	-.26	0.6	0.6	2.4	2.32
81	-.26	0.9	1.0	2.4	2.32
92	-.26	0.6	0.6	2.4	2.32
104	-.26	0.5	0.5	2.4	2.32
109	-.26	0.4	0.5	2.4	2.32
120	-.26	0.3	0.3	2.4	2.32
133	-.26	0.4	0.4	2.4	2.32
137	-.26	0.6	0.6	2.4	2.32
138	-.26	0.5	0.6	2.4	2.32
143	-.26	0.4	0.4	2.4	2.32
186	-.26	0.5	0.8	2.4	2.32
191	-.26	0.6	0.6	2.4	2.32
202	-.26	0.5	0.5	2.4	2.32
205	-.26	0.9	1.0	2.4	2.32
225	-.26	0.2	0.2	2.4	2.32
258	-.26	0.2	0.2	2.4	2.32
262	-.26	0.5	0.5	2.4	2.32
280	-.26	0.7	0.8	2.4	2.32
19	-.40	0.6	0.6	2.3	2.12

ตาราง 16 (ต่อ)

นักเรียน คนที่	ความสามารถ (Logit)	INFIT (MnSq)	OUTFIT (MnSq)	Mean Rating From 2 Rater	Expected Rater from Rasch Model
57	-.40	0.5	0.5	2.3	2.12
58	-.40	0.7	0.7	2.3	2.12
60	-.40	0.7	0.7	2.3	2.12
88	-.40	0.8	0.9	2.3	2.12
134	-.40	0.2	0.2	2.3	2.12
165	-.40	1.0	1.0	2.3	2.12
172	-.40	0.5	0.5	2.3	2.12
174	-.40	0.3	0.3	2.3	2.12
185	-.40	0.5	0.5	2.3	2.12
187	-.40	0.6	0.6	2.3	2.12
190	-.40	0.5	0.5	2.3	2.12
197	-.40	0.8	0.8	2.3	2.12
213	-.40	0.8	0.8	2.3	2.12
221	-.40	0.6	0.7	2.3	2.12
238	-.40	1.1	1.2	2.3	2.12
247	-.40	0.6	0.7	2.3	2.12
255	-.40	0.6	0.6	2.3	2.12
270	-.40	0.5	0.5	2.3	2.12
271	-.40	0.6	0.6	2.3	2.12
276	-.40	0.4	0.4	2.3	2.12
278	-.40	0.4	0.4	2.3	2.12
36	-.54	0.5	0.5	2.2	1.93
48	-.54	0.6	0.6	2.2	1.93
61	-.54	0.5	0.5	2.2	1.93
64	-.54	0.8	0.9	2.2	1.93
68	-.54	1.2	1.2	2.2	1.93
79	-.54	0.6	0.6	2.2	1.93
80	-.54	0.3	0.3	2.2	1.93
96	-.54	0.5	0.5	2.2	1.93

ตาราง 16 (ต่อ)

นักเรียน คนที่	ความสามารถ (Logit)	INFIT (MnSq)	OUTFIT (MnSq)	Mean Rating From 2 Rater	Expected Rater from Rasch Model
97	-.54	0.7	0.7	2.2	1.93
135	-.54	0.3	0.3	2.2	1.93
199	-.54	0.8	0.8	2.2	1.93
234	-.54	0.4	0.4	2.2	1.93
90	-.68	0.6	0.7	2.1	1.76
157	-.68	0.6	0.6	2.1	1.76
183	-.68	1.2	1.8	2.1	1.76
223	-.68	0.4	0.4	2.1	1.76
228	-.68	1.1	1.9	2.1	1.76
253	-.68	0.9	0.8	2.1	1.76
9	-.82	0.8	0.8	2.0	1.61
31	-.82	0.2	0.2	2.0	1.61
34	-.82	0.4	0.4	2.0	1.61
62	-.82	0.8	0.9	2.0	1.61
112	-.82	0.8	0.9	2.0	1.61
130	-.82	0.6	0.6	2.0	1.61
154	-.82	0.7	0.7	2.0	1.61
178	-.82	1.2	2.3	2.0	1.61
188	-.82	1.2	2.3	2.0	1.61
222	-.82	0.6	0.6	2.0	1.61
236	-.82	0.5	0.7	2.0	1.61
268	-.82	0.5	0.6	2.0	1.61
274	-.82	0.8	0.8	2.0	1.61
1	-.98	1.1	1.1	1.9	1.48
17	-.98	0.5	0.4	1.9	1.48
44	-.98	1.1	0.9	1.9	1.48
70	-.98	1.1	1.1	1.9	1.48
152	-.98	0.3	0.3	1.9	1.48
168	-.98	1.4	1.7	1.9	1.48

