

การพัฒนาเครื่องมือวัด ในการวิจัยพฤติกรรมศาสตร์

DEVELOPMENT OF MEASURING INSTRUMENTS
IN BEHAVIORAL SCIENCE RESEARCH



ผู้ช่วยศาสตราจารย์ ดร.ศรัณย์ พิมพ์ทอง

สถาบันวิจัยพฤติกรรมศาสตร์
มหาวิทยาลัยศรีนครินทรวิโรฒ

การพัฒนาเครื่องมือวัดในการวิจัยพฤติกรรมศาสตร์

Development of Measuring Instruments
in Behavioral Science Research



ผู้ช่วยศาสตราจารย์ ดร. ศรัณย์ พิมพ์ทอง

สถาบันวิจัยพฤติกรรมศาสตร์

มหาวิทยาลัยศรีนครินทรวิโรฒ

2564

การพัฒนาเครื่องมือวัดในการวิจัยพฤติกรรมศาสตร์

Development of Measuring Instruments in Behavioral Science Research

ศรัณย์ พิมพ์ทอง

ผู้เขียน

ผู้ช่วยศาสตราจารย์ ดร. ศรัณย์ พิมพ์ทอง

สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ

114 ถนนสุขุมวิท 23 แขวงคลองเตยเหนือ เขตวัฒนา กรุงเทพมหานคร 10110

E-mail: saranpimthong@gmail.com

ออกแบบปก

ดร.เก ประเสริฐสังข์

ออกแบบรูปเล่ม

ภาณุพงศ์ แก้วด้วง

พิมพ์ครั้งที่ 1

กันยายน พ.ศ. 2564 จำนวน 300 เล่ม

ข้อมูลทางบรรณานุกรมของสำนักหอสมุดแห่งชาติ

National Library of Thailand Cataloging in Publication Data

ศรัณย์ พิมพ์ทอง

การพัฒนาเครื่องมือวัดในการวิจัยพฤติกรรมศาสตร์.-- กรุงเทพมหานคร:

บริษัท บุ๊คพลัส พับลิชชิ่ง จำกัด, 2564.

206 หน้า.

1. การวัด. 2. เครื่องมือวัด. 3. พฤติกรรมศาสตร์. 4. การวิจัย. I. ชื่อเรื่อง.

ISBN : 978-616-586-432-9

จัดพิมพ์โดย

ศรัณย์ พิมพ์ทอง

88/16 ซ.สุขุมวิท 97/1

แขวงบางจาก เขตพระโขนง กรุงเทพมหานคร 10260

โทร. 098-469-3624

พิมพ์ที่

บริษัท บุ๊คพลัส พับลิชชิ่ง จำกัด

1111/16 หมู่บ้านกลางเมือง แยกรัชดา-ลาดพร้าว ซอย 2

แขวงจันทระเกษม เขตจตุจักร กรุงเทพมหานคร 10900

โทร. 0-2938-0950-9

คำนิยม

หนังสือเรื่องการพัฒนาเครื่องมือวัดในการวิจัยพฤติกรรมศาสตร์เล่มนี้ เป็นหนังสือที่มีความจำเป็นต่อนักวิจัยทางพฤติกรรมศาสตร์ทุกท่านที่ต้องอ่านทำความเข้าใจ และนำไปใช้ให้ถูกต้อง หนังสือมีเนื้อหา 11 บทที่เรียงลำดับเริ่มจากภาพรวมของการวิจัยทางพฤติกรรมศาสตร์ ตัวแปรประเภทต่างๆ เครื่องมือที่จะนำไปใช้ในการวัดตัวแปรในศาสตร์แต่ละศาสตร์ที่ครอบคลุมทั้ง ความคิด ความรู้สึก และพฤติกรรม รายละเอียดในหนังสือมีคำอธิบายและตัวอย่างให้นำไปใช้ได้อย่างชัดเจน

หลังจากผู้อ่านได้ทำความเข้าใจในส่วนของตัวแปรและประเภทของเครื่องมือเพื่อการเก็บรวบรวมข้อมูลแล้ว ส่วนต่อมาของหนังสือจะกล่าวถึงการนำข้อมูลที่ได้จากกลุ่มเป้าหมายมาวิเคราะห์เพื่อสร้างคุณภาพให้กับเครื่องมือทั้งในแง่ของความเที่ยงตรงและความเชื่อมั่นแบบต่างๆ ผู้เขียนได้อธิบายวิธีการที่เป็นขั้นตอนและแสดงตัวอย่างที่ชัดเจน นำไปปฏิบัติได้

หนังสือเล่มนี้มีความทันสมัยโดยผู้เขียนค้นคว้างานวิจัย ที่เกี่ยวข้องทั้งที่เป็นงานวิจัยของประเทศไทยและงานวิจัยจากต่างประเทศ ที่จะประโยชน์แก่ผู้สนใจด้านการสร้างเครื่องมือและการวัดทางพฤติกรรมศาสตร์จะนำไปสร้างโจทย์วิจัยให้เกิดเครื่องมือใหม่ๆ ที่มีคุณภาพสำหรับการวัดและเปลี่ยนแปลงบุคคลและสังคมได้ นอกจากนี้หนังสือเล่มนี้จะเหมาะแก่นักวิจัยทางพฤติกรรมศาสตร์แล้ว ยังเหมาะแก่นักวิจัยในศาสตร์สังคมศาสตร์อื่น ๆ อีกด้วย จึงขอให้เป็นหนังสือแนะนำสำหรับนักศึกษาระดับบัณฑิตศึกษาทุกคน ที่ควรมีไว้เพื่อเป็นที่ปรึกษาข้างตัว

รองศาสตราจารย์ ดร. ดุษฎี อินทรประเสริฐ

20 พฤษภาคม 2564

คำนำ

หนังสือ “การพัฒนาเครื่องมือวัดในการวิจัยพฤติกรรมศาสตร์” เล่มนี้ ประกอบด้วยบทต่างๆ ที่ครอบคลุมเนื้อหาสาระเกี่ยวกับหลักการ วิธีการสร้าง และตรวจสอบคุณภาพเครื่องมือวัดทางพฤติกรรมศาสตร์ นอกจากนี้ยังมีเนื้อหาสาระที่เกี่ยวข้องกับการใช้โปรแกรมสำเร็จรูป ประกอบหัวข้อการตรวจสอบคุณภาพเครื่องมือวัดด้วย

ผู้เขียนขอขอบพระคุณผู้ทรงคุณวุฒิ รองศาสตราจารย์ ดร. รัตติกรณ์ จงวิศาล และรองศาสตราจารย์ ดร. ปิ่นกนก วงศ์ปิ่นเพชร ที่ได้ให้คำแนะนำและข้อเสนอแนะ เพื่อให้หนังสือเล่มนี้มีคุณภาพเหมาะสมต่อการเผยแพร่ และขอขอบคุณ ดร. ชาริน สุวรรณวงศ์ ที่ช่วยสนับสนุนข้อมูลทางวิชาการจนทำให้หนังสือเล่มนี้มีความสมบูรณ์ในเนื้อหามากยิ่งขึ้น หากมีข้อบกพร่องใด ผู้เขียนพร้อมน้อมรับข้อคิดเห็นหรือข้อเสนอแนะจากทุกท่านเพื่อนำมาใช้ในการปรับปรุงหนังสือเล่มนี้ในโอกาสต่อไป

ผู้เขียนหวังเป็นอย่างยิ่งว่า หนังสือเล่มนี้จะเป็นประโยชน์ต่อนิสิต นักศึกษา นักวิจัย และผู้สนใจในการนำความรู้จากหนังสือไปใช้ในการพัฒนางานวิจัยไม่มากนัก

ศรัณย์ พิมพ์ทอง

สถาบันวิจัยพฤติกรรมศาสตร์
มหาวิทยาลัยศรีนครินทรวิโรฒ

สารบัญ

	หน้า
บทที่ 1 บทนำ	1
บทนำการพัฒนาเครื่องมือวัดในการวิจัยพฤติกรรมศาสตร์	1
แนวทางการศึกษาในหนังสือ	4
สรุป	5
เอกสารอ้างอิง	6
บทที่ 2 แนวคิดพื้นฐานเกี่ยวกับการวัดและเครื่องมือวัด	7
คำศัพท์ที่เกี่ยวข้องกับการวัด	7
ลักษณะการวัดทางพฤติกรรมศาสตร์	11
ประเภทของข้อมูล	12
ระดับของการวัด (Scale of Measurement)	15
การนิยามตัวแปร	18
ข้อจำกัดในการวัดทางพฤติกรรมศาสตร์	20
คุณลักษณะของเครื่องมือวัดที่ดี	21
ประเภทของเครื่องมือวัดทางพฤติกรรมศาสตร์	23
สรุป	23
เอกสารอ้างอิง	24
บทที่ 3 มาตรวัด	26
ความหมายของมาตรวัด	26
หลักการสร้างมาตรวัด	27
ประเภทของมาตรวัดที่สำคัญและการพัฒนามาตรวัด	29
ความคลาดเคลื่อนในการใช้มาตรวัด	42
สรุป	44
เอกสารอ้างอิง	44
บทที่ 4 การวัดด้านการรู้คิด	46
ความเป็นมาของการวัดทางด้านการรู้คิด	46
แนวคิดเบื้องต้นของการวัดทางด้านการรู้คิด	47

สารบัญ (ต่อ)

	หน้า
แบบทดสอบสติปัญญา (Intelligence Test)	47
แบบทดสอบความถนัด (Aptitude Test)	53
แบบทดสอบผลสัมฤทธิ์	58
สรุป	67
เอกสารอ้างอิง	68
บทที่ 5 การวัดด้านความรู้สึกรู้สึก	70
ความเป็นมาของการวัดทางด้านความรู้สึกรู้สึก	70
แนวคิดเบื้องต้นของการวัดทางด้านความรู้สึกรู้สึก	70
แบบวัดความสนใจ	71
แบบวัดค่านิยม	74
แบบวัดบุคลิกภาพ	77
สรุป	84
เอกสารอ้างอิง	84
บทที่ 6 การวัดด้านพฤติกรรม	88
ความเป็นมาของการวัดทางด้านพฤติกรรม	88
แนวคิดเบื้องต้นของการวัดทางด้านพฤติกรรม	89
วิธีการวัดทางด้านพฤติกรรม	93
สรุป	101
เอกสารอ้างอิง	101
บทที่ 7 การวิเคราะห์รายข้อ	104
แนวคิดเบื้องต้นของการวิเคราะห์รายข้อ	104
อำนาจจำแนกรายข้อ (Item Discrimination)	105
ความยากง่ายรายข้อ (Item Difficulty)	110
อำนาจตัวลวง (Distractor Power)	113
ความสัมพันธ์ระหว่างการวิเคราะห์รายข้อรูปแบบต่าง ๆ	114
การวิเคราะห์รายข้อเชิงคุณภาพ (Qualitative Item Analysis)	115
การปรับปรุงแบบทดสอบ (Test Revision)	117

สารบัญ (ต่อ)

	หน้า
สรุป	118
เอกสารอ้างอิง	119
บทที่ 8 ความเที่ยงตรง	120
แนวคิดเบื้องต้นของความเที่ยงตรง	120
ประเภทของความเที่ยงตรง	121
ปัจจัยที่ส่งผลต่อความเที่ยงตรง	134
สรุป	136
เอกสารอ้างอิง	136
บทที่ 9 ความเชื่อมั่น	139
แนวคิดเบื้องต้นของความเชื่อมั่น	139
วิธีการประมาณค่าความเชื่อมั่น	142
ปัจจัยที่ส่งผลต่อความเชื่อมั่น	155
สรุป	156
เอกสารอ้างอิง	157
บทที่ 10 การวิเคราะห์ห้อยค์ประกอบ	159
แนวคิดเบื้องต้นของการวิเคราะห์ห้อยค์ประกอบ	159
การวิเคราะห์ห้อยค์ประกอบเชิงสำรวจ	160
การวิเคราะห์ห้อยค์ประกอบเชิงยืนยัน	170
สรุป	180
เอกสารอ้างอิง	180
บทที่ 11 ทฤษฎีการตอบสนองรายข้อ	182
ภาพรวมของทฤษฎีการตอบสนองรายข้อ	182
ความแตกต่างระหว่างทฤษฎีการทดสอบดั้งเดิมและทฤษฎีการตอบสนองรายข้อ	184
แนวคิดเบื้องต้นของทฤษฎีการตอบสนองรายข้อ	185
พารามิเตอร์ในทฤษฎีการตอบสนองรายข้อ	186
โค้งแสดงคุณลักษณะรายข้อ	188
โมเดลการตอบสนองรายข้อ	189

สารบัญ (ต่อ)

	หน้า
การประเมินความสอดคล้องของโมเดล	196
สรุป	198
เอกสารอ้างอิง	198
ดัชนีคำศัพท์	201
ประวัติผู้เขียน	206



สารบัญตาราง

ตารางที่	รายการ	หน้า
2-1	คุณลักษณะของมาตรการวัดทั้ง 4 ระดับ	24
3-1	คู่มือคัมภีร์ที่ใช้กันบ่อยในมาตรวัดออสกูด	39
4-1	องค์ประกอบของแบบทดสอบ Stanford-Binet ฉบับที่ 5	49
4-2	รายชื่อการทดสอบและคำอธิบายในแบบทดสอบ Wechsler	50
4-3	องค์ประกอบและคำอธิบายของแบบวัดความถนัดทั่วไปในแต่ละองค์ประกอบ	54
4-4	วัตถุประสงค์การเรียนรู้ด้านการรู้คิดของ Bloom (1956)	62
4-5	การกำหนดลักษณะเฉพาะของแบบทดสอบผลสัมฤทธิ์การวัดการรู้คิด จำนวน 50 ข้อ	62
5-1	บุคลิกภาพ 16 องค์ประกอบ รวมทั้งดัชนีอันดับที่สอง 4 ตัว จาก 16 PF	78
7-1	การวิเคราะห์รายข้อเชิงคุณภาพรูปแบบต่าง ๆ	115
8-1	ลักษณะเมทริกซ์พหุคุณลักษณะพหุวิธี เมื่อวัด 2 ลักษณะด้วยการวัด 2 วิธี	132
9-1	คุณลักษณะของวิธีการประมาณค่าความเชื่อมั่น	142

สารบัญภาพ

ภาพที่	รายการ	หน้า
1-1	พฤติกรรมศาสตร์เป็นแนวคิดสหวิทยาการ	2
1-2	ขั้นตอนการวิจัย	4
2-1	การนิยามตัวแปร	20
10-1	ขั้นตอนที่ 1-3 วิธีการวิเคราะห์หองค์ประกอบเชิงสำรวจ	162
10-2	ขั้นตอนที่ 4-7 วิธีการวิเคราะห์หองค์ประกอบเชิงสำรวจ	165
11-1	ตัวอย่างโค้งแสดงคุณลักษณะรายข้อ (ICC)	188
11-2	ตัวอย่างโค้งแสดงคุณลักษณะรายข้อ (ICC) เปรียบเทียบข้อคำถาม 2 ข้อ	189
11-3	โมเดลโลจิสติกแบบ 1 พารามิเตอร์	190
11-4	โมเดลโลจิสติกแบบ 2 พารามิเตอร์	191
11-5	โมเดลโลจิสติกแบบ 3 พารามิเตอร์	192
11-6	ข้อตกลงเบื้องต้นของโมเดลการตอบสนองรายข้อแบบพหุมิติ	195

บทที่ 1

บทนำ

บทนำการพัฒนาเครื่องมือวัดในการวิจัยพฤติกรรมศาสตร์

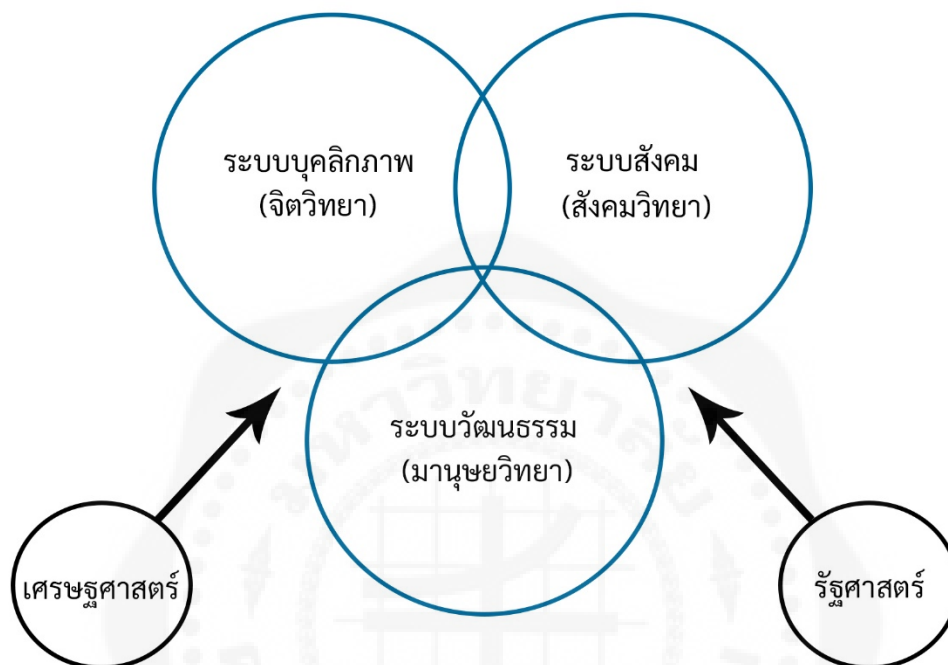
การพัฒนาเครื่องมือวัด ถือเป็นองค์ประกอบหนึ่งที่มีความสำคัญของการวิจัย ไม่ว่าจะเป็นการวิจัยทางด้านวิทยาศาสตร์ การวิจัยทางด้านสังคม รวมทั้งการวิจัยทางด้านพฤติกรรมศาสตร์ด้วยเช่นกัน แม้ว่าโดยทั่วไปแล้ว พฤติกรรมศาสตร์ไม่ได้มีลักษณะเหมือนวิทยาศาสตร์พื้นฐาน เช่น ฟิสิกส์ เคมี ชีววิทยา เป็นต้น แต่พฤติกรรมศาสตร์ถือเป็นศาสตร์ที่มีความท้าทายในการศึกษาวิจัย สาเหตุที่ศาสตร์นี้มีความท้าทาย เนื่องจากมีจุดเน้นที่คุณลักษณะ และพฤติกรรมของมนุษย์มากกว่าลักษณะที่เป็นวัตถุวิสัยเช่นการศึกษาในทางวิทยาศาสตร์ จึงทำให้การวัดมีความซับซ้อนกว่ามาก ดังนั้น แนวคิดการวัดจึงเป็นพื้นฐานที่สำคัญที่ช่วยส่งเสริมการศึกษาและวิจัยในระดับบุคคล นั่นคือ การศึกษาพฤติกรรมของมนุษย์นั่นเอง (Nesselrode & Molenaar, 2016)

สำหรับขอบข่ายเนื้อหาของพฤติกรรมศาสตร์ ประกอบด้วย ศาสตร์ที่เป็นหลัก จำนวน 3 ศาสตร์ นั่นคือ จิตวิทยา (Psychology) สังคมวิทยา (Sociology) และมานุษยวิทยา (Anthropology) และมีศาสตร์อื่น ๆ ที่เป็นศาสตร์รองลงมา ซึ่งขึ้นอยู่กับจุดเน้นของแต่ละหลักสูตรหรือพฤติกรรมที่ต้องการศึกษา ซึ่งอาจจะครอบคลุมถึง รัฐศาสตร์ (Political Science) แพทยศาสตร์ (Medicine) นิติศาสตร์ (Laws) เศรษฐศาสตร์ (Economics) ภูมิศาสตร์ (Geography) ชีววิทยา (Biology) สาธารณสุขศาสตร์ (Public Health) บริหารธุรกิจ (Business) ศึกษาศาสตร์ (Education) และอื่น ๆ (ยุทธนา ไชยจุกุล, 2559) ตัวอย่างเช่น

ศรัณย์ พิมพ์ทอง (2557) ได้ทำการศึกษาปัจจัยเชิงสาเหตุที่มีอิทธิพลต่อการคงอยู่ในองค์กรของพนักงานมหาวิทยาลัยสายวิชาการในประเทศไทย ซึ่งงานวิจัยดังกล่าวเกี่ยวข้องกับจิตวิทยาและสังคมวิทยา รวมทั้งศาสตร์ทางด้านบริหารธุรกิจ

สวรรณยา ทองมั่น และนริสรา พึ่งโพธิ์สภ (2562) ศึกษาปัจจัยทางจิตสังคมที่เกี่ยวข้องกับพฤติกรรมการเรียนแบบสร้างความรู้ด้วยตนเองของนักเรียนระดับชั้นมัธยมศึกษาตอนต้น สำนักงานเขตพื้นที่การศึกษาเขต 33 ซึ่งงานวิจัยดังกล่าวเกี่ยวข้องกับจิตวิทยาและสังคมวิทยา รวมทั้งศาสตร์ที่เกี่ยวข้องทางการศึกษา

ดังนั้น จะเห็นได้ว่า ขอบข่ายเนื้อหาของพฤติกรรมศาสตร์มีความเป็นสหวิทยาการ (Multidisciplinary) (Mullins, 1996) การแก้ไขปัญหาเกี่ยวกับพฤติกรรมใด ๆ ของมนุษย์ จำเป็นต้องอาศัยการทำความเข้าใจจากหลากหลายศาสตร์ ดังภาพที่ 1-1



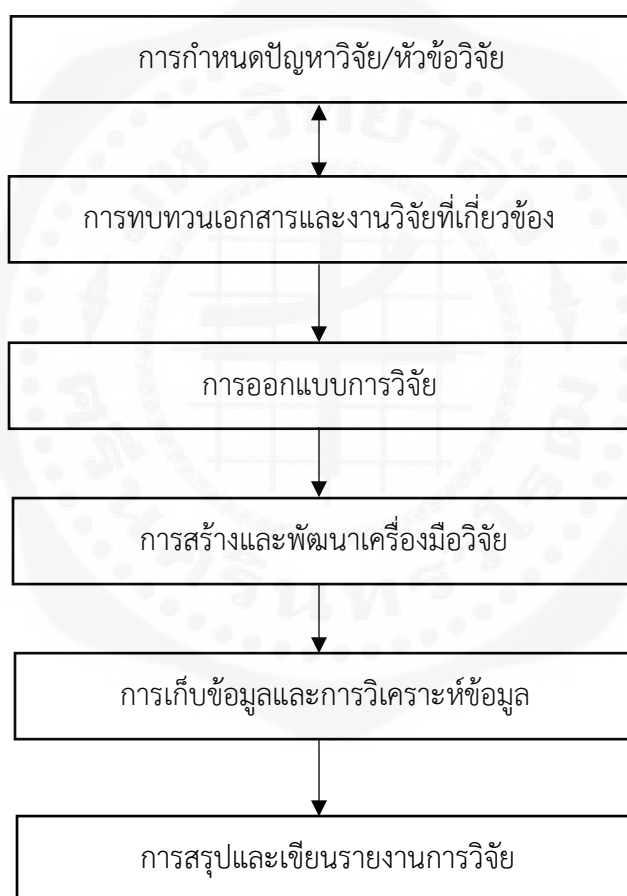
ภาพที่ 1-1 พฤติกรรมศาสตร์เป็นแนวคิดสหวิทยาการ
ที่มา: Mullin (1996)

จากภาพที่ 1 แสดงให้เห็นถึง ศาสตร์ทางจิตวิทยา มองเป็นระบบบุคลิกภาพ (Personality System) มุ่งเน้นคุณลักษณะของบุคคล ได้แก่ การรับรู้ เจตคติ แรงจูงใจ รวมทั้งลักษณะนิสัยของบุคคลหนึ่งๆ ส่วนศาสตร์ทางสังคมวิทยา ให้ความสนใจในระบบสังคม (Social System) มุ่งเน้นไปที่พฤติกรรมทางสังคม ความสัมพันธ์ในกลุ่มทางสังคม โครงสร้างทางสังคม ตำแหน่งสถานภาพ และหน้าที่ต่าง ๆ ในสังคม และศาสตร์ทางมานุษยวิทยา ให้ความสำคัญกับระบบวัฒนธรรม (Cultural System) ประกอบด้วย บริบทพื้นที่ที่หลากหลายภายในกลุ่มหรือสังคม ได้แก่ ความรู้ ความเชื่อ พิธีกรรม ประเพณี ค่านิยม ศิลธรรม บรรทัดฐาน กฎหมาย และความสามารถหรือนิสัยที่ได้มาจากมนุษย์ในฐานะเป็นสมาชิกของสังคม นอกจากนี้ ยังมีศาสตร์อื่น ๆ ที่เกี่ยวข้อง เช่น เศรษฐศาสตร์ รัฐศาสตร์ เป็นต้น (Mullins, 1996)

นอกจาก การพิจารณาว่า พฤติกรรมศาสตร์ ต้องเป็นเรื่องที่เกี่ยวข้องกับพฤติกรรมของมนุษย์แล้ว ยังมีหลักเกณฑ์พิจารณาอื่นๆ ที่สามารถบอกได้ว่าเป็น “พฤติกรรมศาสตร์” (จรรยา สุวรรณทัต, 2559) นั่นคือ

วิธีการศึกษาต้องเป็นวิธีการทางวิทยาศาสตร์ ในการได้มาซึ่งข้อสรุปเกี่ยวกับพฤติกรรมของมนุษย์ จำเป็นต้องมี การศึกษาค้นคว้า การทดลอง การทดสอบ เพื่อให้ได้ข้อมูลต่างๆ มาสนับสนุน มีการเก็บรวบรวมข้อมูลอย่างเป็นระบบ รวมทั้งวัตถุประสงค์ของวิธีการศึกษานั้น ต้องเป็นการทำความเข้าใจ อธิบาย และพยากรณ์พฤติกรรมของมนุษย์ อันเป็นสิ่งสำคัญและจำเป็นสำหรับการแก้ไขปัญหาพฤติกรรมของบุคคลและสังคม สิ่งเหล่านี้จึงถือว่าเป็น การศึกษาทางด้านพฤติกรรมศาสตร์

ในการศึกษาทางด้านพฤติกรรมศาสตร์ มีการวิจัยมากมายหลากหลายประเภท ทั้งการวิจัยเชิงสหสัมพันธ์ การวิจัยเชิงคุณภาพ การวิจัยเชิงทดลอง การวิจัยและพัฒนา การวิจัยผสานวิธี เป็นต้น ซึ่งงานวิจัยเหล่านี้ มีขั้นตอน การวิจัยในรูปแบบที่คล้ายคลึงกัน ดังภาพที่ 2-1



ภาพที่ 1-2 ขั้นตอนการวิจัย

สำหรับขั้นตอนการวิจัยดังกล่าว จะเห็นว่า การวิจัยหรือการศึกษาทางด้านพฤติกรรมศาสตร์เริ่มต้นด้วยการกำหนดปัญหาหรือหัวข้อการวิจัย ไปพร้อม ๆ กับการทบทวนเอกสารและงานวิจัยที่เกี่ยวข้อง จนนำไปสู่ การออกแบบการวิจัยว่าเป็นงานวิจัยรูปแบบใด มีวัตถุประสงค์การวิจัยเป็นเช่นไร มีประชากรและกลุ่มตัวอย่าง

รวมทั้งการสุ่มกลุ่มตัวอย่าง มีการเก็บรวบรวมข้อมูล และการวิเคราะห์ข้อมูลเป็นอย่างไร จากนั้น จึงนำมาสู่การสร้างและพัฒนาเครื่องมือวิจัยให้เหมาะสมกับการออกแบบการวิจัย เมื่อสร้างเครื่องมือเสร็จ จึงนำเครื่องมือไปตรวจสอบคุณภาพเครื่องมือวิจัย แล้วจึงนำไปเก็บรวบรวมข้อมูลจริง ทำการวิเคราะห์ข้อมูล และสุดท้าย จึงนำมาเขียนและสรุปรายงานการวิจัย

จากขั้นตอนการวิจัยดังกล่าว จะเห็นได้ว่า ขั้นตอนการสร้างและพัฒนาเครื่องมือวิจัย เป็นขั้นตอนที่มีความสำคัญเป็นอย่างมาก นั่นคือ หากเครื่องมือวิจัยขาดคุณภาพ จะทำให้ได้ข้อมูลที่ไม่มีคุณภาพด้วย อาจเป็นไปได้ว่า ข้อมูลวัดไม่ตรงตามสิ่งที่นักวิจัยต้องการวัด เมื่อนำข้อมูลมาวิเคราะห์ จะทำให้เกิดการรายงานผลการวิจัยที่คลาดเคลื่อนไปจากความเป็นจริงได้ ซึ่งจากประสบการณ์ของผู้เขียน พบว่า นักวิจัยหน้าใหม่ เช่น กลุ่มนิสิตนักศึกษาที่กำลังพัฒนาปริญญาโท มักประสบปัญหาในการสร้างเครื่องมือและพัฒนาเครื่องมือวิจัย โดยมีปัญหาที่แตกต่างกันไป ได้แก่ การนำเครื่องมือวิจัยที่มีการสร้างไว้แล้ว มาใช้ในการวิจัยของตน โดยไม่ได้ปรับให้เข้ากับบริบทงานวิจัยของตน การสร้างเครื่องมือวิจัยที่ไม่ได้สะท้อนถึงนิยามปฏิบัติการ รวมทั้งไม่มีความรู้ในการตรวจสอบคุณภาพเครื่องมือวิจัย จึงทำให้ได้ผลการวิเคราะห์ที่อาจไม่ตรงกับความเป็นจริง ดังนั้น จึงมีความจำเป็นอย่างยิ่งที่ควรมีความรู้และทักษะในการสร้างและพัฒนาเครื่องมือวิจัยทางด้านพฤติกรรมศาสตร์ เพื่อให้ได้เครื่องมือวิจัยที่มีคุณภาพ สอดคล้องกับงานวิจัย และทำให้ได้งานวิจัยที่มีคุณภาพต่อไป

แนวทางการศึกษาในหนังสือ

การวัดทางพฤติกรรมศาสตร์ จึงเป็นการวัดที่เกี่ยวข้องกับพฤติกรรมของมนุษย์ โดยในหนังสือเล่มนี้ มีแนวทางการสรุปเกี่ยวกับการพัฒนาเครื่องมือวัดในการวิจัยทางด้านพฤติกรรมศาสตร์ เพื่อให้ผู้อ่านได้ทำความเข้าใจเบื้องต้นในรายละเอียดก่อนที่จะเข้าสู่บทถัดไป เพื่อให้เกิดความกระจ่างชัดและความเข้าใจตรงกัน ผู้เขียนจึงได้สรุปแนวทางไว้เป็นเครื่องมือที่สำคัญในการนำไปสู่การสร้างและพัฒนาเครื่องมือวัดทางการวิจัยพฤติกรรมศาสตร์ ได้อย่างถูกต้องและถูกวิธี โดยกระบวนการในการสร้างและพัฒนาเครื่องมือวัดทางด้านพฤติกรรมศาสตร์ สามารถแบ่งได้เป็น 3 ระยะ นั่นคือ 1) การพัฒนาข้อคำถาม (Item Development) 2) การพัฒนาเครื่องมือวัด (Scale Development) และ 3) การประเมินเครื่องมือวัด (Scale Evaluation) (Boateng, Neilands, Frongillo, Melgar-Quinonez, & Young, 2018)

แนวทางในหนังสือเล่มนี้ ผู้เขียนต้องการให้ผู้อ่านได้ทำความเข้าใจในแนวคิดพื้นฐานเกี่ยวกับการวัดและเครื่องมือวัด ทั้งลักษณะของการวัดทางพฤติกรรมศาสตร์ ประเภทของข้อมูล ระดับการวัด การนิยามตัวแปร ข้อจำกัดในการวัด คุณลักษณะของเครื่องมือที่ดี ประโยชน์ของการใช้เครื่องมือวัด ประเภทของเครื่องมือวัด โดยได้อธิบายไว้ในบทที่ 2 เมื่อผู้อ่านทำความเข้าใจบทที่ 2 แล้ว จึงนำไปสู่บทที่ 3 เป็นบทที่เกี่ยวข้องกับมาตรวัด แนวคิดวิธีการสร้างมาตรวัด มาตรวัดประเภทต่าง ๆ รวมทั้งความคลาดเคลื่อนในการใช้มาตรวัด การทำความเข้าใจมาตรวัด จะช่วยทำให้ผู้อ่านเลือกมาตรวัดที่เหมาะสม กับเนื้อหาที่จะทำการวัด เนื้อหาวัดทางการรู้คิด (Cognitive)

ในบทที่ 4 หรือเนื้อหาการวัดอารมณ์ความรู้สึก (Affective) ในบทที่ 5 เป็นเนื้อหาการวัดทางด้านพฤติกรรม (Behavior) ในบทที่ 6

โดยบทที่ 4-6 จะนำเสนอการวัดในรูปแบบต่าง ๆ วิธีการสร้างข้อคำถามที่เหมาะสม ข้อคำถามไม่มีความซับซ้อน กำกวม หรือคลุมเครือ เมื่อสร้างข้อคำถามเสร็จสิ้นแล้ว จนถึงกระบวนการที่นักวิจัยจะนำข้อคำถามเพื่อให้ผู้เชี่ยวชาญหรือผู้ทรงคุณวุฒิตรวจสอบเนื้อหาว่ามีความครบถ้วนเหมาะสมตรงกับนิยามปฏิบัติการ โดยจะอธิบายในรายละเอียด ซึ่งปรากฏในบทที่ 7 ทั้งหมดที่กล่าวมาข้างต้น ถือเป็นกระบวนการสร้างและพัฒนาเครื่องมือวัดในระยะที่ 1 นั่นคือ การพัฒนาข้อคำถาม (Item Development)

เข้าสู่ระยะที่ 2 การพัฒนาเครื่องมือ (Scale Development) โดยรายละเอียดปรากฏในบทที่ 8 อันเป็นการวิเคราะห์ข้อคำถามรายข้อ เพื่อให้ได้ข้อคำถามที่มีคุณภาพ ทั้งค่าอำนาจจำแนก ค่าความยากง่าย ตัวลวง เป็นต้น นอกจากนี้ ยังมีการวิเคราะห์องค์ประกอบ (Factor Analysis) เพื่อสกัดเป็นองค์ประกอบของข้อคำถามทั้งหมด ซึ่งจะปรากฏในบทที่ 9 จะเป็นการวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis: EFA) เป็นการวิเคราะห์เพื่อลดข้อคำถามให้เกิดความชัดเจนเหมาะสม (Item reduction analysis)

จากนั้น จึงเข้าสู่กระบวนการสร้างและพัฒนาเครื่องมือวัดในระยะที่ 3 คือ การประเมินเครื่องมือวัด (Scale Evaluation) ปรากฏในบทที่ 9 ในส่วนการวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) อันเป็นการวิเคราะห์ความเที่ยงตรงเชิงโครงสร้าง ความเที่ยงตรงเชิงเสมือน ความเที่ยงตรงเชิงจำแนก โดยปรากฏรายละเอียดในบทที่ 7 ซึ่งเกี่ยวข้องกับความเที่ยงตรง (Validity) รวมทั้งได้นำเสนอความเชื่อมั่น (Reliability) ในหลายรูปแบบในบทที่ 10 นอกจากนี้ ผู้เขียนได้นำเสนอทฤษฎีการตอบสนองรายข้อ (Item Response Theory) ในบทที่ 11 ซึ่งเป็นวิธีการหนึ่งในการตรวจสอบคุณภาพเครื่องมือด้วยเช่นกัน

สรุป

ในกระบวนการสร้างและพัฒนาเครื่องมือวัดทางพฤติกรรมศาสตร์ อธิบายไว้ 3 ระยะ นั่นคือ ระยะที่ 1 การพัฒนาข้อคำถาม ระยะที่ 2 การพัฒนาเครื่องมือ และระยะที่ 3 การประเมินเครื่องมือวัด ซึ่งมีความแตกต่างกันไปในแต่ละขั้นตอน การศึกษาบางเรื่องที่ยังไม่เคยมีการสร้างเครื่องมือวัดมาก่อน อาจเป็นตัวแปรใหม่ หรือมีการสร้างเครื่องมือมาใช้แล้ว แต่มีความแตกต่างด้านบริบท สมควรที่จะต้องปรับและพัฒนาข้อคำถามใหม่ จึงควรดำเนินการพัฒนาให้ครบทั้งสามระยะ และครบทุกขั้นตอนตามภาพที่ 1-2 แต่เครื่องมือที่ผ่านการพัฒนาและตรวจสอบคุณภาพของเครื่องมือมาแล้ว ก็สามารถนำมาใช้ประเมินเครื่องมือวัดในระยะที่ 3 ได้เลย โดยไม่ต้องผ่านสองระยะก่อนหน้า ผู้เขียนจึงหวังว่า เนื้อหาในหนังสือเล่มนี้จะช่วยให้ผู้อ่านหรือผู้วิจัยสามารถเลือกระยะหรือขั้นตอนของการพัฒนาเครื่องมือวัดได้อย่างเหมาะสม

เอกสารอ้างอิง

- จรรยา สุวรรณทัต. (2559). แนวคิดหลักพฤติกรรมศาสตร์และการประยุกต์ใช้ในการสร้างหลักสูตรการวิจัยพฤติกรรมศาสตร์. ใน ยุทธนา ไชยจุลกุล (บรรณาธิการ), *พฤติกรรมศาสตร์: มุมมองในศาสตร์ที่แตกต่าง* (หน้า 47-58). กรุงเทพฯ: สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- ยุทธนา ไชยจุลกุล. (2559). ความรู้เบื้องต้นเกี่ยวกับพฤติกรรมศาสตร์. ใน ยุทธนา ไชยจุลกุล (บรรณาธิการ), *พฤติกรรมศาสตร์: มุมมองในศาสตร์ที่แตกต่าง* (หน้า 1-36). กรุงเทพฯ: สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- ศรัณย์ พิมพ์ทอง. (2557). ปัจจัยเชิงสาเหตุที่มีอิทธิพลต่อการคงอยู่ในองค์การของพนักงานมหาวิทยาลัยสายวิชาการในประเทศไทย. *วารสารบริหารธุรกิจ*, 37(142), 16-32.
- สวรรยา ทองแมน, และนริศรา พิงโพธิ์สภ. (2562). ปัจจัยทางจิตสังคมที่เกี่ยวข้องกับพฤติกรรมการเรียนแบบสร้างความรู้ด้วยตนเองของนักเรียนระดับชั้นมัธยมศึกษาตอนต้น สำนักงานเขตพื้นที่การศึกษาเขต 33. *วารสารพฤติกรรมศาสตร์เพื่อการพัฒนา*, 11(2), 58-77.
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6, 149.
- Mullins, L. J. (1996). *Management and organisational behavior* (4th ed.). London: Pitman Publishing.
- Nesselrode, J. R., & Molenaar, P. C. M. (2016). Some behavioral science measurement concerns and proposals. *Multivariate Behavioral Research*, 51(2), 396-412.

บทที่ 2

แนวคิดพื้นฐานเกี่ยวกับการวัดและเครื่องมือวัด

บทที่ 2 เป็นการนำเสนอแนวคิดพื้นฐานเกี่ยวกับการวัดและเครื่องมือวัด โดยผู้เขียนต้องการให้ผู้อ่านได้ทราบถึงแนวคิดพื้นฐานที่เกี่ยวข้องกับการวัดและเครื่องมือการวัด เนื่องจากแนวคิดพื้นฐานดังกล่าว ทำให้เกิดความเข้าใจถึงที่มาของการสร้างและพัฒนาเครื่องมือวิจัย จนนำไปสู่การมีเครื่องมือวิจัยที่มีคุณภาพและมีมาตรฐานเป็นที่ยอมรับในการทำงานวิชาการ

คำศัพท์ที่เกี่ยวข้องกับการวัด

เมื่อกล่าวถึงการวัด ผู้อ่านควรทำความเข้าใจคำศัพท์ที่เกี่ยวข้อง ดังนี้

ตัวแปร (Variables) หมายถึง สิ่งที่ผู้วิจัยได้ทำการศึกษาในงานวิจัย หรืออีกนัยหนึ่ง ตัวแปรเป็นคุณลักษณะหรือคุณสมบัติหรือปรากฏการณ์ที่ผู้วิจัยสนใจจะศึกษาค้นหาความจริง อาจจะเป็นสิ่งที่มีชีวิตหรือไม่มีชีวิตก็ได้ เช่น คน สัตว์ สิ่งของ หรือเหตุการณ์ต่าง ๆ เป็นต้น ตัวแปรจะต้องมีค่าเปลี่ยนแปลงได้อย่างน้อยตั้งแต่ 2 ค่าขึ้นไป เช่น

- ตัวแปรเพศ สามารถแปรได้เป็น เพศหญิงและเพศชาย
 - ตัวแปรอายุ สามารถแปรได้เป็นกลุ่มตามที่ผู้วิจัยกำหนด เช่น ต่ำกว่า 20 ปี, 20-29 ปี, 30-39 ปี, 40-49 ปี, 50-59 ปี และ 60 ปีขึ้นไป
 - ตัวแปรอาชีพ สามารถแปรได้เป็น ข้าราชการ รัฐวิสาหกิจ รับจ้างทั่วไป ค้าขาย ผู้ประกอบการ ไม่ได้ประกอบอาชีพ
 - ตัวแปรรายได้ อาจสามารถแปรได้เป็น ต่ำกว่า 10,000 บาท, 10,000-50,000 บาท, 50,001-100,000 บาท และมากกว่า 100,000 บาทขึ้นไป
 - ตัวแปรศาสนา อาจสามารถแปรได้เป็น ศาสนาพุทธ ศาสนาอิสลาม ศาสนาคริสต์ และศาสนาอื่น ๆ
- ตัวแปรที่ศึกษาทางด้านพฤติกรรมศาสตร์ส่วนใหญ่ เป็นตัวแปรที่เกี่ยวข้องกับคุณลักษณะและลักษณะพฤติกรรมต่าง ๆ ได้แบ่งประเภทของตัวแปร ดังนี้

1. ตัวแปรอิสระ (Independent Variable) หมายถึง ตัวแปรที่เกิดขึ้นก่อน หรืออาจกล่าวได้ว่าเป็นตัวแปรเชิงสาเหตุที่ทำให้เกิดผลตามมา

2. ตัวแปรตาม (Dependent Variable) หมายถึง ตัวแปรที่เกิดขึ้นเนื่องจากตัวแปรอิสระ หรืออาจกล่าวได้ว่า เป็นตัวแปรเชิงผลเมื่อมีตัวแปรอิสระเป็นตัวแปรเชิงสาเหตุ

ยกตัวอย่างตัวแปรอิสระและตัวแปรตาม

“ผลการพัฒนาความรอบรู้ด้านสุขภาพด้วยการจัดการเรียนรู้เพื่อการเปลี่ยนแปลงที่มีต่อพฤติกรรม การควบคุมระดับน้ำตาลในเลือดของผู้ป่วยโรคเบาหวาน” (อารยา เขียงของ, 2561)

ตัวแปรอิสระ คือ การพัฒนาความรอบรู้ด้านสุขภาพด้วยการจัดการเรียนรู้เพื่อการเปลี่ยนแปลง

ตัวแปรตาม คือ พฤติกรรมการควบคุมระดับน้ำตาลในเลือด

“ปัจจัยทางจิตสังคมที่เกี่ยวข้องกับพฤติกรรมป้องกันการรับและแพร่โรคในสถานศึกษาของ นักศึกษาระดับปริญญาตรี” (เบญจพร ประณีตวาทกุล และตุจเดือน พันธุมนาวิน, 2561)

ตัวแปรอิสระ คือ ปัจจัยทางจิตสังคม

ตัวแปรตาม คือ พฤติกรรมป้องกันการรับและแพร่โรคในสถานศึกษา

“ปัจจัยทางจิตสังคมที่ส่งผลต่อพฤติกรรมการเป็นพลเมืองดีบนโลกดิจิทัลของนักเรียนระดับชั้น มัธยมศึกษาตอนปลายของโรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ในเขตกรุงเทพมหานคร” (กรกนก แชดวง, ศรีณีย์ พิมพ์ทอง, และพิชญานี พูนพล, 2562)

ตัวแปรอิสระ คือ ปัจจัยทางจิตสังคม

ตัวแปรตาม คือ พฤติกรรมการเป็นพลเมืองดีบนโลกดิจิทัล

“ประสิทธิผลของการฝึกอบรมทางจิตพฤติกรรมศาสตร์เพื่อพัฒนาพฤติกรรมเอื้อสังคมในกลุ่มนักเรียนชั้น มัธยมศึกษา” (ศรีณีย์ พิมพ์ทอง, 2563)

ตัวแปรอิสระ คือ การฝึกอบรมทางจิตพฤติกรรมศาสตร์

ตัวแปรตาม คือ พฤติกรรมเอื้อสังคม

3. ตัวแปรคั่นกลาง (Mediator Variable) หมายถึง ตัวแปรที่แทรกระหว่างตัวแปรอิสระกับตัวแปรตาม เพื่อจะได้ก่อให้เกิดเส้นทางตัวแปรอิสระไปสู่ตัวแปรตามขึ้น หรือเป็นตัวแปรเชื่อมโยงตรงกลางระหว่างตัวแปรอิสระกับตัวแปรตาม นั่นคือ ถ้าไม่มีตัวแปรคั่นกลาง ความสัมพันธ์ตามเส้นทางระหว่างตัวแปรอิสระกับตัวแปรตามก็อาจไม่เกิดขึ้น

4. ตัวแปรกำกับ (Moderator Variable) หมายถึง ตัวแปรที่มีอิทธิพลทำให้ขนาดหรือทิศทางความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามเปลี่ยนแปลงไปตามค่าของตัวแปรนั้น หรือตัวแปรที่ไปปรับเปลี่ยนความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามในลักษณะที่อาจจะเพิ่มหรือลดความสัมพันธ์นั้น

ยกตัวอย่างตัวแปรคั่นกลางและตัวแปรกำกับ

“ปัจจัยเชิงสาเหตุที่มีต่อการพัฒนาหัวข้อวิจัยและการเผยแพร่ผลงานวิจัยของอาจารย์มหาวิทยาลัยศรีนครินทรวิโรฒ โดยส่งผ่านคุณลักษณะที่ดีของการเป็นนักวิจัยที่เลี้ยง” (อังคินันท์ อินทรกำแหง และคณะ, 2561)

ตัวแปรอิสระ คือ ปัจจัยเชิงสาเหตุ

ตัวแปรตาม คือ การพัฒนาหัวข้อวิจัยและการเผยแพร่ผลงานวิจัย

ตัวแปรคั่นกลาง คือ คุณลักษณะที่ดีของการเป็นนักวิจัยที่เลี้ยง

“ความสัมพันธ์ระหว่างการรับรู้ความยุติธรรมในองค์การโดยรวมและความผูกพันที่บุคลากรมีต่อองค์การ หัวหน้างาน และกลุ่มเพื่อนร่วมงาน โดยมีวัฒนธรรมระยะห่างอำนาจเป็นตัวแปรกำกับ” (จารุพรรณ สุวรรณไพโรจน์ และประพิมพา จรัสรัตนกุล, 2560)

ตัวแปรอิสระ คือ การรับรู้ความยุติธรรมในองค์การโดยรวม

ตัวแปรตาม คือ ความผูกพันที่บุคลากรมีต่อองค์การ หัวหน้างาน และกลุ่มเพื่อนร่วมงาน

ตัวแปรกำกับ คือ วัฒนธรรมระยะห่างอำนาจ

“อิทธิพลของบุคลิกภาพแบบเปิดตัวและบุคลิกภาพแบบเป็นมิตร ต่อความพึงพอใจในความสัมพันธ์ของคู่รักโดยมีรูปแบบความผูกพันเป็นตัวแปรส่งผ่านและเพศเป็นตัวแปรกำกับ” (ณัฐชรินทร์ อภิวิษณุโชษิตา, 2556)

ตัวแปรอิสระ คือ บุคลิกภาพแบบเปิดตัวและบุคลิกภาพแบบเป็นมิตร

ตัวแปรตาม คือ ความพึงพอใจในความสัมพันธ์

ตัวแปรคั่นกลาง คือ รูปแบบความผูกพัน

ตัวแปรกำกับ คือ เพศ

5. ตัวแปรแทรกซ้อน (Extraneous Variable) หมายถึง ตัวแปรที่มีผลต่อการเปลี่ยนแปลงค่าของตัวแปรตาม แต่เป็นตัวแปรที่ผู้วิจัยไม่ได้ศึกษา ในกรณีงานวิจัยใด ๆ ที่ไม่ได้ควบคุมตัวแปรแทรกซ้อน และมีผลทำให้ตัวแปรตามเปลี่ยนแปลงไป ซึ่งตัวแปรตามที่เปลี่ยนแปลงไปนั้นไม่สามารถบอกได้ว่ามาจากตัวแปรอิสระ หรือเป็นตัวแปรแทรกซ้อน อาจเป็นตัวแปรอิสระร่วมกับตัวแปรแทรกซ้อน

ยกตัวอย่างตัวแปรแทรกซ้อน

“ผลของกิจกรรมควบคุมตนเองต่อพฤติกรรมการป้องกันการสูบบุหรี่ของนักศึกษามหาวิทยาลัย”

ตัวแปรอิสระ คือ กิจกรรมควบคุมตนเอง

ตัวแปรตาม คือ พฤติกรรมการป้องกันการสูบบุหรี่

ตัวแปรแทรกซ้อน คือ เพศ การสนับสนุนทางสังคม

จากตัวอย่างดังกล่าว พบว่า เพศและการได้รับการสนับสนุนทางสังคมที่แตกต่างกัน มีผลทำให้พฤติกรรมการป้องกันการสูบบุหรี่แตกต่างกัน ดังนั้น จึงมีความเป็นไปได้ที่เพศและการสนับสนุนทางสังคมอาจจะเป็นตัวแปรแทรกซ้อน ที่มีผลทำให้พฤติกรรมการป้องกันการสูบบุหรี่เพิ่มขึ้น จึงไม่สามารถสรุปได้อย่างหนักแน่นว่า กิจกรรมควบคุมตนเองมีผลทำให้พฤติกรรมการป้องกันการสูบบุหรี่เพิ่มขึ้น โดยผู้วิจัยสามารถควบคุมได้ โดยควบคุมเพศและการสนับสนุนทางสังคม ของกลุ่มตัวอย่างในกลุ่มทดลองและกลุ่มควบคุมให้มีความเท่าเทียมกัน กระทำได้โดยกำหนดให้เพศชายและเพศหญิง รวมทั้งให้คะแนนการสนับสนุนทางสังคมมีความเท่าเทียมกันในช่วงก่อนเริ่มให้โปรแกรมจัดกระทำ

6. ตัวแปรสอดแทรก (Intervening Variable) หมายถึง ตัวแปรที่แทรกระหว่างตัวแปรอิสระและตัวแปรตาม และเป็นตัวแปรที่มีผลทำให้ค่าตัวแปรตามเปลี่ยนแปลงไป โดยตัวแปรสอดแทรกมีความแตกต่างจากตัวแปรแทรกซ้อน นั่นคือ ตัวแปรสอดแทรกเป็นตัวแปรที่ผู้วิจัยไม่สามารถควบคุมไม่ให้มีผลต่อตัวแปรตามได้ หรือเป็นตัวแปรที่ยากในการควบคุมไม่ให้เกิดขึ้นกับกลุ่มตัวอย่าง เนื่องจากตัวแปรสอดแทรกเป็นตัวแปรที่เกิดจากกระบวนการทางจิตของบุคคล ไม่สามารถสังเกตเห็นได้ หรือยากที่จะใช้วิธีสังเกตพฤติกรรม เช่น ความหิว ความตั้งใจ เป็นต้น

การวัด (Measurement) หมายถึง กระบวนการของการกำหนดค่าตัวเลขให้กับคุณลักษณะหรือคุณสมบัติที่จะวัดตามกฎเกณฑ์ที่กำหนดไว้ให้ตรงกับความเป็นจริงมากที่สุด ถ้าเป็นการวัดในการวิจัย ก็คือ การวัดตัวแปร โดยค่าตัวเลขที่วัดได้ จะเป็นค่าต่อเนื่อง (Continuous Data) หรือค่าไม่ต่อเนื่อง (Discrete) ขึ้นอยู่กับตัวแปรและเครื่องมือที่ใช้วัด โดยค่าต่อเนื่อง เป็นลักษณะข้อมูลในรูปตัวเลขที่วัดเป็นจุดทศนิยม เช่น ความสูง น้ำหนัก เป็นต้น ส่วนค่าไม่ต่อเนื่อง เป็นลักษณะข้อมูลในรูปตัวเลขจำนวนเต็ม ไม่สามารถวัดค่าเป็นจุดทศนิยมได้ เช่น จำนวนบุตร จำนวนประชาชน เป็นต้น นอกจากนี้ การวัด แบ่งได้เป็น 2 ประเภท (บุญใจ ศรีสถิตย์นรากร, 2555) คือ 1) การวัดโดยตรง เป็นการวัดค่าตัวแปรที่เป็นรูปธรรม สามารถสังเกตได้ สัมผัสได้ ความหมายไม่ซับซ้อน เข้าใจง่าย และแปลความหมายได้ตรงกันทุกคน และการวัดโดยอ้อม โดยการวัดทางพฤติกรรมศาสตร์ส่วนใหญ่เป็นการวัดโดยอ้อม เป็นการวัดค่าตัวแปรโครงสร้าง (Latent Variable) เช่น ความเชื่อ ความรู้ เจตคติ ความเครียด รวมทั้งพฤติกรรมต่าง ๆ โดยเป็นการวัดคุณลักษณะแฝงของบุคคล (latent) ซึ่งตัวแปรจะมีความหมายซับซ้อน เข้าใจยาก ไม่สามารถวัดและสังเกตได้โดยตรง และไม่สามารถวัดได้โดยใช้คำถามเพียงคำถามเดียว

วิธีการวัด (Assessment) หมายถึง กระบวนการในการวัดที่เกี่ยวข้องกับการใช้ตัวเลขหรือไม่ใช้ตัวเลข ซึ่งรวมกิจกรรมในการวัดหลาย ๆ อย่างเข้าด้วยกัน ได้แก่ การสัมภาษณ์ การสังเกตพฤติกรรม การใช้แบบทดสอบ แบบสอบถาม เป็นต้น ซึ่งวิธีการวัดดังกล่าวจะใช้สำหรับการเก็บรวบรวมข้อมูล สำหรับวิธีการวัด ประกอบด้วยวิธีการวัดเชิงปรนัย (Objective component) และวิธีการวัดเชิงอัตนัย (Subjective component) (Matarazzo, 1990) โดยวิธีการวัดเชิงปรนัย เป็นวิธีการวัดที่มีกระบวนการทดสอบ มีชุดของข้อคำถาม มีวิธีการดำเนินการทดสอบที่ได้มาตรฐาน และมีการให้คะแนนคำตอบที่มีความชัดเจน เช่น แบบทดสอบ แบบสอบถาม เป็นต้น

ส่วนวิธีการวัดเชิงอัตนัย เป็นวิธีการวัดที่มีชุดของข้อคำถามเฉพาะ แต่มีความยืดหยุ่นในกระบวนการเก็บรวบรวมข้อมูลมากกว่า และการแปลความหมายของข้อมูลจะใช้วิจารณญาณในการตัดสินใจคำตอบนั้น ๆ เช่น การสัมภาษณ์ การสังเกต เป็นต้น

การประเมิน (Evaluation) หมายถึง กระบวนการตัดสินใจในการตีค่า หรือให้คุณค่าของสิ่งของหรือการกระทำใด ๆ โดยอาศัยข้อมูลที่ได้จากการวัดที่มีความน่าเชื่อถือ แล้วทำการเก็บรวบรวมข้อมูล และการวิเคราะห์ข้อมูลมาพิจารณาเปรียบเทียบกับเกณฑ์หรือมาตรฐานที่ได้ตั้งไว้ การประเมินจะเป็นตัวบอกให้ทราบถึงคุณภาพหรือปริมาณของสิ่งนั้น ๆ ว่ามีคุณค่าเป็นเช่นใด เมื่อเทียบกับเกณฑ์ที่ตั้งไว้ แล้วจึงนำมาทำการปรับปรุงพัฒนาให้มีประสิทธิภาพและมีประสิทธิผลมากขึ้น (Ellington, Percival, & Race, 1993; Patton, 1987)

ลักษณะของการวัดทางพฤติกรรมศาสตร์

การวัดทางพฤติกรรมศาสตร์ (Measurement in Behavioral Science) เป็นการวัดที่แทนค่าคุณลักษณะหรือพฤติกรรมให้อยู่ในรูปของตัวเลข โดยเครื่องมือวัดทางพฤติกรรมศาสตร์ไม่ได้วัดลักษณะของบุคคลโดยรวม แต่เป็นการวัดคุณลักษณะบางอย่างของบุคคลเท่านั้น โดยเครื่องมือวัด มีหลากหลายประเภท ได้แก่ แบบสอบถาม แบบทดสอบ แบบสัมภาษณ์ แบบสังเกต รวมทั้งแบบตรวจสอบรายการ ซึ่งเครื่องมือวัดเหล่านี้ใช้สำหรับการวัดทั้งทางด้านความรู้คิด ความรู้สึก บุคลิกภาพ และพฤติกรรมด้านต่าง ๆ ของบุคคล เครื่องมือวัดแต่ละประเภทมีคุณลักษณะที่แตกต่างกัน และมีความเหมาะสมในการใช้วัดที่ไม่เหมือนกัน ดังนั้น ผู้วิจัยที่จะนำเครื่องมือวัดทางพฤติกรรมศาสตร์ไปใช้ ควรต้องทำการศึกษาและทำความเข้าใจลักษณะการวัดทางพฤติกรรมศาสตร์ เพื่อให้เกิดความเหมาะสมในการเลือกใช้เครื่องมือวัด และวิธีการเก็บรวบรวมข้อมูล ให้มีความสอดคล้องกับจุดมุ่งหมายของการวัดด้วย โดยลักษณะการวัดทางพฤติกรรมศาสตร์ มีลักษณะดังนี้ (อรพินทร์ ชูชม, 2542; Nesselrode & Molenaar, 2016)

1. การวัดทางพฤติกรรมศาสตร์ เป็นการวัดกลุ่มตัวอย่างพฤติกรรมที่เป็นตัวแทนสิ่งที่ต้องการจะศึกษา การวัดทางพฤติกรรมศาสตร์ไม่สามารถวัดพฤติกรรมที่เป็นไปได้ทั้งหมดของสิ่งที่ต้องการวัด เช่น การวัดความสามารถในการคิดอย่างมีวิจารณญาณ วิธีการวัดทำได้โดยการเก็บข้อมูลและประเมินทุกสิ่งทุกอย่างที่บุคคลนั้นเคยคิดหรือถ่ายทอดความคิดออกมาเป็นคำพูด ซึ่งเป็นสิ่งที่เป็นไปได้ในแง่การปฏิบัติ ดังนั้น การวัดทางพฤติกรรมศาสตร์ทำได้โดยการเก็บรวบรวมข้อมูลกลุ่มตัวอย่างพฤติกรรม โดยอาจสร้างเป็นแบบสอบถาม หรือแบบทดสอบเพื่อวัดความสามารถในการคิดอย่างมีวิจารณญาณ ดังนั้น การวัดทางพฤติกรรมศาสตร์จึงเป็นการอ้างอิงสิ่งที่ได้จากการวัดไปยังคุณลักษณะหรือคุณสมบัติของสิ่งที่ต้องการศึกษา นั่นคือ ตัวแปร นั่นเอง

2. การวัดทางพฤติกรรมศาสตร์ เป็นการวัดพฤติกรรมของบุคคล หน่วยการวัดเป็นระดับบุคคล โดยบุคคลแต่ละบุคคลมีความแตกต่างกันภายในบุคคล (Intra-Individual Differences) ทั้งด้านพฤติกรรม ความสนใจ ความชอบ การรับรู้ และความเชื่อ ดังนั้น การวัดทางพฤติกรรมศาสตร์จึงเป็นการสร้างและพัฒนาเครื่องมือวัดที่มี

ความเป็นระบบที่สามารถแปลงความแตกต่างระหว่างบุคคลให้อยู่ในรูปของตัวเลขได้ โดยตัวเลขจะสะท้อนให้เห็นถึงระดับความแตกต่างของบุคคลนั้น ๆ

3. การวัดทางพฤติกรรมศาสตร์ ใช้บรรยายคุณลักษณะส่วนบุคคล มีลักษณะเป็นตัวแปรแฝงที่มีโครงสร้างซึ่งไม่สามารถวัดหรือสังเกตเห็นได้อย่างชัดเจน โครงสร้างเหล่านั้นต้องมีการนิยามออกมาในรูปของพฤติกรรมหรือคุณลักษณะที่สามารถสังเกตและวัดได้ เช่น แบบสอบถามรูปแบบการเผชิญปัญหา (Coping Styles) วัดปริมาณของการเผชิญปัญหาของบุคคลแต่ละคน ปริมาณของการที่บุคคลมีการเผชิญปัญหาสามารถใช้อ้างอิงได้ถึงระดับการเผชิญปัญหา ว่าบุคคลนั้นมีปริมาณในระดับต่ำ ปานกลาง หรือสูง

ประเภทของข้อมูล

ข้อมูล สามารถแบ่งได้หลายประเภท โดยมีเกณฑ์ในการแบ่ง 3 ประเภท ตามแนวทางในการเก็บข้อมูล สำหรับการวัดทางพฤติกรรมศาสตร์ คือ จำแนกตามแหล่งที่มา จำแนกตามลักษณะของข้อมูล จำแนกตามลักษณะการแสดงออกของตัวแปรทางพฤติกรรมศาสตร์ มีรายละเอียดดังนี้

1. จำแนกตามแหล่งที่มา

ข้อมูลปฐมภูมิ (Primary Data) หมายถึง “ข้อมูลที่ต้องเก็บใหม่” เป็นข้อมูลที่ผู้วิจัยหรือหน่วยงานเป็นผู้เก็บรวบรวมข้อมูลด้วยตนเองเพื่อนำมาใช้ในงานวิจัยที่ศึกษา โดยเก็บรวบรวมข้อมูลหรือบันทึกจากแหล่งข้อมูลโดยตรง หรือจากแหล่งกำเนิดของข้อมูลนั้น ๆ ซึ่งวิธีรวบรวมข้อมูลอาจใช้วิธีสัมภาษณ์ สังเกต สัมภาษณ์ เป็นต้น ข้อดีของข้อมูลปฐมภูมิ คือ เป็นข้อมูลที่มีรายละเอียดตรงตามข้อเท็จจริง ส่วนข้อเสียของข้อมูลปฐมภูมิ คือ การเก็บรวบรวมข้อมูลในบางกรณีอาจต้องใช้เวลายาวนาน ใช้ปริมาณเวลามาก ทำให้ต้องใช้งบประมาณค่าใช้จ่ายในการเก็บรวบรวมข้อมูลค่อนข้างมาก

ยกตัวอย่างงานวิจัยที่ใช้ข้อมูลปฐมภูมิ

“วิธีการสร้างแบรนด์และผลของการสร้างแบรนด์ภายในองค์กรของสถาบันอุดมศึกษาในเขตกรุงเทพมหานคร” (วีรพงษ์ พวงเล็ก, นริศรา พึ่งโพธิ์สภ, และดุษฎี โยเหลา, 2561)

ข้อมูลปฐมภูมิ เป็นข้อมูลที่ผู้วิจัยเก็บรวบรวมด้วยตนเอง โดยเก็บรวบรวมข้อมูลจากสถาบันอุดมศึกษาในเขตกรุงเทพมหานคร ด้วยวิธีการสัมภาษณ์

ข้อมูลทุติยภูมิ (Secondary Data) หมายถึง “ข้อมูลที่ต้องเก็บใหม่” เป็นข้อมูลที่มีผู้เก็บรวบรวมไว้แล้วทั้งในรูปของเอกสาร หลักฐาน สิ่งตีพิมพ์ และสิ่งบันทึกต่าง ๆ และตัวเลขหรือตัวหนังสือที่มีการรายงาน การทำระเบียบประวัติและการทำสำมะโนในลักษณะต่าง ๆ ข้อดีของข้อมูลทุติยภูมิ คือ ประหยัดเวลาและค่าใช้จ่ายในการเก็บรวบรวมข้อมูล ส่วนข้อเสียของข้อมูลทุติยภูมิ คือ ข้อมูลอาจมีรายละเอียดไม่เพียงพอในการนำไปวิเคราะห์หรืออาจมีความคลาดเคลื่อน ไม่ตรงตามข้อเท็จจริง ควรตรวจสอบคุณภาพข้อมูลก่อนที่จะนำมาใช้

ยกตัวอย่างงานวิจัยที่ใช้ข้อมูลทฤษฎี

“ปัจจัยที่ส่งผลต่อพฤติกรรมป้องกันการสูบบุหรี่ในกลุ่มวัยรุ่น: การสังเคราะห์งานวิจัยด้วยการทบทวนวรรณกรรมอย่างเป็นระบบ” (ชาริน สุวรรณวงศ์ และศรัณย์ พิมพ์ทอง, 2561)

ข้อมูลทฤษฎี เป็นข้อมูลที่ผู้วิจัยนำข้อมูลที่เก็บรวบรวมไว้แล้ว โดยเก็บรวบรวมรายงานงานวิจัยเพื่อนำมาสังเคราะห์งานวิจัย

2. จำแนกตามลักษณะของข้อมูล

ข้อมูลเชิงปริมาณ (Quantitative Data) เป็นข้อมูลที่เป็นตัวเลขบอกปริมาณความมากน้อยของข้อมูลที่เก็บได้ สามารถवलบลคูณหารกันได้ สามารถนำมาวิเคราะห์ทางสถิติได้ เช่น รายได้ น้ำหนัก คะแนนผลสัมฤทธิ์ทางการเรียน คะแนนทักษะชีวิต เป็นต้น โดยการเก็บรวบรวมข้อมูลใช้วิธีการสำรวจ แบบสอบถาม แบบสังเกต หรือแบบสัมภาษณ์

ยกตัวอย่างงานวิจัยที่ใช้ข้อมูลเชิงปริมาณ

“ปัจจัยทางจิตสังคมที่เกี่ยวข้องกับพฤติกรรมป้องกันการรับและแพร่โรคในสถานศึกษาของนักศึกษาระดับปริญญาตรี” (เบญจพร ประณีตวาทกุล และดุจดเดือน พันธุมนาวิณ, 2561)

งานวิจัยนี้ เก็บรวบรวมข้อมูลด้วยแบบสอบถามและแบบวัดตัวแปร เพื่อให้ได้ข้อมูลเป็นปริมาณตัวเลขสามารถบอกได้ว่าตัวแปรแต่ละตัวแปรมีระดับสูงและระดับต่ำ และสามารถนำตัวเลขมาวิเคราะห์ทางสถิติได้

ข้อมูลเชิงคุณภาพ (Qualitative Data) เป็นข้อมูลที่เป็นสัญลักษณ์หรือตัวหนังสือที่บอกลักษณะหรืออาการของตัวแปรที่ศึกษา ไม่มีความเป็นปริมาณ วลบลคูณหารกันไม่ได้ เช่น เพศ ศาสนา ประสบการณ์ชีวิต เป็นต้น โดยการเก็บรวบรวมข้อมูลมักใช้วิธีการสังเกต การสัมภาษณ์ หรือการสนทนากลุ่ม

ยกตัวอย่างงานวิจัยที่ใช้ข้อมูลเชิงคุณภาพ

“วิธีการสร้างแบรนด์และผลของการสร้างแบรนด์ภายในองค์กรของสถาบันอุดมศึกษาในเขตกรุงเทพมหานคร” (วีรพงษ์ พวงเล็ก, นริศรา พิงโพธิ์สภ, และดุชฎิ โยเหลา, 2561)

งานวิจัยนี้ เก็บรวบรวมข้อมูลด้วยวิธีการสัมภาษณ์ เพื่อให้ได้ข้อมูลเป็นตัวหนังสือที่บอกถึงวิธีการสร้างแบรนด์ และผลที่เกิดจากการสร้างแบรนด์ภายในองค์กร

3. จำแนกตามลักษณะการแสดงออกของตัวแปรทางพฤติกรรมศาสตร์

ความรู้ความเข้าใจ (Cognitive หรือ Knowledge) เป็นสิ่งที่สั่งสมมาจากการศึกษาเล่าเรียน การค้นคว้าหรือประสบการณ์ รวมทั้งความสามารถเชิงปฏิบัติและทักษะ มีลักษณะเป็นข้อเท็จจริงที่มีทั้งถูกและผิด เป็นไปตาม

หลักวิชาและเหตุผลเชิงวิทยาศาสตร์ที่สามารถพิสูจน์หรือตรวจสอบให้เห็นจริงได้ รวมทั้งยังเป็นความสามารถ (Ability) เชิงพฤติกรรมทางสมอง สามารถวัดได้ด้วยการใช้แบบทดสอบหรือแบบวัด คำตอบที่ได้จากการวัดความรู้ความเข้าใจจะเป็นเนื้อหาที่ผู้ตอบทราบในเรื่องเกี่ยวกับความจำ ความเข้าใจ การนำไปใช้ การวิเคราะห์ การสังเคราะห์ และการประเมิน ตามแนวทางการเรียนรู้ของบลูม (Bloom et al., 1956)

ความรู้ความเข้าใจ มีส่วนประกอบที่สำคัญ 2 ส่วน คือ (1) เป็นความรู้ความเข้าใจในเรื่องอะไร และ (2) เป็นความรู้ความเข้าใจของคนกลุ่มใด

ยกตัวอย่าง ความรู้ความเข้าใจ

“ความรู้ความเข้าใจเกี่ยวกับการรักษาสุขภาพของนักเรียนระดับประถมศึกษา”

“ความรู้ความเข้าใจเกี่ยวกับกฎจราจรของผู้ขับขี่รถจักรยานยนต์”

“ความรู้ความเข้าใจเกี่ยวกับกฎหมายสิ่งแวดล้อมของผู้ประกอบการโรงงานอุตสาหกรรม”

ความรู้สึนึกคิด (Affective) เป็นตัวแปรที่แสดงออกมาในรูปของคุณลักษณะที่เกี่ยวข้องกับความรู้สึก อารมณ์ และจิตใจ เช่น การวัดเจตคติ การวัดความคิดเห็น มักจะเป็นคุณลักษณะแฝง ซึ่งเป็นลักษณะที่เราไม่สามารถสังเกตหรือวัดได้โดยตรง แต่เชื่อได้ว่ามีลักษณะนั้นภายในตัวบุคคล โดยนักวิจัยสามารถสรุปผลได้จากพฤติกรรมเฉพาะที่แสดงออกถึงความรู้สึกและความคิดเห็นของบุคคล รวมทั้ง ลักษณะนี้เป็นสิ่งที่มีอยู่ภายในตัวบุคคล และเป็นสิ่งที่ก่อให้เกิดความแตกต่างภายในบุคคล จากนั้นจึงส่งผลให้เกิดความแตกต่างในลักษณะอื่น ๆ ที่เกี่ยวข้อง

ยกตัวอย่าง ความรู้สึนึกคิด

“เจตคติต่อการทำงานอย่างมีความรับผิดชอบ”

“ลักษณะมุ่งอนาคตควบคุมตน”

“ความเครียด”

พฤติกรรม (Behavior) หมายถึง การปฏิบัติ/การกระทำที่เป็นการใช้ความสามารถที่แสดงออกทางร่างกายต่อสถานการณ์หนึ่ง ๆ ซึ่งเกิดขึ้นแล้ว สังเกตเห็นได้ และรวมพฤติกรรมที่ไม่ได้ปฏิบัติทันที แต่คาดคะเนว่าจะปฏิบัติในโอกาสต่อไปด้วย โดยพฤติกรรมเน้นเฉพาะพฤติกรรมภายนอกเท่านั้น เนื่องจากพฤติกรรมภายในเป็นความรู้ความสามารถ ความรู้สึนึกคิดที่ไม่สามารถมองเห็นได้ โดยพฤติกรรมภายนอก ได้แก่ ทักษะ การปฏิบัติเพื่อจัดกระทำกับกิจกรรมต่าง ๆ ซึ่งบุคคลแสดงออกและสังเกตเห็นได้ด้วยตา เช่น การทำงาน การพูด การเขียน การออกกำลังกาย ผู้วิจัยสามารถวัดพฤติกรรมได้โดยการสังเกต ใช้แบบสอบถาม เป็นต้น ซึ่งทำได้โดยการเก็บรวบรวมข้อมูลกลุ่มตัวอย่างพฤติกรรม และอ้างอิงสิ่งที่ได้จากการวัดไปยังคุณลักษณะหรือคุณสมบัติทั้งหมดของสิ่งที่ต้องการศึกษา

ยกตัวอย่าง พฤติกรรม

“พฤติกรรมการทำงานอย่างมีความรับผิดชอบ”

“พฤติกรรมควบคุมความก้าวร้าว”

“พฤติกรรมป้องกันการสูญบุหรี”

สภาพแวดล้อม (Environment) เป็นตัวแปรที่บุคคลแสดงการรับรู้ต่อสภาพแวดล้อมรอบตัว เป็นสิ่งที่ไม่สามารถสังเกตและวัดได้โดยตรง แต่เชื่อว่าบุคคลนั้นมีปฏิสัมพันธ์กับสิ่งแวดล้อมรอบตัว เช่น ครอบครัว กลุ่มเพื่อน เพื่อนร่วมงาน หัวหน้างาน โรงเรียน มหาวิทยาลัย สถานที่ทำงาน เป็นต้น

ยกตัวอย่าง สภาพแวดล้อม

“การอบรมเลี้ยงดู”

“การสนับสนุนทางสังคมจากหัวหน้างาน”

“บรรทัดฐานทางสังคม”

ระดับของการวัด (Scale of Measurement)

ระดับของการวัดเป็นผลที่ได้จากการวัดคุณลักษณะหรือคุณสมบัติที่ต้องการวัดออกมาเป็นตัวเลขหรือคะแนน และมีการระบุระดับของการวัดที่แทนคะแนนหรือข้อมูลนั้น ๆ ระดับของการวัด จำแนกได้เป็น 4 มาตรฐานวัดตามคุณลักษณะที่ทำการวัดได้ โดยแต่ละมาตรฐานวัดมีคุณสมบัติ กฎเกณฑ์ และความซับซ้อนของการวัดที่แตกต่างกัน มาตรฐานวัดจะเป็นการบอกถึงธรรมชาติของข้อมูลที่ได้จากการวัด ดังนั้น จึงเป็นสิ่งสำคัญในการสร้างและพัฒนาเครื่องมือวัดทางพฤติกรรมศาสตร์ที่จะต้องทราบว่าข้อมูลที่ต้องการนั้นอยู่ในการวัดระดับใด โดยมีรายละเอียดแต่ละระดับของการวัด ดังนี้

1. มาตรฐานวัดนามบัญญัติ (Nominal Scale) เป็นมาตรฐานวัดที่ใช้กับตัวแปรที่มีค่าไม่ต่อเนื่อง เป็นวิธีการวัดที่ง่ายที่สุด โดยจำแนกหรือกำหนดชื่อสิ่งที่จะศึกษาไว้เป็นกลุ่มหรือประเภท ถ้าสิ่งที่ศึกษามีคุณสมบัติหรือคุณลักษณะเหมือนกันก็จัดไว้ในประเภทเดียวกัน ถ้าต่างก็จัดไว้คนละประเภท ดังนั้น คุณสมบัติที่สำคัญของมาตรฐานวัดนี้ คือ ความเท่าเทียมกัน (equivalence) ของคุณลักษณะของสมาชิกที่อยู่ในประเภทเดียวกัน เช่น แบ่งตามเพศ เป็นเพศชายและเพศหญิง แบ่งตามศาสนาที่นับถือ เป็นพุทธ อิสลาม คริสต์ เป็นต้น โดยที่ความแตกต่างของแต่ละประเภทไม่สามารถบอกความแตกต่างในเชิงตัวเลข การกำหนดตัวเลขให้กับประเภทต่าง ๆ ไม่สามารถบอกปริมาณมากน้อย และเปรียบเทียบกันไม่ได้ รวมทั้งใช้คำนวณไม่ได้ด้วย

ยกตัวอย่าง การกำหนดรหัสเป็นตัวเลขตามมาตรวัดนามบัญญัติ

“เพศ” : เพศชาย เพศหญิง

กำหนดรหัสตัวแปร : เพศชาย = 1 เพศหญิง = 2

“ศาสนา” : ศาสนาพุทธ ศาสนาคริสต์ ศาสนาอิสลาม

กำหนดรหัสตัวแปร : ศาสนาพุทธ = 1 ศาสนาคริสต์ = 2 ศาสนาอิสลาม = 3

รหัสตัวเลขที่กำหนดให้เป็นเพศชาย เพศหญิง ศาสนาพุทธ ศาสนาคริสต์ ศาสนาอิสลาม เป็นตัวเลขที่ไม่มี ความหมายใด ๆ ในทางคณิตศาสตร์ ตัวเลขไม่มีความหมายบอกระดับปริมาณหรือระดับคุณภาพ นั่นคือ บอกไม่ได้ว่า เพศหญิงดีกว่าเพศชาย เป็นค่าตัวเลขที่ใช้กำหนดกลุ่มประเภทเพียงเท่านั้น ดังนั้น จึงจะกำหนดรหัสตัวเลขใดก็ได้ เพื่อให้ง่ายต่อการวิเคราะห์ทางสถิติ จึงมักกำหนดเริ่มต้นจาก 1, 2, 3, ... แล้วแต่ประเภทของตัวแปร

2. มาตรการวัดเรียงอันดับ (Ordinal Scale) เป็นมาตรการวัดลักษณะข้อมูลที่สูงกว่ามาตรการวัดนามบัญญัติ คือ นอกจากจะรวมคุณสมบัติของมาตรการวัดแรก ยังเพิ่มคุณสมบัติอีกด้านหนึ่ง คือ ชื่อหรือสัญลักษณ์ที่กำหนดให้ นั้นยังสามารถบอกปริมาณหรือคุณภาพได้โดยการใช้วิธีเรียงลำดับ เช่น มากกว่าหรือน้อยกว่า ดีกว่าหรือเลวกว่า เช่น ตำแหน่งทางวิชาการ แบ่งเป็นผู้ช่วยศาสตราจารย์ รองศาสตราจารย์ ศาสตราจารย์ ซึ่งการเรียงอันดับนี้สามารถบอกได้ว่าอะไรดีหรือมากกว่าอะไร คือ บอกทิศทางของความแตกต่างได้ แต่ยังไม่บอกขนาดหรือระบุค่าที่แน่นอนไม่ได้ว่า ความแตกต่างของแต่ละอันดับมีค่าเป็นอย่างไร นั่นคือ บอกระยะห่างของแต่ละอันดับไม่ได้

ยกตัวอย่าง การกำหนดรหัสเป็นตัวเลขตามมาตรวัดเรียงอันดับ

“ตำแหน่งทางวิชาการ” : ผู้ช่วยศาสตราจารย์ รองศาสตราจารย์ ศาสตราจารย์

กำหนดรหัสตัวแปร : ผู้ช่วยศาสตราจารย์ = 1

รองศาสตราจารย์ = 2

ศาสตราจารย์ = 3

รหัสตัวเลขที่กำหนดให้เป็นผู้ช่วยศาสตราจารย์ รองศาสตราจารย์ ศาสตราจารย์ เป็นตัวเลขที่ใช้เรียงลำดับ จากต่ำสุดไปสูงสุด หรือจากสูงสุดไปต่ำสุด ไม่สามารถคิดคำนวณเป็นตัวเลขที่บ่งบอกปริมาณความแตกต่าง เช่น ตัวเลขบอกได้ว่าศาสตราจารย์มีคุณวุฒิสูงกว่ารองศาสตราจารย์ และรองศาสตราจารย์มีคุณวุฒิสูงกว่าผู้ช่วยศาสตราจารย์ แต่ตัวเลขบอกปริมาณความแตกต่างระหว่างศาสตราจารย์กับรองศาสตราจารย์ไม่ได้แสดงว่ามีความแตกต่างกันเท่าไร

3. มาตรการวัดช่วง/อันตรภาค (Interval Scale) เป็นมาตรการวัดที่มีคุณสมบัติของมาตรการวัดเรียงอันดับ แต่มีคุณสมบัติเพิ่มขึ้นมา คือ สามารถระบุค่าหรือช่วงระยะห่างของความแตกต่างได้ จึงสามารถเอาค่าต่าง ๆ มาทำการคำนวณ บวกหรือลบออกจากกันได้ แต่ไม่สามารถนำตัวเลขมาคูณหรือหารกันได้ นั่นคือเป็นมาตรวัดตัวแปรที่มีค่าต่อเนื่อง ลักษณะที่สำคัญของการวัดในระดับนี้ คือ จุดเริ่มต้นของค่าตัวเลขของข้อมูลนี้ไม่มีจุดเริ่มต้นที่เป็นศูนย์แท้ เช่น อุณหภูมิ 0 องศาเซลเซียส ก็ไม่ได้หมายความว่า อุณหภูมิเท่านี้ไม่มีระดับความร้อนหรือเย็นเลย เช่นเดียวกับคะแนนที่ได้จากแบบทดสอบผลสัมฤทธิ์ทางการเรียน คนที่ได้ 0 คะแนน ไม่ได้หมายความว่าคนนี้ไม่มีความรู้ในเรื่องนั้นเลย

ยกตัวอย่าง “แบบวัดการกำกับทตนเองในการออกกำลังกาย” มีนิตได้คะแนนแบบวัด 0, 10, 20, 30, 40 และ 50 คะแนน คะแนนที่กำหนดดังกล่าว เมื่อพิจารณาตัวเลขตามมาตรวัดช่วง/อันตรภาค พบว่า นิตที่ได้คะแนน 0 คะแนน ไม่ได้หมายความว่า นิตจะไม่มีการกำกับทตนเองในการออกกำลังกาย เนื่องจากมาตรวัดช่วงเป็นมาตรวัดที่ไม่มีศูนย์แท้ ส่วนนิตที่ได้ 50 คะแนน มีคะแนนการกำกับทตนเองในการออกกำลังกายมากกว่านิตที่ได้ 10 คะแนน เป็นจำนวน $50-10 = 40$ คะแนน

4. มาตรการวัดอัตราส่วน (Ratio Scale) เป็นมาตรวัดข้อมูลที่มีลักษณะสมบูรณ์ที่สุด เป็นมาตรการวัดระดับสูงสุด รวมเอาคุณสมบัติของข้อมูลทุกอย่าง แต่มีคุณสมบัติเพิ่มขึ้นมา คือ มีค่าศูนย์แท้ (Absolute Zero) มีจุดเริ่มต้นของค่าศูนย์ที่แท้จริง เป็นศูนย์ที่ไม่มีลักษณะนั้น ๆ เลย สามารถเปรียบเทียบในรูปอัตราส่วนได้ เนื่องจากความแตกต่างของแต่ละหน่วยมีปริมาณเท่ากัน เช่น ข้อมูลที่ได้จากการวัดความยาว ส่วนสูง น้ำหนัก ความเร็ว รายได้ เป็นต้น

ยกตัวอย่างเช่น น้ำหนักมีค่าเท่ากับศูนย์ แสดงว่า ไม่มีน้ำหนักเหลืออยู่เลย หรือ A มีเงิน 200 บาท B มีเงิน 100 บาท สามารถบอกได้ว่า A มีเงินมากกว่า B จำนวน 100 บาท และคิดเป็น 2 เท่าของ B ด้วยข้อมูลที่อยู่ในระดับมาตรการวัดอัตราส่วน สามารถนำมาทำการคำนวณได้ทั้งบวกลบคูณและหาร โดยข้อมูลทางพฤติกรรมศาสตร์ส่วนใหญ่ ยังไม่สามารถวัดได้ถึงระดับมาตรการวัดอัตราส่วน เนื่องจากไม่ทราบจุดเริ่มต้นที่แท้จริงของสิ่งที่ต้องการจะวัด โดยเฉพาะอย่างยิ่งการวัดเกี่ยวกับพฤติกรรมของบุคคล

จะเห็นได้ว่า ระดับของการวัดข้อมูลมีหลายระดับ ข้อมูลของตัวแปรหรือสิ่งที่ต้องการวัดไม่จำเป็นต้องแสดงออกมาเป็นตัวเลขเสมอไป จึงต้องมีการกำหนดรหัสตัวแปรด้วยตัวเลข รวมทั้งข้อมูลบางอย่างอาจเป็นตัวเลขมาแล้ว โดยข้อมูลที่มีมาตรการวัดระดับสูง สามารถจำแนกออกมาเป็นข้อมูลที่มีมาตรการวัดระดับต่ำได้ เนื่องจากข้อมูลระดับสูงจะมีคุณสมบัติของข้อมูลระดับต่ำรวมอยู่ด้วย เช่น คะแนนที่ได้จากการวัด เป็นมาตรการวัดช่วง/อันตรภาค สามารถลดระดับมาเป็นมาตรการวัดเรียงอันดับ หรือมาตรการวัดนามบัญญัติได้ ทั้งนี้ขึ้นอยู่กับความต้องการของผู้สร้างเครื่องมือวัดว่าต้องการข้อมูลในระดับใด แต่พึงระมัดระวังด้วยว่า ข้อมูลในระดับต่ำไม่สามารถแปรสภาพให้อยู่ในระดับสูงได้ และข้อมูลที่อยู่ในระดับการวัดที่ต่างกันมักจะใช้สถิติในการวิเคราะห์ข้อมูลที่แตกต่างกันด้วย โดยคุณลักษณะของมาตรการวัดทั้ง 4 ระดับ ดังตารางที่ 2-1

ตารางที่ 2-1 คุณลักษณะของมาตรการวัดทั้ง 4 ระดับ

มาตรการวัด	ความหมาย	ข้อจำกัด	ตัวอย่าง
นามบัญญัติ	จำแนกวัตถุสิ่งของ คน หรือเหตุการณ์ เข้าไว้ในประเภทที่แยกจากกันได้	ไม่สามารถระบุความแตกต่าง ด้านปริมาณระหว่างประเภทได้	เพศ อาชีพ ศาสนา
เรียงอันดับ	เรียงลำดับวัตถุ สิ่งของ คน คุณลักษณะ โดยไม่คำนึงถึงความ แตกต่างที่เท่าเทียมกัน	จำกัดการเปรียบเทียบความแตกต่าง โดยไม่คำนึงถึงปริมาณความแตกต่างที่ สมบูรณ์	ระดับการศึกษา ตำแหน่งทางวิชาการ
ช่วง/อันตรภาค	บ่งบอกความแตกต่างที่เท่าเทียมกัน ระหว่างหน่วยของการวัดที่อยู่ใน ลำดับต่อเนื่องกัน	ไม่มีศูนย์แท้	อุณหภูมิ คะแนนแบบ วัด
อัตราส่วน	มีศูนย์สมบูรณ์และมีช่วงห่างระหว่าง หน่วยที่เท่ากัน	ตัวแปรพฤติกรรมศาสตร์มักไม่อยู่ใน มาตรการวัดนี้	ระยะทาง น้ำหนัก เวลา รายได้

ที่มา: อรพินทร์ ชูชม (2545, หน้า 5)

การนิยามตัวแปร

ก่อนการสร้างเครื่องมือวัด นักวิจัยต้องทำการนิยามตัวแปร ซึ่งเป็นการให้ความหมายของตัวแปรในการวิจัยเพื่อให้ผู้อ่านหรือผู้เชี่ยวชาญตรวจสอบคุณภาพเครื่องมือมีความเข้าใจตัวแปรได้ตรงกับที่ผู้วิจัยกำหนดหรือเข้าใจว่าผู้วิจัยต้องการศึกษาอะไร โดยเฉพาะตัวแปรคุณลักษณะแฝง (Latent Trait/Construct) ซึ่งเป็นตัวแปรด้านความรู้สึกนึกคิด (Affective) ที่เป็นคุณลักษณะของบุคคลที่จะสังเกตได้ค่อนข้างยาก จะสังเกตได้ก็ต่อเมื่อมีสิ่งเร้ามากระตุ้นให้บุคคลแสดงการตอบสนองด้วยความคิดเห็น พฤติกรรม วิธีการปฏิบัติ หรือผลงาน ได้แก่ ความเชื่อ ความสนใจ ความวิตกกังวล ความกลัว ค่านิยม เจตคติ ความคิดเห็น เป็นต้น โดยตัวแปรเหล่านี้เป็นคุณลักษณะที่แฝงอยู่ในตัวบุคคล แม้ว่าจะไม่สามารถสังเกตได้โดยตรง แต่ผู้วิจัยสามารถสรุปอ้างอิงได้จากพฤติกรรมการแสดงออก ความรู้สึก และความคิดเห็นของบุคคล ดังนั้น การนิยามตัวแปรจึงมีความสำคัญเป็นขั้นตอนแรก ๆ ที่ช่วยให้งานวิจัยมีคุณภาพได้ โดยการนิยามตัวแปร สามารถแบ่งได้เป็น 2 ระดับ ดังนี้

ระดับที่ 1 การนิยามเชิงทฤษฎี (Conceptual Definition: CD) เป็นการใช้ทฤษฎีหรือแนวคิดต่าง ๆ มาทำการอธิบายหรือให้ความหมายของตัวแปร เพื่อให้เกิดความเข้าใจลักษณะของตัวแปรนั้น โดยทั่วไปยังคงมีลักษณะเป็นนามธรรม ซึ่งทำให้ผู้วิจัยไม่สามารถสังเกตหรือวัดได้โดยตรง

ระดับที่ 2 การนิยามเชิงปฏิบัติการ (Operational Definition: OD) เป็นการแปลความหมายของตัวแปรจากนิยามเชิงทฤษฎีที่เป็นนามธรรมให้มีลักษณะเป็นรูปธรรม โดยการระบุถึงชุดลักษณะหรือชุดพฤติกรรมบ่งชี้ (Indicators) ที่แสดงให้เห็นคือ ตัวแปร (Construct) ที่ต้องการศึกษา ซึ่งทำให้สามารถสังเกตและวัดตัวแปรดังกล่าวได้

องค์ประกอบของนิยามเชิงปฏิบัติการ ประกอบด้วย

1. ความหมาย/องค์ประกอบหรือมิติ
2. แนวคิดในการวัดมาจากแนวคิดอะไร/ของใคร
3. วัดด้วยเครื่องมือวัดอะไร/ประเภทใด
4. เกณฑ์ในการแปลความหมายจากคะแนน

ยกตัวอย่างการนิยามเชิงทฤษฎีและการนิยามเชิงปฏิบัติการ

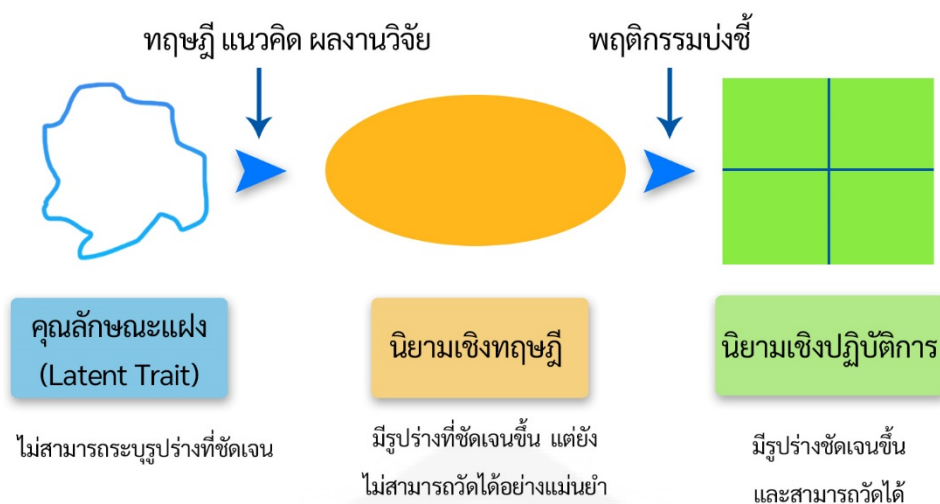
งานวิจัย: “อิทธิพลของปัจจัยเชิงสาเหตุที่มีต่อพฤติกรรมป้องกันการสูบบุหรี่ของนักเรียนระดับมัธยมศึกษาตอนปลายในจังหวัดสตูล: การสังเคราะห์งานวิจัยและการวิเคราะห์พระดัต” (จาริน สุวรรณวงศ์, 2561)

ตัวแปร: การคิดอย่างเป็นเหตุเป็นผลของนักเรียน

นิยามเชิงทฤษฎี: การที่นักเรียนมีการรู้จัก ใช้ความสามารถทางสติปัญญาในการพิจารณาอย่างเป็นเหตุเป็นผลในการรับรู้และประเมินสถานการณ์ต่าง ๆ

นิยามเชิงปฏิบัติการ: การที่นักเรียนมีความสามารถทางด้านความรู้คิด โดยการใช้ความสามารถทางด้านสติปัญญาในการพิจารณาอย่างเป็นเหตุเป็นผลในการรับรู้และประเมินสถานการณ์ต่าง ๆ ทั้งการวิเคราะห์สาเหตุของปัญหา ตลอดจนผลที่เกิดขึ้นตามมาจากการกระทำนั้น ๆ ประกอบด้วย 2 องค์ประกอบ คือ 1) การคิดแบบสาเหตุ-ผล เป็นการคิดหาเหตุผลในลักษณะของการเชื่อมโยงระหว่างสิ่งทีคาดว่าน่าจะเป็นเหตุและสิ่งทีคาดว่าน่าจะเป็นผล มุ่งอธิบายถึงลักษณะโครงสร้างความสัมพันธ์ระหว่างสิ่งที่เป็นสาเหตุและผลนั้น ๆ และ 2) การคิดแบบเหตุผลเชิงวิทยาศาสตร์ เป็นการให้ความสำคัญกับขั้นตอนที่เรียกว่าวิธีการทางวิทยาศาสตร์ ประกอบด้วย การกำหนดนิยามและปัญหา การสร้างสมมติฐาน การเก็บรวบรวม จัดระบบ และวิเคราะห์ข้อมูล ประมวลข้อสรุป และตรวจสอบทดสอบผลในสถานการณ์เฉพาะ

สำหรับการวิจัยในครั้งนี้ ผู้วิจัยใช้แบบวัดจิตเอกภาคด้านความมีเหตุมีผลของโกศล มีคุณ (2551) มี 2 องค์ประกอบ คือ การคิดแบบสาเหตุ-ผล และการคิดแบบเหตุผลเชิงวิทยาศาสตร์ มีลักษณะเป็นมาตรฐานค่า 6 ระดับ ตั้งแต่ “จริงที่สุด” (6 คะแนน) ถึง “ไม่จริงเลย” (1 คะแนน) นักเรียนที่ได้คะแนนสูงกว่า แสดงว่า เป็นผู้ที่มี การคิดอย่างเป็นเหตุเป็นผลสูงกว่านักเรียนที่ได้คะแนนต่ำกว่า



ภาพที่ 2-1 การนิยามตัวแปร

ข้อจำกัดในการวัดทางพฤติกรรมศาสตร์

การวัดทางพฤติกรรมศาสตร์ บางส่วนเป็นการวัดคุณลักษณะทางจิตที่เป็นโครงสร้าง และเป็นนามธรรม ซึ่งยากที่จะวัดได้โดยตรง ดังนั้น การสร้างและการพัฒนาเครื่องมือวัดทางพฤติกรรมศาสตร์ ต้องคำนึงถึงข้อจำกัดในการวัดพฤติกรรมศาสตร์ ดังนี้

1. การวัดทางพฤติกรรมศาสตร์มีความจำกัดในเรื่องกลุ่มตัวอย่างพฤติกรรมที่จะวัด ไม่สามารถวัดคุณสมบัติที่ศึกษาได้อย่างสมบูรณ์ เช่น การวัดความสามารถในการเผชิญปัญหาและฟื้นฝ่าอุปสรรคของนักศึกษา ทำได้โดยการสุ่มกลุ่มตัวอย่าง กำหนดจำนวนข้อคำถามและเนื้อหาที่จำเป็นที่เพียงพอของพฤติกรรมที่ทำการศึกษา คำตอบที่ได้จากข้อคำถามเป็นความจำกัดทางการวัด ใช้อ้างอิงถึงความสามารถในการเผชิญปัญหาและฟื้นฝ่าอุปสรรคของนักศึกษา ซึ่งมีความแตกต่างจากการวัดทางกายภาพที่สามารถวัดคุณลักษณะทางกายภาพที่ต้องการศึกษาได้อย่างสมบูรณ์และถูกต้อง เช่น วัดน้ำหนัก ส่วนสูงของบุคคลได้โดยไม่ต้องอ้างอิงจากกลุ่มตัวอย่างพฤติกรรม

2. การวัดทางพฤติกรรมศาสตร์มีความจำกัดในมาตรวัด ทั้งนี้ เนื่องจากไม่สามารถวัดคุณสมบัติที่ศึกษาได้อย่างสมบูรณ์ หน่วยมาตรวัดยังมีจุดอ่อน และการแปลความหมายค่าที่ได้เป็นปัญหาที่ซับซ้อนที่ต้องพิจารณาในการสร้างเครื่องมือและระบบการให้คะแนน เช่น แบบวัดความฉลาดทางอารมณ์ การที่บุคคลไม่ตอบข้อคำถามใด ๆ หรือตอบคำถามไม่ครบถ้วนสมบูรณ์ในแบบวัดนี้ อาจจะเป็นไปได้ว่ามีความลำบากใจในการตอบ หรือมีความไม่เข้าใจในตัวข้อคำถาม ทำให้ได้ค่าคะแนนต่ำและอาจบ่งชี้ว่าบุคคลนั้นมีความฉลาดทางอารมณ์ต่ำด้วย ในขณะที่อีกบุคคลหนึ่งตอบทุกข้อ จึงมีแนวโน้มที่จะได้คะแนนความฉลาดทางอารมณ์สูงกว่า ดังนั้น คะแนนของทั้งสองคนไม่สามารถนำมาเปรียบเทียบได้ว่ามีความแตกต่างในความฉลาดทางอารมณ์

3. การวัดทางพฤติกรรมศาสตร์มีความคลาดเคลื่อนเกิดขึ้นได้เสมอ เนื่องจากตัวแปรภายนอกต่าง ๆ อาจเกิดขึ้นและส่งผลต่อการวัด เช่น สภาพแวดล้อมของสถานที่ทำแบบทดสอบ/แบบสอบถาม สภาพร่างกายและจิตใจของผู้ที่ทำการวัด ณ ขณะนั้น ไม่ว่าจะเป็น ความเบื่อ ความเหนื่อยล้า การเดา ความสะเพร่า หรือเป็นไปได้ว่าผู้ประเมินมีการให้คะแนนผิดพลาด ทำให้บุคคลที่ทำแบบสอบถามเดียวกันติดต่อกันทั้งสองครั้ง จึงได้คะแนนแบบสอบถามที่ไม่เท่ากันเสมอไป นั่นคือ เกิดความไม่คงเส้นคงวาสม่ำเสมอของคะแนน ถือว่า เป็นความคลาดเคลื่อนที่เกิดขึ้นจากการวัด

4. การวัดทางพฤติกรรมศาสตร์ ที่มีความเป็นโครงสร้างทางจิตของตัวแปร ยังไม่มีการวัดวิธีการใดที่เป็นที่ยอมรับอย่างเป็นทางการ ทั้งนี้ เนื่องจากโครงสร้างทางจิตเป็นการวัดคุณลักษณะทางอ้อมที่เกี่ยวข้องกับโครงสร้างของตัวแปรที่ศึกษา ซึ่งอาจมีแนวคิดทฤษฎีต่าง ๆ ที่กล่าวถึงโครงสร้างเดียวกัน แต่นิยามโครงสร้างนั้นในเชิงคุณลักษณะที่สังเกตและวัดได้แตกต่างกัน เช่น ต้องการวัดทักษะชีวิตของนักเรียน ผู้สร้างแบบวัดต้องกำหนดพฤติกรรมให้นักเรียนแสดงความสามารถที่จะอ้างอิงถึงการมีทักษะชีวิตได้ ผู้สร้างแบบวัดอาจใช้วิธีการต่าง ๆ กัน เช่น ให้นักเรียนรายงานตนเองเกี่ยวกับทักษะชีวิต หรือให้นักเรียนวิเคราะห์สถานการณ์เกี่ยวกับทักษะชีวิต ดังนั้นวิธีการวัดที่แตกต่างกันซึ่งได้มาจากนิยามปฏิบัติการของการวัดที่แตกต่างกัน อาจจะนำไปสู่ข้อสรุปที่แตกต่างกันในระดับทักษะชีวิตของนักเรียน

คุณลักษณะของเครื่องมือวัดที่ดี

เครื่องมือวัดมีหลายชนิด เช่น แบบทดสอบ แบบสำรวจ แบบวัด แบบสอบถาม แบบสัมภาษณ์ แบบสังเกต เป็นต้น ในการใช้เครื่องมือวิจัยเพื่อวัดคุณลักษณะหรือคุณสมบัติของสิ่งที่ต้องการศึกษา หรือเครื่องมือวัดตัวแปร ผลการวัดเป็นอย่างไรรู้ขึ้นอยู่กับเครื่องมือวิจัย ถ้าเครื่องมือวิจัยมีความเที่ยงตรงและเชื่อถือได้ มีคุณสมบัติการวัด (Psychometric properties) ที่ดี ผลการวัดที่ได้จะมีความถูกต้องและมีความน่าเชื่อถือมากยิ่งขึ้น สามารถนำผลการวัดที่ได้ไปใช้ประโยชน์ได้ตามความต้องการของผู้ใช้ แต่ถ้าเครื่องมือวิจัยมีคุณภาพไม่ดี ไม่มีความถูกต้องและความน่าเชื่อถือ ผลการวัดที่ได้ก็จะมีความน่าเชื่อถือลดลง เครื่องมือวิจัยจึงไม่ก่อให้เกิดประโยชน์ใด ๆ และไม่สามารถนำผลการวัดไปใช้ได้อย่างสมบูรณ์ หรือที่เรียกว่า “Garbage in, Garbage out” เหมือนกับการเอาขยะเข้าไป ก็ได้เป็นขยะออกมา เปรียบเปรยว่า หากข้อมูลที่ได้มาจากการวัดที่ไม่มีคุณภาพ ผลที่ได้ออกมาก็ไม่มีคุณภาพเช่นกัน ในการสร้างและพัฒนาเครื่องมือวัดทางพฤติกรรมศาสตร์ให้มีคุณภาพที่ดีและมีความน่าเชื่อถือนั้น เครื่องมือที่ดีควรมีลักษณะที่สำคัญ ดังนี้

1. **ความเที่ยงตรง (Validity)** โดยเครื่องมือวัดที่ดีต้องมีความเที่ยงตรง หมายถึง เครื่องมือวัดมีความสามารถในการวัดได้ตรงกับคุณลักษณะหรือคุณสมบัติของสิ่งที่ต้องการวัดจริง ในการพิจารณาว่าเครื่องมือวัดมีความเที่ยงตรงหรือไม่ จะต้องมีเกณฑ์ที่ใช้เปรียบเทียบว่าจะให้ตรงกับอะไรในสิ่งที่ต้องการวัด เครื่องมือวัดที่สร้างขึ้นมีความเที่ยงตรงต่อเกณฑ์อย่างหนึ่ง แต่อาจจะไม่มีความเที่ยงตรงเลยก็ได้เมื่อใช้เกณฑ์อีกอย่างหนึ่ง เช่น เมื่อนำ

แบบวัดเจตคติต่อการดูแลตนเองมาใช้กับกลุ่มวัยทำงาน พบว่ามีความเที่ยงตรงสูง แสดงว่า แบบวัดเจตคติต่อการดูแลตนเองทำการวัดเจตคติต่อการดูแลตนเองได้ตรงกับกลุ่มตัวอย่างวัยทำงาน แต่เมื่อนำแบบวัดเจตคติต่อการดูแลตนเองของกลุ่มวัยทำงานมาใช้กับกลุ่มเยาวชน อาจจะได้ค่าความเที่ยงตรงไม่สูง มีความเป็นไปได้ว่าแบบวัดเจตคติต่อการดูแลตนเองของกลุ่มวัยทำงานทำการวัดเจตคติต่อการดูแลตนเองได้ไม่ตรงกับกลุ่มเยาวชน

2. ความเชื่อมั่น (Reliability) เครื่องมือวัดต้องมีคุณสมบัติการวัดที่สามารถให้ผลการวัดที่มีความคงเส้นคงวาสม่ำเสมอ ไม่ว่าจะวัดกี่ครั้งก็ได้ผลเท่าเดิมหรือใกล้เคียงกับของเดิมมากที่สุด หรืออีกนัยหนึ่งคือ ความคงที่หรือความคงเส้นคงวาของค่าที่วัดได้จากสถานการณ์ของการวัดหรือการทดสอบ ไม่ว่าจะเป็นการวัดซ้ำ การวัดด้วยแบบวัดหลายข้อความ การวัดด้วยแบบวัดที่แทนกันได้ หรือการวัดโดยผู้ประเมินหรือผู้ตัดสินหลายคน (Kline, 2005) เช่น แบบวัดการปรับตัวของนักศึกษามหาวิทยาลัยมีความเชื่อมั่นสูง นำไปวัดกลุ่มตัวอย่างเดิมก็ได้ผลเท่าเดิม หรือแม้แต่นำแบบวัดการปรับตัวไปใช้ในมหาวิทยาลัยอื่น ก็ได้ผลการวัดเท่าเดิม ความเชื่อมั่นเป็นสิ่งจำเป็นสำหรับคุณสมบัติในด้านความเที่ยงตรง แต่ไม่ได้หมายความว่า เครื่องมือวัดที่มีค่าความเชื่อมั่นสูง จะเป็นเครื่องมือวัดที่มีความเที่ยงตรงเสมอไป เครื่องมือวัดที่ไม่มีค่าความเชื่อมั่น ผลการวัดจะแปรเปลี่ยนไปตลอด มีความคลาดเคลื่อนในการวัดสูง ส่งผลให้การวัดครั้งนั้นไม่น่าเชื่อถือ ไม่สามารถนำผลการวัดมาใช้ประโยชน์ต่อไปได้

3. อำนาจจำแนก (Discrimination) เครื่องมือวัดต้องมีอำนาจจำแนก นั่นคือ เครื่องมือวัดต้องมีคุณสมบัติที่สามารถจำแนก แยกแยะสิ่งที่มีลักษณะต่างกันออกจากกันได้ เครื่องมือวัดที่มีอำนาจจำแนกจะทำให้ได้ข้อมูลที่แสดงถึงความแตกต่างของบุคคลได้อย่างชัดเจน นั่นคือ สามารถจำแนกกลุ่มสูงและกลุ่มต่ำ กลุ่มเก่งและกลุ่มอ่อนได้ เช่น แบบทดสอบผลสัมฤทธิ์ทางการเรียนที่มีค่าอำนาจจำแนกสูง แสดงว่า ผู้เรียนกลุ่มเก่งทำแบบทดสอบผลสัมฤทธิ์ทางการเรียนได้ถูกต้องมากกว่าผู้เรียนกลุ่มอ่อน หรือแบบวัดการรู้เท่าทันสื่อที่มีค่าอำนาจจำแนกสูง แสดงว่า กลุ่มที่ได้คะแนนสูงมีการรู้เท่าทันสื่อที่มากกว่ากลุ่มที่ได้คะแนนต่ำ โดยทั่วไปเครื่องมือวัดที่มีอำนาจจำแนกมีความเกี่ยวข้องกับความเชื่อมั่นและความเที่ยงตรง ถ้าเครื่องมือมีค่าอำนาจจำแนกสูง ก็จะมีค่าความเชื่อมั่นสูงและความเที่ยงตรงสูงด้วย

4. ความเป็นปรนัย (Objectivity) เครื่องมือวัดต้องมีความเป็นปรนัย นั่นคือ เครื่องมือวัดมีการดำเนินการใช้การให้คะแนน และแปลความหมายของคะแนนโดยไม่ขึ้นอยู่กับความคิดเห็นหรือดุลยพินิจส่วนตัว (Subjective Judgment) ของบุคคลใดบุคคลหนึ่ง ทุกคนที่ใช้เครื่องมือวัดจะมีความเข้าใจตรงกันทั้งข้อคำถามมีความชัดเจน การตรวจให้คะแนนชัดเจน และการแปลความหมายของคะแนนได้ตรงกัน ทุกคนรับรู้สิ่งเดียวกัน

5. ความมีประสิทธิภาพ (Efficiency) เครื่องมือวัดต้องมีประสิทธิภาพ นั่นคือ เครื่องมือวัดมีประสิทธิภาพในการประหยัด ใช้ง่าย และสะดวกในการใช้ ถ้ามีเครื่องมือหลายชุดที่วัดในเรื่องเดียวกันที่มีคุณสมบัติทางการวัดที่มีคุณภาพ ก็ควรเลือกเครื่องมือวัดที่มีจำนวนข้อคำถามน้อยกว่า ใช้เวลาไม่มาก ง่ายต่อการนำไปใช้ โดยข้อคำถามมากใช้เวลาในการตอบนาน มีโอกาสมากที่จะตอบได้ไม่ครบถ้วน สำหรับแบบวัดที่มีการพัฒนามาในระยะเวลาหนึ่ง

มักจะพิจารณาทำแบบวัดฉบับปรับปรุงที่มีข้อคำถามน้อยลง เพื่อความกระชับและง่ายต่อการนำไปใช้ เป็นการเพิ่มประสิทธิภาพในการวัด (Polit & Beck, 2008)

ประเภทของเครื่องมือวัดทางพฤติกรรมศาสตร์

เครื่องมือวัดทางพฤติกรรมศาสตร์มีหลายประเภท ซึ่งแตกต่างกันตามความซับซ้อนในการสร้างเครื่องมือวัด รูปแบบการตอบ การนำไปใช้ การวิเคราะห์และการแปลผล โดยเครื่องมือวัดแต่ละประเภทจะเหมาะสมกับการใช้ เก็บรวบรวมข้อมูลที่แตกต่างกัน ผู้วิจัยจะต้องเลือกชนิดและลักษณะของเครื่องมือวัดที่สอดคล้องกับวัตถุประสงค์ของการนำข้อมูลไปใช้ และสอดคล้องกับประเภทและระดับการวัดของข้อมูล ซึ่งจะทำให้สามารถทดสอบสมมติฐานที่กำหนดไว้หรือตอบคำถามการวิจัยได้ถูกต้อง โดยเครื่องมือวัดทางพฤติกรรมศาสตร์ สามารถแบ่งประเภทได้ดังนี้

1. เครื่องมือวัดด้านการรู้คิด
 - แบบทดสอบ (Test)
2. เครื่องมือวัดด้านอารมณ์ ความรู้สึก เจตคติ บุคลิกภาพ
 - แบบสอบถาม (Questionnaire)
 - มาตราวัด (Scaling)
3. เครื่องมือวัดด้านพฤติกรรม
 - แบบสัมภาษณ์ (Interview Form)
 - แบบสังเกต (Observation Form)
 - แบบตรวจสอบรายการ (Check List)

สรุป

แนวคิดพื้นฐานเกี่ยวกับการวัดและเครื่องมือวัด ทำให้ทราบคำศัพท์ที่เกี่ยวข้องกับการวัด ได้แก่ ตัวแปร การวัด วิธีการวัด และการประเมิน รวมทั้งลักษณะการวัดทางพฤติกรรมศาสตร์ ข้อมูลประเภทต่าง ๆ ระดับของการวัด การนิยามตัวแปร นอกจากนี้ ยังมีข้อจำกัดในการวัดทางพฤติกรรมศาสตร์ รวมไปถึงคุณลักษณะของเครื่องมือวัดที่ดี และประเภทของเครื่องมือวัดทางพฤติกรรมศาสตร์ โดยสิ่งเหล่านี้ ทำให้มีความเข้าใจการวัดพฤติกรรมศาสตร์ในเบื้องต้นและในภาพรวม ก่อนจะนำไปสู่กระบวนการในการสร้างข้อคำถามต่อไป

เอกสารอ้างอิง

- กรกนก เขตดวง, ศรีณย์ พิมพ์ทอง, และพิชญานี พูนพล. (2562). ปัจจัยทางจิตสังคมที่ส่งผลต่อพฤติกรรมการเป็นพลเมืองดีบนโลกดิจิทัลของนักเรียนระดับชั้นมัธยมศึกษาตอนปลายของโรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ในเขตกรุงเทพมหานคร. *วารสารพฤติกรรมศาสตร์เพื่อการพัฒนา*, 11(1), 124-143.
- จารุพรรณ สุวรรณไพโรจน์, และประพิมพา จรัสรัตนกุล. (2560). ความสัมพันธ์ระหว่างการรับรู้ความยุติธรรมในองค์การโดยรวมและความผูกพันที่บุคลากรมีต่อองค์การ หัวหน้างาน และกลุ่มเพื่อนร่วมงาน โดยมีวัฒนธรรมระยะห่างอำนาจเป็นตัวแปรกำกับ. *วารสารวิจัย มสส. สาขามนุษยศาสตร์และสังคมศาสตร์*, 13(3), 61-80.
- ชาริน สุวรรณวงศ์. (2561). อิทธิพลของปัจจัยเชิงสาเหตุที่มีต่อพฤติกรรมป้องกันการสูบบุหรี่ของนักเรียนระดับมัธยมศึกษาตอนปลายในจังหวัดสตูล: การสังเคราะห์งานวิจัยและการวิเคราะห์หุ้ระดับ (วิทยานิพนธ์ปริญญาโทศึกษาศาสตร์). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
- ชาริน สุวรรณวงศ์, และศรีณย์ พิมพ์ทอง. (2561). ปัจจัยที่ส่งผลต่อพฤติกรรมการป้องกันการสูบบุหรี่ในกลุ่มวัยรุ่น: การสังเคราะห์งานวิจัยด้วยการทบทวนวรรณกรรมอย่างเป็นระบบ. *วารสารพฤติกรรมศาสตร์*, 24(1), 63-82.
- ณัฐชรินทร์ อภิวิชญ์ไชยชิตา. (2556). อิทธิพลของบุคลิกภาพแบบเปิดตัวและบุคลิกภาพแบบเป็นมิตรต่อความพึงพอใจในความสัมพันธ์ของคู่รักโดยมีรูปแบบความผูกพันเป็นตัวแปรส่งผ่านและเพศเป็นตัวแปรกำกับ (วิทยานิพนธ์ปริญญาโทศึกษาศาสตร์). กรุงเทพฯ: คณะจิตวิทยา จุฬาลงกรณ์มหาวิทยาลัย.
- บุญใจ ศรีสถิตย่นรากร. (2555). การพัฒนาและตรวจสอบคุณภาพเครื่องมือวิจัย: คุณสมบัตินวัตกรรมวัดเชิงจิตวิทยา. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- เบญจพร ประณีตวาทกุล, และดุจดเดือน พันธุมนาวัน. (2561). ปัจจัยทางจิตสังคมที่เกี่ยวข้องกับพฤติกรรมการป้องกันการรับและแพร่โรคในสถานที่สาธารณะของนักศึกษาในระดับปริญญาตรี. *วารสารพฤติกรรมศาสตร์เพื่อการพัฒนา*, 10(1), 83-103.
- วีรพงษ์ พวงเล็ก, นริศรา พิงไธสง, และดุชฎิ โยเหลา. (2561). วิธีการสร้างแบรนด์และผลของการสร้างแบรนด์ภายในองค์กรของสถาบันอุดมศึกษาในเขตกรุงเทพมหานคร. *วารสารพฤติกรรมศาสตร์เพื่อการพัฒนา*, 10(1), 104-125.
- ศรีณย์ พิมพ์ทอง. (2563). ประสิทธิภาพของการฝึกอบรมทางจิตพฤติกรรมศาสตร์เพื่อพัฒนาพฤติกรรมเอื้อสังคมในกลุ่มนักเรียนชั้นมัธยมศึกษา. *วารสารพฤติกรรมศาสตร์เพื่อการพัฒนา*, 12(2), 81-100.