ตาราง 16 (ต่อ)

นักเรียน คนที่	ความสามารถ (Logit)	INFIT (MnSq)	OUTFIT (MnSq)	Mean Rating From 2 Rater	Expected Rater from Rasch Model
237	-.98	1.0	1.1	1.9	1.48
45	-1.04	0.6	1.0	1.8	1.43
82	-1.04	0.8	0.6	1.6	1.43
94	-1.04	0.9	0.9	1.8	1.43
108	-1.04	0.9	1.7	1.8	1.43
116	-1.04	1.0	1.1	1.8	1.43
163	-1.04	1.6	1.5	1.6	1.43
173	-1.04	1.0	0.8	1.8	1.43
182	-1.04	0.8	0.7	1.6	1.43
193	-1.04	1.4	1.7	1.8	1.43
208	-1.04	0.8	0.8	1.8	1.43
233	-1.04	1.6	3.4	1.8	1.43
251	-1.04	0.8	0.9	1.8	1.43
265	-1.04	0.9	0.8	1.8	1.43
23	-1.09	0.6	0.8	1.7	1.40
71	-1.09	0.4	0.7	1.7	1.40
170	-1.09	1.2	1.5	1.7	1.40
211	-1.09	0.9	1.1	1.7	1.40
245	-1.09	0.4	0.6	1.7	1.40
Mean	0.00	0.8	0.9	2.6	2.69
S.D.	0.56	0.4	0.5	0.5	0.71

RMSE (Model) .32 Adj S.D. .46 Separation 1.43 Reliability .67

Fixed (all same) chi-square: 710.5 d.f.: 279 significance: .00

Random (normal) chi-square: 257.3 d.f.: 278 significance: .81

Interpretation of kappa values*	
kappa	Interpretation
<0	No agreement
0.0-0.19	Poor agreement
0.20-0.39	Fair agreement
0.40-0.59	Moderate agreement
0.60-0.79	Substantial agreement
0.80-1.00	Almost perfect agreement

*Landis, JR and Koch, GG. The measurement of observer agreement for categorical data. Biometrics 33:159-174, 1977.

<http://www.musc.edu/dc/icrebm/kappa.html>

เกณฑ์การพิจารณาค่าสัมประสิทธิ์แคปปา

>.80	ความเห็นของผู้ตรวจมีความสอดคล้องมากที่สุด
.60 - .79	ความเห็นของผู้ตรวจมีความสอดคล้องมาก
.40 - .59	ความเห็นของผู้ตรวจมีความสอดคล้องปานกลาง
.20 - .39	ความเห็นของผู้ตรวจมีความสอดคล้องน้อย
.00 - .19	ความเห็นของผู้ตรวจมีความสอดคล้องน้อยที่สุด
<0	ผู้ตรวจมีความเห็นไม่สอดคล้องกันเลย

ประวัติย่อผู้วิจัย

ประวัติย่อผู้วิจัย

ชื่อ ชื่อสกุล	นายชาญวิทย์ จรัสสุทธิอิศร
วันเดือนปีเกิด	8 มีนาคม 2502
สถานที่เกิด	จังหวัดศรีสะเกษ
สถานที่อยู่ปัจจุบัน	333 ถนนราษฎร์ประเสริฐ (หมู่บ้านรุ่งนภาวิลล์ 2) ต.ห้วยทะเล อ.เมืองฯ จ. นครราชสีมา 30000
ตำแหน่งหน้าที่การงานปัจจุบัน	ศึกษานิเทศก์
สถานที่ทำงานปัจจุบัน	กลุ่มงานวัดและประเมินผลการศึกษา กลุ่มนิเทศติดตามและประเมินผลการจัดการศึกษา เขตพื้นที่การศึกษานครราชสีมา เขต 1 อ.เมืองฯ จ.นครราชสีมา 30000
ประวัติการศึกษา	
พ.ศ. 2522	ป.กศ. ชั้นสูง (วิทยาศาสตร์) จากวิทยาลัยครูอุบลราชธานี
พ.ศ. 2528	ศษ.บ. (วิทยาศาสตร์) จากมหาวิทยาลัยสุโขทัยธรรมมาธิราช
พ.ศ. 2545	กศ.ม. (การวัดผลการศึกษา) จากมหาวิทยาลัยศรีนครินทรวิโรฒ
พ.ศ. 2550	กศ.ด. (การทดสอบและวัดผลการศึกษา) จากมหาวิทยาลัยศรีนครินทรวิโรฒ