- อรพินทร์ ชูชม. (2545). *เอกสารคำสอนวิชา วป 502 การสร้างและพัฒนาเครื่องมือวัดทางพฤติกรรมศาสตร์*.
กรุงเทพฯ: สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- อังศินันท์ อินทรกำแหง, ฉัตรชัย เอกปัญญาสกุล, วิชดา กิจธรรม, สิทธิพงศ์ วัฒนานนท์สกุล, อรัญ วานิชกร, ศรัณย์ พิมพ์ทอง, อัจฉรา ประเสริฐสิน, สุภารัตน์ ตันติวิวัฒน์, เจนนิเฟอร์ ชวโนวานิช, และพิชชาดา สุทธิแป้น. (2561). ปัจจัยเชิงสาเหตุที่มีต่อการพัฒนาหัวข้อวิจัยและการเผยแพร่ผลงานวิจัยของอาจารย์ มหาวิทยาลัยศรีนครินทรวิโรฒ โดยส่งผ่านคุณลักษณะที่ดีของการเป็นนักวิจัยที่เลี้ยง.
วารสารพฤติกรรมศาสตร์เพื่อการพัฒนา, 10(1), 42-62.
- อารยา เชียงของ. (2561). *ผลการพัฒนาความรอบรู้ด้านสุขภาพด้วยการจัดการเรียนรู้เพื่อการเปลี่ยนแปลงที่มีต่อพฤติกรรมควบคุมระดับน้ำตาลในเลือดของผู้ป่วยโรคเบาหวาน (วิทยานิพนธ์ปริญญาโทศึกษาศาสตร์)*.
กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York: David McKay Company.
- Ellington, H. I., Percival, F., & Race, P. (1993). *Handbook of educational technology* (3rd ed.).
London: Kogan Page.
- Kline, T. J. B. (2005). *Psychological testing: A practical approach to design and evaluation*.
Thousand Oaks, CA: Sage Publications.
- Matarazzo, J. D. (1990). Psychological assessment versus psychological testing. *American Psychologist*, 45(9), 999-1017.
- Nesselrode, J. R., & Molenaar, P. C. M. (2016). Some behavioral science measurement concerns and proposals. *Multivariate Behavioral Research*, 51(2), 396-412.
- Patton, M. Q. (1987). *Qualitative research evaluation methods*. Thousand Oaks, CA: Sage Publishers.
- Polit, D. F., & Beck, C. T. (2008). *Nursing research: Generating and assessing evidence for nursing practice* (8th ed.). Philadelphia: Lippincott Williams & Wilkins.

บทที่ 3

มาตรวัด

บทที่ 3 เป็นการนำเสนอมาตรวัด โดยผู้เขียนต้องการให้ผู้อ่านได้ทราบถึงแนวคิดเกี่ยวกับมาตรวัด หลักการสร้างมาตรวัด การใช้มาตรวัดประเภทต่าง ๆ และการแปลความหมายข้อมูลจากมาตรวัด มาตรวัดเป็นส่วนหนึ่งของกระบวนการพัฒนาเครื่องมือวิจัยทางพฤติกรรมศาสตร์ ที่ช่วยให้ผู้ตอบแบบสอบถามหรือแบบวัดสามารถให้ระดับความคิดเห็นของข้อความแต่ละข้อได้ นั่นคือ ช่วยประเมินลักษณะของข้อความที่นักวิจัยไม่สามารถวัดค่าออกมาเป็นตัวเลขในเชิงปริมาณ

ความหมายของมาตรวัด

มาตรวัด (Scale) หมายถึง เครื่องมือที่ช่วยกำหนดคุณลักษณะของสิ่งที่ต้องการวัดหรือตัวแปร ให้มีค่าตัวเลขประจำมาตร (Scale Value) อย่างมีระบบ

จุดเริ่มต้นของการใช้มาตรวัดเริ่มแรกส่วนใหญ่ใช้ในการวัดทัศนคติ แต่ปัจจุบันมาตรวัดสามารถใช้ได้อย่างแพร่หลายมากขึ้น นั่นคือ ขยายขอบเขตให้สามารถวัดตัวแปรทางด้านพฤติกรรมศาสตร์ได้ เช่น ความคิดเห็น ความเชื่อ การรับรู้ อารมณ์ บุคลิกภาพ พฤติกรรม ตลอดจนสภาพแวดล้อมของบุคคล

เครื่องมือการวัดทางพฤติกรรมศาสตร์ที่สามารถสร้างเป็นมาตรได้ (Scalability) ต้องมีข้อตกลงเบื้องต้นของมาตรวัด (อรพินทร์ ชูชม, 2545) ดังนี้

- โครงสร้างที่ต้องการวัดมีคุณสมบัติที่แปรเปลี่ยนในเชิงปริมาณที่สามารถวัดได้อย่างต่อเนื่อง โดยวัดบนความต่อเนื่องทางจิต (Psychological Continuum)
- ค่าตัวเลขประจำมาตรที่อยู่บนความต่อเนื่องทางจิต มีคุณสมบัติเป็นจำนวนตัวเลขที่แท้จริง อยู่ในมาตรของการวัดแบบใดแบบหนึ่ง เช่น มาตรเรียงอันดับ มาตรช่วง หรือมาตรวัดอัตราส่วน เป็นต้น

ความต่อเนื่องทางจิต (Psychological Continuum) เป็นความรู้สึกที่ต่อเนื่อง ไม่ได้รู้สึกเป็นท่อน ๆ แต่มีทิศทางและมีความเข้มข้นแตกต่างกันในแต่ละขั้นของความรู้สึกจึงเกี่ยวข้องเชื่อมโยงต่อเนื่องกัน ระดับความรู้สึกจะเริ่มด้วยความเข้มข้นน้อยไปสู่ความเข้มข้นมาก

ยกตัวอย่างการสร้างมาตรวัดโดยคำนึงถึงคนเป็นศูนย์กลาง

“มาตรวัดความซึมเศร้า” ประกอบด้วย ข้อความ 20 ข้อ และทำการจัดประเภทให้กับข้อความนั้นว่าเป็นตัวบ่งชี้ หรือไม่เป็นตัวบ่งชี้ความซึมเศร้า โดยหากคำตอบที่เลือกเป็นตัวบ่งชี้ จะให้ 1 คะแนน และหากคำตอบที่ไม่ใช่ตัวบ่งชี้ให้ 0 คะแนน

“มาตรวัดความซึมเศร้า” ประกอบด้วย ข้อความ 20 ข้อ ที่เลือกตอบจากเห็นด้วยอย่างยิ่ง ถึง ไม่เห็นด้วยอย่างยิ่ง มีคะแนนตั้งแต่ 1-5 คะแนน

ทั้งสองรูปแบบดังกล่าว เมื่อคำนวณหาคะแนนรวมของแต่ละคนแล้ว บุคคลที่ได้คะแนนสูงจะมีตำแหน่งอยู่ตรงปลายสุดทางบวกบนเส้นความต่อเนื่องทางจิตมากกว่าบุคคลที่ได้คะแนนน้อยกว่า

ผู้สร้างมาตรวัดเป็นผู้กำหนดน้ำหนักการให้คะแนน โดยทั่วไปมักจะทำให้น้ำหนักทุกข้อเท่ากัน ไม่สนใจความแตกต่างในข้อที่บุคคลต้องตอบ เช่น รูปแบบมาตรวัดของ Likert ผู้ตอบจะเลือกตอบจาก “เห็นด้วยอย่างยิ่ง” ถึง “ไม่เห็นด้วยอย่างยิ่ง” วิธีการดังกล่าวนี้เป็นการสร้างมาตรวัดที่คำนึงถึงคนเป็นศูนย์กลาง

2. วิธีการสร้างมาตรวัด โดยคำนึงถึงสิ่งเร้าเป็นศูนย์กลาง (Stimulus-centered method) เป็นวิธีการที่ผู้สร้างมาตรวัดสนใจที่จะจัดวางตำแหน่งของสิ่งเร้า (หรือข้อความ) ตามเส้นความต่อเนื่องทางจิต

นักจิตกายภาพ (Psychophysicists) สนใจที่จะระบุความแตกต่างที่น้อยที่สุดของสิ่งเร้าทางกายภาพดังกล่าว เช่น แสง เสียง ที่ประสาทสัมผัสของมนุษย์สามารถตรวจจับได้ เช่น การวัดการรับรู้ความสว่าง ต้องการค้นหาค่าความสว่างที่แตกต่างกันปริมาณน้อยที่สุดเท่าไร ที่มนุษย์จะสามารถบอกได้ว่าสิ่งไหนสว่างกว่า เป็นต้น แต่ปัญหาอยู่ที่การพัฒนาวิธีการสร้างมาตราวัดคำตอบของผู้ตอบว่า ความสว่างสามารถสร้างมาตรให้อยู่บนเส้นความต่อเนื่องทางจิตได้อย่างไร

การสร้างมาตรวัดจะเป็นวิธีทำให้ความแตกต่างของสิ่งเร้าทางกายภาพมาอยู่บนเส้นความต่อเนื่องทางจิต ซึ่งมีอยู่หลายวิธีด้วยกัน แต่มีวิธีหนึ่งที่รู้จักดีที่สุด นั่นคือ ให้คนคนเดียวกันหรือกลุ่มตัวอย่างกลุ่มเดียวกันเปรียบเทียบของ 2 อันซ้ำ ๆ กัน ถ้าสิ่งเร้าแต่ละอันได้รับเลือกด้วยความถี่ที่เท่ากัน (50% ของจำนวนครั้ง) ก็จะถือว่าสิ่งเร้าทั้งสองเท่าเทียมกันได้ตามมาตราประสาทสัมผัส (Sensory Scale) แต่ถ้าสิ่งเร้าหนึ่งถูกเลือกมากกว่าอีกสิ่งหนึ่งถึง 75 % ของจำนวนครั้ง จะถือว่าเกิดความแตกต่างตามมาตรทางจิตระหว่างสิ่งเร้าสองสิ่ง (Gescheider, 1997) เรียกว่า “ความแตกต่างที่สังเกตเห็นได้” (a just noticeable difference: JND) ซึ่ง JND นี้ใช้เป็นหน่วยมาตรวัดความต่อเนื่องทางจิตของนักจิตวิทยาในยุคเริ่มแรก

ต่อมา Thurstone ได้กำหนดวิธีการสร้างมาตรวัดสำหรับการวัดทัศนคติ โดยมีคุณสมบัติของมาตรของสิ่งเร้าที่ไม่เกี่ยวข้องกับารวัดทางกายภาพ เช่น การรับรู้ความร้ายแรงของอาชญากรรม สามารถสร้างมาตรได้โดยการถามผู้ตัดสินให้ระบุความร้ายแรงของอาชญากรรมที่มากที่สุดจากการเปรียบเทียบรายชื่อของอาชญากรรมที่

เป็นไปได้ เช่น สอบวางเพลิงกับขโมย สอบวางเพลิงกับฆาตกรรม นอกจากนี้ได้จัดวางข้อความแต่ละข้อตามเส้นความต่อเนื่องทางจิตที่แบ่งเป็นช่วงเท่ากัน และเมื่อได้ค่าประจำมาตรแล้ว ข้อความต่าง ๆ อาจนำไปให้กลุ่มผู้ตอบในรูปแบบเห็นด้วย-ไม่เห็นด้วย และคำตอบของผู้ตอบแต่ละคนจะได้จากการคำนวณค่าประจำมาตร

3. วิธีการสร้างมาตรวัด โดยคำนึงถึงการตอบเป็นศูนย์กลาง (Response-centered method) เป็นวิธีการสร้างมาตรที่ซับซ้อนที่สุด เป็นการสร้างมาตรวัดที่ผู้สร้างมาตรวัดสนใจที่จะจัดวางทั้งบุคคลและสิ่งเร้าตามตำแหน่งที่แตกต่างกันตามเส้นความต่อเนื่องทางจิต โดยอิงคำตอบที่เข้มข้น (หรือคำตอบที่ถูกต้อง) และในเวลาเดียวกันข้อความถูกวัดในรูปปริมาณหรือความเข้มข้นของคุณลักษณะของกลุ่มตัวอย่างที่มีอยู่

วิธีการสร้างมาตรวัดแบบนี้ที่เป็นที่รู้จักกันดี คือ มาตรของ Guttman หรือการวิเคราะห์ความเป็นมาตร (Scalogram analysis) นอกจากนี้ ยังมีวิธีการสร้างมาตรสำหรับคนและข้อความ เรียกว่า วิธีการแบบเปิด (Unfolding technique) เป็นการวัดเรียงลำดับวัตถุสิ่งเร้า ตามความรู้สึกพึงพอใจของผู้ตอบ

ประเภทของมาตรวัดที่สำคัญและการพัฒนามาตรวัด

มาตรวัดที่ใช้ในการพัฒนาเครื่องมือวัดทางพฤติกรรมศาสตร์ มีหลากหลายประเภท ซึ่งมีรายละเอียดดังนี้

1. มาตรวัดแบบเธอร์สโตน (Thurstone Scale)

มาตรวัดเธอร์สโตน พัฒนาโดยหลุยส์ แอล เธอร์สโตน (Louis L. Thurstone) ใช้สำหรับวัดเจตคติ โดยได้ใช้วิธีการพัฒนามาตรวัดที่เรียกว่า วิธีการกำหนดช่วงให้เท่ากัน (Method of equal-appearing intervals) เป็นวิธีการพัฒนามาตรวัดแบบมิติเดียว หรือความเป็นเอกมิติ (Unidimensional scale) กล่าวคือ ตัวแปรที่ต้องการวัดมีคุณลักษณะเป็นมิติเดียว เช่น เจตคติต่อการเรียนวิชาการะเบียบวิธีวิจัยทางพฤติกรรมศาสตร์

วิธีการกำหนดช่วงให้เท่ากัน ได้ให้มาตรวัดเจตคตินั้นอยู่ในมาตราการวัดช่วง/อันตรภาค (Interval scale) โดยมาตรวัดเธอร์สโตน มีขอบเขตการพัฒนามาตรวัดที่สำคัญอยู่ 2 ประเด็น ได้แก่

- 1) ข้อความแสดงความคิดเห็น/ความรู้สึกเกี่ยวกับเรื่องที่กำหนดประกอบกันหลายข้อความ
- 2) แต่ละข้อความจะมีค่าตัวเลขประจำมาตรในแต่ละข้อความ (Scale value)

ขั้นตอนการสร้างมาตรวัดเธอร์สโตนด้วยวิธีการกำหนดช่วงให้เท่ากัน

1. กำหนดโครงสร้างและขอบเขตของสิ่งที่ต้องการจะวัด เนื่องจากการวัดทางพฤติกรรมศาสตร์เป็นโครงสร้างที่ไม่สามารถทำการวัดได้โดยตรง ต้องอ้างอิงจากพฤติกรรมที่สามารถสังเกตและวัดได้ ซึ่งได้จากการตอบชุดข้อคำถาม ดังนั้น ผู้สร้างจึงต้องกำหนดขอบเขตสิ่งที่จะวัดให้ชัดเจน โดยต้องให้ครอบคลุมพฤติกรรมและคุณลักษณะทั้งหมดที่สังเกตเห็นได้และวัดได้ ที่เป็นตัวบ่งชี้ถึงตัวแปรหรือโครงสร้างที่ต้องการศึกษา

2. สร้างข้อความตามโครงสร้างที่กำหนดไว้ โดยมีขั้นตอนดังนี้

2.1 ทำการรวบรวมจากข้อความหรือความคิดเห็นที่เกี่ยวข้องกับสิ่งที่ต้องการวัดโดยอาศัยวิธีการสร้างนิยามเชิงปฏิบัติการ

2.2 เขียนข้อความทั้งหมดที่มีความครอบคลุมระดับต่าง ๆ ของตัวแปรที่ต้องการศึกษา

(1) ข้อความที่แสดงเจตคติในทางบวก (Positive Sentiment) ต่อสิ่งนั้น ได้แก่ ข้อความที่แสดงทิศทาง คุณค่า และลักษณะที่ดีต่อเรื่องนั้น

(2) ข้อความที่เป็นกลางต่อสิ่งนั้น

(3) ข้อความที่แสดงเจตคติในทางลบ (Negative Sentiment) ต่อสิ่งนั้น ได้แก่ ข้อความที่แสดงทิศทาง คุณค่า และลักษณะที่ไม่ดีต่อเรื่องนั้น

2.3 ควรรวบรวมข้อความที่สร้างขึ้นไว้จำนวนมากพอสมควร (Kline, 2000) หรือควรมีข้อความประมาณ 80-100 ข้อ (Trochim & Donnelly, 2006) หรือ 100-200 ข้อ (Streiner, Norman, & Cairney, 2015) ให้มีเนื้อหาสาระที่ครบถ้วนตามขอบเขตของสิ่งที่ต้องการศึกษา

3. กำหนดกลุ่มผู้เชี่ยวชาญในเรื่องนั้นและในการสร้างเครื่องมือวัดมาเป็นผู้ตัดสินน้ำหนัก โดยใช้จำนวนผู้ตัดสินที่มากพอสมควร (Kline, 2000) หรือจำนวนผู้ตัดสินที่แนะนำควรอยู่ที่ประมาณ 100 คน (Nunnally & Bernstein, 1994) โดยผู้ตัดสินต้องมีความเป็นกลาง ไม่ใช่อคติส่วนบุคคลในการประเมินข้อความเหล่านั้น และผู้ตัดสินแต่ละคนจะมีการพิจารณาของแต่ละคนเป็นอิสระต่อกัน

4. วิธีการในการตัดสินน้ำหนักของข้อความ ผู้ตัดสินแต่ละคนจะประเมินค่าข้อความแต่ละข้อว่าควรอยู่ในตำแหน่งใดของมาตรวัดของเรื่องที่ศึกษา โดยพิจารณาข้อความทั้งหมดแล้วเรียงลำดับข้อความไว้เป็นกลุ่มจำนวน 11 กลุ่ม จากกลุ่ม 1 ถึง 11 โดยทั้ง 11 กลุ่มนั้น มีมาตรการวัดช่วงที่เท่ากัน ซึ่งผู้ตัดสินจะประเมินว่าข้อความแต่ละข้อแสดงคุณลักษณะที่ดีหรือไม่ดีต่อเรื่องนั้นมากน้อยเพียงใด แต่ต้องไม่ให้ความรู้สึกส่วนตัวเข้าไปมีผลต่อการตัดสินเด็ดขาด ทั้งนี้หมายเลขของแต่ละกลุ่มมีความหมายดังนี้

1	2	3	4	5	6	7	8	9	10	11
ข้อความที่สนับสนุนน้อยที่สุด หรือ ข้อความที่แสดงถึงลักษณะที่ไม่ดีที่สุด ต่อเรื่องนั้น				ข้อความที่เป็นกลาง			ข้อความที่สนับสนุนมากที่สุด หรือ ข้อความที่แสดงถึงลักษณะที่ดีที่สุดต่อเรื่องนั้น			

ข้อความกลุ่มที่ 1

คือ ข้อความที่ผู้ตัดสินเห็นว่า ข้อความนั้นมีความเข้มข้นในลักษณะที่แสดงเจตคติทางลบต่อสิ่งนั้นมากที่สุด หรือผู้ตัดสินสนับสนุนน้อยที่สุดต่อเจตคติในสิ่งนั้น

ข้อความกลุ่มที่ 6

คือ ข้อความที่ผู้ตัดสินเห็นว่า ข้อความนั้นมีความเข้มข้นในลักษณะที่แสดงเจตคติที่เป็นกลาง ในลักษณะที่ไม่สนับสนุนหรือไม่คัดค้าน

- ข้อความกลุ่มที่ 11 คือ ข้อความที่ผู้ตัดสินเห็นว่า ข้อความนั้นมีความเข้มข้นในลักษณะที่แสดง เจตคติทางบวกต่อสิ่งนั้นมากที่สุด หรือผู้ตัดสินสนับสนุนมากที่สุดต่อเจตคติ ในสิ่งนั้น
- ข้อความกลุ่มที่ 2, 3, 4 และ 5 คือ ข้อความที่ผู้ตัดสินเห็นว่า ข้อความนั้นมีความเข้มข้นในลักษณะที่แสดง เจตคติทางลบต่อสิ่งนั้นลดลงตามลำดับ
- ข้อความกลุ่มที่ 7, 8, 9 และ 10 คือ ข้อความที่ผู้ตัดสินเห็นว่า ข้อความนั้นมีความเข้มข้นในลักษณะที่แสดง เจตคติทางลบต่อสิ่งนั้นเพิ่มขึ้นตามลำดับ

5. กำหนดค่าประจำมาตรา (Scale Value) โดยนำผลที่ได้จากการประเมินของผู้ตัดสินทุกคนมา คำนวณหาค่าสถิติรายข้อ โดยใช้ค่าสถิติสองประเภทได้แก่

5.1 ค่ามัธยฐาน (Median) หรือ “s”

- ใช้เป็นค่าประจำมาตราของแต่ละข้อความ
- ทำให้ทราบว่าผู้ตัดสินได้ประเมินว่าข้อความนั้นเป็นข้อความที่วัดเจตคติได้ในระดับใด และใช้เป็นคะแนนเจตคติของแต่ละข้อ

5.2 ค่าพิสัยระหว่างควอไทล์ (Interquatile Range) หรือ “Q”

- เป็นการกระจายของผู้ตัดสิน (ความแตกต่างระหว่างควอไทล์ที่ 3 กับ 1)
- ค่า Q ที่คำนวณได้ ใช้เป็นเกณฑ์ในการคัดเลือกข้อความได้
- ยิ่งค่า Q มีค่ามาก แปลว่า ข้อความข้อนั้นทำให้ผู้ตัดสินประเมินค่าได้ไม่สอดคล้องกันเป็นส่วนใหญ่ ทำให้ทราบว่าข้อความนั้นคลุมเครือและกำกวม ไม่น่าจะนำมาใช้ในมาตรวัด
- ถ้าค่า Q มีค่าน้อยมาก แปลว่า ข้อความนั้นทำให้ผู้ตัดสินประเมินค่าได้สอดคล้องกัน ข้อความนั้นจึงเหมาะแก่การถูกคัดเลือกมาใช้ในมาตรวัด

6. คัดเลือกข้อความขั้นสุดท้ายเพื่อเป็นมาตรวัด โดยปกติมาตรวัด ข้อความที่คัดเลือกควรมีประมาณ 10-20 ข้อ (Kline, 2000; Nunnally & Bernstein, 1994) หรือประมาณ 25 ข้อ (Streiner et al., 2015) โดยมีเกณฑ์ในการคัดเลือกคือ

6.1 ให้ข้อความทั้งหมดมีค่ามัธยฐานกระจายอยู่ครบทุกช่วง เพราะจะทำให้มั่นใจว่าข้อความทั้งหมดในมาตรวัดครอบคลุมความต่อเนื่องของเรื่องที่ศึกษาโดยรวม

6.2 คัดเลือกข้อความที่มีค่า Q น้อยไว้ และตัดข้อความที่มีค่า Q มากทิ้งไป หากมีข้อความที่มีค่ามัธยฐานใกล้เคียงกันในช่วงหนึ่ง ๆ หลายข้อ ให้เลือกเอาข้อความที่มีค่า Q น้อยที่สุดไว้

6.3 จัดเรียงข้อความที่คัดเลือกได้ทั้งหมดแบบสุ่ม โดยไม่ให้ค่าประจำมาตราของแต่ละข้อความ เรียงลำดับกัน

ยกตัวอย่างมาตรวัดเธอร์สโตน

“เจตคติต่อวิชาพุทธศาสนาของนักเรียนชั้นมัธยมศึกษา”

คำแนะนำ: โปรดกาเครื่องหมาย X ในช่อง เฉพาะข้อที่เห็นด้วย

- 1. วิชาพุทธศาสนาเป็นเรื่องล้าหลัง (1.4)
- 2. วิชาพุทธศาสนามีเนื้อหาที่ซับซ้อน ยากที่จะเข้าใจ (3.2)
- 3. เป็นการยากที่จะตัดสินใจว่าวิชาพุทธศาสนามีประโยชน์ต่อการใช้ชีวิต (5.5)
- 4. วิชาพุทธศาสนามีหลากหลายเรื่องที่น่าเรียนรู้ (7.5)
- 5. หน้าที่อันสูงสุดของเราชาวพุทธคือ การศึกษาวิชาพุทธศาสนา (10.8)

**ค่าที่อยู่ในวงเล็บ คือ ค่าตัวเลขประจำข้อ ซึ่งจะปรากฏในแบบวัด

จากตัวอย่างดังกล่าว สามารถแปลความหมายของเจตคติต่อวิชาพุทธศาสนาของนักเรียนชั้นมัธยมศึกษาได้ว่า

ข้อ 1 มีค่าใกล้ 1 มากที่สุด แสดงถึง เจตคติทางลบ/ไม่ดีต่อสิ่งนั้น

ข้อ 5 มีค่าใกล้ 11 มากที่สุด แสดงถึง เจตคติทางบวก/ดีต่อสิ่งนั้น

ถ้าคนตอบมาตรวัดนี้ ตอบเห็นด้วยกับข้อ 1 และ 2 แสดงว่า บุคคลมีคะแนนเจตคติต่อวิชาพุทธศาสนา = $(1.4+3.2)/2 = 2.3$ แสดงว่า ผู้ตอบมีเจตคติทางลบ หรือเจตคติที่ไม่ดีต่อวิชาพุทธศาสนา นั่นคือ ไม่ชอบวิชาพุทธศาสนา

สำหรับคนตอบมาตรวัดนี้ ตอบเห็นด้วยกับข้อ 3, 4 และ 5 แสดงว่า บุคคลมีคะแนนเจตคติต่อวิชาพุทธศาสนา = $(5.5+7.5+10.8)/3 = 7.9$ แสดงว่า ผู้ตอบมีเจตคติทางบวก หรือเจตคติที่ดีต่อวิชาพุทธศาสนา นั่นคือ ยอมรับและเห็นประโยชน์ของวิชาพุทธศาสนา

การนำมาตรวัดเธอร์สโตนไปใช้

การนำมาตรวัดเธอร์สโตนแบบกำหนดช่วงให้เท่ากันไปใช้วัดเจตคติก็จะมีแต่ข้อความเพียงอย่างเดียว ไม่มีค่าประจำมาตร และค่า Q กำกับอยู่ ให้ถามแต่เพียงว่าเห็นด้วยหรือไม่เห็นด้วย กับข้อความแต่ละข้อเท่านั้น ไม่ต้องระบุความเห็นหรือให้ทำเครื่องหมายว่าเห็นด้วยหรือไม่เห็นด้วยในข้อใด แล้วนำเอาค่าประจำมาตร ได้แก่ ค่ามัธยฐานของข้อความที่บุคคลนั้นตอบว่าเห็นด้วยมารวมกัน แล้วหาค่าเฉลี่ย เพื่อวัดเจตคติของบุคคลที่มีต่อสิ่งนั้นว่ามีในระดับใด และหากต้องการทราบเจตคติของคนกลุ่มหนึ่ง ๆ ก็ให้นำเอาค่าเฉลี่ยประจำมาตรของแต่ละคนมาเฉลี่ยรวมกันอีกครั้ง เพื่อให้ทราบเจตคติของคนกลุ่มนั้นต่อสิ่งที่ศึกษา (อรพินทร์ ชูชม, 2545)

ข้อจำกัดของมาตรวัดเธอร์สโตน

1. มีความเป็นไปได้ยากที่จะทำให้การตัดสินใจหรือประเมินค่าประจำข้อความของผู้เชี่ยวชาญเป็นไปโดยปราศจากความรู้สึกส่วนตัวของผู้ประเมิน มีโอกาสที่ผู้ตัดสินอาจประเมินความสำคัญของแต่ละข้อความที่แตกต่างกัน

นั่นคือ ให้ความสำคัญกับความคิดเห็นของผู้ตัดสินที่มีต่อข้อความนั้น ๆ ทำให้เจตคติต่อสิ่งนั้นอาจไม่ได้สะท้อนในวิธีการที่เชื่อถือได้มากนัก (Kite & Whitley, Jr., 2018; Mangal & Mangal, 2013)

2. มาตรการได้รับการวิพากษ์วิจารณ์ว่าเป็นวิธีการวัดที่ไม่มีประสิทธิภาพในการวัดความแตกต่างระหว่างบุคคลทางด้านเจตคติ เพราะแม้จะมีวิธีการสร้างมาตรการวัดที่ซับซ้อน แต่มีผู้ตอบจำนวนมากที่มีรูปแบบการตอบแตกต่างกัน แต่กลับได้คะแนนเจตคติในแบบเดียวกัน

3. มาตรการเรอร์สโตนกำหนดให้วัดมิติเดียวหรือความเป็นเอกมิติ โดยวัดความรู้สึกที่เกี่ยวกับการวัดเจตคติในรูปแบบของการเห็นด้วยหรือไม่เห็นด้วยกับข้อความแต่ละข้อความ (Kite & Whitley, Jr., 2018; Mangal & Mangal, 2013) แต่ยังขาดหลักฐานที่แสดงความเป็นเอกมิติของมาตรการวัด

4. นำมาใช้ได้ยากในทางปฏิบัติเพราะมีวิธีการสร้างมาตรการวัดที่ซับซ้อนและต้องใช้ผู้เชี่ยวชาญจำนวนมาก จึงเป็นการสิ้นเปลืองเวลาและค่าใช้จ่าย จึงไม่เป็นที่นิยมใช้ในปัจจุบัน (Kline, 2000)

2. มาตรการลิเคิร์ต (Likert Scale)

มาตรการลิเคิร์ต พัฒนาโดยเร็นซิส ลิเคิร์ต (Rensis Likert) เพื่อใช้ในการวัดเจตคติ มาตรการนี้มีการรวมคะแนนทุกข้อที่ใช้ในมาตร เพื่อนำมาแปลผล จึงมีชื่อเรียกอีกอย่างหนึ่งว่า มาตรการประมาณรวมค่า (Summated Rating Scale) ปัจจุบัน มาตรการลิเคิร์ตเป็นที่นิยมใช้กันอย่างแพร่หลาย เพราะมีกระบวนการสร้างที่ง่ายและสะดวก ลักษณะที่สำคัญของมาตรการนี้ คือ (Lewis-Beck, 1994; Spector, 1992)

1) ผลที่ได้จากการมาตรการวัดนี้ เป็นการรวมคะแนนจากหลายข้อความไว้ด้วยกัน โดยคะแนนการวัดของทุกข้อความจะนำมารวมกันเป็นคะแนนเดียว

2) ข้อความแต่ละข้อวัดคุณลักษณะ (Attribute) ของตัวแปรที่ต้องการวัด ซึ่งมีคุณสมบัติการวัดต้องที่มีความต่อเนื่องทางการวัดเชิงปริมาณ (Quantitative measurement continuum) เช่น เจตคติแปรเปลี่ยนจากชอบมากที่สุด ไปยังไม่ชอบมากที่สุดได้อย่างต่อเนื่อง

3) ข้อความแต่ละข้อไม่มีคำตอบที่ “ถูกต้อง” ทำให้มาตรการนี้แตกต่างไปจากแบบทดสอบแบบเลือกตอบที่มีหลายตัวเลือก (multiple-choice test) ทำให้มาตรการเหล่านี้ไม่สามารถใช้เพื่อทดสอบความรู้หรือความสามารถได้

4) ในการตอบมาตรการวัดนี้ ผู้ตอบต้องประเมินค่าข้อความแต่ละข้อ แล้วจึงระบุว่าจะเลือกระดับใดที่สะท้อนคำตอบที่ตรงกับความรู้สึกนึกคิดและเจตคติที่แท้จริงของผู้ตอบ

ขั้นตอนการสร้างมาตรการลิเคิร์ต

1. กำหนดโครงสร้างและขอบเขตของสิ่งที่ต้องการจะวัด เนื่องจากการวัดทางพฤติกรรมศาสตร์เป็นโครงสร้างที่ไม่สามารถทำการวัดได้โดยตรง ต้องอ้างอิงจากพฤติกรรมที่สามารถสังเกตและวัดได้ ซึ่งได้จากการตอบชุดข้อความ ดังนั้น ผู้สร้างจึงต้องกำหนดขอบเขตสิ่งที่จะวัดให้ชัดเจน โดยต้องให้ครอบคลุมพฤติกรรมและคุณลักษณะทั้งหมดที่สังเกตเห็นได้และวัดได้ ที่เป็นตัวบ่งชี้ถึงตัวแปรหรือโครงสร้างที่ต้องการศึกษา

2. สร้างข้อความให้ครอบคลุมตามโครงสร้างที่กำหนดไว้ โดยเขียนข้อความทั้งหมดที่มีความครอบคลุมระดับต่าง ๆ ของตัวแปรที่ต้องการศึกษา นั่นคือ

(1) ข้อความแสดงความคิด/การกระทำทางบวก แสดงทิศทาง คุณค่าและลักษณะที่ดีต่อสิ่งนั้น

(2) ข้อความแสดงความคิด/การกระทำทางลบ แสดงทิศทาง คุณค่าและลักษณะที่ไม่ดีต่อสิ่งนั้น

แต่จุดสำคัญที่แตกต่างจากการสร้างมาตรวัดแบบเรอร์สโตน คือ ไม่ควรมีข้อความเป็นกลางต่อเรื่องนั้น เพราะจะทำให้เกิดการแปลความหมายที่คลุมเครือได้

3. กำหนดมาตรประเมินค่า มาตรวัดแบบลิเคิร์ตดั้งเดิมใช้มาตรประเมินค่า 5 ระดับ เพื่อให้คนเลือกตอบระดับใดระดับหนึ่งเพียงระดับเดียวในแต่ละข้อความ โดยทั่วไปใช้มาตรประเมินค่า 5 หรือ 7 ระดับ (Kline, 2000) แต่ในทางปฏิบัติอาจกำหนดมากกว่าหรือน้อยกว่า 5 ระดับก็ได้ เช่น 3 ระดับ 7 ระดับ หรือ 9 ระดับ โดยอาจเลือกใช้กับกลุ่มผู้ตอบที่แตกต่างกัน

- มาตรวัดลิเคิร์ตจะให้ผู้ตอบแสดงการเห็นพ้องต้องกัน (Agreement) กับข้อความ ตัวเลือกจะเป็นไปในลักษณะ “เห็นด้วย” และ “ไม่เห็นด้วย” โดยมีลักษณะเป็นแบบสองขั้ว (bipolar) ที่สมมาตรกัน หรือมีสัดส่วนสองข้างที่สมดุลกัน นั่นคือ ระดับที่เห็นด้วยและระดับที่ไม่เห็นด้วยมีจำนวนเท่ากัน และใช้คำที่แสดงระดับคุณลักษณะเหมือนกัน ส่วนระดับตรงกลางอาจมีหรือไม่มีก็ได้ ถ้ามี ควรใช้ตัวเลือกว่า “ไม่ใช่ทั้งสอง” (neither agree nor disagree) หรือ “ไม่มีความเห็น” (no opinion) (Spector, 1992; Streiner et al., 2015)

ยกตัวอย่างมาตรวัดลิเคิร์ตในรูปแบบการเห็นพ้องต้องกัน

“แบบวัดการเปิดกว้างต่ออนาคต (Openness to the Future Scale)” (Botella et al., 2018)

ข้อความ	ตัวเลือกคำตอบ				
	เห็นด้วย อย่างยิ่ง	เห็นด้วย	ไม่แน่ใจ	ไม่เห็นด้วย	ไม่เห็นด้วย อย่างยิ่ง
ฉันยอมรับทั้งสิ่งดีและไม่ดีที่เกิดขึ้นกับฉัน					
ฉันรู้สึกมีความหวังเกี่ยวกับอนาคต					
ฉันคิดว่าฉันควบคุมการดำเนินชีวิตได้					

- สำหรับคำตอบรูปแบบการประเมินค่า นั้น ก็มีลักษณะเป็นแบบสองขั้ว โดยการประเมินค่าเป็นการพิจารณาความดี-ความไม่ดี มีทั้งทางบวกและทางลบ ซึ่งอาจจะเริ่มจาก “แย่มากหรือไม่เป็นที่ยอมรับ” ไปจนถึง “ดีเลิศ” (สุจินดา เทียนสวัสดิ์, 2562)

ยกตัวอย่างมาตรวัดลิเคิร์ทรูปแบบการประเมินค่า

“แบบวัดสถานะทางสุขภาพ (The SF-36 Health Survey)” (Ware, Jr. & Gandek, 1998)

ข้อความ	ตัวเลือกคำตอบ				
	ดีเลิศ	ดีมาก	ดี	พอใช้	แย่
โดยทั่วไป สุขภาพของฉัน					

- สำหรับตัวเลือกที่เป็นลักษณะแบบขั้วเดียว (unipolar) โดยตัวเลือกมักเริ่มตั้งแต่ไม่มีคุณลักษณะนั้นเลยหรือมีน้อยที่สุด จนถึงมีคุณลักษณะนั้นมากที่สุด เช่น มาตรวัดคำคุณศัพท์ (Adjectival scale) เป็นมาตรวัดที่มีคำคุณศัพท์กำกับทุกระดับ นอกจากนี้ ตัวเลือกที่เป็นความถี่ ก็เป็นแบบขั้วเดียวเช่นกัน (สุจินดา เทียนสวัสดิ์, 2562; Streiner et al., 2015)

ยกตัวอย่างมาตรวัดลิเคิร์ทรูปแบบคำคุณศัพท์

“แบบวัดแรงกดดันทางสังคมวัฒนธรรมให้มีรูปร่างผอมบาง ทั้งจากครอบครัว กลุ่มเพื่อน และสื่อ (สุพัตรา สกุศลศรีประเสริฐ, 2560)

ข้อความ	ตัวเลือกคำตอบ				
	จริงที่สุด	จริง	ไม่ใช่ทั้งสอง	ไม่จริง	ไม่จริงที่สุด
สมาชิกในครอบครัวเคยหยอกล้อฉันเกี่ยวกับน้ำหนักตัวหรือรูปร่างของฉัน					
ฉันเคยรู้สึกกดดันจากกลุ่มเพื่อนของฉันในการลดน้ำหนักเพื่อให้มีรูปร่างผอมบาง					
เมื่อดูโฆษณา ฉันเชื่อว่าการมีรูปร่างดีและสวยงามคือการมีรูปร่างที่ผอมบางเหมือนนางแบบในโฆษณา					

ยกตัวอย่างมาตรวัดลิเคิร์ทรูปแบบความถี่

“แบบวัดพฤติกรรมการกินอาหารเพื่อสุขภาพ” (สุพัตรา สกุศลศรีประเสริฐ, 2560)

ข้อความ	ตัวเลือกคำตอบ				
	เป็นประจำ	บ่อยครั้ง	บางครั้ง	นาน ๆ ครั้ง	ไม่เคยเลย
ฉันกินอาหารหลากหลายชนิดเพื่อให้ได้สารอาหารครบถ้วน					
ฉันเลือกซื้ออาหารอย่างระมัดระวัง เพื่อให้ได้กินอาหารที่ปราศจากสารเคมีและสารตกค้างที่อาจเป็นอันตราย					
ฉันกินขนมจุบจิบระหว่างมื้ออาหาร					

ปัจจุบัน เครื่องมือที่ใช้มาตรวัดลิเคิร์ท สามารถใช้ตัวเลือกแบบสองขั้วหรือแบบขั้วเดียวก็ได้ แต่สิ่งสำคัญคือ คำตอบที่กำหนดให้เลือกจะต้องเหมาะสมกับข้อความ เช่น ข้อความที่เกี่ยวกับการปฏิบัติพฤติกรรม ควรจะใช้คำตอบที่เป็นความถี่จึงจะเหมาะสม เป็นต้น และตัวเลือกที่อยู่ตำแหน่งตรงกลางต้องเหมาะสมตามลำดับความต่อเนื่อง โดยควรสะท้อนการมีคุณลักษณะตรงกลาง และไม่ใช่ไม่มีความสามารถที่จะตอบคำถาม เช่น ตัวเลือก “ไม่สามารถตอบได้” อาจสะท้อนให้เห็นถึง ข้อความนั้นไม่เกี่ยวข้อง ผู้ตอบไม่เข้าใจข้อความ หรือผู้ตอบไม่สามารถตัดสินใจได้ ดังนั้น จึงควรใช้คำว่า “ไม่ทั้งสอง” หรือ “ปานกลาง” (Streiner et al., 2015)

ยกตัวอย่างมาตรวัดลิเคิร์ทที่ไม่มีตำแหน่งตรงกลาง

“แบบวัดการคงอยู่ในองค์กรของพนักงานมหาวิทยาลัยสายวิชาการ (ศรัณย์ พิมพ์ทอง, 2557)

ข้อความ	ตัวเลือกคำตอบ					
	จริงที่สุด	จริง	ค่อนข้างจริง	ค่อนข้างไม่จริง	ไม่จริง	ไม่จริงเลย
ท่านมีแผนจะเปลี่ยนงานใหม่ภายใน 3 ปีข้างหน้า						
ท่านสนุกกับการทำงานในมหาวิทยาลัย						
ท่านรู้สึกเบื่อหน่าย เหนื่อยล้ากับการทำงานวิชาการ						

นอกจากนี้ ยังแล้วแต่ว่าผู้สร้างจะเห็นว่าสิ่งที่ต้องการวัดนั้นมีความเป็นกลางจริง ๆ หรือไม่ หากให้ประเมินสิ่งที่เกี่ยวข้องกับความคิดเห็นหรือความรู้สึกที่มีต่อสิ่งหนึ่งสิ่งใด แล้วไม่สามารถตัดสินใจได้ ก็ควรตัดคำตอบที่แสดงความเป็นกลางออก แต่อาจเป็นการเพิ่มโอกาสให้เกิดหลีกเลี่ยงการไม่เป็นที่ต้องการของสังคม โดยเลือกตอบค่อนข้างไปทางลบ (Chyung, Roberts, Swanson, & Hankinson, 2017)

4. กำหนดน้ำหนักหรือคะแนนในแต่ละระดับมาตราประเมินค่าของข้อความในมาตรวัด โดยยึดเนื้อความของข้อความเป็นหลัก คือแบ่งออกเป็น

- ข้อความทางบวก (คือ ข้อความที่มีเนื้อความเป็นไปตามความประสงค์ตรงกับที่ต้องการศึกษา)
- ข้อความทางลบ (คือข้อความที่มีเนื้อความตรงกันข้ามกับสิ่งที่ต้องการศึกษา) ซึ่งจะให้คะแนนกลับกัน

ข้อความทางบวก		ข้อความทางลบ	
เห็นด้วยอย่างยิ่ง	= 5 หรือ 4	เห็นด้วยอย่างยิ่ง	= 1 หรือ 0
เห็นด้วย	= 4 หรือ 3	เห็นด้วย	= 2 หรือ 1
ไม่แน่ใจ	= 3 หรือ 2	ไม่แน่ใจ	= 3 หรือ 2
ไม่เห็นด้วย	= 2 หรือ 1	ไม่เห็นด้วย	= 4 หรือ 3
ไม่เห็นด้วยอย่างยิ่ง	= 1 หรือ 0	ไม่เห็นด้วยอย่างยิ่ง	= 5 หรือ 4

5. ทดลองใช้มาตรวัดลิเคิร์ตที่สร้างขึ้น โดยเป็นการนำข้อความทั้งหมดที่สร้างขึ้นไปทดลองใช้กับกลุ่มตัวอย่างที่คล้ายคลึงกับกลุ่มประชากรที่ต้องการศึกษา โดยสุ่มเลือกมาจำนวนหนึ่ง แล้วตรวจให้คะแนน

- ถ้าคะแนนรวมจากการตอบข้อความทั้งหมดมีค่าสูง แสดงว่าบุคคลมีความคิด/การกระทำในเรื่องนั้นมาก (หรือในทิศทางบวก)

- ถ้าคะแนนรวมจากการตอบข้อความทั้งหมดมีค่าต่ำ แสดงว่าบุคคลมีความคิด/การกระทำในเรื่องนั้นน้อย (หรือในทิศทางลบ)

6. นำผลที่ได้จากการทดลองใช้มาทำการตรวจสอบคุณภาพรายข้อหรือวิเคราะห์รายข้อของมาตรวัดฉบับนั้นโดยทำเพื่อคัดเลือกข้อคำถามหรือข้อความในขั้นสุดท้าย ซึ่งถือว่าเป็นขั้นตอนที่สำคัญ เพราะมาตรวัดแบบลิเคิร์ตนี้ได้กำหนดว่าต้องประกอบขึ้นมาจากข้อความต่าง ๆ ที่มีความสัมพันธ์ระหว่างกัน และรายข้อแต่ละข้อต้องมีความสัมพันธ์ทางบวกกับคะแนนรวมด้วย เพื่อเป็นการแสดงว่ามาตรวัดดังกล่าวมีมิติเดียวหรือความเป็นเอกมิติ (คือ วัดสิ่งที่เป็นโครงสร้างเดียวกัน) โดยมีวิธีการที่ใช้เป็นเกณฑ์สำหรับการคัดเลือกข้อความ ดังนี้

6.1 การวิเคราะห์ความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวม (Item-Total Correlation) โดยทำการคัดเลือกเฉพาะข้อที่มีค่าสัมประสิทธิ์สหสัมพันธ์ทางบวกกับคะแนนรวมสูงอย่างมีนัยสำคัญทางสถิติไว้ แสดงถึงว่า แต่ละข้อคำถามและแบบวัดทั้งฉบับสามารถวัดได้ในสิ่งเดียวกัน

6.2 การวิเคราะห์หาค่าอำนาจจำแนกของข้อความแต่ละข้อ โดยทั่วไปนิยมใช้เทคนิค 25% เป็นเกณฑ์ในการแบ่งกลุ่มสูง-กลุ่มต่ำ หรืออาจใช้เทคนิคอื่น ๆ ก็ได้ เช่น 33%, 27% หรือ 20% เป็นต้น

- ในกรณีที่ใช้เทคนิค 25% ทำได้โดยจัดเรียงคะแนนรวมของผู้ตอบแต่ละคนจากคะแนนสูงสุดไปยังคะแนนต่ำสุด และคัดเอาเฉพาะกลุ่มคนที่ได้คะแนนสูงไว้จำนวน 25% และกลุ่มที่คะแนนต่ำไว้อีก 25% ของจำนวนผู้ตอบมาตรวัดทั้งหมด

- จากนั้นทำการเปรียบเทียบความแตกต่างของคะแนนเฉลี่ยรายข้อของกลุ่มสูงและกลุ่มต่ำ โดยใช้สถิติทดสอบที (t-test) เป็นรายข้อ แล้วจึงทำการคัดเลือกข้อคำถาม โดยคัดเอาเฉพาะข้อที่สามารถจำแนกกลุ่มสูงและกลุ่มต่ำออกจากกันได้อย่างมีนัยสำคัญทางสถิติ

การนำมาตรวัดแบบลิเคิร์ตไปใช้

การนำมาตรวัดลิเคิร์ตไปใช้นั้น ในมาตรวัดฉบับหนึ่งๆ ควรประกอบด้วยข้อความประมาณ 20-25 ข้อที่ได้รับการตรวจสอบคุณภาพของเครื่องมือแล้ว และควรมีข้อความทางบวกและข้อความทางลบในจำนวนพอ ๆ กัน สำหรับการเรียงลำดับข้อความควรอยู่ในลักษณะสุ่ม โดยคะแนนของแต่ละบุคคล คือ คะแนนรวมทุกข้อ โดยทำการตรวจให้คะแนนตามเกณฑ์หรือ Likert Scale ที่กำหนดไว้ โดยหากเป็นข้อความทางลบ ต้องทำการกลับคะแนนก่อนที่จะรวมคะแนนของข้อความทั้งหมด สำหรับคนที่ได้คะแนนรวมในตัวแปรนั้น ๆ สูง หมายถึง คนที่มีคุณลักษณะของตัวแปรนั้นในระดับสูง ส่วนคนที่ได้คะแนนต่ำ หมายถึง คนที่มีคุณลักษณะของตัวแปรนั้นในระดับต่ำ

แม้ว่า ผู้เชี่ยวชาญทางด้านการทดสอบจำนวนมาก มองว่า มาตรฐานวัดแบบลิเคิร์ตยังคงเป็นข้อมูลการวัดในมาตราเรียงลำดับ (Ordinal Scale) เนื่องจากระดับของมาตราในแต่ละช่วงอาจไม่เท่ากัน แต่เมื่อนำคะแนนมารวมกันแล้วได้เป็นหนึ่งค่า ก็ถือว่าเป็นข้อมูลการวัดในมาตราช่วง/อันตรภาค (Interval Scale) (Burns & Grove, 2009)

ข้อจำกัดของมาตรวัดลิเคิร์ต

1. คะแนนรวมที่ได้ของแต่ละบุคคลที่ใช้มาตรวัดลิเคิร์ตที่มีรูปแบบการตอบหลายแบบที่สามารถทำได้ คะแนนรวมแบบเดียวกัน ดังนั้น จึงเป็นข้อจำกัดในการแปลความหมายคะแนนให้เป็นแบบเดียวกัน เช่น บุคคลที่ได้คะแนนตัวแปรในระดับปานกลาง อาจมีรูปแบบการตอบส่วนใหญ่ที่เลือกคำตอบอยู่ในช่วงกึ่งกลางของมาตรวัด หรืออาจมีรูปแบบการตอบที่เลือกคำตอบไปในทางบวกหรือทางลบปะปนกันไป การแปลความหมายของคะแนนที่มีความจำกัดเช่นนี้ ทำให้มาตรวัดลิเคิร์ตมีปัญหาขาดความเที่ยงตรงและความเชื่อมั่นได้ นอกจากนี้ อคติจากการตอบ ได้แก่ การตอบเพื่อให้เป็นที่ยอมรับจากผู้อื่น (Social desirability) และการตอบเน้นไปทางบวก (Acquiescent) ก็ทำให้เกิดปัญหาทางด้านคุณภาพของเครื่องมือวัดเช่นกัน (Kreitchmann, Abad, Ponsoda, Nieto, & Morillo, 2019)

2. มาตรวัดลิเคิร์ตที่ใช้การวิเคราะห์รายข้อเป็นเกณฑ์ในการคัดเลือกข้อความ ข้อความที่ได้รับ การคัดเลือกต้องสามารถจำแนกเป็นกลุ่มสูงและกลุ่มต่ำออกจากกันได้ ดังนั้น ข้อความที่แสดงความเป็นกลางต่อสิ่งหนึ่ง ๆ มักจะมีค่าอำนาจจำแนกต่ำ ทำให้ข้อความเหล่านั้นถูกคัดออกไป ดังนั้น เมื่อจัดเรียงข้อความลงบนมาตรวัดที่แสดงความต่อเนื่องทางจิต ข้อความตรงกลางจะขาดหายไป จึงมีเฉพาะข้อความทางบวกหรือทางลบที่อยู่ตรงปลายหรือค่อนข้างปลายของมาตรวัดเท่านั้น

3. มาตรวัดออสกู๊ด (Osgood Scale)

มาตรวัดออสกู๊ด หรือเรียกอีกอย่างหนึ่งว่า “มาตรวัดความแตกต่างทางความหมายของภาษา” (Semantic Differential Scale) พัฒนาโดยชาร์ลส์ อี ออสกู๊ด (Charles E. Osgood) และคณะ โดยใช้กลุ่มภาษา (Semantic space) ได้แก่ คำคุณศัพท์ที่ใช้แสดงความหมายเกี่ยวกับคุณลักษณะของสิ่งที่ต้องการศึกษา

ออสกู๊ดเชื่อว่ามโนทัศน์ (Concept) สามารถอธิบายได้ด้วยคำคุณศัพท์ที่ตรงข้ามกันที่เป็นขั้วของมาตรวัด (Bipolar Adjective) ได้แก่ คำคุณศัพท์ที่มีความหมายทั้งในทางที่ดีและทางที่ไม่ดีต่อสิ่งนั้น หรือเป็นคำตรงกันข้ามที่ถูกกำกับไว้ที่ปลายของแต่ละด้านของเส้นตรงที่เป็นมาตรวัด โดยคำคุณศัพท์ต่าง ๆ ที่นำมาใช้ในมาตรวัดแบบออสกู๊ดจะสามารถบอกให้ทราบถึงคุณลักษณะของบุคคลได้

โดยกำหนดความหมายใน 3 องค์ประกอบ (Rosenberg & Navarro, 2018) ได้แก่

1) องค์ประกอบด้านการประเมินค่า (Evaluative) เป็นคำคุณศัพท์ที่แสดงออกถึงการมุ่งเน้นคุณค่าและการประเมินค่าของสิ่งหนึ่ง ๆ เช่น ดี-เลว รัก-เกลียด สำเร็จ-ล้มเหลว สะอาด-สกปรก เป็นที่พึงพอใจ-ไม่เป็นที่พึงพอใจ เป็นต้น

2) องค์ประกอบด้านศักยภาพ (Potential) เป็นคำคุณศัพท์ที่แสดงออกถึงกำลังหรือพลังอำนาจของสิ่งหนึ่ง ๆ เช่น แข็งแรง-อ่อนแอ หนัก-เบา ใหญ่-เล็ก ดั่ง-ค้อย ทนทาน-บอบบาง ยาว-สั้น เป็นต้น

3) องค์ประกอบด้านกิจกรรม (Activity) เป็นคำคุณศัพท์ที่แสดงออกถึงลักษณะการเคลื่อนไหวของสิ่งหนึ่ง ๆ เช่น เร็ว-ช้า กระตือรือร้น-เฉื่อยชา ร้อน-เย็น ขยัน-ขี้เกียจ เป็นต้น

ยกตัวอย่างคำคุณศัพท์ที่ใช้กันบ่อย ในแต่ละองค์ประกอบ/มิติ ดังตารางที่ 3-1

ตารางที่ 3-1 คู่คำคุณศัพท์ที่ใช้กันบ่อยในมาตรวัดออสกูด

การประเมินค่า	ศักยภาพ	กิจกรรม
ดี/เลว	อ่อนแอ/แข็งแรง	เฉื่อยชา/กระตือรือร้น
ถูก/แพง	ไม่ตัดสินใจ/ตัดสินใจ	ขี้เกียจ/ขยัน
โง่/ฉลาด	อ่อน/แข็ง	ไร้จุดหมาย/มีแรงจูงใจ
ขี้เหร่/สวย	ไร้ศักยภาพ/มีศักยภาพ	สงบ/ตื่นเต้น
ไม่ซื่อสัตย์/ซื่อสัตย์	รุนแรง/อ่อนโยน	ช้า/เร็ว

ยกตัวอย่าง แบบวัดเจตคติทางเพศของผู้สูงอายุ (Park & Shin, 2015)

1	ไม่มีความสุข	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	มีความสุข
2	ไม่น่ายินดี	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	น่ายินดี
3	ไม่จำเป็น	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	จำเป็น

ขั้นตอนการสร้างมาตรวัดออสกูด

1. เลือกมโนทัศน์หรือโครงสร้างที่ต้องการวัด โดยมโนทัศน์นี้จะกำหนดให้คงที่ และให้ผู้ตอบจัดอันดับคุณภาพตามมาตรคำคุณศัพท์ที่จะกำหนดให้เป็นตัวเลือก มโนทัศน์นี้ควรเป็นสิ่งที่ผู้ตอบมีความคุ้นเคยรู้จักกันทั่วไป มีความเข้าใจตรงกัน แต่ผู้ตอบสามารถแสดงความรู้สึกนึกคิดต่อสิ่งนั้นได้แตกต่างกันออกไป และมโนทัศน์นั้นควรเลือกให้ครอบคลุมสิ่งที่ต้องการศึกษา เช่น ถ้าศึกษาเจตคติต่อการดูแลรักษาผู้สูงอายุ อาจประกอบด้วยมโนทัศน์หลายอย่าง เช่น ร่างกาย จิตใจ สังคม จิตวิญญาณ เป็นต้น

2. เลือกคำคุณศัพท์ที่ใช้เป็นขั้วของมาตรวัด โดยเป็นการเลือกคู่คำคุณศัพท์ที่มีความหมายตรงข้ามที่เกี่ยวข้องตรงประเด็นกับมโนทัศน์ที่ต้องการศึกษาให้มากที่สุด โดยสามารถอธิบายมโนทัศน์ของสิ่งที่ศึกษาได้

ข้อจำกัดของมาตรวัดแบบออสกูด

- เป็นที่นิยมใช้กันอย่างกว้างขวาง เนื่องจากขั้นตอนการสร้างมาตรวัดที่ง่าย อีกทั้งสามารถใช้วัดได้ทั้งทิศทางและความเข้มของทิศทางของคุณลักษณะของมโนทัศน์ที่ต้องการศึกษา

- มีข้อจำกัดในการเลือกคู่ของคำคุณศัพท์ที่มีความหมายตรงกันข้ามมาใช้ ซึ่งบางคนก็อาจตีความหมายของคำเดียวกันได้แตกต่างกัน ซึ่งในภาษาไทย อาจแก้ปัญหาได้โดยการเติมคำว่า “ไม่” ลงไปหน้าคำคุณศัพท์หนึ่งก็ได้ หากไม่สามารถหาคำคุณศัพท์ที่มีความหมายตรงข้ามอย่างชัดเจนได้ เช่น เป็นประโยชน์-ไม่เป็นที่ประโยชน์ สบายใจ-ไม่สบายใจ คืบเคย-ไม่คืบเคย เหมาะสม-ไม่เหมาะสม เป็นต้น

4. มาตรฐานประเมินค่า (Rating Scale)

มาตรฐานประเมินค่า ได้มีการนำมาใช้ในทางจิตวิทยา และเป็นมาตรวัดนิยมใช้กันมากในงานวิจัยทางด้านพฤติกรรมศาสตร์เพื่อประเมินพฤติกรรม การปฏิบัติงาน บุคลิกภาพ รวมทั้งเจตคติของบุคคล โดยมาตรฐานประเมินค่า (Burns & Grove, 2009) ประกอบด้วย

- 1) ข้อความเกี่ยวกับคุณลักษณะของสิ่งหนึ่งหรือตัวแปร
- 2) ชุดตัวเลือกของคำตอบที่มีความต่อเนื่องทางจิต
- 3) มีการกำหนดค่าตัวเลขให้กับแต่ละตัวเลือก

มาตรฐานประเมินค่า มีลักษณะที่คล้ายคลึงกับมาตรวัดลิเคิร์ท ในแง่ของชุดข้อความและชุดคำตอบตัวเลือก แต่มีความแตกต่างกันตรงที่ลักษณะของตัวเลือกที่มาตรวัดประมาณค่าไม่มีรูปแบบจำเพาะ อาจเป็นแบบหลายตัวเลือกที่ใช้คำคุณศัพท์หรือข้อความที่แตกต่างกันตามความเหมาะสม กับข้อความแต่ละข้อคละกันไป หรือเป็นมาตรวัดหลายระดับ ซึ่งมักใช้ตัวเลือกแบบขั้วเดียวเท่านั้น

ยกตัวอย่างมาตรฐานประเมินค่า

แบบวัดความรู้ด้านสุขภาพ (อังศินันท์ อินทรกำแหง และแอนน์ แมคคาสกิล, 2563)

ที่	ความรู้ด้านสุขภาพ	ระดับความเป็นจริงตรงกับตัวท่าน				
		มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด
1	ฉันเปิดรับข้อมูลสุขภาพอยู่เสมอ เพื่อฉันจะได้เป็นผู้มีสุขภาพดี	5	4	3	2	1
2	ฉันอ่านและสามารถเข้าใจข้อมูลด้านสุขภาพจากสื่อต่าง ๆ เช่น เอกสาร สิ่งพิมพ์ แผ่นพับ โปสเตอร์ ใบสั่งแพทย์ เป็นต้น	5	4	3	2	1
3	ฉันหาข้อมูลด้านสุขภาพที่ถูกต้องจากแหล่งต่าง ๆ เพื่อนำมาใช้ประกอบการยืนยันข้อมูลนั้น ก่อนที่จะเชื่อและทำตาม	5	4	3	2	1

ขั้นตอนการสร้างมาตรฐานประเมินค่า

1. กำหนดโครงสร้างและขอบเขตของสิ่งที่ต้องการจะวัด ผู้สร้างจึงต้องกำหนดขอบเขตสิ่งที่วัดให้ชัดเจน โดยต้องให้ครอบคลุมพฤติกรรมและคุณลักษณะทั้งหมดที่สังเกตเห็นได้และวัดได้ ที่เป็นตัวบ่งชี้ถึงตัวแปรหรือโครงสร้างที่ต้องการศึกษา

2. กำหนดประเภทของมาตรฐานประเมินค่าที่จะใช้วัด รวมทั้งกำหนดลักษณะรูปแบบคำถาม และชุดคำตอบที่ใช้ ว่ามีกี่ระดับ ในแต่ละข้อ ตลอดจนกำหนดจำนวนข้อที่จะใช้วัด รวมทั้งระยะเวลาในการวัด

3. เขียนข้อความที่จะใช้ในมาตรฐานประเมินค่าตามรูปแบบที่กำหนดไว้ โดยให้ครอบคลุมโครงสร้างและเนื้อหาสาระนั้น

4. ตรวจสอบคุณภาพของมาตรฐานประเมินค่า ผู้สร้างต้องพิจารณามาตรฐานประเมินค่าที่สร้างขึ้นมานี้ว่าสามารถวัดในสิ่งที่ต้องการวัดจริง ๆ โดยให้ผู้เชี่ยวชาญหรือผู้ทรงคุณวุฒิในเรื่องที่จะวัดเป็นผู้พิจารณาคุณภาพของมาตรฐานประเมินค่าในประเด็นต่าง ๆ ได้แก่ มีความสอดคล้องกับนิยามปฏิบัติการ ความเหมาะสมของภาษาและรูปแบบ และความเหมาะสมของการให้คะแนน

5. มาตรฐานประเมินค่าที่ได้รับการตรวจสอบคุณภาพเบื้องต้นจากผู้เชี่ยวชาญแล้ว และได้รับการปรับปรุงแก้ไขให้นำไปทดลองใช้กับกลุ่มตัวอย่างที่เป็นตัวแทนของประชากร เพื่อตรวจสอบอำนาจจำแนกรายข้อ ความเชื่อมั่น และความเที่ยงตรง แล้วทำการปรับปรุงแก้ไขและทดสอบจนกว่าจะได้มาตรฐานประเมินค่าที่มีคุณสมบัติทางการวัดที่มีคุณภาพดี ทั้งความเที่ยงและเชื่อถือได้

ความคลาดเคลื่อนในการใช้มาตรวัด

สำหรับการนำมาตรวัดแบบต่าง ๆ ไปใช้นั้น อาจเกิดความคลาดเคลื่อนในการใช้มาตรวัดได้ ซึ่งส่งผลกระทบต่อความเชื่อถือได้ของผลการนำไปใช้ ดังนี้ (อรพินทร์ ชูชม, 2545)

1. ความคลาดเคลื่อนที่คงที่ (Constant or Range Restriction Error) แบ่งได้เป็น 3 ประเภท ได้แก่

1.1 ความคลาดเคลื่อนแบบใจกว้าง (Generosity Error) เป็นความคลาดเคลื่อนที่เกิดขึ้นเมื่อผู้ประเมินมีแนวโน้มที่จะประเมินสิ่งต่าง ๆ ไปในทางคะแนนสูงหรือดีมากเกินไป และหลีกเลี่ยงการประเมินไปในทางคะแนนต่ำหรือไม่ดี

1.2 ความคลาดเคลื่อนแบบรุนแรง (Severity Error) เป็นความคลาดเคลื่อนที่เกิดขึ้นเมื่อผู้ประเมินมีแนวโน้มที่จะประเมินสิ่งต่าง ๆ ไปในทางคะแนนต่ำหรือไม่ดีมากเกินไป และหลีกเลี่ยงการประเมินไปในทางคะแนนสูงหรือดีมาก

1.3 ความคลาดเคลื่อนแบบโน้มหว่าส่วนกลาง (Central-tendency Error) เป็นความคลาดเคลื่อนที่เกิดขึ้นเมื่อผู้ประเมินมีแนวโน้มที่จะประเมินสิ่งต่าง ๆ อยู่ตรงกลาง และหลีกเลี่ยงการประเมินไปในทางคะแนนสูงและต่ำ

2. **ความคลาดเคลื่อนแบบฮาโล (Halo Effect)** เป็นความคลาดเคลื่อนที่เกิดขึ้นเมื่อผู้ประเมินมีความประทับใจในผู้ถูกประเมิน ทำให้ประเมินคุณลักษณะอื่น ๆ ตามความประทับใจอันนั้น เช่น ถ้าผู้ประเมินประทับใจบุคคลนั้นในเชิงบวก ผู้ประเมินมีแนวโน้มประเมินคุณลักษณะอื่น ๆ ให้เป็นไปในทางบวกด้วย

3. **ความคลาดเคลื่อนแบบตรงข้าม (Contrast Error)** เป็นความคลาดเคลื่อนที่เกิดขึ้นเมื่อผู้ประเมินมีแนวโน้มประเมินค่าสูงกว่าที่ควรจะเป็น ถ้าผู้ถูกประเมินก่อนหน้านี้ได้รับการประเมินค่าต่ำมาก ๆ หรือเป็นการประเมินค่าต่ำกว่าที่ควรจะเป็น ถ้าผู้ถูกประเมินก่อนหน้านี้ได้รับการประเมินค่าสูงมาก ๆ

4. **ความคลาดเคลื่อนแบบใกล้เคียง (Proximity Error)** เป็นความคลาดเคลื่อนที่เกิดขึ้นเมื่อผู้ประเมินมีแนวโน้มประเมินบุคคลในลักษณะคล้ายคลึงกันในข้อต่าง ๆ ที่อยู่ใกล้เคียงกัน เช่น ถ้าบุคคลได้รับการประเมินอย่างคงเส้นคงวาว่าสูง กลาง หรือต่ำ ในชุดของข้อต่าง ๆ มีอยู่ใกล้เคียงกัน ส่วนใหญ่ข้ออื่น ๆ ที่อยู่ใกล้เคียงก็จะถูกประเมินไปในแนวทางเดียวกัน

5. **ความคลาดเคลื่อนจากผลการปฏิบัติงานหรือพฤติกรรมล่าสุด (Most-recent Performance Error or Latest Behavior Effect)** เป็นความคลาดเคลื่อนที่เกิดขึ้นเมื่อผู้ประเมินพิจารณาเฉพาะผลการปฏิบัติงานหรือพฤติกรรมล่าสุดที่เกิดขึ้นในช่วงระยะหลังใกล้เคียงกับเวลาที่จะประเมิน ผู้ประเมินไม่ได้ให้ความสนใจต่อพฤติกรรมทั้งหมดของผู้ถูกประเมิน

6. **ความคลาดเคลื่อนเนื่องจากความกำกวม (Ambiguity Error)** เป็นความคลาดเคลื่อนที่เกิดขึ้นเนื่องจากการบรรยายลักษณะต่าง ๆ ที่ไม่ชัดเจนในตำแหน่งของมาตร หรือมาตรที่ประกอบด้วยข้อความต่าง ๆ ที่ให้ข้อมูลไม่เพียงพอ ความคลุมเครือของภาษาในระดับต่าง ๆ ของมาตรวัดที่เป็นตัวเลือกต่าง ๆ เป็นต้น

7. **ความคลาดเคลื่อนเนื่องจากมีข้อมูลไม่เพียงพอ (Inadequate Information Error)** เป็นความคลาดเคลื่อนที่เกิดขึ้นในการประเมินผลการปฏิบัติงานที่ไม่มีข้อมูลผู้ถูกประเมินอย่างเพียงพอ โดยเฉพาะเมื่อผู้ประเมินจะต้องประเมินบุคคลหลายคนที่แตกต่างกันและไม่มีโอกาสที่จะสังเกตบุคคลแต่ละคนในช่วงเวลานั้น การขาดข้อมูลที่เพียงพอทำให้ผู้ประเมินอาจได้รับอิทธิพลที่เกินจริงจากการสื่อสารที่ไม่ถูกต้องได้ หรือประเด็นต่าง ๆ ที่ไม่เกี่ยวข้องถูกนำมาประเมินและได้รับความสนใจมากกว่า

8. **ความคลาดเคลื่อนเนื่องจากการใช้ดุลยพินิจของผู้ประเมิน (Error due to Certain Cognitive or Thought Processes Underlying Raters' Judgments)** เป็นความคลาดเคลื่อนที่เกิดขึ้นจากการที่ผู้ประเมินได้รับอิทธิพลในการประเมินความรู้สึกรู้สึก/ความเชื่อที่มีต่อผู้ถูกประเมิน มากกว่าสิ่งที่ผู้ประเมินรู้อย่างแท้จริง ๆ เกี่ยวกับบุคคลนั้น โดยผู้ประเมินอาจนึกถึงเฉพาะข้อมูลที่ยืนยันเกี่ยวกับความเชื่อที่มีต่อผู้ถูกประเมิน

สรุป

มาตรวัดเป็นกระบวนการหนึ่งที่สำคัญในการพัฒนาเครื่องมือวิจัย เป็นวิธีที่ช่วยในการกำหนดค่าตัวเลข เพื่อให้สามารถวิเคราะห์ข้อมูลในเชิงปริมาณได้ โดยทำความเข้าใจตั้งแต่ความหมายของมาตรวัด หลักการสร้างมาตรวัด ประเภทของมาตรวัดที่สำคัญและการพัฒนามาตรวัด เช่น มาตรวัดแบบ Thurstone มาตรวัดแบบ Likert มาตรวัดแบบ Osgood รวมทั้งมาตรประเมินค่า (Rating Scale) และความคลาดเคลื่อนในการใช้มาตรวัด โดยสิ่งเหล่านี้ ทำให้เกิดความเข้าใจในกระบวนการสร้างและพัฒนาเครื่องมือวัดทางด้านพฤติกรรมศาสตร์ได้ต่อไป

เอกสารอ้างอิง

- ศรัณย์ พิมพ์ทอง. (2557). รายงานการวิจัยเรื่องปัจจัยเชิงเหตุที่เกี่ยวข้องกับการคงอยู่ในองค์กรของพนักงาน มหาวิทยาลัยสายวิชาการ. กรุงเทพฯ: สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- สุจินดา เทียนสวัสดิ์. (2562). การพัฒนาเครื่องมือสำหรับการวิจัยทางการพยาบาล. เชียงใหม่: สยามพิมพ์นานา.
- สุพัตรา สกฤตศรีประเสริฐ. (2560). ปัจจัยเชิงเหตุและผลของรูปแบบการกินอาหารของนักเรียนหญิงมัธยมศึกษาตอนปลาย: การวิเคราะห์แบบจำลองโครงสร้างความสัมพันธ์พหุระดับ (วิทยานิพนธ์ปริญญาคุศภศึกษบัณฑิต). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
- อรพินทร์ ชูชม. (2545). เอกสารคำสอนวิชา วป 502 การสร้างและพัฒนาเครื่องมือวัดทางพฤติกรรมศาสตร์. กรุงเทพฯ: สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- อังศินันท์ อินทรกำแหง และแอนน์ แมคคาสกิล. (2563). รายงานฉบับสมบูรณ์เรื่องการประเมินและการสร้างความรอบรู้ด้านสุขภาพและสุขภาวะครอบครัวกลุ่มเลี้ยงโรคไม่ติดต่อเรื้อรัง: การพัฒนาโมเดลเชิงสาเหตุทางวัฒนธรรมของพฤติกรรมสุขภาพและโปรแกรมจิตวิทยาเชิงบวกและความรอบรู้ด้านสุขภาพ. กรุงเทพฯ: สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- Botella, C., Molinari, G., Fernández-Álvarez, J., Guillén, V., García-Palacios, A., Baños, R. M., & Tomás, J. M. (2018). Development and validation of the openness to the future scale: A prospective protector factor. *Health and Quality of Life Outcomes*, 16, 72.
- Burns, N., & Grove, S. (2009). *The practice of nursing research: Appraisal, synthesis and generation of evidence* (6th ed.). St. Louise, MO: Saunders/Elsevier.
- Chyung, S. Y., Roberts, K., Swanwon, I., & Hankinson, A. (2017). Evidence-based survey design: The use of a midpoint on the Likert scale. *Performance Improvement*, 56(10), 15-23.
- Deci, E. L., & Ryan, R. M. (2000). The “What” and “Why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227-268.

- Gescheider, G. A. (1997). *Psychophysics: The fundamentals* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kite, M. E., & Whitley, Jr., B. E. (2018). *Principles of research in behavioral science* (4th ed.). New York: Routledge.
- Kline, P. (2000). *Handbook of psychological testing* (2nd ed.). London: Routledge.
- Kreitchmann, R. S., Abad, F. J., Ponsoda, V., Nieto, M. D., & Morillo, D. (2019). Controlling for response for response biases in self-report scales: Forced-choice vs. psychometric modeling of Likert items. *Frontier in Psychology, 10*, 2309.
- Lewis-Beck, M. S. (1994). *Basic measurement*. London: Sage.
- Mangal, S. K., & Mangal, S. (2013). *Research methodology in behavioural sciences*. New Delhi: PHI Learning Private Limited.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Park, H., & Shin, S. (2015). Development and psychometric testing of a semantic differential scale of sexual attitude for the older person. *Nursing & Health Sciences, 17*(4), 526-532.
- Rosenberg, B. D., & Navarro, M. A. (2018). Semantic differential scaling. In *the SAGE encyclopedia of educational research, measurement, and evaluation*. Edited by B. B. Frey (Ed.), pp. 1504-1507. Thousand Oaks, CA: Sage.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.
- Spector, P. E. (1992). *Summated rating scale construction: An introduction*. Newbury Park, CA: Sage.
- Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use* (5th ed.). Oxford: Oxford University Press.
- Trochim, W. M., & Donnelly, J. P. (2006). *The research methods knowledge base* (3rd ed.). Cincinnati, OH: Atomic Dog.
- Waltz, C. F., Strickland, O. L., & Lenz, E. R. (2005). *Measurement in nursing and health research* (3rd ed.). New York: Springer Publishing Company.
- Ware, J. E., Jr., & Gandek, B. (1998). Overview of the SF-36 health survey and the International Quality of Life Assessment (IQOLA) project. *Journal of Clinical Epidemiology, 51*(11), 903-912.

บทที่ 4

การวัดด้านการรู้คิด

บทที่ 4 เป็นการนำเสนอการวัดด้านการรู้คิด โดยผู้เขียนต้องการให้ผู้อ่านได้ทราบและทำความเข้าใจเกี่ยวกับการวัดด้านการรู้คิด ความเป็นมาของการวัดทางด้านการรู้คิด แนวคิดเบื้องต้นของการวัดด้านการรู้คิด สำหรับการวัดทางด้านการรู้คิดมีการสร้างแบบทดสอบและวิธีการวัดอย่างมากมาย ส่วนใหญ่เน้นการพัฒนาแบบทดสอบสติปัญญา แบบทดสอบวัดความถนัด และแบบทดสอบผลสัมฤทธิ์ โดยบทนี้ จะช่วยให้ผู้วิจัยและนักวิชาการสามารถสร้างแบบสอบถามด้านการรู้คิดได้อย่างถูกต้อง เพื่อช่วยให้กระบวนการพัฒนาเครื่องมือวิจัยทางพฤติกรรมศาสตร์มีความเหมาะสมและสมบูรณ์มากยิ่งขึ้น

ความเป็นมาของการวัดทางด้านการรู้คิด

บุคคลแรก ๆ ที่กล่าวถึงการวัดทางด้านการรู้คิดอย่างเป็นระบบ นั่นคือ Sir Francis Galton ในช่วงปลายศตวรรษที่ 19 โดย Galton พัฒนาแบบทดสอบการทำหน้าที่ของสมองพื้นฐาน เช่น การแบ่งแยกความรู้สึก การรับรู้ความเร็ว เป็นต้น โดยมีสมมติฐานว่า ความแตกต่างในความสามารถทางสติปัญญาเป็นผลมาจากประสิทธิภาพของระบบประสาทส่วนกลางที่แตกต่างกันไป การทดสอบดังกล่าวได้รับความสนใจเป็นอย่างมาก สามารถเก็บรวบรวมข้อมูลได้มากกว่า 9,000 คน แต่ความเที่ยงตรงในการวัดทางด้านการรู้คิดไม่ได้ดีอย่างที่ควรจะเป็น ผลการทดสอบพบว่า มีความสัมพันธ์กันต่ำระหว่างความสามารถทางสติปัญญาและความสำเร็จด้านการศึกษา

บุคคลต่อมา นั่นคือ Alfred Binet ในช่วงต้นศตวรรษที่ 20 โดย Binet ได้พัฒนาวิธีการที่รวดเร็วและเชื่อถือได้ในการระบุเด็กปัญญาอ่อน ซึ่งเป็นผู้ที่ไม่ได้คาดหวังผลกำไรจากการเรียนในระบบปกติในโรงเรียน Binet และ Simon ได้ทำการปรับปรุงสภาพแวดล้อมต่าง ๆ ในการทดสอบความรู้และทักษะเชิงปฏิบัติมากกว่าการทำหน้าที่ของสมองเหมือน Galton โดยเด็กจะได้รับคำสั่งให้ชี้ไปที่ส่วนต่าง ๆ ของร่างกาย ให้บอกชื่อวัตถุที่เห็นในรูปภาพ ให้นิยามความหมาย ให้พุดชุดตัวเลขซ้ำ ๆ หรือเติมประโยคให้สมบูรณ์ การบอกเวลาในนาฬิกา และอื่น ๆ ซึ่งแนวคิดของ Binet และ Simon เป็นแนวคิดใหม่ที่มีการใช้อายุของเด็ก มาเป็นเกณฑ์ภายนอกสำหรับการวัดทางด้านการรู้คิด ด้วยการมอบหมายงานให้กลุ่มอายุที่แตกต่างกัน แบบวัดสติปัญญาจึงเป็นการวัดเชิงปรนัยว่าเด็กมีความก้าวหน้าหรือถอยหลังสำหรับเด็กในช่วงอายุนั้น ๆ และเป็นการแบ่งแยกเด็กปัญญาอ่อนออกจากเด็กคนอื่น ๆ ด้วยการเปรียบเทียบสมองกับอายุตามเวลาจริงถือเป็นพื้นฐานสำหรับแบบวัดสติปัญญา (Intelligence Quotient: IQ) และกลายเป็นแบบทดสอบสติปัญญาที่ใช้กันทุกวันนี้ นั่นคือ แบบทดสอบ IQ ของ Wechler

ในปัจจุบัน รูปแบบการวัดทางด้านการรู้คิด มีลักษณะเป็นขั้น ประกอบด้วย 3 ระดับ ได้แก่ 1) องค์ประกอบทั่วไป (g factor) สะท้อนความจริงที่ว่า การรู้คิดที่หลากหลายมีความใกล้เคียงกับความเป็นสากล 2) ความฉลาดที่ติดตัวมา (Fluid intelligence) เป็นความสามารถในการแก้ปัญหา ประเมินโดยใช้ภาพที่มีต่อการให้เหตุผลเชิงนามธรรม หรือการแก้ปัญหาที่ซับซ้อน รวมทั้งความฉลาดแบบตกผลึก (Crystallized intelligence) สะท้อนความรู้ทั่วไป วัดโดยใช้แบบทดสอบวัดคำศัพท์ การสะกดคำ หรือข้อมูลทั่วไป และ 3) การรู้คิดที่เฉพาะเจาะจง (Specific cognitive abilities) ได้แก่ การให้เหตุผลเชิงปริมาณ โดยเป็นสิ่งที่ไม่อยู่ในสองระดับแรก โดยทั้งสามระดับนี้ อาจอธิบายโครงสร้างความแตกต่างระหว่างบุคคลทางด้านการรู้คิดได้ดี (Grabner & Stern, 2011)

แนวคิดเบื้องต้นของการวัดทางด้านการรู้คิด

ความสามารถด้านการรู้คิด (Cognitive Ability) เป็นความสามารถทางสมองโดยทั่วไป ซึ่งเกี่ยวข้องกับ การให้เหตุผล การแก้ปัญหา การวางแผน การคิดเชิงนามธรรม การทำความเข้าใจความคิดซับซ้อน การเรียนรู้ อย่างฉับไว และการเรียนรู้จากประสบการณ์ (Gottfredson, 1997) โดยตัวแปรทางด้านการรู้คิด เป็นตัวแปรหนึ่งที่ใช้ประเมินความแตกต่างระหว่างบุคคลที่เกี่ยวข้องกับความสามารถทางสมอง ทั้งคัดเลือกพนักงานเข้าทำงาน (Ones, Dilchert, & Viswesvaran, 2012) การเข้าเรียน เลื่อนระดับชั้น การศึกษาต่อ นอกจากนี้ ยังมีประโยชน์อื่น ๆ เช่น ช่วยพัฒนาความจำ ช่วยลดการลืม ช่วยเพิ่มความสามารถในการถ่ายโอนข้อมูล ช่วยลดความสับสนในการเรียนรู้ รวมทั้งยังสามารถบอกได้ว่าเทคนิคการเรียนรู้แบบใดที่ได้ผลหรือไม่ได้ผล (Benjamin & Pashler, 2015) แม้แต่คัดเลือกเด็กที่มีความสามารถพิเศษ (Lohman, 2005)

สำหรับการวัดทางด้านการรู้คิด มีการสร้างแบบทดสอบและวิธีการวัดอย่างมากมาย (Hale, 1987) ส่วนใหญ่ เน้นการพัฒนาแบบทดสอบสติปัญญา (Intelligence test) แบบทดสอบวัดความถนัด (Aptitude test) และแบบทดสอบผลสัมฤทธิ์ (Achievement test) เป็นต้น ซึ่งมีรายละเอียดแต่ละแบบทดสอบดังนี้

แบบทดสอบสติปัญญา (Intelligence Test)

Gottfredson and Saklofske (2009) ได้กล่าวถึง การวัดสติปัญญา เป็นการวัดหนึ่งที่มีการถกเถียงมากที่สุด ในด้านจิตวิทยา และสาขาการศึกษา นักจิตวิทยาบางส่วนไม่ได้สนใจการสร้างแบบทดสอบสติปัญญา มีกลุ่ม พฤติกรรมนิยม กลุ่มการสร้างความรู้ด้านสังคม กลุ่มการตีความ และกลุ่มอื่น ๆ ทั้งภายในและภายนอก สาขาจิตวิทยาเห็นความสำคัญเพียงเล็กน้อยต่อสติปัญญา ว่าเป็นพื้นฐานที่ช่วยอธิบายพฤติกรรมมนุษย์ การใช้แบบทดสอบสติปัญญานำมาใช้ในการแบ่งกลุ่มและจัดที่นั่งให้นักเรียนในชั้นเรียนพิเศษ ซึ่งเป็นยุคสมัยแรก ๆ จนถึงยุคสมัยปัจจุบัน ได้มีการเปลี่ยนแปลงเกี่ยวกับแนวคิดทางด้านสติปัญญา ว่าควรมองอย่างเป็นบูรณาการมากขึ้น เพื่ออธิบายความเป็นบุคคลในภาพรวม (whole person) ในการนำมาใช้ในทางวิชาการ การคัดเลือกงาน หรือการทดสอบศักยภาพในการฟื้นฟูการบาดเจ็บทางสมอง เป็นต้น โดยในปัจจุบัน แนวคิดจิตวิทยาเชิงบวก

(Positive psychology) ได้ให้ความสำคัญเกี่ยวกับสติปัญญา ว่าเป็นปัจจัยหยุนตัวที่สำคัญ และเป็นปัจจัยในการสร้างขีดความสามารถ (Capacity) ในสภาวะทางจิตใจ (Psychological well-being) โดยแบบทดสอบสติปัญญาที่ใช้กันบ่อย เช่น WISC-IV เป็นต้น

ในประเทศไทย ได้มีการสร้างและพัฒนาแบบทดสอบสติปัญญา โดยนำแนวคิดเกี่ยวกับสติปัญญา มาจากต่างประเทศ เช่น แบบทดสอบ Stanford-Binet แบบทดสอบ Wechsler หรือแม้แต่แบบทดสอบจากแนวคิดพหุสติปัญญา (Multiple Intelligence) ของ Howard Gardner เป็นต้น (ได้แก่ ปาริชาติ บัวเจริญ, นพรัตน์ เตชะพันธ์รัตนกุล และรัตนากาล คำสอน, 2560; Paimpuech, 2013) นอกจากนี้ แบบทดสอบสติปัญญาที่ใช้ในไทยในบางแบบทดสอบ ไม่ได้มีการสร้าง แต่เป็นการนำแบบทดสอบมาตรฐานของต่างประเทศมาใช้ทั้งฉบับ เนื่องจากเป็นแบบทดสอบที่ไม่ได้ใช้ภาษาและไม่มียุทธศาสตร์วัฒนธรรมเข้ามาเกี่ยวข้อง ได้แก่ แบบทดสอบ Standard Progressive Matrices ซึ่งแบบทดสอบมีการใช้กันหลากหลาย โดยเฉพาะหน่วยงานด้านกรมสุขภาพจิต รวมทั้งสถาบันสุขภาพจิตเด็กและวัยรุ่นราชนครินทร์ใช้ในการทดสอบด้านสติปัญญา (นพวรรณ ศรีวงศ์พานิช, จันท์ธรรมา สุขทรัพย์, วนิดา ชนินทุยทรวงศ์ และอัมพร เบญจพลพิทักษ์, 2562; อภิชัย มงคล, ทวีศิลป์ วิษณุโยธิน, ปราณี ชาญณรงค์, โขจิตา ภาวสุทธิไพศิฐ และพรทิพย์ วชิรดิถ, 2555) และสำหรับแบบทดสอบต่างประเทศที่นำมาใช้ในประเทศไทย ได้มีการตรวจสอบคุณภาพเครื่องมือวัด รวมทั้งหาเกณฑ์ปกติสำหรับคนไทย แบบทดสอบสติปัญญาในปัจจุบันมีทั้งแบบทดสอบสติปัญญาเป็นรายบุคคลและเป็นรายกลุ่ม (อรพินทร์ ชูชม, 2545)

1. แบบทดสอบสติปัญญารายบุคคล (Individual Test of Intelligence)

แบบทดสอบสติปัญญารายบุคคล ออกแบบมาเพื่อทำการทดสอบทีละคน โดยมุ่งวัดภาพรวมของสติปัญญาหรือคุณลักษณะทั่วไปของแต่ละบุคคล ใช้เวลาในการสอบเป็นเวลานาน ผู้ดำเนินการสอบสามารถสังเกตและเอาใจใส่ผู้สอบได้อย่างใกล้ชิด (อรพินทร์ ชูชม, 2545) รวมทั้งแบบทดสอบสติปัญญาที่มาจากพื้นฐานแนวคิดที่แตกต่างกัน จะทำให้ได้คะแนนสติปัญญาที่แตกต่างกัน ดังนั้น การเลือกแบบทดสอบสติปัญญาชนิดใด ต้องทราบเกี่ยวกับจุดแข็งและจุดอ่อนในการเลือกเครื่องมือวัดให้เหมาะสมกับสิ่งที่ต้องการวัด (Gregory, 1996) ได้แก่

1.1 Stanford-Binet Test

แบบทดสอบนี้ได้ทำการสร้างและพัฒนา โดยในปัจจุบัน เป็นแบบทดสอบฉบับที่ 5 (Roid, 2003) เป็นแบบทดสอบที่ได้รับการออกแบบเพื่อประเมินสติปัญญาของบุคคลทั่วไปตั้งแต่ 2-85 ปี ได้เป็นคะแนนชวามปัญญา (IQ) ได้มาจากองค์ประกอบย่อย 10 องค์ประกอบ ในแต่ละองค์ประกอบย่อย จะมีคะแนนเฉลี่ยเท่ากับ 10 และค่าเบี่ยงเบนมาตรฐาน เท่ากับ 3 โดยผลรวมของแบบทดสอบ ประกอบด้วย คะแนน IQ โดยรวม คะแนน IQ ด้านวจนภาษา (Verbal) และคะแนน IQ ด้านอวจนภาษา (Non-verbal) ซึ่งคะแนนรวมทั้งหมด มีค่าเฉลี่ยเท่ากับ 100 และค่าเบี่ยงเบนมาตรฐานเท่ากับ 15 (Cohen, Swerdlik, & Sturman, 2013)

แบบทดสอบนี้ ประกอบด้วย คะแนนดัชนี 5 องค์ประกอบ ในแต่ละองค์ประกอบ จะมีการทดสอบย่อยทั้งด้านวจนภาษา และด้านอวจนภาษา รวมเป็นแบบทดสอบย่อย 10 ด้าน ดังตารางที่ 4-1

ตารางที่ 4-1 องค์ประกอบของแบบทดสอบ Stanford-Binet ฉบับที่ 5

องค์ประกอบ	นิยาม	ตัวอย่างด้านย่อย
การให้เหตุผลเชิงลื่นไหล (Fluid Reasoning: FR)	การแก้ปัญหาในรูปแบบใหม่ การทำ ความเข้าใจความสัมพันธ์ต่าง ๆ	- เรียงลำดับวัตถุ (Non-Verbal) - เปรียบเทียบด้วยภาษา (Verbal)
ความรู้ (Knowledge: KN)	ทักษะและความรู้ที่ได้มาจากระบบ การศึกษา	- รูปภาพที่ไม่เป็นเหตุเป็นผล (Non-Verbal) - คำศัพท์ (Verbal)
การให้เหตุผลเชิงปริมาณ (Quantitative Reasoning)	ความรู้เกี่ยวกับการคิดเชิงคณิตศาสตร์	- การให้เหตุผลเชิงปริมาณที่ใช้คำพูด (Verbal) - การให้เหตุผลเชิงปริมาณที่ไม่ใช้คำพูด (Non-Verbal)
การจดจำรูปภาพมิติสัมพันธ์ (Visual-spatial processing)	ความสามารถในการเห็นรูปแบบและ ความสัมพันธ์และมิติสัมพันธ์	- ตำแหน่งและทิศทาง (Verbal) - บอร์ดรูปร่าง (Non-Verbal)
ความจำเพื่อใช้งาน (Working Memory)	กระบวนการรู้คิดของการเก็บจำ แปรเปลี่ยนข้อมูลในรูปความจำ	- ความจำรูปประโยค (Verbal) - การตอบสนองล่าช้า (Non-Verbal)

ความเป็นมาตรฐาน (Standardization) หลังจากการพัฒนาข้อคำถามและการวิเคราะห์รายข้อ มา 5 ปี ได้ศึกษาบนพื้นฐานของอคติทางเพศ เชื้อชาติ/ชาติพันธุ์ วัฒนธรรม และศาสนา จึงทำให้ได้แบบทดสอบที่มีความเป็นมาตรฐาน มีผู้ทดสอบจำนวน 500 คน จากทั้ง 50 รัฐ ได้รับการฝึกฝนเพื่อจัดการการทดสอบ ผู้รับการทดสอบ จำนวน 4,800 คน มีอายุ 2-85 ปี กลุ่มตัวอย่างเป็นตัวแทนระดับชาติ จากหน่วยงานสำมะโนประชากรสหรัฐอเมริกา โดยแบ่งระดับชั้นด้วยอายุ เชื้อชาติ/ชาติพันธุ์ พื้นที่ทางภูมิศาสตร์ และระดับทางเศรษฐกิจและสังคม ไม่มีการอำนวยความสะดวกให้กับบุคคลที่มีความต้องการพิเศษในกลุ่มตัวอย่างที่ทดสอบมาตรฐาน แต่จะมีการอำนวยความสะดวกให้กับบุคคลที่มีความต้องการพิเศษในการศึกษาอื่น ๆ บุคคลที่ถูกคัดออกจากการศึกษาจะมีข้อจำกัดเกี่ยวกับความคล่องแคล่วในการใช้ภาษาอังกฤษ มีปัญหาทางการแพทย์รุนแรง มีความบกพร่องทางการรับรู้ความรู้สึกหรือการสื่อสารรุนแรง หรือมีปัญหาทางด้านอารมณ์และพฤติกรรมรุนแรง (Cohen et al., 2013)

คุณสมบัติทางการวัด (Psychometric Soundness) มีรายละเอียดที่สำคัญ ดังนี้ (Cohen et al., 2013)

1) ความเชื่อมั่น (Reliability) จากการวิเคราะห์ความเชื่อมั่นแบบสอดคล้องภายใน เมื่อจำแนกตามกลุ่มอายุ แบบทดสอบ IQ ฉบับเต็มทั้ง 5 องค์ประกอบ มีค่าความเชื่อมั่นระดับสูง ตั้งแต่ .97-.98 รวมทั้งมีค่าความเชื่อมั่นแบบวัดซ้ำมีค่าสูงเช่นกัน และมีค่าความเชื่อมั่นระหว่างผู้ให้คะแนน มีค่าตั้งแต่ .74-.97

2) ความเที่ยงตรง (Validity) จากการวิเคราะห์ความเที่ยงตรง มีการตรวจสอบความเที่ยงตรงเชิงเนื้อหา ความเที่ยงตรงเชิงเกณฑ์ ทั้งความเที่ยงตรงเชิงสภาพและความเที่ยงตรงเชิงพยากรณ์ ในส่วนความเที่ยงตรงเชิงสภาพ (Concurrent Validity) พบว่า แบบทดสอบฉบับที่ 5 มีความสัมพันธ์กับฉบับที่ 4 รวมทั้งมีความสัมพันธ์กับแบบทดสอบ Wechsler สำหรับความเที่ยงตรงเชิงพยากรณ์ (Predictive Validity) พบว่า มีความสัมพันธ์

กับแบบทดสอบผลสัมฤทธิ์ นอกจากนี้ ยังมีความเที่ยงตรงเชิงโครงสร้าง พบว่า ส่วนใหญ่มี 5 องค์ประกอบ แต่กลุ่มตัวอย่างบางกลุ่ม มี 4 องค์ประกอบ

1.2 The Wechsler Intelligence Test

แบบทดสอบนี้ออกแบบเพื่อประเมินความสามารถทางสติปัญญาของบุคคลตั้งแต่อ่อนวัยเรียนไปจนถึงวัยผู้ใหญ่ โดยแบบทดสอบนี้ได้สร้างขึ้นสำหรับกลุ่มก่อนวัยเรียน (ได้แก่ Wechsler Preschool and Primary Scale of Intelligence, Fourth Edition: WPPSI-IV) กลุ่มเด็ก (Wechsler Intelligence Scale for Children, Fourth Edition: WISC-IV) และกลุ่มผู้ใหญ่ (Wechsler Adult Intelligence Scale, Fourth Edition: WAIS-IV) ซึ่งประกอบด้วยคะแนนเต็ม IQ, คะแนน IQ วจนภาษา (Verbal) และคะแนน IQ การปฏิบัติ (Performance) มีค่าเฉลี่ยเท่ากับ 100 และค่าเบี่ยงเบนมาตรฐานเท่ากับ 15 (Cohen et al., 2013)

แบบทดสอบนี้มีคู่มือที่เขียนไว้อย่างชัดเจนที่ให้คำอธิบายในแต่ละด้านย่อย มีการจัดการทดสอบที่มีทิศทาง และมีจำนวนเวลาที่กำหนดสำหรับคำถาม ข้อคิดเห็น หรือความเป็นไปได้อื่น ๆ มีแนวทางในการเริ่มทำการหยุดทำแบบทดสอบ และมีการให้คะแนนที่ชัดเจน สำหรับการแปลความหมาย ใช้ชาร์ตด้านสถิติที่เป็นประโยชน์สำหรับผู้ประเมินในการให้ข้อเสนอแนะต่อพื้นฐานของการวัดประเมิน โดยลักษณะของแบบทดสอบสติปัญญา Wechsler (Cohen et al., 2013) มีรายละเอียดดังตาราง 4-2

ตารางที่ 4-2 รายชื่อการทดสอบและคำอธิบายในแบบทดสอบ Wechsler

การทดสอบ	คำอธิบาย
ความรู้ข้อเท็จจริง (Information)	บราซิลอยู่ในทวีปใด? คำถามต้องการทราบความรู้ทั่วไป การเรียนรู้ และความจำ โดยความสนใจ การศึกษา วัฒนธรรม และทักษะการอ่านมีผลต่อคะแนนที่ได้รับ
ความเข้าใจ (Comprehension)	โดยทั่วไป คำถามต้องการทราบความเข้าใจทางสังคม ความสามารถในการจัดระบบและประยุกต์ความรู้ และสิ่งที่เรียกว่า “สามัญสำนึก” เช่น “ทำไมเด็กจึงต้องระมัดระวังในการคุยกับคนแปลกหน้า”
ความคล้ายคลึง (Similarities)	ปากกาและดินสอคล้ายกันอย่างไร เป็นคำถามทั่วไปที่ปรากฏในด้านนี้ มีการจับคู่คำ เพื่อประเมินความสามารถในการวิเคราะห์ความสัมพันธ์และการมีส่วนในการคิดเชิงตรรกะและนามธรรม
คำศัพท์ (Vocabulary)	ให้นิยามคำ เป็นการวัดสติปัญญาทั่วไป โอกาสทางการศึกษาและวัฒนธรรมช่วยให้ทำแบบทดสอบด้านนี้ได้
การตั้งชื่อภาพ (Picture Naming)	ให้ตั้งชื่อภาพที่ปรากฏในหนังสือ เป็นการแสดงความสามารถทางภาษาและความสามารถในการสืบค้นคำ
การคิดคำนวณ (Arithmetic)	ปัญหาการคิดคำนวณจะนำเสนอและแก้ปัญหามากภาษา ในระดับต้น ให้นับเลข เป็นการทดสอบการเรียนรู้การคิดคำนวณ การตื่นตัวและการมีสมาธิ และความจำระยะสั้น
การเติมภาพให้สมบูรณ์ (Picture Completion)	ให้ระบุสิ่งที่หายไปจากภาพ เช่น รูปภาพแก้วน้ำ ที่ขาแก้วน้ำหายไป 1 ขา เป็นการวัดความสามารถในการรับรู้ภาพ การตื่นตัว ความจำ สมาธิ ความสนใจในรายละเอียด และความสามารถในการแยกสิ่งสำคัญออกมาจากสิ่งที่ไม่สำคัญ

ตารางที่ 4.2 (ต่อ)

การทดสอบ	คำอธิบาย
การจัดเตรียมภาพ (Picture Arrangement)	ให้ชุดของภาพที่ไม่เรียงลำดับกัน ให้นำภาพมาเรียงลำดับภาพให้เป็นเรื่องราวที่ถูกต้อง เป็นการวัดความสามารถในการเข้าใจสถานการณ์ ความตั้งใจ สมาธิ และความสามารถในการเชื่อมโยงเหตุและผล
การออกแบบบล็อก (Block Design)	ให้ออกแบบบล็อกให้เหมือนกับภาพที่กำหนดไว้ เป็นการวัดทักษะการรับรู้และการเคลื่อนไหว ความเร็วในการเคลื่อนไหว และความสามารถในการวิเคราะห์และสังเคราะห์
การประกอบวัตถุ (Object Assembly)	ให้ประกอบชิ้นส่วนต่าง ๆ เข้าด้วยกันอย่างรวดเร็ว เป็นการวัดความสามารถในการจำรูปแบบ ทักษะการสร้างประกอบร่าง และความเร็วในการเคลื่อนไหว

ความเป็นมาตรฐาน (Standardization) กลุ่มตัวอย่างที่เป็นมาตรฐาน ของแบบทดสอบ WAIS-IV ประกอบด้วย ผู้ใหญ่ อายุ 16-90 ปี จำนวน 2,200 คน แบ่งชั้นบนข้อมูลสำมะโนประชากรสหรัฐอเมริกา จำแนกตามอายุ เพศ เชื้อชาติ/ชาติพันธุ์ ระดับการศึกษา และพื้นที่ทางภูมิศาสตร์ เมื่อเทียบกับข้อมูลสำมะโนแล้ว พบว่ามีเพศหญิงมากกว่าเพศชายในกลุ่มสูงอายุ เมื่อเปรียบเทียบกับกลุ่มตัวอย่างที่เป็นมาตรฐานของแบบทดสอบ WAIS-III พบว่า กลุ่มตัวอย่างมีอายุและมีความหลากหลายมากกว่า และได้ปรับปรุงมาตรฐานการดำรงชีวิต (Cohen et al., 2013)

คุณสมบัติทางการวัด (Psychometric Soundness) (Cohen et al., 2013) โดยคู่มือของแบบทดสอบ มีการนำเสนองานวิจัยที่ทำการทดสอบค่าความเชื่อมั่น ความเที่ยงตรง และคุณสมบัติทางการวัดของแบบทดสอบ มีรายละเอียดที่สำคัญ ดังนี้

1) ความเชื่อมั่น (Reliability) พบว่า มีค่าความเชื่อมั่นแบบสอดคล้องภายในระดับสูงในทุกด้าน ดังนั้น ค่าความเชื่อมั่นจึงมีความเหมาะสม

2) ความเที่ยงตรง (Validity) พบว่า มีการศึกษาที่ทำการตรวจสอบความเที่ยงตรงเชิงสภาพ ความเที่ยงตรงเชิงเสมือน และความเที่ยงตรงเชิงจำแนก นอกจากนี้ มีการศึกษาวิจัยเชิงคุณภาพ เพื่อตรวจสอบว่า ผู้ทดสอบใช้การแก้ปัญหาในการตอบคำถาม เพื่อสนับสนุนถึงกระบวนการที่เป็นเป้าหมายของการวัดด้วย

1.3 Kaufman Brief Intelligence Test (K-BIT)

สำหรับแบบทดสอบนี้ Alan Kaufman ได้ออกแบบเพื่อตอบสนองความต้องการการทดสอบสติปัญญาที่สั้นกระชับ บริหารจัดการทดสอบได้ง่าย และใช้คัดกรองทางสติปัญญา โดยแบบทดสอบนี้เป็นเครื่องมือที่ช่วยระบุจะแข็งและจุดบกพร่องในทักษะกระบวนการทางสมอง มีพื้นฐานมาจากงานวิจัยสมองซีกขวาและสมองซีกซ้าย และการทำหน้าที่ของสมองที่ต่อเนื่องกันและทำหลาย ๆ อย่างพร้อมกัน

แบบทดสอบ ประกอบด้วย คำศัพท์ (Vocabulary) และเมทริกซ์ (Matrices) ใช้กับกลุ่มตัวอย่างอายุ 4-90 ปี และใช้เวลาในการทดสอบ 15-30 นาที มีค่าเฉลี่ยเท่ากับ 100 และมีค่าเบี่ยงเบนมาตรฐานเท่ากับ 15 เป็นการนำคะแนนคำศัพท์และเมทริกซ์รวมเข้าด้วยกันเป็นคะแนน IQ

เมื่อเปรียบเทียบการให้คะแนน กับแบบทดสอบอื่น ๆ ผู้สร้างมีความชัดเจนว่า แบบทดสอบนี้ไม่ได้มาแทนที่แบบทดสอบ Wechsler แต่เป้าหมายหลักของแบบทดสอบเพื่อเป็นแบบทดสอบคัดกรอง ในการส่งสัญญาณต่อความต้องการใช้แบบทดสอบอื่น ๆ ที่เข้มข้นขึ้น โดยความสั้นกระชับของเครื่องมือสามารถเป็นตัวเลือกที่นำมาใช้ในงานวิจัยทางด้านสติปัญญา

คุณสมบัติการวัด (Psychometric Soundness) มีรายละเอียดที่สำคัญ ดังนี้ (Gregory, 1996)

1) ความเชื่อมั่น (Reliability) พบว่า มีค่าความเชื่อมั่นแบบแบ่งครึ่ง และแบบวัดซ้ำ ในกลุ่มตัวอย่างที่หลากหลาย อยู่ระหว่าง .80-.90

2) ความเที่ยงตรง (Validity) พบว่า ในคู่มือได้รายงานการศึกษาต่าง ๆ มีค่าความเที่ยงตรงเชิงสภาพ นั่นคือ แบบทดสอบมีความสัมพันธ์กับแบบทดสอบ WAIS-R ในระดับสูง โดยส่วนคำศัพท์ มีความสัมพันธ์กับคะแนน IQ ด้าน Verbal (WAIS-R) เท่ากับ .83 ส่วนเมทริกซ์ มีความสัมพันธ์กับคะแนน IQ ด้านการปฏิบัติ (WAIS-R) เท่ากับ .77 ส่วนคะแนน IQ โดยรวมของทั้งสองแบบทดสอบ มีความสัมพันธ์กัน เท่ากับ .88 ดังนั้น แบบทดสอบ K-BIT จึงเป็นแบบทดสอบคัดกรองทางด้านสติปัญญาที่มักมีการใช้ในการวิจัย หรือมีข้อจำกัดด้านเวลา จึงควรใช้แบบทดสอบนี้

2. แบบทดสอบสติปัญญารายกลุ่ม (Group Test of Intelligence) ได้แก่

แบบทดสอบสติปัญญารายกลุ่มออกแบบเพื่อดำเนินการทดสอบกับผู้ทดสอบเป็นจำนวนมากพร้อมกัน ทำให้ดำเนินการทดสอบได้อย่างรวดเร็วและมีประสิทธิภาพ โดยมีคุณลักษณะที่สำคัญ (Gregory, 1996) ดังนี้ (1) มักใช้กับข้อสอบที่มีตัวเลือกหลายตัว (2) ตรวจให้คะแนนด้วยคอมพิวเตอร์ ทำให้การตรวจให้คะแนนเป็นไปอย่างรวดเร็วและมีความเป็นปรนัยสูง ขจัดความคลาดเคลื่อนจากการให้ดุลยพินิจของผู้ตรวจให้คะแนน (3) ผู้ดำเนินการสอนมีบทบาทน้อย เฉพาะอ่านคำชี้แจงและควบคุมเวลา ผู้เข้าสอบได้รับข้อสอบที่เรียงลำดับจากง่ายไปหายากเหมือนกันทั้งหมด (4) มักใช้ในการคัดกรอง (Screening) เพื่อการตัดสินใจของหน่วยงาน ได้แก่ การส่งเสริมและแก้ไขทางการเรียน การคัดเลือกเข้าทำงานที่เหมาะสมกับตำแหน่งงาน และ (5) ใช้กลุ่มตัวอย่างที่เป็นมาตรฐานขนาดใหญ่

Raven Progressive Matrices (RPM)

แบบทดสอบสติปัญญารายกลุ่มที่ใช้กันมากในงานวิจัยและใช้เพื่อคัดกรองสติปัญญาของหน่วยงานต่าง ๆ คือ แบบทดสอบ Raven Progressive Matrices เป็นแบบทดสอบที่ไม่ใช้ภาษา (Nonverbal) เกี่ยวกับการให้เหตุผลเชิงอุปนัย (Inductive Reasoning) โดยมีพื้นฐานจากการอาศัยสิ่งเร้าจากรูปภาพ

RPM ได้ออกมาเพื่อวัตถุประสงค์ประกอบ g ของ Spearman จึงได้เลือกรูปแบบการทดสอบเพื่อฝึกฝนองค์ประกอบ g โดยสนใจกระบวนการความสัมพันธ์ของสิ่งที่ขาดหายไป ด้วยพื้นฐานการรับรู้สิ่งที่คล้ายคลึงกัน โดยให้เลือกรูปแบบที่สัมพันธ์กับรูปภาพข้างต้น โดยแต่ละข้อจะเรียงลำดับจากง่ายไปยาก โดยแบบทดสอบ RPM แบ่งเป็น 3 ระดับ ได้แก่

(1) The Coloured Progressive Matrices มีทั้งหมด 36 ข้อ ออกแบบมาสำหรับเด็กที่มีอายุ 5-11 ปี มีการใช้สีในฉบับนี้เพื่อดึงดูดความสนใจของเด็ก

(2) The Standard Progressive Matrices มีทั้งหมด 60 ข้อ จัดกลุ่มเป็น 5 ชุด ชุดละ 12 ข้อ ใช้กับกลุ่มตัวอย่าง 6 ปีขึ้นไป ข้อส่วนใหญ่มีความยาก แบบทดสอบนี้จึงเหมาะสมกับกลุ่มผู้ใหญ่

(3) The Advanced Progressive Matrices มีทั้งหมด 48 ข้อ จัดกลุ่มเป็น 2 ชุด ชุดแรกมี 12 ข้อ และชุดที่สองมี 36 ข้อ แบบทดสอบนี้มักใช้กับบุคคลที่มีสติปัญญาเลิศ

คุณสมบัติการวัด (Psychometric Soundness) มีรายละเอียดที่สำคัญ ดังนี้ (Gregory, 1996)

1) ความเชื่อมั่น (Reliability) พบว่า มีค่าความเชื่อมั่นแบบแบ่งครึ่ง อยู่ระหว่าง .65-.94 ในกลุ่มเด็ก และมีค่าความเชื่อมั่นแบบวัดซ้ำอยู่ระหว่าง .80-.93 ในกลุ่มผู้ใหญ่ และมีค่าเท่ากับ .71 ในกลุ่มเด็ก ดังนั้น ในกลุ่มเด็ก RPM มีค่าความเชื่อมั่นที่เพียงพอสำหรับการนำไปใช้เพื่อการตัดสินใจต่าง ๆ

2) ความเที่ยงตรง (Validity) พบว่า มีการศึกษา Coloured Progressive Matrices พบว่า มี 3 องค์ประกอบ ข้อยากเป็นการให้เหตุผลเชิงนามธรรม ส่วนข้อง่ายเป็นการเติมเต็มแบบแผนที่ยาก โดยข้อที่ง่ายมากและยากมาก เป็นการกระตุ้นกระบวนการทางสติปัญญาที่แตกต่างกัน นอกจากนี้มีความเที่ยงตรงเชิงพยากรณ์ โดยมีความสัมพันธ์กับแบบทดสอบผลสัมฤทธิ์ อยู่ระหว่าง .30-.60 และมีความสัมพันธ์กับแบบทดสอบสติปัญญาอื่น ๆ ด้วย แบบทดสอบสติปัญญาส่วนใหญ่ เป็นแบบทดสอบที่มีความเป็นมาตรฐาน มีเกณฑ์ปกติ และมีคุณสมบัติทางการวัดที่ดี มีค่าความเชื่อมั่นสูง และผ่านการตรวจสอบความเที่ยงตรงหลากหลายประเภท ทั้งความเที่ยงตรงเชิงโครงสร้าง ความเที่ยงตรงเชิงเกณฑ์ ความเที่ยงตรงเชิงพยากรณ์ เป็นต้น สำหรับผู้วิจัยที่ศึกษาตัวแปรทางด้านสติปัญญาในการวิจัยทางด้านพฤติกรรมศาสตร์ แบบทดสอบแต่ละประเภทมีหลักเกณฑ์การใช้และการตีความที่ไม่เหมือนกัน เนื่องจากอยู่บนแนวคิดที่ไม่เหมือนกัน ดังนั้น ผู้เขียนแนะนำให้ศึกษาแบบทดสอบดังกล่าวให้ละเอียดถี่ถ้วนก่อนการนำไปใช้

แบบทดสอบความถนัด (Aptitude Test)

การวัดความถนัด เป็นการวัดศักยภาพในการเรียนรู้หรือการทำงาน ใช้เพื่อทำนายความสำเร็จทางการเรียน (เช่น ภาษาต่างประเทศ ดนตรี) หรือโปรแกรม (เช่น การศึกษาต่อในระดับบัณฑิตศึกษา โปรแกรมคอมพิวเตอร์) ดังนั้น การวัดความถนัดต้องเป็นการวัดทางด้านการรู้คิดหรือทักษะกลไก คือ มุ่งเน้นความรู้และความสามารถ นอกจากนี้ การวัดความถนัดยังมุ่งวัดประสบการณ์ที่หลากหลายของบุคคล มุ่งเน้นความสามารถที่เลื่อนไหล (Fluid abilities) ซึ่งเป็นการเรียนรู้ที่สั่งสมมายาวนาน (อรพินทร์ ชูชม, 2545; Sax, 1997)

แบบทดสอบความถนัด แบ่งเป็น 2 ประเภท (อรพินทร์ ชูชม, 2545) ได้แก่

1. แบบทดสอบความถนัดทางการเรียน (Scholastic Aptitude Test) เป็นแบบทดสอบที่มุ่งทำนายความสามารถของผู้เรียนว่าผู้เรียนเรียนรู้สิ่งต่าง ๆ หรือทักษะต่าง ๆ ได้มากน้อยเพียงใด อาศัยความสามารถพื้นฐานของแต่ละบุคคล

2. แบบทดสอบความถนัดเฉพาะทาง (Specific Aptitude Test) เป็นแบบทดสอบที่มุ่งเน้นศักยภาพของแต่ละบุคคลว่าบุคคลหนึ่ง ๆ มีความสามารถพิเศษมากน้อยเพียงใด เช่น การวัดความถนัดทางด้านเครื่องยนต์กลไก กีฬา ภาษา การคำนวณ เป็นต้น

ความเป็นมาตรฐาน (Standardization) แบบทดสอบความถนัดส่วนใหญ่ จะมีค่าเปอร์เซ็นต์ไทล์และเกณฑ์ปกติ บางแบบทดสอบได้จำแนกตามกลุ่มเพศ และบางแบบทดสอบมีเกณฑ์ปกติสำหรับผู้สมัครงานและบุคคลที่ต้องการทำงานในตำแหน่งงานนั้น ๆ

คุณสมบัติการวัด (Psychometric Soundness) มีรายละเอียดที่สำคัญ ดังนี้ (Sax, 1997)

1) ความเชื่อมั่น (Reliability) พบว่า โดยทั่วไป ค่าความเชื่อมั่นแบบทดสอบคล้อยภายในของแบบทดสอบความถนัดมีค่าค่อนข้างสูง อยู่ระหว่าง .79-.93

2) ความเที่ยงตรง (Validity) พบว่า แบบทดสอบความถนัดทุกแบบทดสอบมีความเที่ยงตรงเชิงพยากรณ์ นั่นคือ เนื้อหาการวัดสามารถทำนายเชิงเกณฑ์ได้ โดยผลการเรียนเฉลี่ยสะสม (GPA) มักถูกนำมาใช้ในด้านการศึกษา พบว่า ความถนัดทางการเรียนเป็นตัวทำนายที่ดีสำหรับคะแนนเฉลี่ยสะสม นอกจากนี้ ยังพบว่าแบบทดสอบความถนัดยังมีความสัมพันธ์กับแบบทดสอบผลสัมฤทธิ์ทางการเรียน

ยกตัวอย่างแบบทดสอบความถนัด ดังนี้

1. แบบทดสอบความถนัดทั่วไป (General Aptitude Test Battery: GATB)

แบบทดสอบนี้มีประโยชน์ใช้ในการทำนายผลการปฏิบัติงานได้ (Hunter, 1994) มาจากการวิเคราะห์องค์ประกอบ 59 แบบทดสอบ กับกลุ่มตัวอย่างหลายพันคน โดยได้ทำการปรับปรุงและมีการทำให้ทันสมัยมากยิ่งขึ้น ดังนั้น แบบทดสอบนี้จึงเป็นเครื่องมือวัดในสมัยใหม่ แม้จะมีการเปลี่ยนแปลงเนื้อหาเพียงเล็กน้อยก็ตาม

แบบทดสอบนี้ ใช้เวลาในการทำ ประมาณ 2 ชั่วโมง 30 นาที มี 9 องค์ประกอบที่สำคัญ ดังตารางที่ 4-3

ตารางที่ 4-3 องค์ประกอบและคำอธิบายของแบบวัดความถนัดทั่วไปในแต่ละองค์ประกอบ

องค์ประกอบ	คำอธิบาย
ความสามารถด้านการเรียนรู้ทั่วไป (General Learning Ability)	คะแนนรวมของแบบทดสอบคำศัพท์ การให้เหตุผลเชิงคำนวณ และพื้นที่สามมิติ
ความถนัดทางภาษา (Verbal Aptitude)	คะแนนจากแบบทดสอบคำศัพท์ โดยผู้ทดสอบจะต้องบอกได้ว่าคำสองคำที่ปรากฏเป็นคำเหมือนกันหรือต่างกัน

ตารางที่ 4-3 (ต่อ)

องค์ประกอบ	คำอธิบาย
ความถนัดทางตัวเลข (Number Aptitude)	คะแนนจากแบบทดสอบการคำนวณและการให้เหตุผลเชิงคำนวณ
ความถนัดเชิงมิติสัมพันธ์ (Spatial Aptitude)	คะแนนจากแบบทดสอบพื้นที่สามมิติ วัดความสามารถในการรับรู้สองมิติของวัตถุสามมิติ และมองเห็นภาพเคลื่อนไหวสามมิติ
การรับรู้รูปร่าง (Form Perception)	คะแนนจากการจับคู่รูปร่างและการจับคู่เครื่องมือ โดยทั้งสองแบบทดสอบ ผู้ทดสอบต้องจับคู่สิ่งที่เหมือนกัน
การรับรู้ทางเสมียน (Clerical Perception)	คะแนนจากแบบทดสอบพิสูจน์อักษร เรียกว่า การเปรียบเทียบชื่อ ผู้ทดสอบต้องจับคู่ชื่อภายใต้เวลาที่จำกัด
การประสานงานการเคลื่อนไหว (Motor Coordination)	คะแนนจากการวัดความสามารถในการลากจุดเชื่อมอย่างรวดเร็ว
ความคล่องแคล่วของนิ้วมือ (Finger Dexterity)	คะแนนจากแบบทดสอบความคล่องแคล่วในการใช้นิ้วมือต่อเชื่อมหมุดย้าและแหวนรอง
ทักษะการใช้มือ (Manual Dexterity)	คะแนนจากการวางและคืนกลับมา โดยให้ผู้ทดสอบเคลื่อนย้ายและถอดหมุดในบอร์ด

2. แบบทดสอบความถนัดเฉพาะทาง (Specific Aptitude Test)

แบบทดสอบความถนัดเฉพาะทางมักใช้ทำนายแนวโน้มความสำเร็จเฉพาะทาง ได้แก่ ศิลปะ ดนตรี ความคิดสร้างสรรค์ เสมียน และกลไก โดยแบบทดสอบความถนัดเฉพาะทางอาจมีความยาวกว่า และมีเกณฑ์ปกติที่ดีกว่า และอาจเสียค่าใช้จ่ายน้อยการทดสอบความถนัดทั่วไปทั้งฉบับ (Sax, 1997) โดยแบบทดสอบความถนัดเฉพาะที่มักนิยมใช้กัน ดังนี้

2.1 แบบทดสอบความถนัดทางเสมียน (Clerical Aptitude Test)

แบบวัดความถนัดทางเสมียนที่ใช้กันอย่างกว้างขวาง คือ The Minnesota Clerical Test ออกแบบมาสำหรับเด็กนักเรียนอายุ 8-12 ปี แต่มีเกณฑ์ปกติสำหรับผู้ใหญ่ในอาชีพที่ต้องอาศัยทักษะทางเสมียน (เช่น นักบัญชี ผู้ตรวจสอบ พนักงาน เสมียน) แบบทดสอบประกอบด้วย 2 ส่วน คือ การเปรียบเทียบตัวเลข (Number Comparison) และการเปรียบเทียบชื่อ (Name Comparison) ในแต่ละส่วน ให้เปรียบเทียบตัวเลขหรือชื่อประมาณ 200 คู่ และให้ชี้ว่า แต่ละคู่เหมือนกันหรือต่างกัน นอกจากนี้ แบบทดสอบความถนัดทางเสมียนแบบอื่น ๆ จะมีจุดเน้นที่แตกต่างกัน เช่น การสะกดคำ การเรียงลำดับพยัญชนะอักษร การจัดเรียงประโยคให้เรียบร้อย และการคำนวณอย่างง่าย รวมทั้งการพิมพ์ดีด เป็นต้น

2.2 แบบทดสอบความถนัดทางกลไก (Mechanical Aptitude Test)

แบบทดสอบความถนัดทางกลไก เป็นแบบทดสอบที่มุ่งเน้นแนวคิดภาพรวม เข้าใจว่าความถนัดทางกลไกเป็นคุณลักษณะหนึ่งที่มีความแตกต่างกันในตัวบุคคล ส่วนอีกแนวคิดหนึ่งเชื่อว่าความถนัดทางกลไกประกอบด้วยความสามารถที่มีความเป็นอิสระจำนวนหนึ่ง ซึ่งในปัจจุบันให้ความสำคัญกับแนวคิดหลัง เนื่องจากการวัดความถนัดทางกลไกจำเป็นต้องใช้หลายแบบทดสอบมากกว่าการใช้เพียงแบบทดสอบเดียว ได้แก่

- แบบทดสอบความคล่องแคล่วและการสร้าง (Dexterity and Assembly Test) ในส่วนการทดสอบการสร้าง เป็นการนำคำแนะนำจากจำนวนชิ้นส่วนที่สร้างหรือประกอบในช่วงเวลาหนึ่ง เป็นแบบทดสอบที่มีค่าใช้จ่ายสูงและเป็นการทดสอบรายบุคคล ดังนั้น ครูหรือผู้ให้คำปรึกษามักไม่ค่อยใช้แบบทดสอบส่วนนี้ ส่วนการทดสอบความคล่องแคล่ว เพื่อวัดการประสานงานของมือและตาอย่างรวดเร็ว

- แบบทดสอบข้อมูลข่าวสาร (Information Test) แบบทดสอบมักวัดข้อมูลหรือประสบการณ์เกี่ยวกับวัตถุกลไก โดยแบบทดสอบนี้มีประโยชน์ในการคัดเลือกคนที่เคยมีประสบการณ์ในการสัมผัสกับวัตถุและวิธีการทางกลไกมาก่อน

- แบบทดสอบการให้เหตุผลทางกลไก (Mechanical Reasoning Test) เป็นแบบทดสอบที่วัดการให้เหตุผลหรือความเข้าใจในกลไก แบบทดสอบที่ใช้กันอย่างแพร่หลาย คือ the Bennett Mechanical Comprehension Test (BMCT) ออกแบบเบื้องต้นเพื่อนักเรียนมัธยมศึกษาปีที่ 5-6 ผู้สมัครและพนักงานโรงงานอุตสาหกรรม แบบทดสอบประกอบด้วย รูปภาพของวัตถุที่คุ้นเคย โดยให้เลือกว่า อะไรที่ตัดโลหะได้ดีที่สุด และห้องที่มีปริมาณเฟอร์นิเจอร์ที่แตกต่างกัน ห้องใดมีเสียงสะท้อนกว่ากัน

- แบบทดสอบมิติสัมพันธ์ (Spatial Relations Test) เป็นแบบทดสอบที่วัดความถนัดเชิงเสมือนจริง แบบทดสอบสติปัญญาส่วนใหญ่ประกอบด้วยมิติสัมพันธ์ แบบทดสอบนี้เป็นตัวทำนายที่มีประโยชน์ต่อความสำเร็จในหลายอาชีพและการเรียนวิชาต่าง ๆ ในโรงเรียน เช่น ความสามารถทางด้านมิติสัมพันธ์มักเกี่ยวข้องกับวิศวกรรมศาสตร์ การวาดภาพร่าง ศิลปะ ทันตแพทยศาสตร์ และอาชีพอื่น ๆ เป็นต้น

2.3 แบบทดสอบความถนัดทางดนตรี (Musical Aptitude Test)

แบบทดสอบความถนัดทางดนตรีที่ใช้กันแพร่หลายในปัจจุบัน คือ the Seashore Measures of Musical Talents ประกอบด้วย บันทึกแผ่นเสียง และเครื่องบันทึกเทปแบบ Reel-to-reel ที่วัดการจำแนกทางการได้ยิน 6 ด้าน ได้แก่ ระดับเสียง (pitch) ความดัง-เบาของเสียง (loudness) อัตราจังหวะ (time) คุณภาพเสียง (timbre) ลีลาจังหวะ (rhythm) และการจดจำทำนอง (tonal memory) โดยออกแบบมาสำหรับเด็กนักเรียนจนถึงระดับมหาวิทยาลัย บุคคลที่มีความถนัดทางดนตรีสูง จะเป็นผู้ที่ประสบความสำเร็จทางด้านดนตรี

2.4 แบบทดสอบความถนัดทางศิลปะ (Artistic Aptitude Test)

แบบทดสอบความถนัดทางศิลปะ ประกอบด้วย การวัดการตัดสินคุณค่าทางความงาม (Aesthetic judgment) และผลผลิตทางความงาม (Aesthetic production) โดยส่วนแรกอาจมีความจำเป็นสำหรับการวิพากษ์วิจารณ์งานศิลปะ และส่วนหลังเป็นการฝึกฝนการเป็นศิลปิน

- แบบทดสอบการตัดสินคุณค่าทางความงามที่ชื่อเสียงเป็นที่รู้จัก คือ Meier Art Judgment Test ประกอบด้วย ภาพคู่ขาวดำ 100 คู่ โดยอันแรกถือเป็นผลงานศิลปะชิ้นเอก (Masterpiece) ส่วนอีกอันมีการปรับแต่งเล็กน้อย เป็นการทดสอบให้ค้นหาว่า สิ่งที่แตกต่างกันที่ไม่ใช่ภาพคลาสสิก คืออะไร

- แบบทดสอบผลผลิตทางความงาม คือ the Horn Art Aptitude Inventory ประกอบด้วย 2 ส่วน ส่วนแรกให้วาดภาพร่างที่แตกต่างกัน 20 ภาพ แต่ต้องเป็นวัตถุที่คุ้นเคย เช่น ต้นไม้ หนังสือ ส้อม เป็นต้น โดยแต่ละภาพต้องทำให้เสร็จภายใน 3-10 วินาที และส่วนที่สอง ประกอบด้วย สีเหลี่ยมมุมฉาก 12 อันสร้างให้เป็นภาพขึ้นมา

2.5 แบบทดสอบความคิดสร้างสรรค์ (Creativity Measurement Test)

แบบทดสอบความคิดสร้างสรรค์ เป็นความคิดแบบอเนกนัย (Divergent thinking) คือ ความคิดแปลกใหม่ และเป็นความคิดริเริ่ม โดยการวัดมักให้ผู้ทดสอบเขียนวัตถุประสงค์ต่าง ๆ ของการใช้วัสดุ หรือเขียนคำต่าง ๆ ให้มากที่สุดเท่าที่จะเป็นไปได้ ซึ่งคำนั้นจะต้องมีทั้งตัวอักษรและพยัญชนะ เป็นต้น

นอกจากแบบทดสอบความถนัดทั่วไปและแบบทดสอบความถนัดเฉพาะทางแล้ว การวัดความพร้อม (Readiness) ถือว่าเป็นแบบทดสอบความถนัดรูปแบบหนึ่ง ความพร้อมเป็นปัจจัยทางด้านกายภาพ ด้านบุคลิกภาพ และด้านอื่น ๆ ที่เป็นสิ่งจำเป็นสำหรับเด็กที่พร้อมที่จะเรียนรู้ โดยความพร้อมทางการเรียน มักวัดในระดับก่อนวัยเรียนและระดับประถมศึกษา (Preschool and Elementary School) แบบทดสอบความพร้อมช่วยให้เด็กเตรียมพร้อมเริ่มต้นในระบบการเรียนการสอนที่เป็นทางการในวิชาหนึ่ง ๆ หรือทักษะ เช่น การอ่าน (Cohen et al., 2013)

คะแนนจากการวัดความถนัดต่ำ ไม่ได้หมายความว่า เด็กขาดความสามารถในการเรียนรู้ แต่เป็นการบอกว่า เด็กต้องการความช่วยเหลือบางอย่างที่จำเป็นเพื่อเริ่มต้นในระบบการเรียนการสอน เด็กที่ไม่สามารถจำแนกตัวอักษรระหว่าง p และ q ได้ อาจมีคะแนนความพร้อมทางการอ่านต่ำ เป็นสิ่งที่บอกครูว่า ครูควรทำอะไรบางอย่าง เพื่อให้เด็กมีประสบการณ์ในการเรียนรู้วิธีการอ่าน และคะแนนต่ำไม่ได้หมายความว่า ครูควรทอดทิ้งการสอนเด็ก (Sax, 1997)

สำหรับการวัดความพร้อมของเด็กนักเรียนก่อนวัยเรียนและประถมศึกษา พบว่า มีเครื่องมือที่สำคัญคือ รายการตรวจสอบ (Checklist) เป็นแบบสอบถามที่สร้างขึ้น เพื่อทำการตรวจสอบว่าพฤติกรรม ความคิด หรือสถานการณ์ต่าง ๆ ตามรายการที่ปรากฏว่า มีหรือไม่มี โดยบุคคลที่ทำหน้าที่ตรวจสอบตามรายการตรวจสอบ อาจเป็นครู นักจิตวิทยา ผู้ปกครอง หรือผู้ดูแล โดยรายการตรวจสอบดังกล่าวต้องมีเนื้อหาครบถ้วน เป็นวิธีการที่

ประหยัดและสะดวกรวดเร็ว สำหรับรายการตรวจสอบเกี่ยวกับความพร้อมของเด็ก ควรครอบคลุมปัญหาทางอารมณ์ พฤติกรรม และอื่น ๆ ส่วนมาตราประมาณค่า (Rating Scale) อาจใช้ข้อความจากรายการตรวจสอบ แต่มีการให้ระดับคะแนนในเชิงระดับความถี่ เป็นต้น (Cohen et al., 2013)

แบบทดสอบความถนัดส่วนใหญ่ เป็นแบบทดสอบที่มีมาตรฐาน มีเกณฑ์ปกติ มีทั้งความถนัดโดยทั่วไป ความถนัดเฉพาะทาง รวมไปถึงความพร้อมในการเรียน สำหรับผู้วิจัยที่ศึกษาตัวแปรทางด้านความถนัดในการวิจัยทางด้านพฤติกรรมศาสตร์ ยกตัวอย่างเช่น หากต้องการทำนายความสำเร็จด้านดนตรี อาจต้องศึกษาความถนัดทางดนตรี เป็นต้น แบบทดสอบความถนัดแต่ละประเภทไม่เหมือนกัน ดังนั้น ผู้เขียนแนะนำให้ศึกษาแบบทดสอบดังกล่าวให้ละเอียดถี่ถ้วนก่อนการนำไปใช้

แบบทดสอบผลสัมฤทธิ์

ผลสัมฤทธิ์ (Achievement) เป็นความสามารถทางการรู้คิดของบุคคล ซึ่งเป็นผลมาจากประสบการณ์ในการเรียนรู้ที่บุคคลได้รับจากการเรียนการสอนหรือนวัตกรรมทางการศึกษาในช่วงระยะเวลาหนึ่ง ๆ โดยการวัดผลสัมฤทธิ์ สามารถกระทำได้โดยใช้แบบทดสอบผลสัมฤทธิ์ (Achievement Test) โดยแบบทดสอบผลสัมฤทธิ์ดังกล่าว เป็นการทดสอบที่ดูว่าบุคคลมีความสามารถทางการรู้คิดที่เกี่ยวกับประสบการณ์การเรียนรู้ในเรื่องหนึ่ง ๆ มากน้อยเพียงใด ในการเรียนในโรงเรียนและมหาวิทยาลัย เรียกแบบทดสอบผลสัมฤทธิ์ตามเนื้อหาวิชาที่ใช้ในการทดสอบเป็นหลัก เช่น แบบทดสอบผลสัมฤทธิ์วิชาวิทยาศาสตร์ แบบทดสอบผลสัมฤทธิ์วิชาจิตวิทยาการเรียนรู้ เป็นต้น ถ้าการทดสอบวัดการเรียนรู้ในเนื้อหาวิชาที่ผ่านมาแล้วทั้งหมด อย่างเช่นในการศึกษาระดับปริญญาโทและปริญญาเอก เรียกว่า แบบทดสอบประมวลความรู้ (Comprehensive Test) รวมทั้งแบบทดสอบผลสัมฤทธิ์มีชื่อเรียกหลากหลายรูปแบบตามการจำแนกประเภทของแบบทดสอบ (โชติกา ภาชีผล, 2559; ณีภูธรณ์ หลาวทอง, 2561; อรพินทร์ ชูชม, 2545)

นอกจากนี้ ยังมีข้อถกเถียงเกี่ยวกับแบบทดสอบผลสัมฤทธิ์และแบบทดสอบความถนัด (Sax, 1997) โดยทั้งสอง เป็นแบบทดสอบที่มุ่งวัดความสามารถทางการรู้คิดของบุคคล แต่แบบทดสอบผลสัมฤทธิ์มีความแตกต่างจากแบบทดสอบความถนัด ในแง่ของวัตถุประสงค์ (Purpose) และวิธีการตรวจสอบความเที่ยงตรง (Validation) โดยแบบทดสอบความถนัดใช้เพื่อทำนายความสำเร็จ วัดก่อนเริ่มต้น (prior) โครงการ เพื่อคัดเลือกจัดชั้นเรียน หรือฝึกอบรมต่าง ๆ ให้ความสำคัญกับความเที่ยงตรงเชิงพยากรณ์ เนื่องจากเป้าหมายของแบบทดสอบความถนัดคือ เพื่อทำนายระดับความสำเร็จของผู้เรียน เช่น แบบทดสอบความถนัดทางดนตรี ใช้เพื่อทำนายความสำเร็จทางดนตรีในอนาคต ส่วนแบบทดสอบผลสัมฤทธิ์วัดหลังจาก (after) ที่ผู้เรียนได้รับการจัดการเรียนการสอนไปแล้ว โดยมุ่งเน้นความสำคัญกับความเที่ยงตรงเชิงเนื้อหา เพราะแบบทดสอบที่สร้างขึ้นมาจากวัตถุประสงค์ที่ปรากฏอยู่ในหลักสูตรเป็นส่วนใหญ่ โดยเฉพาะอย่างยิ่งทักษะพื้นฐานต่าง ๆ จากประเด็นดังกล่าว

ข้างต้น จึงกล่าวได้ว่า แบบทดสอบที่ใช้วัดเพื่อทำนายความสำเร็จในอนาคต เรียกว่า แบบทดสอบความถนัด ส่วนแบบทดสอบที่วัดความรู้ที่ได้เรียนรู้มาแล้ว เป็นผลมาจากการเรียนการสอน เรียกว่า แบบทดสอบผลสัมฤทธิ์

ประเภทของแบบทดสอบผลสัมฤทธิ์

แบบทดสอบผลสัมฤทธิ์ สามารถแบ่งได้หลายประเภท ตามการจำแนกรูปแบบต่าง ๆ ดังนี้ (อรพินทร์ ชูชม, 2545)

1. จำแนกตามรูปแบบของการสอบ สามารถแบ่งได้เป็น 3 ประเภท คือ

1.1 แบบทดสอบปรนัย (Objective Test) เป็นแบบทดสอบผลสัมฤทธิ์ที่ใช้วัดความสามารถทางการรู้คิดในระดับเบื้องต้น เช่น การวัดความรู้ความเข้าใจในเรื่องหนึ่ง ๆ และสามารถสร้างได้ครอบคลุมตามเนื้อหาที่ต้องการจะวัด แบบทดสอบประเภทนี้กำหนดขอบเขตหรือมีคำตอบไว้ให้เลือกตอบชัดเจน โดยมีวิธีการให้เลือกตอบแบบใดแบบหนึ่ง เช่น แบบหลายตัวเลือก แบบถูกผิด แบบจับคู่ และแบบเติมคำ เป็นต้น แบบทดสอบปรนัย มีคุณสมบัติที่สำคัญ ได้แก่

- ความชัดเจนในคำถาม มีความเข้าใจตรงกันว่าคำถามแต่ละข้อถามอะไร
- การให้คะแนนตรงกัน ไม่ว่าใครตรวจให้คะแนน จะได้คะแนนตรงกัน
- มีการกำหนดเกณฑ์ในการให้คะแนนและแปลความหมายของคะแนนที่ชัดเจน

1.2 แบบทดสอบอัตนัย (Subjective Test) เป็นแบบทดสอบผลสัมฤทธิ์ที่เหมาะสมกับการใช้วัดความสามารถทางการรู้คิดในระดับสูง เช่น การวัดความสามารถในการคิดเชิงสังเคราะห์ การคิดเชิงประเมิน แบบทดสอบอัตนัยอาจไม่จำเป็นต้องใช้ข้อความจำนวนมาก โดยเป็นแบบทดสอบที่ให้ตอบคำถามที่กำหนดไว้ อย่างเป็นอิสระ สามารถแสดงความรู้และความคิดเห็นได้อย่างเต็มที่ภายในขอบเขตระยะเวลาที่กำหนด ไม่ได้มีคำตอบให้เลือกอย่างชัดเจนเหมือนแบบทดสอบปรนัย การตรวจให้คะแนนอาจได้รับอิทธิพลจากความเป็นอัตวิสัย (Subjective) หรือความไม่แน่นอนและอคติของผู้ให้คะแนนพอสมควร

1.3 แบบทดสอบการปฏิบัติ (Performance Test) เป็นแบบทดสอบผลสัมฤทธิ์ที่ใช้วัดความสามารถและการเรียนรู้ของบุคคล โดยมุ่งเน้นให้บุคคลแสดงการกระทำและลงมือปฏิบัติจริง ซึ่งสามารถวัดได้จากความสามารถในขั้นกระบวนการ (process) และผลผลิต (product) ของกระบวนการ หรืออาจทั้งสองอย่าง แบบทดสอบการปฏิบัติสามารถใช้การสังเกตหรือการประเมินค่าผลสัมฤทธิ์ที่ได้จากการปฏิบัติสิ่ง ๆ หนึ่ง ดังนั้นการออกแบบแบบทดสอบการปฏิบัติควรคำนึงถึงการกำหนดประเด็นที่จะทดสอบ การออกแบบบริบทที่วัดโดยการสร้างงาน หรือสถานการณ์จำลอง เพื่อให้บุคคลได้แสดงความรู้ ทักษะ รวมทั้งความสามารถที่ได้เรียนรู้มาแล้ว ภายใต้เงื่อนไขการทดสอบที่ระบุไว้ มักใช้ระบบการให้คะแนนที่ชัดเจน

2. จำแนกตามวัตถุประสงค์ของการสอบ สามารถแบ่งได้เป็น 4 ประเภท คือ

2.1 แบบทดสอบวัดระดับ (Placement Test) เป็นแบบทดสอบผลสัมฤทธิ์ที่ต้องการวัดทักษะที่จำเป็น โดยทั่วไปเป็นข้อสอบที่ไม่ซับซ้อนและเป็นแบบทดสอบอิงเกณฑ์ หรือถ้าใช้เพื่อคัดเลือคนจะเป็นข้อสอบ

ที่มีความยากหรือมีความซับซ้อนมากขึ้น และสำหรับข้อสอบที่ใช้จำแนกความสามารถของผู้เรียนว่าแต่ละคนอยู่ในกลุ่มเด็กเก่งหรือกลุ่มเด็กอ่อน มักเป็นแบบทดสอบอิงกลุ่ม โดยมีระดับความยากง่ายที่หลากหลาย

2.2 แบบทดสอบการวินิจฉัย (Diagnostic Test) เป็นแบบทดสอบผลสัมฤทธิ์ที่ใช้เพื่อค้นหาสาเหตุของข้อบกพร่องหรืออุปสรรคในการเรียนรู้ โดยทั่วไปเป็นข้อสอบที่ง่ายและใช้เพื่อบ่งชี้ถึงสาเหตุที่ทำให้เกิดความบกพร่องทางการเรียนรู้ ข้อสอบมีลักษณะของเนื้อหาและขั้นตอนการเรียนรู้ที่ละเอียด อันเป็นการช่วยระบุได้ว่า บุคคลไม่รู้เรื่องในเนื้อหาตรงส่วนใด เพราะเหตุใด แบบทดสอบประเภทนี้จึงเหมาะกับการใช้เพื่อปรับปรุงการเรียนการสอน และเหมาะที่จะใช้ประเมินผลระหว่างการเรียนการสอนได้ด้วย

2.3 แบบทดสอบผลสรุป (Summative Test) เป็นแบบทดสอบผลสัมฤทธิ์ที่ใช้เพื่อประเมินผลสรุปโดยรวมเมื่อสิ้นสุดการเรียนการสอน แบบทดสอบนี้มีความครอบคลุมในเนื้อหา และวัตถุประสงค์ของรายวิชาหรือเรื่องนั้น ๆ โดยทั่วไปเป็นข้อสอบที่มีระดับความยากง่ายที่หลากหลาย รวมทั้งเป็นแบบทดสอบอิงกลุ่มด้วย

2.4 แบบทดสอบระหว่างเรียน (Formative Test) เป็นแบบทดสอบผลสัมฤทธิ์ที่ใช้ระหว่างที่มีการเรียนการสอน เพื่อเป็นการให้ข้อมูลย้อนกลับ (feedback) แก่ผู้สอน รวมทั้งตัวผู้เรียนได้ทราบความก้าวหน้าในการเรียนรู้ ซึ่งส่วนนี้ใช้เพื่อการปรับปรุงการเรียนการสอนให้ดีขึ้น แบบทดสอบนี้มุ่งตรวจสอบวัตถุประสงค์ที่จำเป็นของเนื้อหาแต่ละเรื่อง โดยทั่วไปข้อสอบมีระดับความยากที่สอดคล้องกับวัตถุประสงค์ในแต่ละหน่วย รวมทั้งเป็นแบบทดสอบอิงเกณฑ์

3. จำแนกตามลักษณะของการสร้าง สามารถแบ่งได้เป็น 2 ประเภท คือ

3.1 แบบทดสอบผลสัมฤทธิ์ที่สร้างโดยผู้สอน (Teacher-Made Achievement Test) เป็นแบบทดสอบที่ใช้ภายในห้องเรียน โดยผู้สอนสร้างขึ้นเพื่อใช้วัดและประเมินผลการเรียนการสอนภายในห้องเรียน ส่วนใหญ่เป็นการวัดผลสัมฤทธิ์ในการเรียนรายวิชาต่าง ๆ เป็นการวัดเนื้อหาและการเรียนรู้ที่ได้เรียนไปแล้ว โดยยึดตามเนื้อหาที่ได้กำหนดไว้ตามโครงสร้างของหลักสูตร แบบทดสอบประเภทนี้อาจมีความเที่ยงตรงตามเนื้อหาสูง เพราะสร้างเองโดยตัวผู้สอนที่มีความเข้าใจในเนื้อหาตามรายวิชาที่ต้องการวัดเป็นอย่างดี แต่อย่างไรก็ตามการตรวจสอบคุณภาพเครื่องมือด้วยเกณฑ์ประเภทอื่น ก็ยังมีความจำเป็นที่ต้องดำเนินการเพื่อให้เป็นแบบทดสอบที่มีความน่าเชื่อถือมากขึ้น

3.2 แบบทดสอบผลสัมฤทธิ์มาตรฐาน (Standard Achievement Test) เป็นแบบทดสอบผลสัมฤทธิ์ที่สร้างขึ้นอย่างเป็นระบบและมีหลักเกณฑ์จากผู้เชี่ยวชาญในสาขาวิชานั้นๆ มีการดำเนินการตรวจสอบคุณภาพของแบบทดสอบ ทำการวิเคราะห์ รวมทั้งปรับปรุงแก้ไขจนแบบทดสอบมีคุณภาพในระดับที่เชื่อถือได้ และมีประสิทธิภาพสูง สามารถนำไปใช้ได้กว้างขวาง โดยแบบทดสอบนี้มีคุณสมบัติที่สำคัญ คือ มีมาตรฐานในวิธีดำเนินการสอบ มีมาตรฐานในการตรวจให้คะแนนที่เป็นระบบ และมีมาตรฐานในการกำหนดเกณฑ์การแปลความหมายของคะแนน นอกจากนี้ แบบทดสอบมาตรฐานยังมีเกณฑ์ปกติ (norm) ของแบบทดสอบ มีทั้งเกณฑ์ปกติระดับกลุ่ม ระดับชั้น ระดับภาค และระดับชาติ ในการใช้เป็นเกณฑ์เพื่อเปรียบเทียบผลเมื่อนำแบบทดสอบนี้

ไปใช้วัดความสามารถในการเรียนรู้ของแต่ละคน เพื่อดูว่าผลสัมฤทธิ์ของผู้เรียนเป็นเช่นใดเมื่อเทียบกับเกณฑ์ปกติของแบบทดสอบ

4. จำแนกตามการแปลความหมาย สามารถแบ่งได้เป็น 2 ประเภท คือ

4.1 แบบทดสอบอิงกลุ่ม (Norm-Referenced Test) เป็นแบบทดสอบผลสัมฤทธิ์ที่มีวัตถุประสงค์ในการวัดความแตกต่างระหว่างบุคคล โดยนำเอาความสามารถของแต่ละคนมาทำการเปรียบเทียบกับบุคคลอื่น ๆ ในกลุ่มที่เป็นกลุ่มอ้างอิงหรือกลุ่มมาตรฐาน เป็นการตรวจสอบว่าความสามารถของแต่ละคนเป็นอย่างไร เมื่อเทียบกับบุคคลอื่นภายในกลุ่ม เช่น ผลการเรียนวิชาชีววิทยาของนาย A อยู่ในตำแหน่งเปอร์เซ็นต์ไทล์ที่ 25 นั่นคือ ในชั้นเรียนนี้มีคนที่มีความสามารถในวิชาชีววิทยาน้อยกว่านาย A อยู่ที่ร้อยละ 25

4.2 แบบทดสอบอิงเกณฑ์ (Criterion-Referenced Test) เป็นแบบทดสอบผลสัมฤทธิ์ที่เปรียบเทียบกับผลการสอบกับค่ามาตรฐานของเกณฑ์การประเมิน มักเน้นไปที่ความสามารถระดับบุคคลมากกว่าเปรียบเทียบผลการสอบกับคนอื่น เป็นการระบุว่าบุคคลนั้นมีความรู้ระดับใดในเนื้อหาที่ทำการสอบ แบบทดสอบประเภทนี้มักนำมาใช้ในด้านการศึกษา โดยให้ความสำคัญเกี่ยวกับกระบวนการศึกษาและธรรมชาติของการเรียนรู้ว่า บุคคลสามารถเรียนรู้เนื้อหาต่าง ๆ ได้ ความแตกต่างระหว่างบุคคลในการเรียนรู้เนื้อหานั้น ๆ ขึ้นอยู่กับระยะเวลา หรือปริมาณการเรียนการสอนที่ให้กับผู้เรียน ทุกคนสามารถมีความรู้ในเนื้อหานั้น ๆ ได้เท่าเทียมกัน ไม่มีความแตกต่างระหว่างบุคคลในแง่ของความรู้และทักษะ ดังนั้น จึงเป็นเครื่องมือที่ใช้สำหรับการวินิจฉัยมากกว่าเป็นเครื่องมือที่วัดผลการสอบเพียงอย่างเดียว

ขั้นตอนการสร้างแบบทดสอบผลสัมฤทธิ์

แบบทดสอบผลสัมฤทธิ์ มีขั้นตอนในการสร้างและพัฒนา (ไซติกา ภาชีผล, 2559; อรพินทร์ ชูชม, 2545) ดังนี้

1. กำหนดวัตถุประสงค์ ในการสร้างแบบทดสอบ ต้องกำหนดวัตถุประสงค์ในการสอบให้ชัดเจนว่าแบบทดสอบที่สร้างขึ้นมีวัตถุประสงค์ใช้เพื่ออะไร จะได้ออกแบบแบบทดสอบผลสัมฤทธิ์ให้ได้ตรงตามวัตถุประสงค์ของการสอบ เช่น เพื่อวัดความรู้เบื้องต้นเกี่ยวกับชีวอนามัยเจริญพันธุ์ทั่วไปของนักเรียนชั้นมัธยมปลาย ข้อสอบที่ใช้วัดควรเป็นข้อสอบแบบปรนัย หรือเพื่อสอบคัดเลือกชิงทุนเข้าศึกษาต่อในระดับปริญญาเอก ข้อสอบที่ใช้ควรมีลักษณะที่เน้นทดสอบการแสดงความสามารถทางความคิด และควรเป็นข้อสอบอัตนัย

2. กำหนดเนื้อหาและสิ่งที่ต้องการวัด เป็นการกำหนดขอบเขตเนื้อหาและสิ่งที่ต้องการวัด โดยให้จัดทำตารางการกำหนดลักษณะเฉพาะ การกำหนดลักษณะเฉพาะช่วยให้ผู้สอนมีความเข้าใจว่าจะสร้างข้อสอบที่วัดผู้เรียนให้เกิดการเรียนรู้และเนื้อหาใด ดังนั้น สิ่งที่ผู้สอนควรคำนึงถึงคือวัตถุประสงค์และเนื้อหาที่ใช้สร้างข้อสอบ ว่าต้องการให้ผู้เรียนแสดงออกถึงการเรียนรู้อะไรบ้าง ซึ่งการกำหนดวัตถุประสงค์ดังกล่าวต้องแปลงเป็นคุณลักษณะที่ต้องการวัดให้เป็นวัตถุประสงค์การเรียนรู้ โดยมุ่งเน้นการเรียนรู้ด้านการรู้คิด (Cognitive) ตามแนวคิด Bloom (1956) โดยการเรียนรู้แบ่งได้เป็น 6 ระดับ ดังตารางที่ 4-4 โดยการออกแบบข้อสอบให้สอดคล้องกับวัตถุประสงค์การเรียนรู้ จึงต้องทำความเข้าใจวัตถุประสงค์การเรียนรู้ นิยาม และค่าบ่งชี้วัตถุประสงค์

ตารางที่ 4-4 วัตถุประสงค์การเรียนรู้ด้านการรู้คิดของ Bloom (1956)

วัตถุประสงค์การเรียนรู้	นิยาม	คำบ่งชี้วัตถุประสงค์การเรียนรู้
ความรู้ความจำ (Knowledge)	ความสามารถในการระลึก จดจำคำศัพท์ ข้อเท็จจริง กระบวนการ ความสัมพันธ์ แนวคิดต่าง ๆ	นับ บอก ระบุ ชี้ บอกชื่อ ชิดเส้นใต้ ให้นิยาม จับคู่ ท่อง เลือก กำหนด เขียน วาด ลอก
ความเข้าใจ (Comprehension)	ความเข้าใจในสิ่งที่เรียนรู้ โดยสามารถแปลความ ตีความ ขยายความด้วยภาษาตนเอง	อธิบาย แปล ปรับแก้ ตีความหมาย ขยาย ความ ยกตัวอย่าง เปรียบเทียบ อภิปราย เรียบเรียงใหม่ แปลงเป็นสมการ
การนำไปใช้ (Application)	ความสามารถในการใช้สิ่งที่ได้เรียนรู้นำมาแก้ปัญหา หรือสร้างแนวทางเลือกให้เหมาะสมกับบริบท	สาธิต จัดกระทำ แก้ไข ใช้ ผลิต คำนวณ ปฏิบัติ ดำเนินการ เปลี่ยน สร้าง ทำนาย แก้ปัญหา
การวิเคราะห์ (Analysis)	ความสามารถจำแนกองค์ประกอบย่อยของสิ่งต่าง ๆ สามารถจัดประเภท แบ่งแยกสิ่งต่าง ๆ และบอก ความสัมพันธ์ระหว่างองค์ประกอบได้	แบ่งแยก บอกความแตกต่าง จัดประเภท แยกย่อย แยกประเด็น จำแนก ทำแผนผัง
การสังเคราะห์ (Synthesis)	ความสามารถในการประมวลหรือรวบรวมความรู้ที่ เป็นส่วน ๆ มาเป็นแนวคิดใหม่ ริเริ่ม สร้างสรรค์ แนวทางใหม่	รวม รวมกลุ่ม กำหนดขั้นตอนใหม่ สร้าง ออกแบบ ทำสูตร สรุปอ้างอิง จัดลำดับ จัด กลุ่ม สรุป ตั้งสมมติฐาน
การประเมินค่า (Evaluation)	ความสามารถตัดสินคุณค่าหรือความเหมาะสมของ สิ่งใด ๆ โดยใช้เกณฑ์ที่เหมาะสมและสมเหตุสมผล	ประเมิน ตัดสิน โต้แย้ง วิพากษ์ วิวิจารณ์ ให้ ระดับคุณภาพ เสนอแนะ

ตารางการกำหนดลักษณะเฉพาะ ควรประกอบด้วย หัวข้อหลักของเนื้อหา วัตถุประสงค์การเรียนรู้ และการให้น้ำหนักแต่ละเนื้อหาและแต่ละการเรียนรู้ ดังตารางที่ 4-5

ตารางที่ 4-5 การกำหนดลักษณะเฉพาะของแบบทดสอบผลสัมฤทธิ์การวัดการรู้คิด จำนวน 50 ข้อ

หัวข้อหลัก	วัตถุประสงค์การเรียนรู้						รวม	ร้อยละ
	ความรู้	เข้าใจ	นำไปใช้	วิเคราะห์	สังเคราะห์	ประเมินค่า		
สติปัญญา	2	2	3	3	-	5	15	30
ความถนัด	2	2	3	3	-	5	15	30
ผลสัมฤทธิ์	3	3	4	4	-	6	20	40
รวม	7	7	10	10	-	16	50	
ร้อยละ	14	14	20	20	-	32		100

3. กำหนดประเภทของแบบทดสอบ แบบทดสอบผลสัมฤทธิ์มีหลายประเภท ต้องกำหนดให้เหมาะสมและสอดคล้องกับวัตถุประสงค์ของการสอบ เนื่องจากแบบทดสอบแต่ละประเภทเหมาะที่จะใช้วัดด้านการรู้คิดที่แตกต่างกัน

4. สร้างข้อสอบ โดยต้องออกข้อสอบให้ตรงตามเนื้อหาและวัตถุประสงค์การเรียนรู้ที่กำหนดไว้ ตามตารางการกำหนดลักษณะเฉพาะ และควรสร้างข้อสอบให้มากกว่าจำนวนข้อสอบที่กำหนดไว้ในตาราง เพื่อใช้ในการคัดเลือกต่อไป นอกจากนี้ ควรมีคู่มือเฉลยข้อสอบด้วย รวมทั้งการกำหนดเกณฑ์การให้คะแนนด้วย

5. ตรวจสอบคุณภาพแบบทดสอบ โดยนำเอาแบบทดสอบที่สร้างขึ้นมาทำการตรวจสอบคุณภาพของแบบทดสอบ เริ่มต้นตรวจสอบความเที่ยงตรงเชิงเนื้อหา โดยให้ผู้เชี่ยวชาญเป็นผู้พิจารณาความเหมาะสมของภาษา รวมทั้งพิจารณาว่าสร้างได้สอดคล้องกับวัตถุประสงค์การเรียนรู้และเนื้อหาที่ต้องการวัด และข้อสอบมีความเหมาะสมกับระดับความสามารถของผู้เรียนด้วย หลังจากนั้น จึงนำไปปรับปรุงแก้ไขตามข้อเสนอแนะของผู้เชี่ยวชาญ เมื่อปรับปรุงเสร็จเรียบร้อยแล้วจึงนำไปทดลองใช้กับกลุ่มตัวอย่างจำนวนหนึ่ง เพื่อวิเคราะห์เป็นรายข้อ ได้แก่ อำนาจจำแนก ความยากง่าย รวมทั้งความเชื่อมั่น หากผลการตรวจสอบยังไม่เป็นไปตามเกณฑ์ที่กำหนด จำเป็นต้องมีการปรับปรุงแก้ไข แล้วนำไปทดลองใช้ใหม่ เพื่อให้ได้ข้อสอบที่มีคุณสมบัติทางการวัดที่เหมาะสมและยอมรับได้

หลักการสร้างแบบทดสอบผลสัมฤทธิ์

แบบทดสอบผลสัมฤทธิ์ที่นิยมใช้กัน นั่นคือ แบบทดสอบปรนัยและแบบทดสอบอัตนัย ซึ่งมีรายละเอียดที่สำคัญ (โชติกา ภาชีผล, 2559; อรพินทร์ ชูชม, 2545) ดังนี้

1. แบบทดสอบปรนัย (Objective Test) เป็นแบบทดสอบผลสัมฤทธิ์ที่ประกอบด้วยข้อสอบที่มีคำตอบให้ผู้ตอบเลือกตอบหรือเว้นช่องว่างให้ตอบ มีข้อดีคือ ข้อสอบครอบคลุมเนื้อหาที่ต้องการวัด มีความชัดเจนของข้อคำถาม ความชัดเจนในการให้คะแนน และความชัดเจนในการแปลความหมายของคะแนน มักใช้วัดความรู้ความเข้าใจ แต่มีข้อจำกัดคือ ถ้าผู้ออกข้อสอบขาดความสามารถในการออกข้อสอบ ข้อสอบจะเป็นเหมือนการคัดลอกข้อความมาจากตำราเอกสาร เป็นเพียงการวัดการท่องจำเท่านั้น แบบทดสอบปรนัยที่นิยมใช้ มี 4 ประเภท ดังนี้

1.1 ข้อสอบแบบถูกผิด (true-false items) เป็นข้อสอบที่ประกอบด้วย ข้อความและมีคำตอบให้ผู้ตอบเลือกตอบอย่างใดอย่างหนึ่ง จากตัวเลือกที่กำหนดไว้สองทาง เช่น ถูก-ผิดเห็นด้วย-ไม่เห็นด้วย

ตัวอย่างข้อสอบแบบถูกผิด

คำชี้แจง จงกาเครื่องหมาย / หน้าข้อความที่คิดว่าถูก และ X หน้าข้อความที่คิดว่าผิด

..... 1. แบบทดสอบผลสัมฤทธิ์เป็นแบบทดสอบใช้เพื่อทำนายความสำเร็จ

..... 2. แบบทดสอบ Raven Progressive Matrices เป็นแบบทดสอบสติปัญญา

หลักในการเขียนข้อสอบแบบถูกผิด

- (1) หลีกเลี่ยงการใช้คำชี้แนะ เช่น เสมอ ทั้งหมด ทุกครั้ง ไม่เคยเลย เป็นต้น เพราะคำเหล่านี้เมื่อใช้แล้วมักจะเป็นข้อที่ผิด
- (2) การจัดเรียงข้อความที่ถูกต้องควรเป็นไปอย่างสุ่ม เพื่อไม่ให้เกิดรูปแบบการตอบแบบเดียวกันตลอด และเพื่อหลีกเลี่ยงการเดาคำตอบ
- (3) ข้อความที่ผิด ควรใช้กับสิ่งที่คุณเรียนมักเข้าใจผิด
- (4) ใช้รูปแบบคำถามที่เหมาะสมเพื่อให้ง่ายต่อการตรวจให้คะแนน
- (5) ข้อความต้องมีความชัดเจนในตัวคำตอบที่แน่นอนว่าถูกหรือผิด
- (6) หลีกเลี่ยงคำถามที่ใช้ข้อความปฏิเสธ ถ้าใช้ ควรระมัดระวังโดยการเน้นให้ชัดเจน
- (7) หลีกเลี่ยงคำถามที่เป็นความคิดเห็นเฉพาะบุคคลหรือต้องตัดสินใจ
- (8) ควรเป็นคำถามที่สำคัญเพียงประเด็นเดียว เพื่อให้เข้าใจง่าย กระชับ และชัดเจน
- (9) วัดเฉพาะแนวคิดหรือประเด็นสำคัญที่ควรทดสอบ

1.2 ข้อสอบแบบตัวเลือก (multiple choice items) เป็นข้อสอบที่ประกอบด้วย ส่วนที่เป็นคำถาม และส่วนที่เป็นตัวเลือก ซึ่งมีทั้งตัวเลือกที่เป็นคำตอบที่ถูก ส่วนตัวเลือกที่เหลือเป็นตัวเลือกที่เป็นคำตอบผิด หรือเป็นตัวลวง

ตัวอย่างข้อสอบแบบตัวเลือก

คำชี้แจง จงเลือกตัวเลือกที่ถูกต้องเพียงตัวเลือกเดียวในแต่ละข้อ

1. แบบทดสอบ Raven Progressive Matrices เป็นแบบทดสอบชนิดใด
 - ก. แบบทดสอบสติปัญญา
 - ข. แบบทดสอบความถนัด
 - ค. แบบทดสอบผลสัมฤทธิ์
 - ง. ถูกทุกข้อ

หลักในการเขียนข้อสอบแบบหลายตัวเลือก

- (1) ไม่ควรลอกประโยคจากตำรามาเขียนเป็นข้อสอบ แต่ควรเรียบเรียงใหม่
- (2) อย่าให้ข้อสอบข้อหนึ่งส่งผลต่อการชี้แนะคำตอบในข้ออื่น
- (3) ไม่ใช่คำถามที่เป็นการชี้แนะคำตอบ หรือข้อความในคำถามไม่ควรซ้ำกับข้อความในตัวเลือกที่เป็นคำตอบที่ถูก

- (4) ใช้ภาษาให้เหมาะสมกับระดับของผู้ตอบ
- (5) ไม่ควรใช้ประโยคปฏิเสธในตัวคำถาม ถ้าใช้ให้ขีดเส้นใต้หรือแสดงเป็นตัวทึบ และไม่ควรถูกใช้ คำปฏิเสธซ้อน คือ เมื่อใช้ปฏิเสธในตัวคำถามแล้ว ไม่ควรมีปฏิเสธในตัวเลือก เพราะอาจเกิดความสับสนได้
- (6) ใช้คำที่มีความหมายชัดเจน สั้น กระชับ ไม่ซับซ้อน
- (7) เขียนตัวเลือกที่ถูกเพียงคำตอบเดียว
- (8) ตัวเลือกทุกตัวมีความเป็นไปได้ที่จะเป็นคำตอบที่ถูก ต้องสมเหตุสมผล
- (9) ข้อความในตัวเลือกควรเป็นอิสระจากกัน ไม่ทับซ้อนกัน
- (10) ใช้ตัวเลือกปลายเปิดให้เหมาะสม ได้แก่ ถูกทุกข้อ ผิดทุกข้อ ไม่มีข้อใดถูก

1.3 ข้อสอบแบบจับคู่ (matching items) เป็นข้อสอบที่ประกอบด้วย รายการตัวคำถามหรือตัวนำเรื่อง (premises) และรายการตัวเลือก (responses) โดยให้ผู้สอบพิจารณาจับคู่ข้อความรายการตัวคำถามกับรายการตัวเลือกให้สอดคล้องเหมาะสมกัน

ตัวอย่างข้อสอบแบบจับคู่

คำชี้แจง คำทางซ้ายมือเป็นคำศัพท์การวัดด้านการรู้คิด ข้อความทางขวาเป็นข้อความเกี่ยวกับคำศัพท์ ให้นำตัวอักษรหน้าข้อความทางขวามือ ใส่ในช่องว่างหน้าตัวเลขซ้ายมือ

- | | |
|-----------------------------|--------------------------------|
| 1. แบบทดสอบสติปัญญา | ก. วัดความสามารถทางดนตรี |
| 2. แบบทดสอบความถนัด | ข. วัดสิ่งที่ได้เรียนรู้มาแล้ว |
| 3. แบบทดสอบผลสัมฤทธิ์ | ค. วัดบุคลิกภาพ |
| | ง. วัดเจตคติ |
| | จ. วัด IQ |

หลักในการเขียนข้อสอบแบบจับคู่

- (1) คำถามและคำตอบควรพิมพ์ไว้ในหน้าเดียวกัน
- (2) ควรจัดเรียงคำตอบตามลำดับโดยยึดหลักเกณฑ์หนึ่ง ๆ
- (3) แต่ละคำตอบสามารถเลือกได้ครั้งเดียวหรืออาจหลายครั้งก็ได้
- (4) ควรเขียนคำชี้แจงให้ชัดเจน ว่าจับคู่ในรูปแบบใด เช่น เขียนตัวอักษร ลากเส้น
- (5) ใช้ข้อความที่สั้น กระชับ กะทัดรัด ชัดเจน ใจความครบถ้วน
- (6) คำถามและคำตอบควรมีจำนวนไม่เท่ากัน และคำตอบทุกคำตอบควรมีโอกาสเลือกมาจับคู่ได้
- (7) คำถามและคำตอบต้องมีความเกี่ยวข้องสัมพันธ์กัน
- (8) เนื้อหาที่นำมาสร้างควรมีลักษณะเป็นเรื่องเดียวกัน หรือคล้ายคลึงกัน

2. แบบทดสอบอัตนัย (Subjective Test) เป็นแบบทดสอบผลสัมฤทธิ์ที่ประกอบด้วยข้อสอบที่ผู้ตอบเป็นผู้เขียนคำตอบด้วยตัวเอง เป็นการเปิดโอกาสให้ผู้ตอบได้แสดงความรู้ความคิดเห็นในสิ่งที่ถามได้อย่างเต็มที่ตามประเด็นในแบบทดสอบ ข้อสอบประเภทนี้แสดงถึงความสามารถในการเขียน ที่สะท้อนถึงการคิดวิเคราะห์ การสังเคราะห์ และการสร้างสรรค์ โดยวัดความสามารถของผู้สอบในด้านการระลึกความรู้ การประเมินความรู้ การจัดระเบียบความคิด และการเสนอแนวคิดที่เกี่ยวข้องกันอย่างสมเหตุสมผล แบบทดสอบอัตนัยมีรูปแบบการตอบอยู่ 2 ลักษณะ ดังนี้

2.1 ข้อสอบแบบไม่จำกัดคำตอบ (Essay-extended) เป็นข้อสอบที่เปิดโอกาสให้มีอิสระในการตอบมากกว่าแบบจำกัดคำตอบ โดยเลือกเอาความรู้ หลักการ แนวคิดทฤษฎีอื่น ๆ มาเรียบเรียงเพื่อสนับสนุนให้ได้คำตอบที่ดีที่สุด ไม่จำกัดขอบเขตคำตอบ แต่ต้องตอบภายในระยะเวลาที่จำกัด ข้อสอบนี้มีความยืดหยุ่นและให้เสรีภาพในการตอบสูง คำตอบจึงกว้างมาก จึงอาจทำให้เกิดการตรวจให้คะแนนที่ไม่สอดคล้องกันได้ ดังนั้นข้อสอบนี้จึงเหมาะใช้วัดการรู้คิดขั้นสูง ได้แก่ การคิดวิเคราะห์ การสังเคราะห์ และการประเมินค่า โดยวัดความสามารถในการจัดระบบ เรียบเรียง ตีความ สรุป สร้างสรรค์สิ่งใหม่ เสนอแนวคิด ประเมินคุณค่า และบูรณาการการเรียนรู้ในเรื่องต่าง ๆ เป็นต้น

ตัวอย่างข้อสอบแบบไม่จำกัดคำตอบ

1. จากสถานการณ์โรคติดเชื้อไวรัสโคโรนาที่แพร่ระบาดในประเทศไทย จงอภิปรายถึงวิธีการป้องกันตนเองว่ามีอะไรบ้าง รวมทั้งจะเสนอแนะแนวทางการป้องกันดังกล่าวให้กับคนในชุมชนได้อย่างไร

2.2 ข้อสอบแบบจำกัดคำตอบ (Essay-Restricted) เป็นข้อสอบที่จำกัดกรอบเนื้อหาและความยาวของคำตอบ หรือมีการกำหนดแนวทางการตอบที่ค่อนข้างตายตัว คำตอบจึงสั้นและแคบกว่า รวมทั้งมีอิสระในการตอบน้อยกว่าแบบไม่จำกัดคำตอบ การตรวจให้คะแนนจึงมีเกณฑ์ที่ชัดเจนกว่า เนื่องจากคำตอบอยู่ภายในกรอบเนื้อหาเดียวกัน ดังนั้น ข้อสอบนี้จึงเหมาะใช้วัดการรู้คิดเพื่ออธิบายความสัมพันธ์เชิงเหตุและผล บรรยายถึงการนำหลักการต่าง ๆ ไปใช้ แสดงข้อโต้แย้งในประเด็นต่าง ๆ ให้ข้อสรุป เป็นต้น

ตัวอย่างข้อสอบแบบจำกัดคำตอบ

1. จงบอกข้อดีและข้อเสียของการกักตัวอยู่ที่บ้านในสถานการณ์การแพร่ระบาดโรคติดเชื้อไวรัสโคโรนาในประเทศไทย มาอย่างละ 3 ข้อ

หลักในการเขียนข้อสอบแบบอัตนัย

- (1) สร้างข้อสอบเมื่อต้องการวัดการรู้คิดขั้นสูง โดยต้องเกี่ยวข้องกับเนื้อหาที่สอน
- (2) คำถามต้องชัดเจนว่าต้องการให้ผู้เรียนตอบอะไรหรือไปในทิศทางใด
- (3) จำกัดขอบเขตการตอบให้ชัดเจน สำหรับข้อสอบแบบจำกัดคำตอบ

(4) ควรใช้คำที่มองเห็นแนวทางการตอบชัดเจน เช่น อธิบาย อภิปราย เสนอแนะ
 (5) ไม่ควรใช้คำถามที่เริ่มต้นด้วย ใคร อะไร ที่ไหน เพราะเป็นการวัดการรู้คิดขั้นต้น
 (6) ควรออกข้อสอบที่ต้องการให้ตอบสั้น จำนวนมากข้อ เพราะสามารถวัดได้ครอบคลุมเนื้อหาและมีความเป็นปรนัยในการให้คะแนนมากกว่าข้อสอบที่ให้ตอบยาว ๆ

(7) ควรกำหนดคะแนนและเวลาให้ตอบแต่ละคำถาม เพื่อที่ผู้ตอบจะได้วางแผนได้เหมาะสม

(8) ไม่ควรให้เลือกรับบางข้อ เพราะเป็นการยากที่จะทำให้ข้อสรุปแต่ละข้อมีความยากง่ายไม่เท่ากัน

หลักในการตรวจให้คะแนนข้อสอบแบบอัตนัย

(1) กำหนดขอบเขตของคำตอบที่คาดหวังไว้ล่วงหน้า เพื่อใช้เป็นเกณฑ์หรือแนวทางในการกำหนดคะแนนแต่ละประเด็น

(2) อ่านคำตอบคร่าว ๆ เพื่อให้เห็นแนวทางในการตรวจ ทำให้สามารถกำหนดเกณฑ์การให้คะแนนที่เหมาะสมยิ่งขึ้น

(3) พยายามปิดชื่อผู้สอบ และไม่เรียงคำตอบตามลำดับเลขที่ผู้สอบ เพื่อไม่ให้เกิดอคติในการตรวจให้คะแนน

(4) ตรวจข้อสอบทีละข้อหรือทีละคำถาม และตรวจให้เสร็จในคราวเดียวกันสำหรับข้อนั้น ๆ เพื่อให้การตรวจให้คะแนนมีความเป็นเอกภาพมากขึ้น

(5) ใช้วิธีการตรวจให้คะแนนที่มีความเหมาะสม โดยมี 2 วิธี คือ

- ตรวจแบบวิเคราะห์ (analytic scoring) การกำหนดประเด็นหรือจุดสำคัญที่จะให้คะแนนหรือหักคะแนนได้อย่างชัดเจน เปรียบเทียบคำตอบกับคำตอบ และให้คะแนนสำหรับคำตอบของผู้เรียนตามที่กำหนดซึ่งเหมาะกับแบบจำกัดคำตอบ

- ตรวจแบบสรุปรวม (holistic scoring) การให้คะแนนคำตอบโดยประมาณค่า หลังจากอ่านคำตอบแล้วจะจัดคำตอบไว้เป็นกลุ่ม แต่ละกลุ่มแสดงคุณภาพของคำตอบ ให้คะแนนรวมโดยพิจารณาความถูกต้องเหตุผลในการตอบ ซึ่งเหมาะกับแบบไม่จำกัดคำตอบ

สรุป

การวัดทางด้านการรู้คิดถือเป็นกระบวนการหนึ่งที่สำคัญในการพัฒนาเครื่องมือวิจัย เป็นวิธีการวัดที่ช่วยในการคัดเลือกและประเมินความสามารถของแต่ละบุคคล ดังนั้น จึงมีความจำเป็นต้องเข้าใจความเป็นมาแนวคิดเบื้องต้น การวัดทางด้านการรู้คิดที่สำคัญ เช่น แบบทดสอบสติปัญญา แบบทดสอบความถนัด และแบบทดสอบผลสัมฤทธิ์ โดยการวัดทางด้านการรู้คิดดังกล่าว เป็นแนวทางในการสร้างและพัฒนาเครื่องมือทางด้านพฤติกรรมศาสตร์ให้มีความเหมาะสม และสามารถวัดความแตกต่างทางด้านความสามารถของแต่ละบุคคล

เอกสารอ้างอิง

- โชติกา ภาชีผล. (2559). *การวัดและประเมินผลการเรียนรู้*. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ณัฐภรณ์ หลาวทอง. (2561). *การสร้างเครื่องมือวิจัยทางการศึกษา* (พิมพ์ครั้งที่ 2). กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- นพวรรณ ศรีวงศ์พานิช, จันทร์อาภา สุขทัพบ, วนิดา ชนินทุยทรวงศ์, และอัมพร เบญจพลพิทักษ์. (2562). ความสัมพันธ์ระหว่างความเสี่ยงต่อโรคสมาธิสั้น ออทิสซึม ภาวะบกพร่องทางการเรียนรู้และภาวะบกพร่องทางสติปัญญาจากการสังเกตพฤติกรรมโดยครูกับระดับสติปัญญา (IQ) ของเด็กไทยชั้นประถมศึกษาปีที่ 1. *วารสารสุขภาพจิตแห่งประเทศไทย*, 27(3), 159-170.
- ปาริชาติ บัวเจริญ, นพรัตน์ เตชะพันธ์รัตนกุล, และรัตนากาล คำสอน (2560). ความสัมพันธ์ระหว่างเขาวนปัญญาเขาวนอารมณ์ และผลสัมฤทธิ์ทางการศึกษาของนักศึกษาระดับปริญญาตรีของมหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เชียงใหม่. *วารสารบัณฑิตวิจัย*, 8(2), 101-117.
- อภิชัย มงคล, ทวีศิลป์ วิษณุโยธิน, ปราณี ชาญณรงค์, โชชิตา ภาวสุทธิไพศิฐ, และพรทิพย์ วชิรดิถ. (2555). สถานการณ์ระดับสติปัญญาเด็กนักเรียนไทย ปี 2554. *วารสารสุขภาพจิตแห่งประเทศไทย*, 20(2), 79-89.
- อรพินทร์ ชูชม. (2545). *เอกสารคำสอนวิชา วป 502 การสร้างและพัฒนาเครื่องมือวัดทางพฤติกรรมศาสตร์*. กรุงเทพฯ: สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- Benjamin, A. S., & Pashler, H. (2015). The value of standardized testing: A perspective from cognitive psychology. *Policy Insights from the Behavioral and Brain Sciences*, 2(1), 13-23.
- Bloom, B. S. (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain*. New York: David McKay Co Inc.
- Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2013). *Psychological testing and assessment: An introduction to tests and measurement* (8th ed.). New York: McGraw-Hill.
- Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, 24(1), 79-132.
- Gottfredson, L. S., & Saklofske, D. H. (2009). Intelligence: Foundations and issues in assessment. *Canadian Psychology*, 50(3), 183-195.
- Grabner, R. H., & Stern, E. (2011). Measuring cognitive abilities. In German Data Forum (RatSWD) (Ed.), *Building on progress: Expanding the research infrastructure for the social, economic, and behavioral sciences – Recommendations of the German data forum (RatSWD)*, pp. 753-768. Opladen & Farmington Hills: Budrich UniPress.

- Gregory, R. J. (1996). *Psychological testing: History, principles, and applications* (2nd ed.). Boston: Allyn and Bacon.
- Hale, R. L. (1987). Evaluation of intelligence, achievement, aptitude, and interest. In V. B. Van Hasselt, & Hersen, M. (Eds.), *psychological evaluation of the developmentally and physically disabled* (pp. 41-67). New York: Springer Science+Business Media.
- Hunter, J. E. (1994). General Aptitude Test Battery. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence*. New York: Macmillan.
- Lohman, D. F. (2005). An aptitude perspective on talent: Implications for identification of academically gifted minority students. *Journal for the Education of the Gifted*, 28, 333-360.
- Ones, D. S., Dilchert, S., & Viswesvaran, C. (2012). Cognitive abilities. In N. Schmitt (Ed.), *Oxford library of psychology. The Oxford handbook of personnel assessment and selection* (pp. 179-224). Oxford: Oxford University Press.
- Paimpuech, R. (2013). The relationship of intelligence quotient scores between the WISC-III-Thai version and Leiter-R in children with learning disabilities. *International Journal of Child Development and Mental Health*, 1(1), 19-28.
- Roid, G. H. (2003). *Stanford-Binet Intelligence Scales, Fifth Edition, Examiner's Manual*. Itasca, IL: Riverside Publishing.
- Sax, G. (1997). *Principles of educational and psychological measurement and evaluation* (4th ed.). Belmont, CA: Wadsworth Publishing Company.

บทที่ 5

การวัดด้านความรู้สึก

บทที่ 5 เป็นการนำเสนอการวัดด้านความรู้สึก โดยผู้เขียนต้องการให้ผู้อ่านได้ทราบและทำความเข้าใจเกี่ยวกับการวัดด้านความรู้สึก ความเป็นมาของการวัดทางด้านความรู้สึก แนวคิดเบื้องต้นของการวัดด้านความรู้สึก สำหรับการวัดทางด้านความรู้สึก มีการสร้างแบบทดสอบและวิธีการวัดอย่างมากมาย มุ่งเน้นการวัดความสนใจ ค่านิยม รวมทั้งบุคลิกภาพ โดยบทนี้ จะช่วยให้ผู้วิจัยและนักวิชาการสามารถสร้างแบบสอบถามด้านความรู้สึก ได้อย่างถูกต้อง เพื่อช่วยให้กระบวนการพัฒนาเครื่องมือวิจัยทางพฤติกรรมศาสตร์มีความเหมาะสมและสมบูรณ์มากยิ่งขึ้น

ความเป็นมาของการวัดทางด้านความรู้สึก

ในช่วงที่ผ่านมา นักวิชาการและนักจิตวิทยาให้ความสำคัญกับการวัดทางด้านความรู้สึก เช่น การวัดสติปัญญา ผลสัมฤทธิ์ และการรู้คิดแบบอื่น ๆ มากกว่าการวัดทางด้านความรู้สึก (Oakland, 1997) เช่นเดียวกับการศึกษาที่ผ่านมา ที่มุ่งวัดการเรียนรู้ด้านการรู้คิด (Cognitive) ตามแนวคิด Bloom (1956) มากกว่าด้านความรู้สึกหรือจิตพิสัย (Affective) ในด้านความรู้สึกมักสนใจการแสดงออกที่เกี่ยวข้องกันทางด้านอารมณ์ เช่น ความรู้สึก ค่านิยม ความซาบซึ้ง ความกระตือรือร้น แรงจูงใจ และเจตคติ เป็นต้น (Krathwohl, Bloom, & Masia, 1973) และที่ผ่านมา การศึกษาด้านบุคลิกภาพ ก็ได้รับความนิยมนำมาใช้มากขึ้นเรื่อย ๆ โดยแบบวัดบุคลิกภาพมักจะวัดโครงสร้างหรือกระบวนการภายในของบุคคล เช่น อารมณ์ ความสนใจ ค่านิยม เป็นต้น ประกอบกัน (Staats, 2003) ดังนั้น ในบทของการวัดทางด้านความรู้สึก ผู้เขียนจึงขอเรียบเรียงการวัด ความสนใจ การวัด ค่านิยม และต่อท้ายด้วยการวัดบุคลิกภาพ เพื่อให้มีความเข้าใจในการวัดพฤติกรรมศาสตร์ โดยเน้นทางด้านความรู้สึก จากการวัดด้านความรู้สึกที่เรียบง่ายไปจนถึงการวัดด้านความรู้สึกที่ซับซ้อนมากขึ้น

แนวคิดเบื้องต้นของการวัดทางด้านความรู้สึก

ความรู้สึก (Affective Characteristics) เป็นคุณลักษณะหนึ่งภายในบุคคล (Anderson, 1981) มีคุณลักษณะดังนี้ 1) มีความเกี่ยวข้องทางด้านความรู้สึกและอารมณ์ 2) เป็นแบบอย่างของความคิดหรือพฤติกรรมของบุคคล 3) มีความเข้มข้น 4) มีทิศทางทั้งไปในทางบวกหรือทางลบ และ 5) มีความรู้สึกต่อเป้าหมายหนึ่ง ๆ นั่นคือความรู้สึกถือว่าเป็นอารมณ์หนึ่งที่สะท้อนมาจากความเชื่อ ค่านิยม ความสนใจ และพฤติกรรมของบุคคล (Gronlund

& Brookhart, 2009) โดยมีการนำแนวคิดของ Bloom ด้านความรู้สึก มาใช้วัดการเรียนรู้ในโรงเรียนมากขึ้น เนื่องจากมีการศึกษาพบว่า นอกจากด้านการรู้คิดแล้ว ด้านความรู้สึกส่งผลต่อความสำเร็จทางการเรียนของเด็กวัยเรียนได้เช่นกัน (Dolan, 1983; Oppong, 2014; Thien, Darmawan, & Ong, 2015) แต่ตัวแปรทางด้านความรู้สึกมีความเป็นนามธรรม ไม่สามารถทำการวัดได้โดยตรง จึงต้องทำการนิยามเชิงปฏิบัติการของตัวแปรนั้น ๆ ก่อน จึงจะทำการวัดได้ (McCoach, Gable, & Madura, 2013)

สำหรับการวัดทางด้านความรู้สึก ยังถูกแผ่ขยายไปใช้วัดได้อย่างหลากหลายตามวัตถุประสงค์ของการวัด โดยการวัดทางพฤติกรรมศาสตร์ มักสนใจวัดความรู้สึก เช่น การวัดความสนใจ การวัดค่านิยม การวัดบุคลิกภาพ เป็นต้น ซึ่งมีรายละเอียดแต่ละแบบวัดดังนี้

แบบวัดความสนใจ

ความสนใจ (Interest) เป็นความชอบของบุคคลที่มีต่อกิจกรรมหนึ่ง ๆ มากกว่ากิจกรรมอื่น ๆ โดยนิยามความสนใจดังกล่าวมีจุดเน้นใน 2 ประเด็น (Sax, 1997) นั่นคือ

1) ความสนใจเกี่ยวข้องข้องกับการเลือกและจัดลำดับกิจกรรมต่าง ๆ ตามความชอบและไม่ชอบ เช่น นักศึกษาคณะหนึ่งอาจชอบฟังเพลงมากกว่าทำวิจัย เป็นต้น

2) ความสนใจเกี่ยวข้องข้องกับกิจกรรมหรือพฤติกรรมที่บุคคลทำอยู่ ความสนใจจึงแสดงออกโดยคำกริยาในการกระทำ เช่น ใช้โทรศัพท์เคลื่อนที่ เล่นกีตาร์ หรือขี่จักรยาน เป็นต้น

ความสนใจยังมีความเกี่ยวข้องข้องกับการประเมินการเลือกอาชีพและให้คำปรึกษาก่อนการจ้างงาน เป็นเครื่องมือที่ออกแบบเพื่อประเมินความชอบ ไม่ชอบ กิจกรรมนั้นหนา การ ความใคร่รู้ และงานอดิเรก ในการเปรียบเทียบกับกลุ่มอาชีพต่าง ๆ (Cohen, Swerdlik, & Sturman, 2013)

ประเภทของความสนใจ

ความสนใจ แบ่งได้เป็น 3 ประเภท (Super & Crites, 1962 cited as Sax, 1997) ดังนี้

1. ความสนใจที่แสดงออก (Expressed Interest) เป็นการแสดงข้อความของความชอบในกิจกรรมหนึ่ง ๆ มากกว่ากิจกรรมอื่น ๆ เป็นวิธีการได้ข้อมูลเกี่ยวกับความสนใจที่ง่ายและวัดโดยตรงมากที่สุด แต่ก็มีข้อด้อยจำนวนมาก บุคคลที่ยังไม่บรรลุนิติภาวะจะมีแนวโน้มแสดงความสนใจตามสิ่งที่สังคมต้องการหรือคาดหวัง เช่น เด็กมีความสนใจที่จะเป็นแพทย์ตามความต้องการของผู้ปกครอง เป็นต้น

2. ความสนใจที่ปรากฏ (Manifest Interest) เป็นการแสดงพฤติกรรมของบุคคลในการเข้าร่วมกิจกรรมหนึ่ง ๆ อย่างสมัครใจ เช่น บุคคลเข้าร่วมในชั้นเรียนโยคะ โดยบางครั้งเป็นการยากที่จะแยกแยะว่าเหตุผลที่แท้จริงที่บุคคลเข้าร่วมกิจกรรมหนึ่ง ๆ เช่น การเข้าร่วมในชั้นเรียนโยคะ อาจเพราะสนใจในการเรียนโยคะ แต่อาจเป็นเพราะต้องการมีสุขภาพที่ดี หรือหลีกเลี่ยงกิจกรรมอื่น หรือความจริงที่ว่าในละแวกบ้าน มีแต่ชั้นเรียนโยคะ ไม่มีกิจกรรมอื่น ๆ นอกจากนี้ ไม่จำเป็นว่าความสนใจที่แสดงออกกับความสนใจที่ปรากฏจะมีความสัมพันธ์กัน

ทุกคนมักทำกิจกรรมที่ตนเองไม่ชอบ เช่นเดียวกันกับที่บุคคลอาจปฏิเสธไม่เข้าร่วมกิจกรรมที่ตนเองชอบ แต่โดยส่วนใหญ่บุคคลมักอุทิศเวลาอย่างสมัครใจกับกิจกรรมที่ตนเองแสดงออกว่าชอบ

3. ความสนใจที่ได้จากการสำรวจ (Inventoried Interest) เป็นการวัดโดยการใช้เครื่องมือวัด โดยการเปรียบเทียบความสนใจในแต่ละกิจกรรมที่แตกต่างกัน บุคคลจะชี้ให้เห็นถึงกิจกรรมต่าง ๆ ที่ใกล้เคียงกับความชอบของตนเองมากที่สุดและน้อยที่สุด ดังนั้น การตอบจึงผสมผสานทั้งความสนใจสูงและความสนใจต่ำ เครื่องมือวัดประเภทนี้สามารถออกแบบให้เหมาะกับกลุ่มอายุและอาชีพได้

การสร้างแบบวัดความสนใจ

แบบวัดความสนใจมีการสร้าง โดยใช้ 3 แนวคิด (Sax, 1997) ดังนี้

1. **แนวคิดเชิงเหตุผล (Rational approach)** เป็นแนวคิดที่เกี่ยวข้องกับการเลือกหลักการพื้นฐานสำหรับการจัดประเภทความสนใจ และเขียนข้อความซึ่งปรากฏในแต่ละความสนใจ เช่น ครูอาจสร้างแบบวัดความสนใจขึ้นมาเอง เพื่อตรวจสอบความสนใจของนักเรียน เมื่อตรวจสอบความสนใจของนักเรียนแล้วพบว่า ผู้ตอบส่วนใหญ่เลือกกิจกรรมที่เกี่ยวข้องกับวิชาศิลปะ แสดงว่า นักเรียนมีความสนใจในวิชาศิลปะ ยกตัวอย่างแบบวัดที่พัฒนาตามแนวคิดนี้ คือ Lee-Thorpe Occupational Interest Inventory

2. **แนวคิดวิธีการความสอดคล้องภายใน (Internal consistency method)** เป็นการพัฒนาข้อความจากข้อพิจารณาเชิงเหตุผล แต่มาตรวัดที่สร้างขึ้นเพื่อวัดความสนใจในแต่ละเรื่องมีความสอดคล้องภายใน โดยส่วนมากมักใช้การวิเคราะห์องค์ประกอบ เพื่อกำหนดว่ารายข้อต่าง ๆ สามารถจัดกลุ่มหรือจัดประเภท เพื่อให้ได้มิติที่มีความเหมือนกัน ข้อคำถามต่าง ๆ ที่มีความสัมพันธ์กันสูง จะวัดในองค์ประกอบหรือความสนใจเดียวกัน ข้อคำถามบางข้อที่ไม่สัมพันธ์กันหรือไม่มีน้ำหนักอยู่บนองค์ประกอบนั้น ๆ จะถูกขจัดออกไป หรืออาจมีข้อคำถามใหม่ ๆ เพิ่มเข้ามา เพื่อพัฒนาให้มาตรวัดมีความสามารถในการวัดสิ่งที่เป็นลักษณะเดียวกันได้

3. **แนวคิดเชิงประจักษ์ (Empirical approach)** มักเริ่มต้นด้วยการทดลองข้อความหลายข้อ แต่ข้อความแต่ละข้อที่ถูกเลือกไว้จำแนกความแตกต่างระหว่างบุคคลที่ประสบความสำเร็จและบุคคลทั่วไป เช่น แบบวัด Strong Interest Inventory (SII) ให้นิยามว่า บุคคลที่ประสบความสำเร็จ คือ บุคคลที่มีอายุ 25-60 ปี (หรือมากกว่า) ประกอบอาชีพเป็นเวลาอย่างน้อย 3 ปี และมีความพึงพอใจในงาน

คุณสมบัติทางการวัดของแบบวัดความสนใจ มีรายละเอียดที่สำคัญดังนี้ (Sax, 1997)

1) **ความเชื่อมั่น (Reliability)** โดยการศึกษาส่วนใหญ่ ได้รับการพิสูจน์ว่า ไม่พบค่าความเชื่อมั่นที่ยอมรับได้ในกลุ่มนักเรียนระดับประถมศึกษาและกลุ่มมัธยมศึกษา แบบวัดความสนใจมีแนวโน้มคงที่ในกลุ่มวัยรุ่นตอนปลาย และแบบแผนหรือมีการเปลี่ยนแปลงความสนใจเล็กน้อยหลังจากอายุ 25 ปี นอกจากนี้ ยังพบว่า แบบวัด SII วัดในระยะยาวมีแนวโน้มคงที่ และมีค่าความเชื่อมั่นสูง (Harmon, Hansen, Borgen, & Hammer, 1994)

2) **ความเที่ยงตรง (Validity)** คะแนนที่ได้จากแบบวัดความสนใจ ไม่มีประโยชน์ในแง่การทำนายผลสัมฤทธิ์รายวิชา เนื่องจากค่าสัมประสิทธิ์ความเที่ยงตรงมีค่าต่ำ แต่ถ้าเป็นผลสัมฤทธิ์ทางการเรียนที่เป็นฉบับ

มาตรฐานจะเชื่อถือได้มากขึ้น แบบวัดความสนใจดูเหมือนจะทำนายได้ดีในกลุ่มบุคคลที่สำเร็จการฝึกอบรมมากกว่า ผลสัมฤทธิ์ของนักเรียน นอกจากนี้ยังสามารถทำนายความตั้งใจและความพึงพอใจในงาน และมีความเที่ยงตรงเชิงสภาพกับตัวแปรทางด้านอาชีพ รวมทั้งความมั่นใจในทักษะทางอาชีพ (Flores, Spanierman, Armstrong, & Velez, 2006)

ยกตัวอย่าง แบบวัดความสนใจ

แบบวัด *Self-Directed Search* (Holland, Fritzsche, & Powell, 1994)

แบบวัด *Self-Directed Search* ด้านกิจกรรม

ข้อความ	ตัวเลือกคำตอบ	
	ชอบ	ไม่ชอบ
เรียนวิชาช่างไม้		
อ่านบทละคร		
ทำธุรกิจส่วนตัว		
ไปงานสังคม		

แบบวัด *Self-Directed Search* ด้านความสามารถ

ข้อความ	ตัวเลือกคำตอบ	
	ทำได้ดี	ทำได้ไม่ดี
ฉันรู้จักวิธีการใช้งานเครื่องคิดเลข		
ฉันเล่นละครได้		
ฉันทำงานร่วมกับผู้อื่นได้ดี		
ฉันทำรายรับรายจ่ายของตนเองได้		

แบบวัด *Self-Directed Search* ด้านอาชีพ

ข้อความ	ตัวเลือกคำตอบ	
	สนใจ	ไม่สนใจ
นักสำรวจ		
นักเทคนิคการแพทย์		
นักเขียน		
ผู้จัดการรายการโทรทัศน์		

แบบวัดค่านิยม

ค่านิยม หมายถึง ความเชื่อที่แต่ละบุคคลยึดถือว่าควรปฏิบัติอย่างไร และยอมรับว่าเป็นสิ่งที่มีความหมาย มีความสำคัญสำหรับตนเองและคนอื่น ๆ ในสังคม สิ่งใดมีคุณค่าหรือไม่มีคุณค่า โดยทั่วไปแล้ว ค่านิยมจะมีอิทธิพล ต่อพฤติกรรมของบุคคล ค่านิยมเป็นสิ่งที่เป็นมาตรฐานของแต่ละบุคคลในการตัดสินว่าสิ่งใดเป็นสิ่งที่ดี สิ่งใดเป็นสิ่งที่ไม่ดี และมักใช้ในการตัดสินพฤติกรรมของบุคคลหนึ่ง ๆ ด้วย (Rokeach, 1968) นอกจากนี้ วิลลาสลักซ์ (2551) ได้ขยายเพิ่มเติมว่า ค่านิยมทางบวกเป็นค่านิยมที่ควรค่าแก่การส่งเสริม ส่วนค่านิยมทางลบเป็นค่านิยมที่ ต้องมีการปรับปรุงแก้ไข แต่ค่านิยมทางลบหรือไม่ดีมาก ก็ควรตัดทิ้ง โดยค่านิยมทางสังคมผนวกเข้ากับการถ่ายทอดทางสังคม และคุณลักษณะภายในจิตใจของบุคคล ร่วมกันส่งผลให้เกิดเป็นค่านิยมส่วนบุคคล โดยค่านิยมที่บุคคลยึดถือนี้ จะส่งผลต่อความคิด จิตใจ และพฤติกรรมของบุคคลนั้น รวมทั้ง ความคิด จิตใจ และพฤติกรรม ก็สามารถทำให้ค่านิยมส่วนบุคคลเปลี่ยนแปลงไปได้เช่นเดียวกัน

ประเภทของค่านิยม

Rokeach (1973) ได้มีการแบ่งค่านิยมออกเป็น 2 ลักษณะ ได้แก่

1. ค่านิยมเชิงเครื่องมือ/วิธีการ (Instrumental values) เป็นค่านิยมที่แสดงถึงวิธีการปฏิบัติ เป็นวิธีการ ที่ช่วยให้ได้รับค่านิยมปลายทาง แบ่งเป็น 2 ชนิด ดังนี้ (1) ค่านิยมเชิงจริยธรรม (Moral values) มุ่งเน้นจริยธรรม และความสัมพันธ์ เช่น การให้อภัย การช่วยเหลือ ความซื่อสัตย์ เป็นต้น และ (2) ค่านิยมเชิงความสามารถ (Competence values) มุ่งเน้นความสามารถของบุคคล เช่น ความกล้าหาญ การควบคุมตนเอง การมีสติปัญญา เป็นต้น

2. ค่านิยมปลายทาง (Terminal values) เป็นค่านิยมที่เป็นจุดหมายปลายทางของชีวิต และเป็นสิ่งที่ บุคคลปรารถนา เป้าหมายที่บุคคลจะต้องได้รับในช่วงเวลาหนึ่ง ๆ และอาจมีความแตกต่างกันไปตามกลุ่มบุคคลใน วัฒนธรรมที่แตกต่างกัน แบ่งเป็น 2 ชนิด ดังนี้ (1) ค่านิยมส่วนบุคคล (Personal values) เป็นค่านิยมที่มีจุดหมาย ปลายทางที่ยึดตนเองเป็นศูนย์กลาง เช่น ความสุข การมีชีวิตสะดวกสบาย ความภาคภูมิใจในตนเอง เป็นต้น และ (2) ค่านิยมทางสังคม (Social values) เป็นค่านิยมที่มีจุดหมายปลายทางที่ยึดสังคมเป็นศูนย์กลาง เช่น ความเป็น ระเบียบของพื้นที่สาธารณะ ความเท่าเทียมกันในสังคม ความมั่นคงของประเทศชาติ เป็นต้น

Rokeach ยังได้นำแนวคิดการแบ่งประเภทค่านิยมดังกล่าว มาสร้างแบบวัดที่รู้จักกันอย่างแพร่หลายว่า “The Rokeach Value Survey” เป็นเครื่องมือวัดที่ออกแบบมาเพื่อจัดอันดับค่านิยมตามความสำคัญ จำนวน 36 ค่านิยม แบ่งเป็นค่านิยมเชิงเครื่องมือ/วิธีการ และค่านิยมปลายทาง อย่างละ 18 ค่านิยม โดยพบคุณสมบัติ ทางการวัด นั่นคือ มีค่าความเชื่อมั่นของการวัดแบบวัดซ้ำสูง และเป็นแบบวัดที่มีการแปลเป็นภาษาต่าง ๆ เช่น ญี่ปุ่น สโลวีเนีย ไทย เป็นต้น (Murphy, Jr., Mujtaba, Manyak, Sungkhawan, & Greenwood, 2010; Musil, Rus, & Musek, 2009)

การสร้างแบบวัดค่านิยม

การวัดค่านิยมเป็นการวัดประเภทหนึ่งที่มีโครงสร้าง และมักเป็นรูปแบบการรายงานตนเอง (Self-report) การวัดค่านิยมอยู่ในรูปมิติของความสำเร็จ และมีเนื้อหาการวัดที่แสดงถึงเป้าหมายหนึ่ง ๆ โดยการสร้างและพัฒนาแบบวัดค่านิยม (อรพินทร์ ชูชม, 2545) มีแนวทางในการสร้างดังนี้

1. กำหนดค่านิยมที่ต้องการวัด ผู้สร้างแบบวัดต้องกำหนดให้ชัดเจนว่าต้องการวัดค่านิยมอะไร โดยทั่วไป เครื่องมือที่ใช้วัดค่านิยมมีอยู่ 3 ประเภท ดังนี้

(1) เครื่องมือวัดค่านิยมทั่วไป มีขอบเขตกว้างและครอบคลุมสภาพชีวิตประจำวันทั่วไปของมนุษย์ เช่น ค่านิยมการใช้ชีวิต

(2) เครื่องมือวัดค่านิยมที่เกี่ยวข้องกับอาชีพและการทำงาน เป็นเครื่องมือวัดค่านิยมที่ใช้ในสภาพของการทำงานเท่านั้น เช่น ค่านิยมของนักการเมือง

(3) เครื่องมือวัดค่านิยมเฉพาะด้าน เป็นเครื่องมือวัดค่านิยมที่วัดเพียงด้านเดียว เช่น ค่านิยมด้านความรับผิดชอบต่อโรงเรียน

2. ศึกษาแนวคิดทฤษฎี รวมทั้งเอกสารที่เกี่ยวข้อง การสร้างเครื่องมือวัดค่านิยมต้องศึกษาว่าค่านิยมที่ต้องการวัด มีแนวคิดทฤษฎีเป็นอย่างไร ตลอดจนศึกษาเครื่องมือวิจัย และงานวิจัยที่เกี่ยวข้อง เพื่อใช้ในการสร้างนิยามเชิงปฏิบัติการ รวมทั้งใช้เป็นแนวทางในการเขียนข้อความหรือข้อความสำหรับการวัดค่านิยม เกณฑ์การให้คะแนน ตลอดจนการตรวจสอบคุณภาพเครื่องมือวัด

3. นิยามเชิงปฏิบัติการตัวแปรค่านิยมที่ต้องการวัด โดยให้ความหมายที่มีความชัดเจน โครงสร้างที่ต้องการวัดให้มีความชัดเจน รวมทั้งกำหนดเกี่ยวกับวิธีการวัดค่านิยมด้วย

4. กำหนดรูปแบบการเขียนข้อความวัดค่านิยม โดยมีวิธีการกำหนดรูปแบบการเขียนหลายวิธี เช่น

4.1 การเปรียบเทียบเป็นรายคู่ (Paired comparison) เป็นวิธีการจับคู่ข้อความค่านิยมแต่ละเรื่อง โดยแต่ละคู่ให้ผู้ตอบเลือกค่านิยมที่ตนเองเห็นว่าสำคัญหรือชอบอันใดอันหนึ่งเพียงค่านิยมเดียว หรืออาจให้เรียงลำดับในคู่ตามความสำคัญ สัดส่วนของจำนวนครั้งของค่านิยมที่เลือก แปลงเป็นค่ามาตรวัดที่แสดงถึงตำแหน่งความแตกต่างระหว่างชุดข้อความ แสดงให้เห็นถึงความแตกต่างในเรื่องความสำคัญของค่านิมนั้น ๆ

4.2 การเรียงลำดับหลายลำดับ (Multiple rank orders) เป็นวิธีการนำเสนอชุดข้อความวัดค่านิยมมากกว่า 2 เรื่องขึ้นไป ในแต่ละข้อ ผู้ตอบเรียงลำดับชุดข้อความของค่านิยมในแต่ละข้อ

4.3 การจัดเรียงลำดับ (Ranking) เป็นวิธีการนำเสนอชุดข้อความวัดค่านิยมทั้งหมดไว้ด้วยกัน แล้วให้ผู้ตอบจัดลำดับของค่านิยมแต่ละเรื่องตามเกณฑ์ที่กำหนดไว้ เช่น ตามความสำคัญ หรือตามความชอบ วิธีการจัดลำดับที่เชื่อถือได้มากที่สุด อยู่ตรงค่าสุดโต่งของการกระจาย และเชื่อถือได้น้อยตรงค่าตรงกลางของการกระจาย

4.4 มาตรฐานประเมินค่า (Rating scale) เป็นวิธีการนำเสนอชุดข้อความวัดค่านิยมในแต่ละข้อ แล้วให้ผู้ตอบประเมินข้อความวัดค่านิยมนั้น ๆ ตามมาตรฐานประเมินค่าที่กำหนดไว้ เช่น 4 ระดับ 5 ระดับ หรือ 6 ระดับ มักกำหนดระดับความคิดเห็น ดังนี้

- “สำคัญมาก” ถึง “ไม่สำคัญ”
- “เห็นด้วยอย่างยิ่ง” ถึง “ไม่เห็นด้วยอย่างยิ่ง”
- “จริงที่สุด” ถึง “ไม่จริงเลย”

สำหรับการวิจัยทางด้านพฤติกรรมศาสตร์ในปัจจุบัน นักวิชาการและนักวิจัยทางด้านพฤติกรรมศาสตร์หลายท่านมักทำการวัดค่านิยม โดยใช้แบบวัดค่านิยม วิธีมาตรฐานประเมินค่า 5-7 ระดับ เนื่องจากเป็นรูปแบบที่ใช้ง่าย และมีการวัดที่ชัดเจน

ยกตัวอย่าง แบบวัดค่านิยมที่ใช้มาตรฐานประเมินค่า

“แบบวัดค่านิยมกล้ายืนหยัดทำในสิ่งที่ถูกต้อง” (ทิพนพพร เกื้อเกต, 2558)

นิยามปฏิบัติการ: ค่านิยมกล้ายืนหยัดทำในสิ่งที่ถูกต้อง หมายถึง ความเชื่อความคิดเห็นที่ยึดมั่นว่าในการปฏิบัติหน้าที่จะต้องรักษาผลประโยชน์ของทางราชการเป็นหลัก พร้อมกล้าที่จะปฏิเสธไม่กระทำการใด ๆ หากขัดต่อบรรทัดฐานทางสังคม และกฎหมาย ระเบียบที่เกี่ยวข้องในการปฏิบัติหน้าที่ของตน แม้จะอยู่ในสถานการณ์บีบคั้น เช่น มีคำสั่งจากผู้บังคับบัญชา หรือโดนข่มขู่จากผู้มีอิทธิพล ก็ตามแต่ไม่ทำให้บุคคลนั้นโอนอ่อนต่อสถานการณ์ สำหรับแบบวัดได้ทำการปรับปรุงจากผู้อื่น โดยให้เนื้อหาสอดคล้องกับนิยามปฏิบัติการ เป็นแบบวัดมาตรฐานประเมินค่า 6 ระดับ ตั้งแต่ “จริงที่สุด” ถึง “ไม่จริงเลย” บุคคลที่ได้คะแนนสูง แสดงว่าเป็นผู้ที่มีค่านิยมกล้ายืนหยัดทำในสิ่งที่ถูกต้องมากกว่าบุคคลที่ได้คะแนนต่ำกว่า

ข้อความ	ตัวเลือกคำตอบ					
	จริงที่สุด	จริง	ค่อนข้างจริง	ค่อนข้างไม่จริง	ไม่จริง	ไม่จริงเลย
แม้ผู้ใต้บังคับบัญชากระทำผิด ฉันไม่กล้าลงโทษเขา						
ฉันไม่เปิดโอกาสให้ผู้ใดมาแสวงหาผลประโยชน์โดยมิชอบในหน่วยงานของฉัน						

“แบบวัดค่านิยมทางวัตถุ” (ทรรคนันท์ ชินศิริพันธ์, 2555)

นิยามปฏิบัติการ: ค่านิยมทางวัตถุ หมายถึง การที่บุคคลให้ความสำคัญกับวัตถุสิ่งของ ทรัพย์สินเงินทอง และสิ่งอำนวยความสะดวกทางกายเป็นหลัก โดยผู้ชายยกย่องเงิน วัตถุสิ่งของที่มีค่าหรือทันสมัยมากกว่าจิตใจ สำหรับแบบวัดได้ทำการปรับปรุงจากผู้อื่น โดยให้เนื้อหาสอดคล้องกับนิยามปฏิบัติการ เป็นแบบวัดมาตรฐานประเมินค่า 6 ระดับ ตั้งแต่ “จริงที่สุด” ถึง “ไม่จริงเลย” บุคคลที่ได้คะแนนสูง แสดงว่าเป็นผู้ที่มีค่านิยมทางวัตถุมากกว่าบุคคลที่ได้คะแนนต่ำกว่า

ข้อความ	ตัวเลือกคำตอบ					
	จริงที่สุด	จริง	ค่อนข้างจริง	ค่อนข้างไม่จริง	ไม่จริง	ไม่จริงเลย
ความสุขไม่สามารถซื้อหาได้ด้วยเงิน						
สำหรับฉัน การมีทรัพย์สินสมบัติจำนวนมากไม่ใช่เครื่องแสดงถึงความสำเร็จ						

แบบวัดบุคลิกภาพ

บุคลิกภาพ หมายถึง กลุ่มลักษณะนิสัยของบุคคลที่มีความเป็นเอกลักษณ์ โดยมีความคงเส้นคงวาของลักษณะนิสัยเมื่อเวลาผ่านไประยะเวลาหนึ่ง และสำหรับการวัดบุคลิกภาพ หมายถึง เครื่องมือวัดและการประเมินลักษณะนิสัย สภาวะ ค่านิยม ความสนใจ เจตคติ โลกทัศน์ การรับเอาวัฒนธรรม แบบแผนการรู้คิดและพฤติกรรม ซึ่งเกี่ยวข้องกับคุณลักษณะส่วนบุคคล ด้วยวิธีการต่าง ๆ ที่หลากหลาย (Cohen et al., 2013) แสดงให้เห็นว่าบุคลิกภาพ เป็นส่วนผสมระหว่างคุณลักษณะการรู้คิด และคุณลักษณะด้านอารมณ์ความรู้สึก รวมทั้งเป็นตัวแปรที่มีความแตกต่างระหว่างบุคคล ดังนั้น จึงเป็นการยากที่จะนิยามบุคลิกภาพได้ จึงจำเป็นต้องศึกษาแนวคิดทฤษฎีและทำการสร้างนิยามปฏิบัติการให้ชัดเจน จึงจะสามารถวัดบุคลิกภาพออกมาได้ ซึ่งจะได้กล่าวถึงในลำดับถัดไป

แนวคิดพื้นฐานเกี่ยวกับบุคลิกภาพ

พื้นฐานสำคัญที่จำเป็นต้องรู้เกี่ยวกับบุคลิกภาพ เนื่องจากมีการใช้กันบ่อย เพื่อไม่ให้เกิดความสับสน ได้แก่ ลักษณะนิสัย (Personality traits) รูปแบบบุคลิกภาพ (Personality types) และบุคลิกภาพตามสถานการณ์ (Personality states) ซึ่งมีรายละเอียดดังนี้ (Cohen et al., 2013)

1. ลักษณะนิสัย (Personality traits) เป็นคุณลักษณะที่สร้างขึ้นในตัวบุคคลที่มีความคงเส้นคงวาในแบบแผนของพฤติกรรม เป็นสิ่งที่บอกได้ว่าบุคคลหนึ่งมีความแตกต่างกับคนอื่น ๆ แม้สถานการณ์จะเปลี่ยนแปลงไป บุคคลยังคงแสดงพฤติกรรมนั้น ๆ อยู่ เช่น พื้นฐานเป็นคนที่เป็นมิตร ไม่ว่าจะอยู่ในร้านอาหาร ชั้นเรียน หรือสถานที่ต่าง ๆ ก็ยังคงเป็นมิตร ทักทายพูดคุยกับทุกคน เป็นต้น

โดยแนวคิดลักษณะนิสัย (Personality Traits) ที่เป็นที่นิยมใช้กันมาก ได้แก่

(1) แนวคิด 16 Personality Factor (16PF) ของ Raymond B. Cattell โดยเป็นเครื่องมือการประเมินบุคลิกภาพของบุคคล Cattell ได้ทำการวิเคราะห์คำศัพท์ที่เกี่ยวข้องกับบุคลิกภาพ โดยใช้การวิเคราะห์องค์ประกอบจนได้ปัจจัยบุคลิกภาพที่สำคัญ จำนวนทั้งหมด 16 องค์ประกอบ ลักษณะนิสัยทั้ง 16 ลักษณะนิสัย มีความต่อเนื่องกัน แต่ละบุคคลจะมีลักษณะนิสัยทั้ง 16 ลักษณะนิสัย ขึ้นอยู่กับระดับของลักษณะนิสัยนั้น ๆ ดังตารางที่ 5-1

ตารางที่ 5-1 บุคลิกภาพ 16 องค์ประกอบ รวมทั้งดัชนีอันดับที่สอง 4 ตัว จาก 16 PF

ชื่อองค์ประกอบ	การแปลผลคะแนนต่ำ	การแปลผลคะแนนสูง
Warmth	Reserved, detached, cool, impersonal	Warm, outgoing, likes people
Intelligence	Concrete thinking	Abstract thinking, bright
Emotional Stability	Emotionally less stable, changeable	Emotionally stable, calm, mature
Dominance	Submissive, conforming, mild	Dominant, assertive, competitive
Impulsivity	Serious, prudent, sober, taciturn	Enthusiastic, cheerful, heedless
Conformity	Expedient, disregards rules	Conforming, persevering, moralistic
Boldness	Shy, timid, restrained	Bold, uninhibited, spontaneous
Sensitivity	Tough-minded, self-reliant	Tender-minded, sensitive
Suspiciousness	Trusting, adaptable	Suspicious, hard to fool, opinionated
Imagination	Practical, conventional	Impractical, absent-minded, unconventional
Shrewdness	Forthright, genuine, unpretentious	Calculating, polished, socially alert
Insecurity	Confident, self-satisfied, secure	Self-blaming, worrying, troubled
Radicalism	Conservative, resisting change	Liberal, analytical, innovative
Self-sufficiency	Group-oriented, sociable	Resourceful, self-sufficient
Self-discipline	Undisciplined, impulsive	Compulsive, socially precise
Tension	Relaxed, tranquil, low drive	Frustrated, driven, tense
Extraversion (Q ₁)	Introversion	Extraversion
Anxiety (Q ₂)	Low anxiety	High anxiety
Tough Poise (Q ₃)	Sensitivity, emotionalism	Tough Poise
Independence (Q ₄)	Dependence	Independence

ที่มา: Cattell, R. B. (1986). *The Handbook for the 16 Personality Factor questionnaire*. Champaign, IL: Institute for Personality and Ability Testing.

(2) แนวคิด Big Five Personality ของ Costa & McCrae ได้พัฒนาแบบวัด NEO Personality Inventory Revised (NEO-PI-R) โดยเป็นแบบวัดที่นิยมใช้กันอย่างแพร่หลาย โดยเฉพาะทางการวิจัยพฤติกรรมศาสตร์ เพราะเป็นแบบวัดที่ใช้สะดวกและมีความเป็นสากล โดยบุคลิกภาพ สามารถแบ่งออกเป็นองค์ประกอบ 5 ด้าน ได้แก่ บุคลิกภาพหวั่นไหว (Neuroticism) บุคลิกภาพแบบแสดงตัว (Extraversion) บุคลิกภาพแบบเปิดรับประสบการณ์ (Openness to experience) บุคลิกภาพแบบประนีประนอม (Agreeableness) และบุคลิกภาพแบบมีจิตสำนึก (Conscientiousness) (Costa & McCrae, 1992)

(2.1) บุคลิกภาพแบบหวั่นไหว (Neuroticism) แสดงถึงความสามารถในการปรับอารมณ์และความสามารถในการทนต่อสิ่งกระตุ้นภายนอกที่ทำให้บุคลิกภาพเกิดอารมณ์ทางลบ บุคคลที่ปรับและควบคุมความหวั่นไหวได้ดี จะถูกรบกวนจากสิ่งกระตุ้นในสิ่งแวดล้อมได้ยากประกอบด้วยลักษณะนิสัยย่อย 6 ด้าน คือ ความวิตกกังวล (Worry) ความโกรธ (Angry) ความท้อแท้ (Discouragement) การคำนึงถึงแต่ตนเอง (Self-Consideration) การถูกกระตุ้น (Impulsiveness) และความเปราะบาง (Vulnerability)

(2.2) บุคลิกภาพแบบแสดงตัว (Extraversion) แสดงถึงปริมาณความสัมพันธ์ที่มีกับคนอื่นและความสามารถในการปรับตัวทางสังคม ถ้าเป็นแบบแสดงตัวจะชอบแสดงการสร้างความสัมพันธ์กับผู้อื่นและใช้เวลาอยู่กับคนอื่นมาก แต่ถ้าเป็นคนที่ไม่ชอบแสดงตัวก็จะมี การสร้างความสัมพันธ์กับผู้อื่นน้อยและใช้เวลาอยู่กับผู้อื่นน้อย ประกอบด้วยลักษณะนิสัยย่อย 6 ด้าน คือ ความอบอุ่น (Warmth) การชอบอยู่ร่วมกับผู้อื่น (Gregariousness) ความกล้าแสดงออก (Assertiveness) การชอบทำกิจกรรม (Activity) การชอบแสวงหาความตื่นเต้น (Excitement Seeking) และการมีอารมณ์ด้านบวก (Positive-Emotions)

(2.3) บุคลิกภาพแบบเปิดรับประสบการณ์ (Openness to Experience) แสดงถึงความสามารถในการปรับความคิด ความเชื่อ และการปฏิบัติตัวให้เข้ากับการเปลี่ยนแปลงทางความคิดและสถานการณ์ ประกอบด้วยลักษณะย่อย 6 ด้าน คือ การเป็นคนช่างฝัน (Fantasy) การซาบซึ้งในความงาม (Aesthetics) การเปิดเผยความรู้สึก (Feelings) การปฏิบัติตัว (Actions) การมีความคิด (Ideas) และการยอมรับค่านิยม (Values)

(2.4) บุคลิกภาพแบบประนีประนอม (Agreeableness) แสดงถึงปริมาณการรับฟังผู้อื่นและความสามารถในการยอมรับผู้อื่น บุคลิกภาพแบบเห็นด้วยกับผู้อื่นสูงจะเป็นผู้ที่ยอมรับฟังคำแนะนำจากคนอื่น ๆ ได้ง่าย ประกอบด้วยลักษณะย่อย 6 ด้าน คือ การเป็นผู้ไว้ใจผู้อื่น (Trust) ความตรงไปตรงมา (Straight forwardness) ความเอื้อเฟื้อ (Altruism) การยอมตามผู้อื่น (Compliance) ความสุภาพ (Modesty) และการมีจิตใจอ่อนโยน (Tender-Mindedness)

(2.5) บุคลิกภาพแบบมีจิตสำนึก (Conscientiousness) แสดงถึงความขยันและความมานะน้อยของจุดมุ่งหมายที่แต่ละบุคคลกำหนดไว้ รวมทั้งความสามารถในการบังคับตัวเองและการมีวินัยในตนเองให้พยายามทำตามเป้าหมายที่วางไว้ได้ครบตามที่ตั้งใจไว้ ถ้าบุคคลใดมีบุคลิกภาพแบบมีจิตสำนึกสูง แสดงว่าเป็นคนที่เป้าหมายที่ชัดเจนและเป็นคนที่มีการมีระเบียบวินัยในตนเองสูงในการทำให้เป้าหมายที่ตั้งใจไว้ประสบความสำเร็จ

ประกอบด้วยลักษณะนิสัยย่อย 6 ด้าน คือ การมีความสามารถ (Competency) ความมีระเบียบ (Order) การมีความรับผิดชอบต่อหน้าที่ (Dutifulness) ความต้องการสัมฤทธิ์ผล (Achievement Striving) ความมีวินัยในตนเอง (Self-Discipline) และความสุ่มรอบคอบ (Deliberation)

จากแนวคิดบุคลิกภาพห้าองค์ประกอบ มีนักวิชาการได้นำแบบทดสอบบุคลิกภาพห้าองค์ประกอบไปใช้ โดยเป็นแบบทดสอบที่มีคำถามให้ผู้ตอบประเมินตนเอง (Self-Report Scales) จากมาตรวัด 5 ระดับ ตั้งแต่ “เห็นด้วยอย่างยิ่ง” ถึง “ไม่เห็นด้วยอย่างยิ่ง”

ยกตัวอย่าง แบบวัดบุคลิกภาพชนิด Trait

“แบบวัดบุคลิกภาพห้าองค์ประกอบ” (Costa & McCrae, 1992)

ข้อความ	ตัวเลือกคำตอบ				
	เห็นด้วย อย่างยิ่ง	เห็นด้วย	ปานกลาง	ไม่เห็นด้วย	ไม่เห็นด้วย อย่างยิ่ง
ฉันชอบคุยกับผู้คน					
ฉันทำงานหนักเพื่อให้สำเร็จตามเป้าหมาย					
ฉันคิดว่าฉันสนใจที่จะเรียนรู้งานอดิเรกใหม่ ๆ					

2. รูปแบบบุคลิกภาพ (Personality types) เป็นกลุ่มลักษณะนิสัยที่มีความใกล้เคียงกันในแบบแผนนำไปสู่การระบุประเภทของบุคลิกภาพได้ ยกตัวอย่างเช่น การระบุว่าบุคคลนั้นมีลักษณะซิมเศร่า มีความแตกต่างกับการอธิบายบุคคลนั้น ว่ามีรูปแบบความซิมเศร่า โดยรูปแบบความซิมเศร่า ไม่ได้มาจากลักษณะนิสัยเพียงอย่างเดียว แต่ประกอบด้วยระดับกิจกรรม ความสามารถในการใช้ชีวิต และระดับของความสนใจสังคม

โดยแนวคิดรูปแบบบุคลิกภาพ (Personality Types) ที่นิยมใช้กันมาก ได้แก่

(1) Myers-Briggs Type Indicator (MBTI) มีข้อตกลงเบื้องต้นในการพัฒนาแบบวัดคือ บุคคลแสดงความชอบในวิธีการที่เขาเรียนรู้หรือตระหนักต่อสิ่งนั้น และการตัดสินใจของบุคคลเกี่ยวกับผู้คน เหตุการณ์ สถานการณ์ และความคิด โดยแบ่งออกเป็น 4 มิติ แต่ละมิติจะมีความโน้มเอียงไปใน 2 ลักษณะ ที่เป็นขั้วตรงข้ามกัน (Gregory, 1996) ได้แก่ มิติ แสดงตัว-เก็บตัว (Extraversion-Introversion) มิติ สัมผัส-หยั่งรู้ (Sensing-Intuition) มิติความคิด-ความรู้สึก (Thinking-Feeling) และมิติตัดสินใจ-ยืดหยุ่น (Judging-Perceptive)

โดยบุคคลจะมีตัวเลือกที่โน้มเอียงในแต่ละมิติ และเมื่อนำตัวเลือกของแต่ละมิติมารวมกัน ก็จะได้เป็นลักษณะบุคลิกภาพเฉพาะของแต่ละบุคคล เช่น บุคคลที่ได้ ENFP นั่นคือ มีข้อมูลโดยรวมของบุคคลที่เป็นลักษณะบุคลิกภาพ ได้แก่ มุ่งสนใจโลกภายนอก (E) สนใจความเป็นไปได้มากกว่าข้อเท็จจริงที่มีอยู่แล้ว (N) ให้ความสำคัญ

กับความรู้สึกมากกว่าหลักการและเหตุผล (F) และมีความยืดหยุ่นปรับเปลี่ยนได้ (P) นี้คือภาพรวมบุคลิกภาพของบุคคลหนึ่ง ๆ ดังนั้น การผสมผสานบุคลิกภาพในมิติต่าง ๆ จะทำให้ได้รูปแบบบุคลิกภาพ จำนวน 16 รูปภาพ

3. บุคลิกภาพตามสถานการณ์ (Personality states) สำหรับคำว่า “States” ที่นิยมใช้กันในการวัดบุคลิกภาพ หมายถึง การถ่ายโยงลักษณะนิสัยบางประการ โดยลักษณะนิสัย (Trait) มีความแตกต่างกับ States ตรงที่ลักษณะนิสัย (Trait) มีความคงเส้นคงวาของพฤติกรรม ขณะที่ States มีความคงเส้นคงวาของพฤติกรรมในลักษณะชั่วคราว (Chaplin, John, & Goldberg, 1988) ดังนั้น การวัดบุคลิกภาพตามสถานการณ์ ควรทำการวัดความทนทานของลักษณะนิสัยในสถานการณ์ที่มีความเฉพาะเจาะจง เช่น การวัดความวิตกกังวลในลักษณะ Trait เป็นการวัดคุณลักษณะของความวิตกกังวลของบุคคลทั่วไป ส่วนการวัดความวิตกกังวลในลักษณะ State เป็นการวัดความวิตกกังวล ภายใต้ประสบการณ์ความตึงเครียดในสถานการณ์ต่าง ๆ

ยกตัวอย่าง แบบวัดบุคลิกภาพชนิด States

“แบบวัดความทุกข์ทางใจ (Psychological distress)” (Kessler et al., 2003)

ข้อความ	ตัวเลือกคำตอบ				
	มากที่สุด	มาก	บางครั้ง	น้อย	น้อยที่สุด
ในช่วง 4 เดือนที่ผ่านมา ฉันรู้สึกกังวลบ่อยแค่ไหน					
ในช่วง 4 เดือนที่ผ่านมา ฉันรู้สึกเหนื่อยง่ายโดยไม่มีเหตุผลบ่อยแค่ไหน					

“แบบวัดทุนจิตวิทยาเชิงบวก” (ณัฐวุฒิ อรินทร์, 2555)

ข้อความ	ตัวเลือกคำตอบ					
	จริงที่สุด	จริง	ค่อนข้างจริง	ค่อนข้างไม่จริง	ไม่จริง	ไม่จริงเลย
ฉันเป็นคนที่มีมุ่งมั่นต่อสื่อกับอุปสรรคในการทำงาน แม้คนอื่นจะมองว่าเป็นงานที่ยาก						
ฉันเป็นคนที่เชื่อว่าโชคช่วยให้ชีวิตดีขึ้นได้จริง						
ฉันเป็นคนที่เชื่อว่าความสำเร็จต้องมาจากการลงมือทำด้วยตัวเราเอง						

วิธีการวัดบุคลิกภาพ

การวัดบุคลิกภาพ อาจใช้เครื่องมือวัดได้หลายประเภท (อรพินทร์ ชูชม, 2545; Sax, 1997) ดังนี้

1. วิธีการวัดแบบไม่ให้รู้ตัว (Unobtrusive Measures) แบบทดสอบโดยทั่วไปแล้วเป็นแบบรู้ตัว มีการอาศัยความร่วมมือจากผู้ตอบ ซึ่งรับรู้ว่าตนเองกำลังทำแบบทดสอบ แต่วิธีการวัดแบบไม่ให้รู้ตัว เป็นวิธีการวัดที่ไม่ต้องอาศัยความร่วมมือจากกลุ่มตัวอย่าง โดยกลุ่มตัวอย่างจะไม่รู้ตัวเลยว่าพฤติกรรมของตนเองกำลังถูกวัดอยู่ โดยมีวิธีการวัด ได้แก่ (Webb et al., 1981)

1.1 หลักฐานทางกายภาพ (Physical Traces) ใช้วัดลักษณะบุคลิกภาพ เช่น รูปภาพกราฟฟิกบนฝาผนังสามารถวัดเจตคติ ความกลัว แรงบันดาลใจของนักเรียนได้ บันทึกการยืมหนังสือจากห้องสมุดสามารถตรวจสอบได้โดยดูจากประเภทของหนังสือที่ยืมออกไป สามารถบอกได้ถึงการใช้ชีวิตและความสนใจ รอยน้ที่ซับกก็สะท้อนลักษณะนิสัยได้เช่นกัน แม้แต่ปริมาณขยะและของเสียสามารถใช้วัดความพึงพอใจของนักเรียนต่ออาหารกลางวันของโรงเรียนได้ อย่างไรก็ตามหลักฐานทางกายภาพเหล่านี้สามารถบอกบุคลิกภาพได้ แต่ควรตรวจสอบด้วยวิธีการแบบรู้ตัวเพิ่มเติมด้วย

1.2 บันทึกต่าง ๆ (Records) ได้แก่ เอกสาร รายงาน จดหมายต่าง ๆ สะท้อนถึงคุณลักษณะเฉพาะ และพฤติกรรมของผู้บันทึกได้ นอกจากนี้ยังมีบันทึกประจำวัน การเมือง กฎหมาย หรือจดหมายเหตุต่าง ๆ ที่มีความต่อเนื่อง มีวัตถุประสงค์ในการใช้บอกลักษณะบางอย่างได้ เช่น ข้อมูลด้านภาษีเงินได้บุคคลธรรมดาที่บุคคลแสดงเจตนาในการอุดหนุนพรรคการเมืองสามารถใช้เป็นการวัดความพึงพอใจของบุคคลที่มีพรรคการเมือง เป็นต้น

2. เทคนิคการสังเกตแบบควบคุมได้ (Controlled Observational Techniques) เกิดขึ้นเมื่อบุคคลเข้าไปอยู่ในสถานการณ์ที่เป็นโครงสร้างที่ช่วยให้เกิดการตอบสนองหรือพฤติกรรมบางอย่าง เช่น ถ้าครูต้องการดูการตอบสนองต่อความคับข้องใจของเด็ก อาจต้องสร้างสถานการณ์ตั้งเครียดบางอย่าง เพื่อสังเกตการตอบสนองของเด็ก โดยมีวิธีการวัด ได้แก่

2.1 การสัมภาษณ์ (Interview) ประกอบด้วย การมีปฏิสัมพันธ์ทางการสนทนาพูดคุยระหว่างผู้สัมภาษณ์และผู้ให้สัมภาษณ์ ลักษณะคำถามอาจมีโครงสร้างหรือไม่มีโครงสร้างก็ได้ แต่ต้องเป็นไปตามความต้องการของผู้สัมภาษณ์ คำถามที่มีโครงสร้างอาจจัดประเภทและสรุปคำตอบได้ง่าย แต่ถูกจำกัดในสิ่งที่กลุ่มตัวอย่างต้องการจะพูด แต่คำถามแบบไม่มีโครงสร้างอาจมีอิสระในการแสดงความคิดเห็นมากกว่า แต่จัดประเภทคำตอบได้ค่อนข้างยาก ดังนั้น การสัมภาษณ์จึงควรมีทั้งสองประเภท

2.2 การสร้างสถานการณ์ (Situational Tests) ออกแบบมาเพื่อสังเกตพฤติกรรม บุคลิกภาพที่แสดงออกมามีในสถานการณ์ที่เสมือนจริงที่ผู้สร้างกำหนดขึ้นมา การสร้างสถานการณ์ดังกล่าวมีลักษณะซ่อนเร้น อาจไม่รู้ตัวว่ากำลังถูกทดสอบอยู่ บางครั้งสถานการณ์ที่จัดเตรียมไว้อาจไม่แสดงออกมาตามที่คาดการณ์ไว้ วิธีนี้มีการเตรียมตัวมาก ยากและเสียเวลา และในบางสถานการณ์ต้องคำนึงถึงจริยธรรมด้วย

3. การรายงานตนเอง (Self-Reports) เป็นเครื่องมือวัดแบบรายงานตนเองที่สร้างขึ้นมา มักประกอบด้วย ข้อความที่เป็นข้อคำถาม ที่มีรูปแบบการตอบให้เลือก อาจมีลักษณะเป็น “ใช่” “ไม่ใช่” หรือ “ไม่แน่ใจ” หรืออาจ อยู่ในรูปแบบการตอบที่เป็นมาตรวัดลิเคิร์ทหรือมาตรวัดประเมินค่า โดยในปัจจุบัน การวัดทางด้านพฤติกรรมศาสตร์ มักใช้การรายงานตนเองในการวัดบุคลิกภาพ มีรูปแบบข้อคำถามและคำตอบ ดังนี้

3.1 แบบข้อความเดียวและมีคำตอบให้เลือก เป็นการรายงานบุคลิกภาพตนเองที่อาศัยข้อคำถามที่เป็นประโยคสั้น กระชับ และได้ใจความ โดยมีคำตอบให้เลือกตอบหลายรูปแบบ ดังนี้

- 2 ตัวเลือก เช่น ใช่-ไม่ใช่, จริง-ไม่จริง, ชอบ-ไม่ชอบ
- 3 ตัวเลือก เช่น ใช่-ไม่แน่ใจ-ไม่ใช่, จริง-ไม่แน่ใจ-ไม่จริง
- ตัวเลือกมีหลายระดับ ตามมาตรประเมินค่า อาจกำหนดเป็น 3-7 ระดับ

3.2 แบบข้อความคู่และให้เลือกตอบข้อความใดข้อความหนึ่ง เป็นการรายงานบุคลิกภาพตนเอง โดยอาศัยข้อความที่อธิบายคุณลักษณะบุคลิกภาพที่แตกต่างกันมาจับคู่กัน เพื่อให้เลือกตอบข้อความใดข้อความหนึ่ง

3.3 แบบสถานการณ์ เป็นการรายงานบุคลิกภาพตนเอง โดยให้สถานการณ์จำลองมาเป็นสิ่งเร้า อาจประกอบด้วยภาพหรือเหตุการณ์เรื่องราวสมมติที่เขียนขึ้น แล้วให้ผู้ตอบสมมติตนเองว่า ถ้าตนเองอยู่ในเหตุการณ์นั้น จะมีการตอบสนองต่อเหตุการณ์อย่างไร โดยคำตอบที่ให้เลือก มีดังนี้

- ตัวเลือกเป็นคุณลักษณะหรือพฤติกรรมที่ตอบสนองสถานการณ์ ซึ่งมีหลายตัวเลือก โดยให้เลือกข้อความคุณลักษณะหรือพฤติกรรมที่อธิบายถึงตนเองมากที่สุด
- 2 ตัวเลือก เช่น ใช่-ไม่ใช่, มี-ไม่มี, จริง-ไม่จริง
- ตัวเลือกมีหลายระดับ ตามมาตรประเมินค่า อาจกำหนดเป็น 3-7 ระดับ

4. เทคนิคการฉายออก (Projective Technique) วิธีการนี้เกี่ยวข้องกับการนำสิ่งเร้าหรืองานที่ไม่มีโครงสร้างและคลุมเครือ และให้ผู้ตอบแปลความหมายสิ่งเร้า นั้น หรือตอบสนองต่อสิ่งเร้าตามประสบการณ์ เป็นการวัดพลังจิตของบุคลิกภาพและแรงจูงใจ ส่วนใหญ่วิธีนี้เอามาใช้วินิจฉัยปัญหาทางด้านบุคลิกภาพ โดยมีวิธีการฉายภาพ (Lindzey, 1959 cited as Sax, 1997) ได้แก่

4.1 เทคนิคการเชื่อมโยง (Association Techniques) เป็นเทคนิคการวัดบุคลิกภาพที่ต้องการให้กลุ่มตัวอย่างตอบสนองอย่างรวดเร็วเท่าที่จะเป็นไปได้ต่อสิ่งเร้าประเภทคำหรือรูปภาพ โดยให้คำที่อาจเป็นกลางหรือ ก่อให้เกิดอารมณ์ แล้วให้ผู้ตอบตอบสนองถึงคำแรกที่เกิดขึ้นในใจอย่างรวดเร็ว มีการบันทึกเวลาในการตอบสนอง และสิ่งที่เกิดจากการตอบสนอง ตัวอย่างแบบทดสอบ คือ The Rorschach Inkblot Test

4.2 เทคนิคการสร้าง (Construction Techniques) เป็นเทคนิคการวัดบุคลิกภาพที่ให้กลุ่มตัวอย่างตอบสนองต่อสิ่งเร้าที่เป็นรูปภาพที่มีวิวิวิทัศน์ คน และสถานการณ์ทางสังคม และให้กลุ่มตัวอย่างเล่าเรื่องราวจากภาพนั้น สิ่งสำคัญจากเทคนิคการวัดนี้ คือ ใจความเนื้อเรื่อง (theme) ในการตอบ และรูปแบบการตอบ ไม่มีการบันทึกเวลาเหมือนกับเทคนิคการเชื่อมโยง ตัวอย่างแบบทดสอบคือ The Thematic Apperception Test (TAT)

4.3 เทคนิคความสมบูรณ์ (Completion Techniques) เป็นเทคนิคการวัดบุคลิกภาพที่ให้กลุ่มตัวอย่างตอบสนองต่อสิ่งเร้าที่เป็นข้อความ เรื่องราว ภาพ หรือสิ่งเร้าอื่น ๆ ที่ไม่สมบูรณ์ เหมือนมีสิ่งขาดหายไปจากสิ่งเร้านั้น และให้กลุ่มตัวอย่างเติมให้มีความสมบูรณ์ ตัวอย่างแบบทดสอบคือ Sentence-Completion Tests (SCT)

4.4 เทคนิคการเลือกหรือเรียงลำดับ (Choice or Ordering Techniques) เป็นเทคนิคการวัดบุคลิกภาพที่ให้กลุ่มตัวอย่างเลือกตัวเลือกที่กำหนดให้ หรือจัดลำดับจากตัวเลือกที่กำหนดให้ แล้วให้อธิบายว่าทำไมถึงเลือกตัวเลือกนั้น มักไม่นิยมใช้เทคนิคนี้แล้ว ตัวอย่างแบบทดสอบนี้ คือ The Szondi Test

4.5 เทคนิคการแสดงออก (Expressive Techniques) เป็นเทคนิคการวัดบุคลิกภาพที่ให้ผู้ตอบแสดงบทบาทในการวาดรูป ระบายสี ละคร หรือการเล่น ซึ่งสิ่งที่แสดงออกมาเป็นสิ่งที่สะท้อนการแสดงออกทางด้านบุคลิกภาพ สามารถให้คะแนนได้หลายมิติ เช่น ความสมบูรณ์ การขาดหายไปบางส่วน การไม่ได้สัดส่วน ตัวอย่างแบบทดสอบนี้ คือ Draw A Person Test

สรุป

การวัดทางด้านความรู้สึกถือเป็นกระบวนการหนึ่งที่สำคัญในการพัฒนาเครื่องมือวิจัย เพราะเป็นการวัดคุณลักษณะภายในจิตใจของบุคคล ซึ่งมีความเป็นนามธรรม ไม่สามารถสังเกตหรือสัมผัสได้โดยตรง จำเป็นต้องมีการนิยามให้ชัดเจน โดยมีวิธีการวัดหรือเครื่องมือวัดมากมาย โดยส่วนใหญ่การวัดทางด้านความรู้สึกมักใช้แบบรายงานตนเอง (Self-report) ให้บุคคลรายงานตนเองเกี่ยวกับความสนใจ ค่านิยม รวมทั้งบุคลิกภาพ โดยการวัดทางด้านความรู้สึกดังกล่าว เป็นแนวทางในการสร้างและพัฒนาเครื่องมือทางด้านพฤติกรรมศาสตร์ให้มีความเหมาะสมและสามารถใช้วัดคุณลักษณะทางจิตวิทยาของแต่ละบุคคลได้

เอกสารอ้างอิง

- ณัฐวุฒิ อรินทร์. (2555). การศึกษาปัจจัยเชิงเหตุและผลของทุนทางจิตวิทยาเชิงบวกที่มีต่อสุขภาวะทางจิตและพฤติกรรมการปฏิบัติงานภายใต้สถานการณ์ความไม่สงบของบุคลากรสาธารณสุขในจังหวัดชายแดนใต้ (วิทยานิพนธ์ปริญญาคุชฎีบัณฑิต. กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
- ทรงศนันท์ ชินศิริพันธุ์. (2555). ปัจจัยทางจิตสังคมที่เกี่ยวข้องกับการปฏิบัติตามหลักปรัชญาเศรษฐกิจพอเพียง ด้านความประมาทของนักเรียนมัธยมศึกษาปีที่ 3 ในโรงเรียนสถานศึกษาพอเพียงแบบอย่าง (วิทยานิพนธ์ปริญญามหาบัณฑิต. กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.

- ทินนพพร เกื้อเกตุ. (2558). *ปัจจัยทางจิตสังคมที่เกี่ยวข้องกับพฤติกรรมการทำงานตามหลักธรรมาภิบาล ด้านความโปร่งใสของเจ้าหน้าที่พัสดุสังกัดองค์การบริหารส่วนตำบล ในเขตภาคกลาง* (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
- วิลาสลักษณ์ ชวัลลสี. (2551). *กรณีศึกษาบุคคลตัวอย่างและเครื่องชี้วัดเพื่อนำไปสู่การพัฒนาพฤติกรรมข้าราชการใหม่*. กรุงเทพฯ: สำนักงานคณะกรรมการข้าราชการพลเรือน.
- อรพินทร์ ชูชม. (2545). *เอกสารคำสอนวิชา วป 502 การสร้างและพัฒนาเครื่องมือวัดทางพฤติกรรมศาสตร์*. กรุงเทพฯ: สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- Anderson, L. W. (1981). *Assessing affective characteristics in schools*. Boston: Allyn & Bacon.
- Bloom, B. S. (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain*. New York: David McKay Co., Inc.
- Cattell, R. B. (1986). *The Handbook for the 16 Personality Factor questionnaire*. Champaign, IL: Institute for Personality and Ability Testing.
- Chaplin, W. F., John, O. P., & Goldberg, L. R. (1988). Conceptions of state and traits: Dimensional attributes with ideals as prototypes. *Journal of Personality and Social Psychology*, 54, 541-577.
- Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2013). *Psychological testing and assessment: An introduction to tests and measurement* (8th ed.). New York: McGraw-Hill.
- Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- Dolan, L. (1983). Affective characteristics of the adult learner: A study of mastery based instruction. *Community Junior College Research Quarterly of Research and Practice*, 7(4), 367-378.
- Flores, L. Y., Spanierman, L. B., Armstrong, P. I., & Velez, A. D. (2006). Validity of the strong interest inventory and skills confidence inventory with Mexican American high school students. *Journal of Career Assessment*, 14(2), 183-202.
- Gregory, R. J. (1996). *Psychological testing* (2nd ed.). Boston: Allyn and Bacon.
- Gronlund, N. E., & Brookhart, S. M. (2009). *Writing instructional objectives* (8th ed.). Upper Saddle River, NJ: Pearson Education.

- Harmon, L. W., Hansen, J. C., Borgen, F. H., & Hammer, A. C. (1994). *Strong interest inventory applications and technical guide*. Palo Alto, CA: Consulting Psychological Press.
- Holland, J. L., Fritzsche, B. A., & Powell, A. B. (1994). *The Self-Directed Search technical manual*. Odessa, FL: Psychological Assessment Resources.
- Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., Howes, M. J., Normand, S. T., Manderscheid, R. W., Walters, E. E., & Zaslavsky, A. M. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry*, *60*(2), 184-189.
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1973). *Taxonomy of educational objectives, the classification of educational goals, Handbook II: Affective domain*. New York: David McKay Co., Inc.
- McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument development in the affective domain: School and corporate applications* (3rd ed.). New York: Springer Science+Business Media New.
- Murphy, Jr., E. F., Mujtaba, B. G., Manyak, T., Sungkhawan, J., & Greenwood, R. (2010). Generational value differences of baby boomers in Thailand. *Asia Pacific Business Review*, *16*(4), 545-566.
- Musil, B., Rus, V. S., & Musek, J. (2009). The Rokeach value survey in comparative study of Japanese and Slovenian students: Toward the underlying structure. *Studia Psychologica*, *51*(1), 53-68.
- Oakland, T. (1997). Affective assessment. *Psicologia escolar e Educacional*, *1*, <https://doi.org/10.1590/S1413-85571997000100002>
- Opong, C. A. (2014). Cognitive and affective characteristics of history students of the University of Cape Coast. *International Journal of Scientific and Research Publications*, *4*(10), 1-7.
- Rokeach, M. (1968). *Beliefs, attitudes and values: A theory of organization and change*. San Francisco: Jossey-Bass.
- Rokeach, M. (1973). *The nature of human values*. New York: The Free Press.
- Sax, G. (1997). *Principles of educational and psychological measurement and evaluation* (4th ed.). Belmont, CA: Wadsworth Publishing Company.

- Staats, A. W. (2003). A psychological behaviorism theory of personality. In T. Millon, & M. J. Lerner (Eds.), *Handbook of Psychology, vol. 5: Personality and social psychology* (pp. 135-158). Hoboken, NJ: John Wiley & Sons.
- Thien, L. M., Darmawan, I. G. N., & Ong, M. Y. (2015). Affective characteristics and mathematics performance in Indonesia, Malaysia, and Thailand: What can PISA 2012 data tell us? *Large-scale Assessment in Education, 3, 3*.
- Webb, E. T., Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. B. (1981). *Nonreactive measures in the social sciences* (2nd ed.). Boston: Houghton Mifflin.



บทที่ 6

การวัดด้านพฤติกรรม

บทที่ 6 เป็นการนำเสนอการวัดด้านพฤติกรรม โดยผู้เขียนต้องการให้ผู้อ่านได้ทราบและทำความเข้าใจเกี่ยวกับการวัดด้านพฤติกรรม ความเป็นมาของการวัดทางด้านพฤติกรรม แนวคิดเบื้องต้นของการวัดด้านพฤติกรรม สำหรับการวัดทางด้านพฤติกรรม มักเริ่มต้นจากการวัดและประเมินผลการเรียนรู้ทางด้านทักษะพิสัย ในปัจจุบันมีการสร้างวิธีการวัดทางด้านพฤติกรรมที่มีมากมายหลากหลายรูปแบบ โดยบทนี้ จะช่วยให้ผู้วิจัยและนักวิชาการสามารถสร้างเครื่องมือวัดด้านพฤติกรรมได้อย่างถูกต้อง เพื่อช่วยให้กระบวนการพัฒนาเครื่องมือวิจัยทางพฤติกรรมศาสตร์มีความเหมาะสมและสมบูรณ์มากยิ่งขึ้น

ความเป็นมาของการวัดทางด้านพฤติกรรม

จุดเริ่มต้นของการวัดทางด้านพฤติกรรมช่วงแรก ๆ เป็นการวัดที่มีความเกี่ยวข้องกับการปรับเปลี่ยนพฤติกรรม (Behavior Modification) และพฤติกรรมบำบัด (Behavioral Therapy) ในบริบทด้านการรักษาเป็นส่วนใหญ่ ซึ่งการวัดทางด้านพฤติกรรมนี้เป็นกระบวนการวินิจฉัยและประเมินปัญหาทางด้านพฤติกรรมและปัจจัยควบคุม (เช่น สภาพแวดล้อม ชีววิทยา หรือส่วนบุคคล) ซึ่งมีวัตถุประสงค์เพื่อจัดกระทำตัวแปรที่เกี่ยวข้องกับชุดเป้าหมาย และทำการประเมินและวัดการเปลี่ยนแปลงพฤติกรรมเหล่านั้น นั่นคือ การวัดและประเมินด้านพฤติกรรมถือเป็นรูปแบบทางทฤษฎีของการประเมินทางจิตวิทยา ความสำคัญของการวัด/ประเมินทางพฤติกรรมไม่ได้มีไว้สำหรับการบำบัด/ปรับเปลี่ยนพฤติกรรม แต่ยังไว้ใช้จำแนกอาการทางจิต และใช้ประเมินทางจิตวิทยาด้วย ดังนั้นจุดเริ่มต้นของการวัด/ประเมินทางด้านพฤติกรรม เกิดจากปัญหาทางด้านพฤติกรรมที่นำไปสู่บริบทของการรักษาทางด้านคลินิก จากนั้นจึงรวมแนวคิดเข้าด้วยกันเป็นแนวคิดจิตวิทยาคลินิก (Fernández-Ballesteros, 2001) ซึ่งในปัจจุบัน การประเมินทางด้านพฤติกรรมมักถูกนำมาใช้เป็นส่วนหนึ่งของการประเมินทางคลินิกสำหรับเด็กและวัยรุ่น (Eckert, & Lovett, 2013) และในช่วงที่ผ่านมา นักวิชาการและนักจิตวิทยาให้ความสำคัญกับการวัดทางด้านพฤติกรรมมากขึ้น เช่น การประเมินการเรียนรู้ โดยการประเมินการเรียนรู้ เริ่มต้นมุ่งเน้นการเรียนรู้ด้านการรู้คิด (Cognitive) ตามแนวคิด Bloom (1956) ต่อมาจึงให้ความสนใจในการวัดความรู้สึกหรือจิตพิสัย (Affective) (Krathwohl, Bloom, & Masia, 1973) จนกระทั่ง เริ่มให้ความสนใจในด้านทักษะพิสัย (Simpson, 1972) ซึ่งเป็นการเรียนรู้เกี่ยวกับทักษะ การปฏิบัติ ซึ่งเป็นการแสดงพฤติกรรมหรือการกระทำ โดยการศึกษาส่วนใหญ่ให้ความสนใจโดยมุ่งการวัดทางด้านการรู้คิดและความรู้สึก ดังนั้น จึงควรมุ่งประเมินการเรียนรู้ด้านทักษะพิสัยด้วย เป็นการวัดทักษะการเคลื่อนไหว การปฏิบัติงาน โดยเริ่มจากทักษะง่าย ๆ ไปจนถึงทักษะที่ต้องใช้

ความชำนาญ(โชติกา ภาชีผล, 2559) ดังนั้น ในบทของการวัดทางด้านพฤติกรรม ผู้เขียนจึงขอเรียบเรียงการวัดทางด้านพฤติกรรมศาสตร์เกี่ยวกับวิธีการวัดพฤติกรรมในรูปแบบต่าง ๆ ที่มีความหลากหลาย เพื่อให้เกิดความเข้าใจในการวัดพฤติกรรมศาสตร์ โดยมุ่งเน้นทางด้านพฤติกรรม

แนวคิดเบื้องต้นของการวัดทางด้านพฤติกรรม

Heiby and Haynes (2003) กล่าวว่า การวัดประเมินทางด้านพฤติกรรม เป็นแนวคิดการวัดประเมินทางด้านจิตวิทยาที่มีหลักการทางวิทยาศาสตร์ มีจุดเน้นที่สำคัญอยู่ 4 ประการ ดังนี้

1. ใช้การวัดอนุมานเป็นอย่างน้อยที่สุด การวัดทางด้านพฤติกรรมเน้นการอนุมานจากข้อมูลการวัด ยกตัวอย่างเช่น ถ้าเด็กวัยรุ่นหญิงถูกส่งมาเพื่อประเมิน บอกกับผู้ประเมินว่า ไม่สามารถนอนหลับตอนกลางคืนได้ ในช่วงเดือนที่ผ่านมา อนุมานได้ว่า วัยรุ่นหญิงมีประสบการณ์ปัญหาด้านการนอนหลับ ซึ่งต้องการเพียงแค่ปริมาณการอนุมานเป็นอย่างน้อย การอนุมานว่าวัยรุ่นหญิงมีความกังวลบางสิ่งบางอย่าง เพียงแค่ความคิดเกี่ยวกับการนอนหลับ ทำให้เกิดความเสี่ยงในการอนุมาน การอนุมานว่าวัยรุ่นหญิงได้รับความบอบช้ำทางจิตใจอย่างรุนแรง เมื่อเดือนที่แล้ว ก็ถือว่าเป็นการตีความที่มีความเสี่ยง แนวคิดการวัดทางด้านพฤติกรรมจึงต้องการการตีความที่ต้องการเพียงแค่ปริมาณการอนุมานเป็นอย่างน้อย ความต้องการนำไปสู่เครื่องมือการประเมิน เช่น ข้อมูลจากมาตรวัดให้คะแนนด้านพฤติกรรม จากแม่เป็นผู้ประเมิน จะให้ข้อมูลที่ดีกว่าข้อมูลที่ได้จากการหยดหมึกด้วยแบบทดสอบ Rorschach inkblot

2. ตรวจสอบความถูกต้องด้วยวิธีการที่เหมาะสมกับบริบทของการประเมิน ยกตัวอย่างเช่น ถ้านักจิตวิทยาต้องการประเมินความวิตกกังวลภายในบริบทห้องเรียน การประเมินให้คะแนนโดยครู ย่อมได้ผลลัพธ์ที่ดีกว่าให้ผู้ปกครองเป็นผู้ประเมิน แต่ถ้าต้องการประเมินความวิตกกังวลภายในบ้าน ควรให้ผู้ปกครองเป็นผู้ประเมิน ซึ่งบริบทภายในบ้าน แต่ให้ครูประเมิน จึงมีความถูกต้องน้อยกว่าบริบทภายในห้องเรียน

3. การประเมินความสัมพันธ์เชิงหน้าที่ นั่นคือ ความสัมพันธ์เชิงเหตุผลระหว่างสภาพแวดล้อมและพฤติกรรม การวัดทางด้านพฤติกรรมใช้หลักการเรียนรู้ และขึ้นอยู่กับหลักฐานสำหรับการอธิบายว่า มีความต้องการการอนุมานเป็นอย่างน้อยสำหรับหลักฐานนั้น ๆ ยกตัวอย่างเช่น การประเมินทางคลินิกของเด็กก่อนวัยเรียนที่ถูกส่งมาด้วยพฤติกรรมที่เป็นปัญหา อาจอธิบายพฤติกรรมเช่นนี้ด้วยเงื่อนไขทางสรีรจิตวิทยา เช่น โรค ADHD แต่สำหรับมุมมองการวัดทางด้านพฤติกรรม มองว่าการอธิบายด้วยสรีรจิตวิทยายังไม่สมบูรณ์ และแนวคิดการวัดทางด้านพฤติกรรมของเด็กจะค้นหาหลักฐานจากปัจจัยสภาพแวดล้อม (เช่น ความสนใจของกลุ่มเพื่อน) อาจเป็นสิ่งที่ทำให้เกิดพฤติกรรมที่เป็นปัญหาได้

4. การวัดตามบริบท แหล่งข้อมูล เครื่องมือ และช่วงเวลา การวัดทางด้านพฤติกรรมมีแนวคิดที่แตกต่างกันในด้านการแปลความหมาย ยกตัวอย่างเช่น ถ้าแหล่งข้อมูล 2 แหล่ง (บิดา และมารดา) ก็จะให้ข้อมูลที่แตกต่างกัน ซึ่งการวัดทางด้านพฤติกรรมจะไม่ได้สรุปลงไปทันทีว่าแหล่งข้อมูลใด ๆ หรือแหล่งข้อมูลทั้งสองไม่ถูกต้อง แต่อาจ

เป็นไปได้ว่าเด็กแสดงพฤติกรรมแตกต่างกันตอนที่อยู่กับมารดา หรือถ้าการประเมินทักษะทางวิชาการของเด็กมีข้อมูลแตกต่างกันในแต่ละสัปดาห์ แนวคิดการวัดทางด้านพฤติกรรมจะสรุปได้ว่าระดับทักษะทางวิชาการของเด็กเปลี่ยนแปลงไป

ดังนั้น สำหรับการวัดทางด้านพฤติกรรม จึงควรให้ความสนใจกับการวัดเกี่ยวกับพฤติกรรมโดยตรง และปัจจัยสภาพแวดล้อมที่เกี่ยวข้อง รวมทั้งการอนุมานเป็นอย่างน้อยจากข้อมูลที่ได้จากการวัด

Eckert and Lovett (2013) ได้กล่าวถึงการวัดทางด้านพฤติกรรม ว่าวิธีการวัดทางด้านพฤติกรรมได้รับอิทธิพลมาจากแนวคิดความสัมพันธ์เชิงเหตุผลและการทำหน้าที่ (Causation and Functional Relationships) และหลักการทางเครื่องมือวัดทางจิตวิทยา (Psychometric Principles) ซึ่งมีรายละเอียดดังนี้

1. ความสัมพันธ์เชิงเหตุผลและการทำหน้าที่ โดยคุณลักษณะที่สำคัญของการวัดทางด้านพฤติกรรมเน้นปัจจัยด้านสภาพแวดล้อมของพฤติกรรมตามช่วงเวลาและสภาพแวดล้อม นั่นคือ พฤติกรรมต่าง ๆ และเหตุการณ์แวดล้อม จะถูกวัดซ้ำในความพยายามที่ได้รับผลลัพธ์ที่แน่นอนและอนุมานเป็นอย่างน้อย วิธีการวัดทางด้านพฤติกรรมมีตั้งแต่วิธีการวัดโดยตรงไปจนถึงวิธีการวัดโดยอ้อม ซึ่งวิธีการวัดทางด้านพฤติกรรมโดยอ้อมมักขึ้นอยู่กับกรรายงานพฤติกรรมและเหตุการณ์รอบ ๆ ที่เกิดขึ้นเพื่อสร้างสมมติฐานที่เกี่ยวข้องกับความสัมพันธ์ระหว่างพฤติกรรมและสภาพแวดล้อม ส่วนวิธีการวัดทางด้านพฤติกรรมโดยตรง เช่น การสังเกตโดยตรง การประเมินการปฏิบัติ สะท้อนถึงความสัมพันธ์เชิงสหสัมพันธ์ เมื่อได้ตรวจสอบพร้อมกับเหตุการณ์รอบตัว

2. หลักการทางเครื่องมือวัดทางจิตวิทยา แนวคิดเบื้องต้นของเครื่องมือวัดทางจิตวิทยา (เช่น ความเชื่อมั่น ความเที่ยงตรง) ควรประยุกต์ในวิธีการวัดทางด้านพฤติกรรมด้วย ยกตัวอย่างเช่น มาตรฐานค่าเชิงพฤติกรรม (Behavior rating scales) และการตรวจสอบรายการ (Checklist) ในเครื่องมือวัดทางด้านพฤติกรรม ควรมีการตรวจสอบหลักฐานทั้งความเชื่อมั่นและความเที่ยงตรง เพื่อให้เกิดเครื่องมือวัดทางด้านพฤติกรรมที่เป็นมาตรฐาน แต่อย่างไรก็ตาม มีวิธีการวัดทางด้านพฤติกรรมอื่น ๆ เช่น การสังเกตโดยตรง ที่ใช้แนวคิดเครื่องมือวัดทางจิตวิทยาที่แตกต่างกัน แม้ว่าคุณภาพของเครื่องมือวัดทางจิตวิทยาของการสังเกตจะถูกบรรจุในคู่มือการตรวจสอบคุณภาพเครื่องมือวัดทางจิตวิทยาอื่น ๆ สามารถทำได้ เช่น การเห็นพ้องต้องกันระหว่างผู้สังเกต (Interobserver agreement) เป็นต้น

นอกจากนี้ แนวคิดเบื้องต้นเกี่ยวกับการวัดทางด้านพฤติกรรม ยังได้กล่าวถึงแนวคิดการวัดและประเมินผลการเรียนด้านทักษะพิสัย โชติกา ภาชีผล (2559) ได้ให้รายละเอียดเกี่ยวกับทักษะพิสัยว่าเป็นพฤติกรรมทางการเคลื่อนไหว เช่น วิ่ง กระโดด ขับรถ เป็นต้น ซึ่งได้แบ่งระดับของทักษะพิสัย โดยเรียงจากระดับต่ำสุดไปจนถึงระดับสูงสุด 6 ระดับ (Harrow, 1972) มีรายละเอียดดังนี้

1. การเคลื่อนไหวในการปฏิริยาสะท้อนกลับ (Reflex movements) ได้แก่

1.1 ปฏิริยาสะท้อนกลับที่สั่งมาจากเส้นประสาทไขสันหลังส่วนหนึ่ง (Segmental reflexes) เช่น การเคลื่อนไหวแขนหรือขา เป็นต้น

1.2 ปฏิกริยาสะท้อนกลับที่สั่งมาจากเส้นประสาทไขสันหลังมากกว่าหนึ่งส่วน (Intersegmental reflexes) เช่น การเคลื่อนไหวแขนและขาในเวลาที่มีการเดินหรือวิ่ง เป็นต้น

1.3 ปฏิกริยาสะท้อนกลับที่สั่งมาจากเส้นประสาทไขสันหลังและสมองร่วมกัน (Segmental reflexes) เช่น การทรงตัวของร่างกายให้อยู่ในสภาพที่สมดุลในขณะที่มีการเคลื่อนไหว เป็นต้น

2. การเคลื่อนไหวในขั้นพื้นฐาน (Basic-fundamental movements) ได้แก่

2.1 การเคลื่อนไหวจากจุดหนึ่งไปยังอีกจุดหนึ่ง (Locomotor movements) เช่น เดิน วิ่ง กระโดด เป็นต้น

2.2 การเคลื่อนไหวอยู่กับที่ (Non-locomotor movements) เช่น การเคลื่อนไหวของนิ้วมือหรือนิ้วเท้า เป็นต้น

2.3 การเคลื่อนไหวเชิงบังคับ โดยปฏิกริยาสะท้อนกลับหลายอย่างร่วมกัน (Manipulative movement) เช่น เล่นเปียโน พิมพ์ดีด เป็นต้น

3. ความสามารถในการรับรู้ (Perceptual abilities) ได้แก่

3.1 การรับรู้ความแตกต่างด้วยการเคลื่อนไหวของร่างกาย (Kinesthetic discrimination) เช่น การกำมือ การกระพริบตา การงอเข่า เป็นต้น

3.2 การรับรู้ความแตกต่างด้วยการมองเห็น (Visual discrimination) เช่น การมองเห็นความแตกต่างของวัตถุที่สังเกตได้ เป็นต้น

3.3 การรับรู้ความแตกต่างด้วยการได้ยิน (Auditory discrimination) เช่น การจำแนกความแตกต่างของระดับเสียง เป็นต้น

3.4 การรับรู้ความแตกต่างด้วยการสัมผัส (Tactile discrimination) เช่น การจำแนกลักษณะของวัตถุเกี่ยวกับพื้นผิว ว่าหยาบ เรียบ แข็ง หรืออ่อน เป็นต้น

3.5 ความสามารถในการใช้ประสาทรับรู้ร่วมกัน (Coordinated abilities) เช่น การใช้ประสาทหลายส่วนร่วมมือกัน เช่น ค้นหาวัตถุสิ่งหนึ่ง ประกอบด้วย การเดินหา สายตาสอดส่อง ไข่มือหยิบค้น เป็นต้น

4. สมรรถภาพทางกาย (Physical abilities) ได้แก่

4.1 ความทนทาน (Endurance) เช่น ความทนทานของร่างกายในการแข่งไตรกีฬา เป็นต้น

4.2 ความแข็งแรง (Strength) เช่น ความแข็งแรงของแขนในการยกน้ำหนัก ทุ่มน้ำหนัก เป็นต้น

4.3 ความยืดหยุ่น (Flexibility) เช่น ความยืดหยุ่นของกล้ามเนื้อแขนขาในการเต้นลีลาศ เป็นต้น

4.4 ความคล่องตัว (Agility) เช่น ความคล่องตัวในการเปลี่ยนทิศทางการเคลื่อนไหว เป็นต้น

5. การเคลื่อนไหวที่ต้องอาศัยทักษะ (Skilled movements) ได้แก่

5.1 ทักษะการปรับตัวทางการเคลื่อนไหวแบบง่าย (Simple adaptive skill) เช่น ทักษะการเลี้ยวไม้ เป็นต้น

5.2 ทักษะการปรับตัวทางการเคลื่อนไหวที่นำไปพร้อมกัน (Compound adaptive skill) เช่น ทักษะในการเล่นบาสเก็ตบอล เป็นต้น

5.3 ทักษะการปรับตัวทางการเคลื่อนไหวที่มีลักษณะซับซ้อนมาก (Complex adaptive skill) เช่น ทักษะการเล่นสกี ทักษะการเล่นไอซ์สเก็ต เป็นต้น

6. การสื่อสารที่ต้องใช้ทักษะในการแสดงออก (Non-discursive communication) ได้แก่

6.1 การเคลื่อนไหวทางการแสดงออก (Expressive movement) เช่น การแสดงออกทางสีหน้า เป็นต้น

6.1 การเคลื่อนไหวทางการตีความ (Interpretative movement) เช่น การเคลื่อนไหวในเชิงสุนทรียภาพหรือเชิงสร้างสรรค์ เป็นต้น

ในขณะที่ Dave (1975) ได้พัฒนาระดับของทักษะพิสัย ออกเป็น 5 ระดับ มีขั้นตอนจากระดับน้อยไปสู่ระดับมาก ได้แก่

1. การเลียนแบบ (Imitation) เป็นขั้นตอนที่ทำตามผู้อื่น หรือทำตามตัวแบบ ซึ่งการทำตามอาจทำให้เกิดคุณภาพงานต่ำ เช่น ลอกเลียนแบบศิลปะ เป็นต้น

2. การทำได้เอง (Manipulation) เป็นขั้นตอนที่สามารถกระทำตัวเอง โดยอาศัยการเรียนรู้และการฝึกฝน เช่น การสร้างผลงานขึ้นมาหลังจากที่ได้ศึกษาเล่าเรียนค้นคว้าเกี่ยวกับสิ่งนั้น เป็นต้น

3. การทำได้อย่างถูกต้องด้วยตนเอง (Precision) เป็นขั้นตอนที่ทำได้ดี มีความผิดพลาดเกิดขึ้นน้อยมาก

4. การทำได้อย่างเชื่อมโยง (Articulation) เป็นขั้นตอนที่มีการประสานสอดคล้องการกระทำได้เป็นหนึ่งเดียวกัน เช่น การผลิตเพลงที่ต้องอาศัยความกลมกลืนของเสียง จังหวะ หรือแม้แต่การผลิตภาพยนตร์ที่ต้องอาศัยความกลมกลืนของดนตรี ภาพ สี เสียง เป็นต้น

5. การทำได้อย่างอัตโนมัติ (Naturalization) เป็นขั้นตอนที่ทำได้อย่างอัตโนมัติ มีความเป็นธรรมชาติ มีความคล่องแคล่ว

แต่ในปัจจุบันการประเมินผลแนวทางใหม่ จะเน้นการประเมินตามสภาพที่เป็นจริง เพื่อให้ผู้เรียนสามารถเชื่อมโยงชีวิตจริงกับการเรียนได้อย่างกลมกลืน ดังนั้น จึงมีการวัดและประเมินผลการเรียนรู้ทางด้านทักษะพิสัยในรูปแบบใหม่ที่ไม่ได้เน้นเฉพาะการใช้กล้ามเนื้อในการเคลื่อนไหวเพียงเท่านั้น แต่จะรวมทั้งการได้ปฏิบัติที่ต้องใช้ความรู้ ความเข้าใจ ความคิดสร้างสรรค์ และความสามารถอื่น ๆ อีกด้วย เพื่อที่จะได้พัฒนาผู้เรียนให้มีความสมบูรณ์ทั้งทางด้านร่างกาย สติปัญญา และจิตใจ (โชติกา ภาชีผล, 2559)

สำหรับการวัดทางด้านพฤติกรรม ควรมีการพิจารณาให้นำไปใช้ในบริบทสำหรับการแก้ปัญหาและการตัดสินใจต่าง ๆ (Barnett, Lentz, & Macmann, 2000) เช่น การพัฒนาโปรแกรมจัดการกระทำ การวินิจฉัย การประเมินผลการเรียนรู้ เป็นต้น โดยมีการนำการวัดทางด้านพฤติกรรมไปใช้ได้อย่างหลากหลายตามวัตถุประสงค์ของการวัด โดยการวัดทางด้านพฤติกรรม มีวิธีการต่าง ๆ มากมาย แตกต่างกันไปตามความเหมาะสมและวัตถุประสงค์ของการวัด โดยจะกล่าวถึงวิธีการวัดทางด้านพฤติกรรมในลำดับถัดไป

วิธีการวัดทางด้านพฤติกรรม

เทคนิควิธีการวัดทางด้านพฤติกรรมมีวิธีการที่หลากหลาย มีความแตกต่างกันไปตามวัตถุประสงค์ของการวัด ซึ่งมีรายละเอียดดังนี้ (โชติกา ภาชีผล, 2559; อรพินทร์ ชูชม, 2545; Cohen, Swerdlik, & Sturman, 2013; Eckert, & Lovett, 2013; Kaplan, & Saccuzzo, 2013; Kubiszyn, & Borich, 2000; Ollendick, Alvarez, & Greene, 2004; Reynolds, Livingston, & Willson, 2010)

1. การสังเกตพฤติกรรม

การสังเกตพฤติกรรม (Behavioral Observation) เป็นวิธีการวัดทางด้านพฤติกรรมวิธีหนึ่งที่ใช้ในการวัดและบันทึกพฤติกรรม สามารถนำมาใช้ในบริบทสภาพแวดล้อมที่หลากหลาย (เช่น โรงเรียน บ้าน เป็นต้น) เพื่อระบุพฤติกรรมที่เป็นปัญหาของเด็กและวัยรุ่น หรือเพื่อระบุปัจจัยควบคุมของพฤติกรรมที่เป็นปัญหา นอกจากนี้คุณลักษณะของความเป็นปรนัยของพฤติกรรม ได้แก่ ความถี่ (frequency) ลักษณะชั่วคราว (temporality) และความเข้มข้น (intensity) ซึ่งสามารถวัดและประเมินพฤติกรรมได้ด้วยวิธีการบันทึกการสังเกต เมื่อทำการเก็บรวบรวมข้อมูลจากการสังเกตพฤติกรรมที่บอกจำนวนได้ มีแนวโน้มและการเปลี่ยนแปลงของพฤติกรรม สามารถที่จะทำการอธิบายได้

วิธีการสังเกตโดยตรง (Direct observation) เป็นการวัดทางด้านพฤติกรรม โดยใช้วิธีการสังเกตพฤติกรรมที่ต้องการศึกษาโดยตรง ผู้ที่สังเกตอาจรู้ตัวหรือไม่รู้ตัวว่าถูกสังเกตก็ได้ วิธีนี้อาจทำการกำหนดเวลาหรือไม่กำหนดเวลาก็ได้ รวมทั้งจะสังเกตเวลาใดก็ได้ที่เกิดพฤติกรรมนั้น ๆ เกิดขึ้น โดยการสังเกตพฤติกรรมจะต้องมีการจดบันทึกพฤติกรรมที่สังเกตไว้ ดังนี้

- การบันทึกการเล่าเรื่อง (narrative recording) เป็นลักษณะการบันทึกพฤติกรรมที่เกิดขึ้นในช่วงเวลาการสังเกตที่ได้กำหนดไว้ โดยใช้การเขียนเล่าเรื่องบรรยายถึงพฤติกรรมต่าง ๆ โดยอาจมีวัสดุอุปกรณ์ช่วยในการบันทึก เช่น เครื่องมือบันทึกเสียงและภาพ การบันทึกรูปแบบนี้เรียกอีกอย่างว่า การบันทึกอย่างต่อเนื่อง (continuous recording) วิธีนี้จะช่วยในการบ่งชี้พฤติกรรมที่สำคัญ ยกตัวอย่างเช่น งานวิจัยของ Lanovaz, Argumedes, Roy, Duquette and Watkins (2013) ได้ใช้การบันทึกการเล่าเรื่อง โดยให้บันทึกในแง่ของปัจจัยเชิงสาเหตุ พฤติกรรม และผลลัพธ์ที่ตามมาของพฤติกรรมที่เป็นปัญหา โดยวิธีวิธีนี้ช่วยสร้างสมมติฐานที่เกิดขึ้นของเด็กที่มีพฤติกรรมที่เป็นปัญหา

- การบันทึกเหตุการณ์ (event recording) เป็นลักษณะการบันทึกพฤติกรรมตามเหตุการณ์ที่ต้องการเท่านั้น เช่น อาจใช้วิธีการแจกนับความถี่ที่เกิดพฤติกรรมนั้น ๆ ขึ้นมา เช่น การแสดงพฤติกรรมรุนแรง การพูดโต้เถียง เป็นต้น ยกตัวอย่าง งานวิจัยของ Juberg, Alfano, Coughlin and Thompson (2001) ได้ใช้การบันทึกเหตุการณ์เกี่ยวกับพฤติกรรมการเอาของเข้าปาก โดยมีแบบฟอร์มบันทึกเหตุการณ์ ให้ผู้ปกครองบันทึกความถี่ แต่อาจมีปัญหาเพราะเป็นพฤติกรรมที่ค่อนข้างรวดเร็ว ผู้ปกครองอาจจะมองไม่เห็น

“แบบสังเกตพฤติกรรมความสุขของเด็กปฐมวัย” (ศรีวิไล เชาวน์ปรีชา, 2550) ซึ่งเป็นแบบสังเกตพฤติกรรมแบบมาตราประมาณค่า (Rating scale)

คนที่	รายการพฤติกรรม																		รวม	
	เพิ่มพูนความรู้เกี่ยวกับความสุข									เสริมสร้างทักษะทางสังคม										
	ตั้งใจร่วมกิจกรรม			มีความคิดเห็นเป็นของตนเอง			กล้าแสดงออก			ปฏิบัติตามข้อตกลง			สนทนาโต้ตอบ			รู้จักการให้อภัย				
	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2		
1																				
2																				

* หมายเหตุ: 0 คือ ไม่แสดงพฤติกรรมเลย, 1 คือ แสดงพฤติกรรมบางเวลา, 2 คือ แสดงพฤติกรรมตลอดเวลา

คุณสมบัติทางการวัดของแบบสังเกตพฤติกรรม ซึ่งมีรายละเอียดดังนี้

1) ความเที่ยงตรง (Validity) โดยความเที่ยงตรงของแบบสังเกตพฤติกรรม เป็นความเที่ยงตรงเชิงเนื้อหา ตรวจสอบว่าพฤติกรรมที่ทำการสังเกตเป็นตัวแทนของพฤติกรรมที่ศึกษาหรือไม่ นอกจากนี้ แบบสังเกตพฤติกรรมอาจมีความเที่ยงตรงเชิงโครงสร้างได้ พิจารณาจากผลของแบบสังเกตพฤติกรรมวัดได้ครบตามเกณฑ์หรือโครงสร้างที่วางไว้หรือไม่ ถ้าผลของการสังเกตพฤติกรรมครั้งนั้นวัดตามพฤติกรรมครบตามโครงสร้างที่วางไว้ก็แสดงว่า แบบสังเกตพฤติกรรมมีความเที่ยงตรงเชิงโครงสร้าง

2) ความเชื่อมั่น (Reliability) โดยมีการหาความเชื่อมั่นของการสังเกต ได้แก่ ความเชื่อมั่นภายในผู้สังเกต (Intra-observer reliability) เป็นวิธีการที่ตรวจสอบว่าผู้สังเกตคนเดียวกันมีความคงเส้นคงวาในการบันทึกผลการสังเกตหรือไม่ โดยให้สังเกตซ้ำในเวลาที่แตกต่างกัน แล้วนำผลที่ได้มาหาสัมประสิทธิ์สหสัมพันธ์ระหว่างการสังเกตทั้ง 2 ครั้ง นอกจากนี้ ยังมีความเชื่อมั่นระหว่างผู้สังเกต (Inter-observer reliability) เป็นวิธีการที่ตรวจสอบว่าผู้สังเกตหลายคนมีความคงเส้นคงวาในการสังเกตพฤติกรรมหรือไม่ หากสังเกตของผู้สังเกตแต่ละคนจะต้องเป็นอิสระต่อกัน แล้วนำผลมาหาความสัมพันธ์หรือความสอดคล้องของผู้สังเกตดังกล่าว

2. การสัมภาษณ์เชิงพฤติกรรม

การสัมภาษณ์เชิงพฤติกรรม (Behavioral Interviewing) เป็นวิธีการวัดทางด้านพฤติกรรมวิธีหนึ่งที่ใช้กันเป็นประจำ ซึ่งเป็นการให้ข้อมูลรายละเอียดเกี่ยวกับมุมมองของบุคคลหรือผู้ถูกสัมภาษณ์ การสัมภาษณ์สามารถประเมินคุณภาพของเครื่องมือวัดทางจิตวิทยาได้เช่นเดียวกัน เช่น ความเชื่อมั่นและความเที่ยงตรง นอกจากนี้วิธีการสัมภาษณ์เชิงพฤติกรรมนั้นมีความหลากหลาย การเลือกใช้วิธีใดวิธีหนึ่งจึงขึ้นอยู่กับประเภทของข้อมูลพฤติกรรมที่ทำการค้นหา รวมทั้งเป้าหมายของผู้ให้สัมภาษณ์ด้วย

ประเภทของการสัมภาษณ์ แบ่งได้เป็น 2 ประเภท ดังนี้

1. การสัมภาษณ์แบบมีโครงสร้าง (Structured interview) เป็นวิธีการสัมภาษณ์ที่อยู่ในรูปแบบที่กำหนดไว้ ทั้งในแง่ของคำถาม การให้คะแนน หรือแม้แต่การเรียงลำดับคำถาม มีระบบที่ชัดเจน นั่นคือมีแบบสัมภาษณ์เตรียมคำถามไว้ล่วงหน้าเรียบร้อยแล้ว ถามคำถามตามลำดับ เมื่อหมดคำถามก็ถือว่าเสร็จสิ้นกระบวนการสัมภาษณ์ ลักษณะการสัมภาษณ์แบบนี้ใช้คำถามเดียวกันกับผู้ถูกสัมภาษณ์ทุกคนเหมือนกันหมด เช่น การสัมภาษณ์เพื่อวินิจฉัยอาการทางจิตเวชของเด็กและวัยรุ่น โดยการวินิจฉัยไม่ได้ขึ้นอยู่กับการสังเกตพฤติกรรมของเด็กและวัยรุ่น แต่จะใช้แบบสัมภาษณ์เชิงโครงสร้าง เพื่อถามคำถามเกี่ยวกับอาการจิตเวชต่าง ๆ ซึ่งแบบสัมภาษณ์นี้ช่วยให้เกิดการวินิจฉัยเด็กและวัยรุ่นทุกคนได้อย่างเท่าเทียมกัน

2. การสัมภาษณ์แบบกึ่งโครงสร้าง (Semi-structured interview) เป็นวิธีการสัมภาษณ์ที่ไม่ได้กำหนดหรือรูปแบบตายตัว นั่นคือ ไม่ได้เตรียมแบบสัมภาษณ์ล่วงหน้าไว้เหมือนแบบสัมภาษณ์แบบมีโครงสร้าง รูปแบบการสัมภาษณ์สามารถยืดหยุ่นได้ สามารถเปลี่ยนคำถามและลำดับของการถามได้ มีอิสระในการถามคำถาม แต่คำถามเหล่านั้นควรอยู่ในกรอบตามวัตถุประสงค์ของการศึกษา เพื่อให้ได้ข้อมูลตามที่ต้องการ เช่น การสัมภาษณ์เชิงคลินิกแบบกึ่งโครงสร้างสำหรับเด็กและวัยรุ่น โดยมีแบบฟอร์มเกี่ยวกับข้อมูลการแนะนำการสัมภาษณ์ ตัวอย่างคำถามแบบเปิด หัวข้อที่จะทำการถามคำถาม (เช่น ความสัมพันธ์กับครอบครัว การรับรู้และความรู้สึกต่อตนเอง) โดยอาการทางคลินิกจะวินิจฉัยออกมาตามคำตอบที่ได้รับจากผู้ถูกสัมภาษณ์

คุณสมบัติทางการวัดของการสัมภาษณ์ ซึ่งมีรายละเอียดดังนี้

1) ความเที่ยงตรง (Validity) โดยความเที่ยงตรงของการสัมภาษณ์มีวิธีการประเมินความเที่ยงตรง เช่นเดียวกับเครื่องมือวัดอื่น ๆ นั่นคือ ความเที่ยงตรงเชิงเนื้อหา โดยพิจารณาว่าการสัมภาษณ์นั้นใช้คำถามได้สอดคล้องตรงตามวัตถุประสงค์ของการสัมภาษณ์ซึ่งครอบคลุมประเด็นต่าง ๆ ที่ศึกษาหรือไม่ ถามในเรื่องที่ผู้ตอบให้ข้อมูลได้อย่างถูกต้องหรือไม่ ตลอดจนพิจารณาว่าคำถามนั้น ๆ ถามได้ชัดเจนหรือไม่ โดยความเที่ยงตรงของการสัมภาษณ์จะมีมากหรือน้อยเพียงใด ขึ้นอยู่กับผลของการสัมภาษณ์ในครั้งนั้นว่าวัดได้ครบตามประเด็นต่าง ๆ หรือไม่

2) ความเชื่อมั่น (Reliability) โดยความเชื่อมั่นสามารถหาได้จากค่าสัมประสิทธิ์ความคงที่ (coefficient of stability) เพื่อดูความคงที่แน่นอนของผลการสัมภาษณ์เมื่อใช้ผู้สัมภาษณ์คนเดียวกัน ทำการสัมภาษณ์ซ้ำกับบุคคลกลุ่มเดิมในเวลาที่แตกต่างกัน แล้วนำมาคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างผลการสัมภาษณ์ทั้ง 2 ครั้ง นอกจากนี้ ความเชื่อมั่นยังสามารถพิจารณาจากสัมประสิทธิ์ความเท่าเทียมกัน (coefficient of equivalence) โดยมาจากการใช้รูปแบบคำถามการสัมภาษณ์ที่มีหลายรูปแบบ แต่มีเนื้อหาแบบเดียวกันและเปรียบเทียบผลได้สำหรับความเชื่อมั่นระหว่างผู้สัมภาษณ์ โดยนำผลการสัมภาษณ์จากผู้สัมภาษณ์หลายคนมาวิเคราะห์หาความสอดคล้องระหว่างผู้สัมภาษณ์ด้วยการหาค่าสัมประสิทธิ์สหสัมพันธ์ดังกล่าว

3. มาตรฐานค่าเชิงพฤติกรรม

มาตรฐานค่าเชิงพฤติกรรม (Behavioral Rating Scale) เป็นวิธีการวัดทางด้านพฤติกรรมวิธีหนึ่ง ที่ประเมินพฤติกรรมของบุคคลหนึ่ง ๆ โดยบุคคลอื่นที่มีความสำคัญกับบุคคลนั้น ๆ ซึ่งเป็นวิธีการที่มีประโยชน์ในการวัดทางด้านพฤติกรรมของเด็กและวัยรุ่น เนื่องจากบุคคลที่มีความสำคัญกับเด็กและวัยรุ่นจะมีความเข้าใจในการประเมินพฤติกรรม สามารถระบุจุดแข็งและจุดอ่อนของพฤติกรรมของเด็กและวัยรุ่นได้ มีลักษณะเป็นการให้คะแนนของผู้ให้คะแนนในการประเมินพฤติกรรมของเด็กและวัยรุ่น

ยกตัวอย่าง มาตรฐานค่าเชิงพฤติกรรม

“The IOWA Conners Teacher’s Rating Scale” (Loney, & Milich, 1982)

คำชี้แจง: ให้ครูเป็นผู้ประเมินเด็กเกี่ยวกับพฤติกรรมต่าง ๆ

ข้อความ	ตัวเลือกคำตอบ			
	ทำมาก	ทำค่อนข้างมาก	ทำเล็กน้อย	ไม่ทำเลย
ยุกยิก อยู่ไม่สุข				
ชอบส่งเสียงรบกวนคนอื่น				
ตื่นตัวง่าย หุนหันพลันแล่น วู่วาม				

4. การรายงานตนเอง

การรายงานตนเอง (Self-report) เป็นวิธีการวัดทางด้านพฤติกรรมวิธีหนึ่งที่มีความสะดวกในทางปฏิบัติที่ต้องการวัดพฤติกรรมของบุคคลจำนวนมาก โดยแบบรายงานตนเอง เป็นเครื่องมือวัดที่บุคคลจะเป็นผู้รายงานตนเองเพื่ออธิบายประสบการณ์เชิงอัตวิสัยของตนเอง ในแง่ของอารมณ์ แรงจูงใจ ความสัมพันธ์ระหว่างบุคคล และลักษณะทางเจตคติ (Anastasi & Urbina, 1997)

แบบรายงานตนเอง ประกอบด้วยข้อความต่าง ๆ ที่เป็นข้อความแสดงพฤติกรรม ที่มีรูปแบบการตอบที่คำตอบ “ใช่” “ไม่แน่ใจ” และ “ไม่ใช่” นอกจากนี้ ยังมีรูปแบบการตอบที่เป็นมาตรวัดลิเคิร์ตหรือมาตรวัดประเมินค่า โดยในปัจจุบัน การวัดทางด้านพฤติกรรมศาสตร์ มักใช้แบบรายงานตนเองในการวัดพฤติกรรม วิธีการรายงานตนเองมีข้อตกลงเบื้องต้นว่า คำตอบของแต่ละบุคคล สะท้อนให้เห็นถึงความแตกต่างระหว่างบุคคล รวมทั้งปรากฏการณ์ โดยมีรูปแบบข้อความและคำตอบ ดังนี้

ยกตัวอย่าง แบบรายงานตนเอง

“แบบวัดพฤติกรรมป้องกันการสูบบุหรี่ในกลุ่มวัยรุ่น” (ชาริน สุวรรณวงศ์ และศรัณย์ พิมพ์ทอง, 2563)

ข้อความ	ตัวเลือกคำตอบ					
	ปฏิบัติมากที่สุด	ปฏิบัติมาก	ปฏิบัติค่อนข้างมาก	ปฏิบัติค่อนข้างน้อย	ปฏิบัติน้อย	ปฏิบัติน้อยที่สุด
ฉันพยายามไม่คบเพื่อนที่ชวนสูบบุหรี่						
ฉันจะไม่สูบบุหรี่ เมื่อรู้สึกเครียด ทุกข์หรือมีเรื่องกลุ้มใจ						
เมื่อเห็นคนใกล้ชิดชดภายในบ้านสูบบุหรี่ฉันเดินเลี่ยงไปที่อื่นทันที						

5. การวัดทางจิตสรีรวิทยา

การวัดทางจิตสรีรวิทยา (Psychophysiological assessment) เป็นวิธีการวัดทางด้านพฤติกรรมวิธีหนึ่ง ที่เมื่อสมัยก่อนพฤติกรรมได้รับการนิยามว่าเป็นการกระทำของส่วนต่าง ๆ ของร่างกาย ซึ่งอยู่ภายใต้การควบคุมของระบบประสาทและระบบต่อมไร้ท่อ ดังนั้น การวัดเกี่ยวกับกิจกรรมทางกล้ามเนื้อ ระบบประสาท และระบบต่อมไร้ท่อ จึงถือว่าเป็นรูปแบบหนึ่งของการวัดทางด้านพฤติกรรม

แม้ว่าการวัดทางด้านจิตสรีรวิทยาจะไม่มีการใช้ในทุกคลินิก แต่ถ้าหากทำงานร่วมกับเด็ก การวัดทางด้านจิตสรีรวิทยาควรจะถูกนำมาพิจารณาใช้ด้วย บางครั้งการวัดทางจิตสรีรวิทยาให้ข้อมูลว่าการวัดทางด้านพฤติกรรมอื่น ๆ ให้ไม่ได้ นั่นคือ บุคคลไม่สามารถติดตามพฤติกรรมได้ดีเท่ากับการสัมภาษณ์เชิงพฤติกรรมหรือใช้มาตรประมาณค่า การสังเกตเชิงพฤติกรรมก็ช่วยแก้ปัญหาเหล่านี้ แต่ก็ยังมีการประเมินเชิงอัตวิสัยของผู้สังเกต ดังนั้น การวัดทางจิตสรีรวิทยาช่วยแก้ปัญหาดังกล่าว โดยเป็นการวัดที่สามารถวัดได้โดยตรงและมีความเป็นปรนัย

ยกตัวอย่างการวัดทางด้านจิตสรีรวิทยา เช่น

- การวัดคลื่นไฟฟ้ากล้ามเนื้อ (Electromyography: EMG) ซึ่งเป็นการวัดทางจิตสรีรวิทยาเกี่ยวกับการทำงานของกล้ามเนื้อ

- การวัดคลื่นไฟฟ้าจอตา (Electrooculography: EOG) ซึ่งเป็นการวัดทางจิตสรีรวิทยาเกี่ยวกับการเคลื่อนไหวของตา

6. การประเมินการปฏิบัติ

การประเมินการปฏิบัติ (Performance Assessment) เป็นวิธีการวัดทางด้านพฤติกรรมวิธีหนึ่ง ซึ่งเป็นการใช้สถานการณ์ในการทดสอบการปฏิบัติงานของบุคคล (Performance Task) สามารถประเมินการปฏิบัติได้

3 ประเภท ได้แก่ 1) การประเมินการปฏิบัติเน้นกระบวนการ เป็นการประเมินสำหรับขั้นตอนการปฏิบัติงาน เช่น การนำเสนองาน/การบ้านหน้าชั้นเรียน การอ่านหนังสือ การเดินรำ การเล่นเกม เป็นต้น 2) การประเมินการปฏิบัติเน้นผลงาน เป็นการประเมินผลลัพธ์ของการปฏิบัติซึ่งเห็นเป็นรูปธรรมอย่างชัดเจน ได้แก่ ภาพวาด รายงาน ผลงานวิจัย เป็นต้น และ 3) การประเมินการปฏิบัติเน้นทั้งกระบวนการและผลงาน เป็นการประเมินการปฏิบัติที่สามารถแบ่งแยกประเมินกระบวนการและผลงานได้อย่างชัดเจน เช่น การทำวิจัย ประกอบด้วย กระบวนการ การทำวิจัย และผลงานวิจัยตีพิมพ์ เป็นต้น

กระบวนการประเมินการปฏิบัติ มีขั้นตอนที่สำคัญดังนี้ (สุวิมล ว่องวานิช, 2546)

1. กำหนดงานที่จะให้ปฏิบัติ ต้องมีความสอดคล้องกับตัวชี้วัด/ความมุ่งหมายของการประเมินการปฏิบัติ และกำหนดรูปแบบของงานว่าเป็นงานในระยะสั้น (เช่น สอบภาคปฏิบัติ) งานที่เป็นเหตุการณ์ (เช่น การเล่าเรื่อง) หรืองานในระยะยาว (เช่น การทำโครงการ) โดยการกำหนดงานมีหลักในการพิจารณางานที่จะปฏิบัติ คือ

- 1) งานที่กำหนดต้องเหมาะสมกับความสามารถของบุคคล และสอดคล้องกับเนื้อหาในบทเรียน
- 2) ระบุลักษณะงานให้ชัดเจนว่าเป็นการปฏิบัติงานกลุ่มหรืองานเดี่ยว
- 3) งานที่กำหนด บุคคลควรสามารถทำได้ภายในระยะเวลาที่กำหนด
- 4) การมอบหมายงานต้องมีคำชี้แจงที่ชัดเจน ทั้งในแง่ความมุ่งหมาย ขอบข่ายงาน เครื่องมือ

เวลา ผลงาน รวมถึงเกณฑ์การให้คะแนน

2. กำหนดคุณลักษณะที่ใช้ในการประเมินการปฏิบัติ เป็นการกำหนดคุณลักษณะที่ต้องการวัดในการปฏิบัติ โดยมีกระบวนการในการจำแนกขั้นตอนในการทำงาน ประกอบด้วยรายละเอียดย่อย ๆ เรียกว่า การวิเคราะห์งาน (job analysis) ซึ่งเป็นกระบวนการได้สารสนเทศเกี่ยวกับงานและการกำหนดน้ำหนักความสำคัญของลักษณะงาน ดังนี้

กระบวนการได้สารสนเทศเกี่ยวกับงาน มีขั้นตอนดังนี้

- 1) รวบรวมรายละเอียดข้อมูลเกี่ยวกับงานที่ปฏิบัติ
- 2) ตรวจสอบความถูกต้องของข้อมูลเกี่ยวกับงานที่ปฏิบัติ
- 3) เขียนรายละเอียดคำบรรยายเกี่ยวกับงานที่ปฏิบัติ
- 4) ระบุทักษะ ความสามารถ และความรู้ที่ใช้ในการทำงาน
- 5) ปรับข้อมูลเกี่ยวกับงานที่ปฏิบัติให้ทันสมัยตลอดเวลา

การกำหนดน้ำหนักความสำคัญของลักษณะงาน มีการกำหนดน้ำหนักความสำคัญของกระบวนการทำงาน (process) รวมทั้งผลงาน (product) ว่าในแต่ละส่วนให้มีความสำคัญเพื่อใช้ในการประเมินปริมาณเท่าใด

3. กำหนดวิธีการประเมินการปฏิบัติ กระทำได้หลายวิธี ดังนี้

1) ประเมินโดยการให้เขียนตอบ การประเมินด้วยวิธีนี้เหมาะสำหรับการปฏิบัติงานบางประเภท เช่น การออกแบบเสื้อผ้า การเขียนแปลนบ้าน เป็นต้น นอกจากนี้ งานบางประเภทที่ต้องการวัดความรู้เกี่ยวกับ

การปฏิบัติก่อนที่จะทำการปฏิบัติจริง เนื่องจากเป็นงานที่มีความเสี่ยงและมีผลกระทบต่อตนเองและผู้อื่น เช่น การก่อสร้าง การผ่าตัด เป็นต้น

2) ประเมินโดยบุคคลปฏิบัติงานจริงในสถานการณ์จริงหรือสถานการณ์จำลอง การประเมินด้วยวิธีนี้เน้นกระบวนการปฏิบัติ ซึ่งจะต้องสังเกตพฤติกรรมของบุคคลในสถานการณ์ที่ได้จัดไว้ให้ หรือในสถานการณ์จริง บุคคลนั้นอาจจะรู้ตัวหรือไม่รู้ตัวก็ได้

3) ประเมินตัวอย่างของงานที่ได้จากการปฏิบัติ การประเมินด้วยวิธีนี้ใช้สำหรับวัดผลงานเป็นส่วนใหญ่ พิจารณาจากชิ้นงาน เช่น รายงาน งานฝีมือ เป็นต้น

4. กำหนดเครื่องมือที่ใช้ประเมินการปฏิบัติ หลังจากที่ได้เลือกวิธีการประเมินการปฏิบัติแล้ว ต้องกำหนดเครื่องมือที่ใช้ประเมิน รวมทั้งเวลาที่ใช้ประเมิน โดยเครื่องมือมีหลายประเภท ได้แก่ แบบสอบภาคปฏิบัติ แบบตรวจสอบรายการ มาตรฐานค่า เป็นต้น โดยการประเมินการปฏิบัติบางเรื่องอาจต้องใช้เครื่องมือมากกว่า 1 ประเภท มีรายละเอียดดังนี้

- แบบสอบภาคปฏิบัติ อาจเป็นสถานการณ์จำลองหรือสถานการณ์จริงที่ให้ผู้คนได้ปฏิบัติการประเมินสามารถประเมินจากกระบวนการที่เป็นวิธีการในขั้นตอนต่าง ๆ ของการปฏิบัติ หรือจากผลงานที่เป็นผลงานของกระบวนการก็ได้

- แบบตรวจสอบรายการ เป็นรายการที่เกี่ยวข้องกับการปฏิบัติ ซึ่งอาจแยกเป็นรายการย่อยของการปฏิบัติ ได้แก่ ขั้นเตรียม ขั้นปฏิบัติ ผลงาน โดยอาศัยการสังเกตในการประเมินแต่ละรายการว่า “มี” หรือ “ไม่มี”

- มาตรฐานค่า เป็นรายการที่เกี่ยวข้องกับการปฏิบัติ และมาตรฐานค่าต้องการทราบรายละเอียดที่ชัดเจนขึ้นว่ามีมากน้อย อยู่ในระดับใด การประเมินโดยให้คะแนนเป็นมาตรฐานค่า มีจุดอ่อนคือเกณฑ์การให้คะแนนขาดความเป็นปรนัย นั่นคือ ผู้ให้คะแนนคนเดียวกันอาจมีโอกาสให้คะแนนบุคคลที่มีความสามารถเท่ากันได้รับคะแนนไม่เท่ากัน ดังนั้น แนวทางการแก้ไข คือ การให้คะแนนที่มีความเป็นปรนัย โดยสร้างเกณฑ์รูบริก (Rubric scoring)

5. กำหนดวิธีการประเมินผลและการรายงานผล ข้อมูลที่ได้จากการประเมินการปฏิบัติต้องนำมาตัดสินคุณภาพ อาจทำได้ด้วยการเปรียบเทียบกับกลุ่มหรือเกณฑ์ที่กำหนด หรืออาจเทียบกับตัวผู้เรียนเอง เพื่อให้เห็นการเปลี่ยนแปลงหรือพัฒนาการของบุคคลแต่ละคนได้ชัดเจนยิ่งขึ้น

สรุป

การวัดทางด้านพฤติกรรมถือเป็นกระบวนการหนึ่งที่สำคัญในการพัฒนาเครื่องมือวิจัย เพราะเป็นการวัด การแสดงออกหรือการปฏิบัติของบุคคล ซึ่งเป็นสิ่งที่วัดได้โดยยาก มีหลากหลายวิธีในการวัด ดังนั้น จำเป็นต้องมีการนิยามให้ชัดเจน โดยต้องเลือกวิธีการวัดหรือเครื่องมือวัดให้เหมาะสมกับกลุ่มตัวอย่างและพฤติกรรมที่ต้องการจะศึกษา โดยการวัดทางด้านพฤติกรรมมีวิธีการวัดที่หลากหลาย ได้แก่ การสังเกตเชิงพฤติกรรม การสัมภาษณ์เชิงพฤติกรรม มาตรฐานค่าเชิงพฤติกรรม การรายงานตนเอง การวัดทางจิตสรีรวิทยา และการประเมินการปฏิบัติ เป็นแนวทางในการสร้างและพัฒนาเครื่องมือทางด้านพฤติกรรมศาสตร์ให้มีความเหมาะสม และสามารถใช้อัตการ แสดงออกหรือการปฏิบัติของแต่ละบุคคลได้

เอกสารอ้างอิง

- ชาริน สุวรรณวงศ์, และศรัณย์ พิมพ์ทอง. (2563). การสร้างและตรวจสอบคุณภาพแบบวัดพฤติกรรมป้องกัน การสูบบุหรี่ในกลุ่มวัยรุ่น. *วารสารเกื้อการุณย์*, 27(1), 46-58.
- โชติกา ภาชีผล. (2559). *การวัดและประเมินผลการเรียนรู้*. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ศรีวิไล เขาวนบุรีชา. (2550). *การศึกษาพฤติกรรมความสุขของเด็กปฐมวัยโดยใช้กิจกรรมคุณค่าเพื่อชีวิต* (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
- สาวิตรี จุ้ยทอง. (2554). *การตรวจสอบคุณภาพรูปแบบของชุดเครื่องมือการประเมินผลการเรียนรู้ตามสภาพจริง กลุ่มสาระการเรียนรู้คณิตศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 1 โดยการประยุกต์ใช้ทฤษฎีการสรุป อ้างอิงความน่าเชื่อถือของผลการวัด* (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
- สุวิมล ว่องวานิช. (2546). *การประเมินผลการเรียนรู้แนวใหม่*. กรุงเทพฯ: ศูนย์ตำราและเอกสารทางวิชาการ คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- อรพินทร์ ชูชม. (2545). *เอกสารคำสอนวิชา วป 502 การสร้างและพัฒนาเครื่องมือวัดทางพฤติกรรมศาสตร์*. กรุงเทพฯ: สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

- Barnett, D. W., Lentz, F. E., & Macmann, G. (2000). Psychometric qualities of professional practice. In E. S. Shapiro, & T. R. Kratochwill (Eds.), *Behavioral assessment in schools: Theory, research, and clinical foundations* (2nd ed.) (pp. 355-386). New York: Guilford Press.
- Bloom, B. S. (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain*. New York: David McKay Co.
- Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2013). *Psychological testing and Assessment: An introduction to tests and measurement* (8th ed.). New York: McGraw-Hill.
- Dave, R. H. (1975). Psychomotor levels. In R. J. Armstrong (Ed.), *Developing and writing behavioural objectives* (pp. 33-34). Tucson, AZ: Educational Innovators Press.
- Eckert, T. L., & Lovett, B. J. (2013). Principles of behavioral assessment. In D. H. Saklofske, C. R. Reynolds, & V. L. Schwann (Eds.), *The oxford handbook of child psychological assessment* (pp. 366-384). New York: Oxford University Press.
- Fernández-Ballesteros, R. (2001). Behavioral assessment. In N. J. Smelser, & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 1090-1094). New York: Pergamon.
- Hanley, G. P., Cammilleri, A. P., Tiger, J. H., & Ingvarsson, E. T. (2007). A method for describing preschoolers' activity preferences. *Journal of Applied Behavior Analysis*, *40*, 603-618.
- Harrow, A. J. (1972). *A taxonomy of the psychomotor domain: A guide for developing behavioral objectives*. New York: David McKay Co.
- Heiby, E. M., & Haynes, S. N. (2003). Introduction to behavioral assessment. In S. P. Haynes, & E. M. Heiby (Eds.), *Behavioral assessment* (pp. 3-18). Hoboken, NJ: Wiley.
- Juberg, D. R., Alfano, K., Coughlin, R. J., & Thompson, K. M. (2001). An observational study of object mouthing behavior by young children. *Pediatrics*, *107*, 135-142.
- Kaplan, R. M., & Saccuzzo, D. P. (2013). *Psychological assessment and theory: Creating and using psychological tests* (8th ed.). Boston, MA: Cengage Learning.
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1973). *Taxonomy of educational objectives, the classification of educational goals, Handbook II: Affective domain*. New York: David McKay Co.

- Kubiszyn, T., & Borich, G. (2000). *Educational testing and measurement: Classroom application and practice* (6th ed.). New York: John Wiley & Sons.
- Lanovaz, M. J., Argumedes, M., Roy, D., Duquette, J. R., & Watkins, N. (2013). Using ABC narrative recording to identify the function of problem behavior: A pilot study. *Research in Developmental Disabilities, 34*, 2734-2742.
- Loney, J., & Milich, R. (1982). Hyperactivity, inattention, and aggression in clinical practice. In M. Wolraich, & D. K. Routh (Eds.), *Advances in developmental and behavioral pediatrics* (Vol. 3, pp. 113-147). Greenwich, CT: JAI.
- Ollendick, T. H., Alvarez, H. K., & Greene, R. W. (2004). Behavioral assessment: History of underlying concepts and methods. In S. N. Haynes, & E. M. Heiby (Eds.), *Comprehensive handbook of psychological assessment: Vol. 3. Behavioral assessment* (pp. 19-36). Hoboken, NJ: John Wiley & Sons.
- Reynolds, C. R., Livingston, R. B., & Willson, V. (2010). *Measurement and assessment in education* (2nd ed.). Upper Saddle River, NJ: Pearson Education.
- Simpson, E. J. (1972). *The classification of educational objectives in the psychomotor domain*. Washington, DC: Gryphon House.

บทที่ 7

การวิเคราะห์รายข้อ

บทที่ 7 เป็นการนำเสนอเกี่ยวกับการวิเคราะห์รายข้อ โดยผู้เขียนต้องการให้ผู้อ่านได้ทราบและทำความเข้าใจเกี่ยวกับการวิเคราะห์รายข้อ (Item analysis) ในรูปแบบต่าง ๆ เนื่องจากการวิเคราะห์รายข้อแต่ละรูปแบบใช้กับข้อความในเครื่องมือวัดที่แตกต่างกัน การวิเคราะห์รายข้อถือว่าเป็นวิธีการหนึ่งในการตรวจสอบคุณภาพของข้อคำถามเป็นรายข้อ ทำให้ทราบว่าข้อคำถามแต่ละข้อมีคุณภาพมากน้อยเพียงใด และหากข้อที่มีคุณภาพน้อยจะทำการปรับปรุงข้อคำถามอย่างไร เพื่อให้ข้อคำถามมีคุณภาพมากยิ่งขึ้น โดยบทนี้ จะช่วยให้ผู้วิจัยและนักวิชาการสามารถสร้างเครื่องมือวัดได้อย่างมีคุณภาพ เพื่อช่วยให้กระบวนการพัฒนาเครื่องมือวิจัยทางพฤติกรรมศาสตร์มีความเหมาะสมและสมบูรณ์มากยิ่งขึ้น

แนวคิดเบื้องต้นของการวิเคราะห์รายข้อ

สำหรับแบบสอบถาม/แบบทดสอบที่ดีและมีคุณภาพนั้น มีความจำเป็นที่จะต้องมีความหรือข้อคำถามที่มีลักษณะที่สำคัญ 3 ข้อ ดังนี้ (Fishman & Galguera, 2003; สุจิตรา เทียนสวัสดิ์, 2562)

1. ความสามารถในการจำแนกข้อคำถาม (item discriminability) นั่นคือ ข้อคำถามสามารถจำแนกผู้ตอบที่มีความแตกต่างกันได้ เช่น กลุ่มเรียนดี-กลุ่มเรียนด้อย หรือ กลุ่มคะแนนสูง-กลุ่มคะแนนต่ำ เป็นต้น รวมทั้งสามารถจำแนกผู้ตอบได้สอดคล้องกับการจำแนกจากแบบวัดทั้งหมด

2. ความสอดคล้องกันของข้อคำถาม (item consistency) นั่นคือ คะแนนจากการวัดโดยข้อคำถามแต่ละข้อจะต้องมีความสอดคล้องหรือสัมพันธ์กับคะแนนจากแบบวัดทั้งหมด เพื่อแสดงให้เห็นว่า ข้อคำถามแต่ละข้อวัดเรื่องเดียวกัน

3. ความยากง่ายของข้อคำถาม (item difficulty) นั่นคือ ความยากง่ายของข้อคำถามที่มีความเหมาะสม ข้อมูลจากการวัดมีการกระจายตัวของข้อมูลน้อย กล่าวคือ จำนวนของผู้ตอบไม่กระจุกอยู่ที่ตัวเลือกใดตัวเลือกหนึ่งมากเกินไป เช่น หากผู้ตอบส่วนใหญ่ตอบตัวเลือกที่ถูก แสดงว่าข้อสอบข้อนั้นง่าย แต่ถ้ามีผู้ตอบถูกน้อย แสดงว่าข้อสอบข้อนั้นยาก

การตรวจสอบคุณภาพของเครื่องมือเป็นรายข้อ เป็นวิธีการหนึ่งที่มีความสำคัญเป็นอย่างมาก มักเกิดขึ้นหลังจากที่แบบสอบถาม/แบบทดสอบผ่านการตรวจสอบความเที่ยงตรงเชิงเนื้อหาจากผู้เชี่ยวชาญ/ผู้ทรงคุณวุฒิ (ซึ่งจะกล่าวถึงในบทลำดับถัดไป) และนำไปทดลองใช้ (try out) โดยแบบสอบถาม/แบบทดสอบควรนำไปทดลองใช้กับบุคคลที่มีลักษณะที่คล้ายคลึงกับกลุ่มตัวอย่างที่เราต้องการเก็บข้อมูล สำหรับจำนวนกลุ่มตัวอย่างที่นำไป

ทดลองใช้มีจำนวนขั้นต่ำอย่างน้อย 30 คน (Browne, 1995; Johanson & Brooks, 2010) ดังนั้น ผู้เขียนจึงเห็นว่า กลุ่มตัวอย่างขั้นต่ำที่ควรใช้สำหรับการทดลองใช้เพื่อตรวจสอบคุณภาพของเครื่องมือวัด ประมาณอย่างน้อย 30 คน

การวิเคราะห์รายข้อ (Item analysis) หมายถึง การตรวจสอบคุณสมบัติการวัดด้วยกระบวนการทางสถิติของคำตอบของผู้ตอบเป็นรายข้อ การวิเคราะห์รายข้อมีค่าสถิติหลายรูปแบบที่ใช้ในการแสดงคุณภาพของข้อสอบหรืออีกนัยหนึ่งคือ เครื่องมือวัดมีคุณภาพมากน้อยเพียงใด สามารถแสดงเป็นหลักฐานในเชิงปริมาณด้วยค่าตัวเลขทางสถิติ (ณัฐภรณ์ หลาวทอง, 2561; อรพินทร์ ชูชม, 2545) นอกจากนี้ ยังกล่าวได้ว่า การวิเคราะห์รายข้อเป็นการใช้วิธีการทางด้านสถิติที่ช่วยให้ข้อมูลเกี่ยวกับความยากง่ายของแต่ละคำถาม ความสามารถในการจำแนกระหว่างบุคคล และแนวโน้มของการเกิดอคติหรือความคลาดเคลื่อนได้ (Miller, Lovler, & McIntire, 2013)

ในการวัดของเครื่องมือทางด้านพฤติกรรมศาสตร์ มีค่าสถิติที่นิยมใช้ในการวิเคราะห์รายข้อ ได้แก่ ค่าอำนาจจำแนก ค่าความเที่ยงตรง ค่าความเชื่อมั่น ค่าความยากง่าย และค่าอำนาจตัวลง โดยทั่วไปแล้วแบบทดสอบที่เป็นการวัดด้านารรู้คิด (Cognitive) แบบชนิดเลือกตอบหลายคำตอบ จะมีการใช้สถิติหลายตัว ได้แก่ ค่าอำนาจจำแนก ค่าความเที่ยงตรง ค่าความเชื่อมั่น ค่าความยากง่าย และค่าอำนาจตัวลง ส่วนแบบสอบถามที่เป็นการวัดทางด้านอารมณ์และความรู้สึก (Affective) รวมทั้งการวัดทางด้านพฤติกรรม (Behavior) ที่มีลักษณะการตอบเป็นมาตรวัดประมาณค่า จะใช้สถิติทุกตัว ยกเว้นค่าความยากง่าย และค่าอำนาจตัวลง สำหรับในบทนี้ ผู้เขียนจึงขอแนะนำเฉพาะค่าอำนาจจำแนก ค่าความยากง่าย และอำนาจตัวลง ส่วนค่าความเที่ยงตรงและความเชื่อมั่นจะกล่าวถึงในบทลำดับถัดไป

อำนาจจำแนกรายข้อ (Item Discrimination)

อำนาจจำแนกรายข้อ เป็นคุณลักษณะของการวัดของข้อคำถามแต่ละข้อว่าสามารถแยกแยะหรือจำแนกบุคคลที่มีคะแนนสูงและบุคคลที่มีคะแนนต่ำออกจากกันได้ เช่น ข้อคำถามจากแบบทดสอบ/แบบสอบถามแต่ละข้อสามารถจำแนกกลุ่มที่มีระดับทักษะ ความรู้ เจตคติ หรือแม้แต่ลักษณะบุคลิกภาพสูงออกจากกลุ่มที่มีระดับทักษะ ความรู้ เจตคติ หรือลักษณะบุคลิกภาพต่ำ กล่าวคือ การหาค่าอำนาจจำแนกเป็นการตรวจสอบคุณภาพเครื่องมือวัดว่าเครื่องมือวัดนั้นมีประสิทธิผลในการจำแนกบุคคลได้ตามสิ่งที่สนใจศึกษาได้มากน้อยเพียงใด โดยวิธีการหาค่าอำนาจจำแนกรายข้อ มีหลายวิธีซึ่งมีรายละเอียดดังนี้ (ณัฐภรณ์ หลาวทอง, 2561; สุจิตรา เทียนสวัสดิ์, 2562; อรพินทร์ ชูชม, 2545; Cohen, Swerdlik, & Sturman, 2013; Gregory, 1996; Kaplan, & Saccuzzo, 2013; Miller et al., 2013)

1. หาค่าอำนาจจำแนกรายข้อ ด้วยดัชนีค่าอำนาจจำแนกรายข้อ (item discrimination index: D) เป็นค่าสถิติที่ได้จากการวิเคราะห์ความแตกต่างของสัดส่วนของผู้ที่ตอบถูกระหว่างกลุ่มที่ได้คะแนนสูงกับกลุ่มที่ได้คะแนนต่ำ ซึ่งระบบการให้คะแนน เป็นแบบ 0 กับ 1 นั่นคือ ถ้าตอบถูกได้ 1 คะแนน ส่วนตอบผิดได้ 0 คะแนน โดยมีวิธีการวิเคราะห์ดังนี้

- 1) ทำการตรวจให้คะแนน แล้วนำคะแนนมารวมกัน แล้วเรียงลำดับคะแนนจากมากไปหาน้อย
- 2) แบ่งกลุ่มผู้ตอบออกเป็น 2 กลุ่มเท่า ๆ กัน นั่นคือ กลุ่มบนเป็นกลุ่มที่ได้คะแนนสูง และกลุ่มล่างเป็นกลุ่มที่ได้คะแนนต่ำ โดย Murphy and Davidshofer (2005) ให้ข้อเสนอแนะว่า ควรแบ่งจำนวนกลุ่มผู้ตอบเป็นร้อยละ 25-35 นอกจากนี้ Waltz, Strickland and Lenz (2005) กล่าวว่า ควรแบ่งจำนวนกลุ่มผู้ตอบที่ร้อยละ 25
- 3) แจงนับจำนวนคนที่ตอบคำถามถูกในแต่ละกลุ่ม ทั้งกลุ่มบนและกลุ่มล่าง แล้วทำการหาสัดส่วนของผู้ตอบในกลุ่มบน (U) และกลุ่มล่าง (L) ที่ตอบคำถามได้ถูกต้อง นั่นคือ

$$U = \frac{\text{จำนวนผู้ตอบในกลุ่มบนที่ตอบถูก}}{\text{จำนวนผู้ตอบในกลุ่มบนทั้งหมด}} \times 100$$

$$L = \frac{\text{จำนวนผู้ตอบในกลุ่มล่างที่ตอบถูก}}{\text{จำนวนผู้ตอบในกลุ่มล่างทั้งหมด}} \times 100$$

- 4) จากนั้น จึงทำการคำนวณค่าดัชนีอำนาจจำแนกรายข้อ (D) โดยคำนวณจากสูตร

$$D = U - L$$

- 5) ทำตามข้อ 3-4 ในทุกข้อคำถาม

ค่าอำนาจจำแนก (D) มีค่าอยู่ระหว่าง -1 ถึง +1 ถ้าค่าอำนาจจำแนกที่คำนวณได้มีค่าเป็นบวกและมีค่าเข้าใกล้ +1 แสดงว่า ข้อคำถามข้อนั้นมีค่าอำนาจจำแนกดี แต่ถ้ามีค่าเป็น 0 หรือเข้าใกล้ -1 แสดงว่า ข้อคำถามข้อนั้นไม่มีค่าอำนาจจำแนก เป็นข้อที่ต้องตัดทิ้ง เนื่องจากไม่สามารถจำแนกกลุ่มสูงออกจากกลุ่มต่ำได้ ส่วนค่าอำนาจจำแนกที่สามารถจำแนกกลุ่มสูงและกลุ่มต่ำได้ จะต้องมีความอำนาจจำแนกตั้งแต่ .20 ขึ้นไป

ตัวอย่าง การคำนวณดัชนีค่าอำนาจจำแนกรายข้อ

แบบทดสอบฉบับหนึ่ง ทดสอบกลุ่มตัวอย่าง 400 คน ใช้เกณฑ์ร้อยละ 25 ในการแบ่งกลุ่มสูงและกลุ่มต่ำ จึงทำให้มีกลุ่มสูงและกลุ่มต่ำ กลุ่มละ 100 คน (Gregory, 1996)

ข้อที่	U	L	D = U - L	แปลผล
1	.49	.23	.26	เป็นข้อที่ดี และมีความยาก
2	.79	.19	.60	เป็นข้อที่ดีมาก แต่หายาก
3	.52	.52	.00	เป็นข้อที่ไม่ดี ควรปรับข้อคำถาม
4	1.00	.00	1.00	เป็นข้อในอุดมคติ
5	.20	.80	-.60	เป็นข้อที่แย่มาก ควรตัดทิ้ง
6	.00	1.00	-1.00	เป็นข้อที่แย่มากที่สุด

จากตัวอย่างข้างต้น เห็นได้ว่า ข้อที่ 1 กลุ่มสูง จำนวน 49 คน ใน 100 คน ตอบถูก ส่วนกลุ่มต่ำ จำนวน 23 คน ใน 100 คน ตอบถูก โดยเมื่อคำนวณแล้ว เป็นข้อที่มีค่าอำนาจจำแนกที่ดี สามารถจำแนกกลุ่มสูง และกลุ่มต่ำได้ แต่ข้อที่ 3 ไม่สามารถจำแนกความแตกต่างของกลุ่มสูงและกลุ่มต่ำได้ เนื่องจากกลุ่มสูงและกลุ่มต่ำ ตอบถูกในจำนวนพอ ๆ กัน อาจทำการปรับปรุงข้อคำถามใหม่ และข้อที่ 5 ค่าอำนาจจำแนกเป็นลบ แสดงว่า กลุ่มต่ำมีจำนวนคนที่ตอบถูกมากกว่ากลุ่มสูง ซึ่งเป็นทิศทางที่ไม่ถูกต้อง ดังนั้น ข้อนี้จึงควรตัดทิ้ง

2. หาค่าอำนาจจำแนกรายข้อ ด้วยการทำการทดสอบค่าที (t-test) เป็นการหาค่าอำนาจจำแนกรายข้อ ของเครื่องมือวัดที่เป็นมาตราประมาณค่า (rating scale) ด้วยการทำการทดสอบค่าที โดยมีวิธีการวิเคราะห์ดังนี้

- 1) ทำการตรวจให้คะแนน แล้วนำคะแนนมารวมกัน แล้วเรียงลำดับคะแนนจากมากไปหาน้อย
- 2) แบ่งกลุ่มผู้ตอบออกเป็น 2 กลุ่มเท่า ๆ กัน เช่นเดียวกับวิธีการหาค่าอำนาจจำแนกรายข้อวิธีแรก
- 3) จากนั้น จึงทำการคำนวณค่าอำนาจจำแนกรายข้อด้วยการทดสอบค่าที (t-test) ด้วยโปรแกรมทางสถิติ

ตัวอย่าง การคำนวณค่าอำนาจจำแนกรายข้อด้วยการทดสอบค่าที

แบบวัดพฤติกรรมป้องกันความก้าวร้าว ทดสอบกลุ่มตัวอย่าง 32 คน มีวิธีการวิเคราะห์ตามโปรแกรมทางสถิติ ดังนี้

(1) คำนวณหาคะแนนรวม โดยใช้คำสั่ง Compute

(2) แบ่งกลุ่มสูงและกลุ่มต่ำ จากคะแนนรวม โดยใช้วิธีแบ่งกลุ่มสูง ร้อยละ 25 และกลุ่มต่ำ ร้อยละ 25 โดยใช้คำสั่ง: Analyze > Descriptive Statistics > Frequencies นำตัวแปรที่ศึกษาใส่เข้าไปในช่อง 'Variables' แล้วกด Statistics เลือก Percentile แล้วกด Add 25 และ 75 จึงกด Continue และ OK

Behavior		
N	Valid	32
	Missing	0
Percentiles	25	52.5000
	75	67.5000

สำหรับการหา Percentile ที่ 25 มีคะแนนรวมเท่ากับ 52.5 หมายความว่า มีจำนวนร้อยละ 25 ที่ได้คะแนนรวมต่ำกว่า 52.5 และ Percentile ที่ 75 มีคะแนนรวมเท่ากับ 67.5 หมายความว่า มีจำนวนร้อยละ 75 ที่ได้คะแนนรวมต่ำกว่า 67.5 หรือมีจำนวนร้อยละ 25 ที่ได้คะแนนรวมมากกว่า 67.5 โดยให้แปลงข้อมูลใหม่ให้เป็นกลุ่มสูงและกลุ่มต่ำ นั่นคือ กลุ่มต่ำ ให้รหัสเป็น 1 และกลุ่มสูงให้รหัสเป็น 2 โดยใช้คำสั่ง: Transform > Recode into Different Variables

(3) คำนวณหาค่าอำนาจจำแนกรายข้อ ด้วยการทดสอบค่าที โดยใช้คำสั่ง Analyze > Compare Means > Independent Samples T-Test แล้วใส่ข้อคำถามลงในช่อง Test Variable และใส่ Group ลงไป

ในช่อง Grouping variable และกำหนดค่า (Define Groups...) ให้ Group 1 เป็นกลุ่มสูง ใส่เลข 2 และ Group 2 เป็นกลุ่มต่ำ ใส่เลข 1 ทำให้ได้ผลการวิเคราะห์ดังนี้

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
b1	2.00	8	5.00	.000	.000
	1.00	8	4.00	.926	.327
b2	2.00	8	5.00	.000	.000
	1.00	8	4.00	.535	.189
b3	2.00	8	5.00	.000	.000
	1.00	8	4.25	.707	.250
b4	2.00	8	5.00	.000	.000
	1.00	8	3.75	.463	.164
b5	2.00	8	4.88	.354	.125
	1.00	8	3.50	.535	.189

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
b1	Equal variances assumed	21.000	.000	3.055	14	.009	1.000	.327	.298	1.702
	Equal variances not assumed			3.055	7.000	.018	1.000	.327	.226	1.774
b2	Equal variances assumed	2.333	.149	5.292	14	.000	1.000	.189	.595	1.405
	Equal variances not assumed			5.292	7.000	.001	1.000	.189	.553	1.447
b3	Equal variances assumed	18.290	.001	3.000	14	.010	.750	.250	.214	1.286
	Equal variances not assumed			3.000	7.000	.020	.750	.250	.159	1.341
b4	Equal variances assumed	21.000	.000	7.638	14	.000	1.250	.164	.899	1.601
	Equal variances not assumed			7.638	7.000	.000	1.250	.164	.863	1.637
b5	Equal variances assumed	9.000	.010	6.068	14	.000	1.375	.227	.889	1.861
	Equal variances not assumed			6.068	12.141	.000	1.375	.227	.882	1.868

จากผลการวิเคราะห์ค่าอำนาจจำแนกรายข้อด้วยการทดสอบค่าที กลุ่มสูง (group 2) มีค่าเฉลี่ยสูงกว่ากลุ่มต่ำ (group 1) ทั้ง 5 ข้อ และเมื่อพิจารณาค่า independent samples t-test หลักการคือ เลือกใช้ข้อที่ค่า t มีค่ามากกว่าหรือเท่ากับ 2.0 ขึ้นไป (ดูจเดือน พันธุนาวัน, 2551) พบว่า ค่า t ของทุกข้อมีค่าเกิน 2.0 นั้นหมายความว่า กลุ่มสูงมีคะแนนเฉลี่ยพฤติกรรมป้องกันความก้าวร้าวที่แตกต่างจากกลุ่มต่ำ แสดงให้เห็นว่าข้อคำถามทั้ง 5 ข้อ มีอำนาจจำแนก สามารถจำแนกกลุ่มคนที่มีความประพฤติป้องกันความก้าวร้าวสูงออกจากกลุ่มคนที่มีความประพฤติป้องกันความก้าวร้าวต่ำได้

3. หาค่าอำนาจจำแนกรายข้อ ด้วยการหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างรายข้อกับคะแนนรวม (Item-total correlation) เป็นการหาค่าอำนาจจำแนกรายข้อของเครื่องมือวัดที่เป็นมาตราประมาณค่า (rating scale) ด้วยการหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน โดยมีวิธีการวิเคราะห์ดังนี้

- 1) ทำการตรวจให้คะแนน ในแต่ละข้อ
- 2) จากนั้น จึงทำการคำนวณค่าอำนาจจำแนกรายข้อด้วยการหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างรายข้อกับคะแนนรวม ด้วยโปรแกรมทางสถิติ

ตัวอย่าง การคำนวณค่าอำนาจจำแนกรายข้อด้วยการหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างรายข้อกับคะแนนรวม

แบบวัดพฤติกรรมป้องกันความก้าวร้าว ทดสอบกลุ่มตัวอย่าง 32 คน มีวิธีการวิเคราะห์ตามโปรแกรมทางสถิติ ดังนี้

(1) คำนวณหาค่าอำนาจจำแนกรายข้อ ด้วยการหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างรายข้อกับคะแนนรวม โดยใช้คำสั่ง Analyze > Scale > Reliability Analysis แล้วใส่ข้อคำถามลงไปเป็นช่อง Items และกด Statistics จากนั้นเลือก Item, Scale และ Scale if item delete แล้วกด Continue และ OK ทำให้ได้ผลการวิเคราะห์ดังนี้

Case Processing Summary

		N	%
Cases	Valid	32	100.0
	Excluded ^a	0	.0
	Total	32	100.0

a. Listwise deletion based on all variables in the procedure.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
b1	55.19	57.964	.332	.928
b2	55.31	57.448	.427	.925
b3	55.13	57.468	.502	.923
b4	55.28	54.273	.761	.916
b5	55.53	54.193	.664	.919

จากผลการวิเคราะห์ค่าอำนาจจำแนกรายข้อด้วยหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างรายข้อกับคะแนนรวม ให้พิจารณาตรงค่า Corrected Item-Total Correlation ควรมีค่าตั้งแต่ .30 ขึ้นไป (Aiken, 2003; Nunnally & Bernstein, 1994; Streiner, Norman, & Cairney, 2015) หากมีค่าต่ำกว่าเกณฑ์ที่กำหนด นั่นคือต่ำกว่า .30 ให้พิจารณาปรับข้อคำถามหรือตัดทิ้ง จากผลการวิเคราะห์ดังกล่าว แสดงให้เห็นว่า ข้อคำถามทั้ง 5 ข้อ มีอำนาจจำแนกที่ดี สามารถจำแนกกลุ่มคนที่มีความสามารถสูงออกจากกลุ่มคนที่มีความสามารถต่ำได้

สำหรับการแปลความหมายค่าสหสัมพันธ์ระหว่างรายข้อกับคะแนนรวม (ค่า r) มีการแปลความหมายไว้ดังนี้

1) เมื่อค่า r เป็นบวก แสดงว่า ข้อคำถามมีอำนาจจำแนกที่ดี แต่ต้องมีค่ามากกว่า .30 จึงจะถือว่าเป็นอำนาจจำแนกดี กล่าวคือ บุคคลที่ได้คะแนนสูงในข้อนั้น จะได้คะแนนรวมสูงตามด้วย และในทางตรงกันข้าม บุคคลที่ได้คะแนนต่ำข้อนั้น จะได้คะแนนรวมต่ำด้วย แสดงให้เห็นว่า ข้อคำถามข้อนั้นสามารถจำแนกบุคคลที่ได้คะแนนสูงออกจากบุคคลที่ได้คะแนนต่ำ

2) เมื่อค่า r เป็นลบ แสดงว่า ข้อคำถามไม่สามารถจำแนกได้ นั่นคือ จำแนกไปในทางตรงกันข้ามกับข้อที่ 1 บุคคลที่ได้คะแนนสูงในข้อนั้น จะได้คะแนนรวมต่ำ หรือบุคคลที่ได้คะแนนต่ำข้อนั้น จะมีคะแนนรวมสูง และข้อคำถามไม่สอดคล้องกับคะแนนรวม

ความยากง่ายรายข้อ (Item Difficulty)

ความยากง่ายรายข้อ เป็นสัดส่วนหรือร้อยละของผู้ตอบในกลุ่มที่ตอบข้อคำถามข้อนั้นถูก ความยากง่ายรายข้อเหมาะกับแบบทดสอบที่วัดทางด้านความรู้คิด (cognitive) ที่เป็นการให้คะแนนรายข้อในลักษณะของการตอบถูกและผิด อาจจะเป็นรูปแบบข้อคำถามถูกผิด หรือข้อคำถามหลายตัวเลือกก็ได้ กล่าวคือ การหาค่าความยากง่ายรายข้อเป็นเครื่องมือที่มีประโยชน์ในการปรับเปลี่ยนหรือตัดข้อคำถามทิ้ง โดยมีวิธีการหาค่าความยาก

ง่ายรายข้อ ซึ่งมีรายละเอียดดังนี้ (ณัฐภรณ์ หลาวทอง, 2561; อรพินทร์ ชูชม, 2545; Cohen et al., 2013; Gregory, 1996; Kaplan, & Saccuzzo, 2013; Miller et al., 2013)

ดัชนีความยากง่ายรายข้อ (Item difficulty index) สามารถคำนวณได้ดังนี้

$$P = \frac{\text{จำนวนผู้ตอบที่ตอบข้อนั้นถูก}}{\text{จำนวนคนทั้งหมด}}$$

คุณลักษณะที่สำคัญของความยากง่ายรายข้อ มีดังนี้

- (1) ค่าดัชนีความยากง่ายรายข้อ มีค่าตั้งแต่ 0 ถึง 1 และเป็นค่าบวกเท่านั้น
 - (2) ค่า P ขึ้นอยู่กับจำนวนผู้ตอบที่ตอบข้อนั้นถูก
 - ค่า P สูง แสดงว่า มีจำนวนผู้ตอบจำนวนมากที่ตอบถูก แสดงว่า ข้อคำถามเป็นข้อที่ง่าย
 - ค่า P ต่ำ แสดงว่า มีจำนวนผู้ตอบจำนวนน้อยที่ตอบถูก หรือผู้ตอบส่วนใหญ่ตอบข้อนั้นผิด แสดงว่า ข้อคำถามเป็นข้อที่ยาก
 - ค่า P เท่ากับ 1 แสดงว่า ผู้ตอบทุกคนตอบข้อนั้นถูกหมด แสดงว่า ข้อคำถามนั้นง่ายมาก
 - ส่วนค่า P เท่ากับ 0 แสดงว่า ผู้ตอบทุกคนตอบข้อนั้นผิดหมด แสดงว่า ข้อคำถามนั้นยากมาก
 - (3) ค่า P ไม่สามารถบอกได้ว่าข้อคำถามแต่ละข้อดีหรือไม่ดี แต่สามารถบอกได้ว่าข้อคำถามแต่ละข้อง่ายหรือยาก และค่า P เป็นเครื่องมือที่ช่วยคัดเลือกข้อคำถามให้เป็นไปตามจุดมุ่งหมายของแบบทดสอบ
- สำหรับการแปลความหมายค่าความยากง่ายรายข้อ (ค่า P) มีการแปลความหมายไว้ดังนี้

- ค่า P = 0.00 – 0.20 หมายถึง ข้อคำถามยากมาก
- ค่า P = 0.21 – 0.40 หมายถึง ข้อคำถามค่อนข้างยาก
- ค่า P = 0.41 – 0.60 หมายถึง ข้อคำถามยากปานกลาง
- ค่า P = 0.61 – 0.80 หมายถึง ข้อคำถามค่อนข้างง่าย
- ค่า P = 0.81 – 1.00 หมายถึง ข้อคำถามง่ายมาก

โดยทั่วไป ค่าความยากง่ายรายข้อ ควรมีค่า P อยู่ระหว่าง .30 - .70 (Gregory, 1996; Kaplan, & Saccuzzo, 2013) นอกจากนี้ Cohen et al. (2013) กล่าวว่า ควรมีค่า P อยู่ระหว่าง .30 - .80 และณัฐภรณ์ หลาวทอง (2561) กล่าวว่า ในการทดสอบแบบอิงกลุ่มนิยมคัดเลือกข้อคำถามที่มีความยากง่ายรายข้ออยู่ระหว่าง .20 - .80 เกือบไว้ใช้ แต่จะตัดข้อที่มีความยากมากและง่ายมากออกไป เนื่องจากข้อคำถามเหล่านี้มีโอกาสที่ค่าอำนาจจำแนกจะมีค่าน้อยได้ ซึ่งมีผลให้คุณภาพของเครื่องมือด้อยลงไปได้

Gregory (1996) กล่าวถึง แบบทดสอบที่ใช้ในการคัดเลือกนิสิตนักศึกษาเข้าศึกษาต่อในมหาวิทยาลัยที่ต้องการรับนิสิตนักศึกษาเข้าศึกษาต่อในจำนวนน้อย จึงควรมีข้อคำถามที่ยากมากรวมอยู่ด้วย หรือแม้แต่แบบทดสอบที่ออกแบบมาเพื่อเด็กนักเรียนเข้ามาเรียนในโปรแกรมซ่อมเสริมทางการศึกษา จึงควรมีข้อคำถามที่ง่ายมากรวมอยู่ด้วยเช่นกัน ดังนั้น ในการคัดเลือกข้อคำถามที่ใช้ว่ามีความยากง่ายรายข้อในระดับใด จึงควรพิจารณาไปถึงจุดมุ่งหมายของการทดสอบด้วย ความยากง่ายรายข้อนอกจากจะใช้ในการคัดเลือกข้อคำถามแล้วยังมีประโยชน์ในการนำความยากง่ายของข้อคำถามมาจัดลำดับข้อคำถาม อาจจะเริ่มจากข้อคำถามที่ง่ายไปจนถึงข้อคำถามที่ยาก เป็นต้น

ตัวอย่าง การคำนวณดัชนีค่าความยากง่ายรายข้อ

แบบทดสอบฉบับหนึ่ง เป็นลักษณะแบบเลือกตอบ จำนวน 20 ข้อ ทดสอบกลุ่มตัวอย่าง 100 คน

ข้อที่	คนที่						จำนวนคำตอบ ที่ถูกต้อง	P
	1	2	3	4	...	100		
1	1	0	1	1	...	1	70	.70
2	1	0	0	0	...	0	10	.10
3	1	1	1	1	...	1	100	1.00
4	1	1	0	0	...	1	45	.45
5	1	0	0	1	...	1	56	.56
...
20	0	1	1	1	...	1	85	.85

จากตัวอย่างข้างต้น เห็นได้ว่า ข้อที่ 1 มีจำนวนคนที่ตอบถูก จำนวน 70 คน จากทั้งหมด 100 คน ดังนั้น ค่า P ของข้อที่ 1 = $70/100 = .70$ โดยเกณฑ์ค่าความยากง่ายรายข้อที่เหมาะสมควรอยู่ในช่วง .30 - .70 สำหรับข้อที่ 1 ถือว่า มีความยากง่ายในระดับที่เหมาะสม เช่นเดียวกับข้อ 4 และ 5 ส่วนข้อที่ 3 และ 20 ซึ่งมีค่า P สูง คือ มากกว่า .85 จะเกิดปัญหาในการจำแนกผู้ตอบ เนื่องจากผู้ตอบส่วนใหญ่ตอบข้อนี้ถูก เป็นข้อที่ง่ายมาก และข้อที่ 2 มีค่า P ต่ำ คือ มีค่าเท่ากับ .10 ข้อนี้ไม่สามารถบอกความแตกต่างของผู้ตอบได้มากนัก เนื่องจากคนส่วนใหญ่มักตอบข้อนี้ผิด กล่าวโดยสรุปคือ ข้อคำถามที่มีความยากง่ายรายข้อที่เหมาะสม คือ ข้อที่ 1, 4 และ 5 ส่วนข้อคำถามที่มีความยากง่ายรายข้อไม่เป็นไปตามเกณฑ์ คือ ข้อคำถามที่ยากเกินไป หรือง่ายเกินไป ได้แก่ ข้อที่ 2, 3 และ 20 ให้พิจารณาตัดข้อคำถามทิ้ง เนื่องจากเป็นข้อที่ไม่สามารถบอกความแตกต่างของบุคคลได้ดีนัก

อำนาจตัวหลง (Distractor Power)

ตัวหลง หมายถึง ตัวเลือกอื่น ๆ ที่ไม่ใช่ตัวเลือกที่ถูกต้องในแบบทดสอบที่วัดด้านความรู้คิด (cognitive) ที่มีหลายตัวเลือก (multiple choice) เช่น แบบทดสอบวัดผลสัมฤทธิ์ เป็นต้น

การวิเคราะห์อำนาจตัวหลง มีความสำคัญ ได้แก่ (อรพินทร์ ชูชม, 2545)

1. เพื่อตรวจสอบจำนวนผู้ตอบที่เลือกตอบในแต่ละตัวเลือก
 2. ทำให้ทราบว่าตัวหลงนั้น ๆ มีประโยชน์มากน้อยเพียงใด เช่น ในข้อสอบปรนัยที่มีตัวเลือก 4 ตัวเลือก ถ้าในข้อคำถามนั้น ๆ มีตัวเลือกที่ผิดชัดเจนอยู่ 1-2 ตัวเลือก แม้ผู้ตอบจะไม่มีความรู้เกี่ยวกับข้อคำถาม ก็ยังคงเดาคำตอบได้ว่าตัวเลือกนั้นผิด ตัวเลือกนั้นจะไม่ได้รับการเลือก ถูกขจัดออกไป เป็นการเพิ่มโอกาสการเดาคำตอบที่ถูกต้องได้มากยิ่งขึ้น
 3. ช่วยระบุว่าตัวเลือกไหนจำเป็นต้องทำการปรับปรุง นั่นคือ ตัวหลงที่เขียนได้ไม่ดี คนได้คะแนนสูงขึ้นซึ่งอาจเกิดจากการเดา การขจัดตัวเลือกออกไปได้ตัวหรือสองตัว ทำให้คนเดาได้ง่ายขึ้น ดังนั้น จึงต้องทำการปรับปรุงตัวเลือกนั้นให้เป็นตัวหลงที่ดีและเหมาะสมมากขึ้น
 4. เอื้อต่อการพัฒนาความเชื่อมั่นและความเที่ยงตรงของแบบทดสอบ ตัวหลงที่ไม่ดี จะไปลดคุณภาพของแบบทดสอบ เนื่องจากทำให้เกิดการเดาได้มากขึ้น รวมทั้งการเดานั้นจะไปลดความน่าเชื่อถือของแบบทดสอบอีกด้วย
- อำนาจตัวหลงหรือตัวหลงที่คาดหวัง สามารถคำนวณได้จาก

$$\text{อำนาจตัวหลงที่คาดหวัง} = \frac{\text{จำนวนผู้ตอบที่ตอบข้อนั้นผิด}}{\text{จำนวนตัวหลง}}$$

อำนาจตัวหลงที่คาดหวังในแบบทดสอบชนิดหลายตัวเลือกที่ดี จำเป็นต้องมีลักษณะที่สำคัญ 2 ข้อ ได้แก่ (อรพินทร์ ชูชม, 2545)

1. บุคคลที่มีความรู้จะเลือกคำตอบที่ถูกต้อง และบุคคลที่ขาดความรู้จะเลือกจากตัวเลือกที่มีอยู่เป็นไปอย่างสุ่ม ดังนั้น บุคคลที่ขาดความรู้ควรเลือกตัวหลงอย่างสุ่มโดยมีสัดส่วนเท่า ๆ กันในแต่ละตัวหลง นั่นคือผู้ตอบที่ตอบผิดในข้อคำถามนั้นควรมีการกระจายในจำนวนเท่ากันในตัวหลงแต่ละตัว
2. ในการวิเคราะห์อำนาจตัวหลงว่าตัวหลงใดใช้ได้หรือไม่ มักจะเกิดขึ้นในกรณีที่ไม่มีคนเลือกตัวหลงนั้น หรือมีผู้ตอบเลือกน้อยกว่าอำนาจตัวหลงที่คาดหวังไว้มาก ตัวหลงนั้น ๆ ควรมีการตรวจสอบและทำการปรับปรุงใหม่ ซึ่งอาจเป็นไปได้ว่าตัวหลงนั้นเป็นตัวเลือกที่เป็นคำตอบผิดที่ชัดเจน หรือกรณีที่ผู้ตอบเลือกตัวหลงนั้นมากกว่าที่คาดหวังไว้มาก ตัวหลงนั้นควรต้องทำการปรับปรุงเช่นเดียวกัน ซึ่งอาจเป็นไปได้ว่าตัวหลงนั้นมีลักษณะคล้ายคำตอบที่ถูกค่อนข้างมาก แม้กระทั่งบุคคลที่มีความรู้ก็ยังเลือกตัวหลงนี้

ตัวอย่าง การวิเคราะห์ตัวลวง

แบบทดสอบฉบับหนึ่ง เป็นลักษณะแบบเลือกตอบ จำนวน 5 ข้อ ทดสอบกลุ่มตัวอย่าง 100 คน

ข้อที่	จำนวนคนที่ตอบถูก	จำนวนคนที่ตอบผิด	จำนวนคนที่คาดว่าจะเลือกตัวลวง	จำนวนคนที่เลือกตัวเลือก			
				A	B	C	D
1	70	30	10	10	70*	11	9
2	10	90	30	50	20	10*	20
3	90	10	3.3	4	3	3	90*
4	45	55	18.3	45*	17	18	20
5	56	44	14.7	30	56*	8	6

จากตัวอย่างข้างต้น แสดงตัวอย่างการวิเคราะห์ตัวลวงจากข้อสอบ 5 ข้อ ซึ่งแต่ละข้อจะมีตัวเลือก 4 ตัว มีคำตอบที่ถูก 1 ตัวเลือก ส่วนอีก 3 ตัวเลือก เป็นตัวลวง พบว่า ข้อที่ 1 และข้อที่ 4 แสดงรูปแบบการตอบที่ดีและเหมาะสม เนื่องจากมีคนเลือกที่ผิดกระจายอยู่เกือบเท่ากันทุกตัวเลือก ในข้อที่ 2 พบว่า ตัวลวง A มีปัญหาในการลวง เนื่องจากมีผู้ตอบจำนวนมากเลือกตัวลวงนี้ ซึ่งเป็นตัวเลือกที่ผิด ดังนั้น จึงควรปรับปรุงตัวลวงนี้ใหม่ ตัวเลือก A อาจทำให้เกิดการเข้าใจผิดว่าเป็นคำตอบที่ถูกต้องอาจเป็นเพราะภาษาหรือถ้อยคำที่ใช้ทำให้เกิดการสับสนได้ ส่วนข้อที่ 3 เป็นข้อที่ง่ายมาก มีคนตอบผิดเพียง 10 คนเท่านั้น คำถามข้อนี้น่าที่จะประเมินความเหมาะสมของตัวลวงได้

ความสัมพันธ์ระหว่างการวิเคราะห์รายข้อรูปแบบต่าง ๆ

สำหรับค่าอำนาจจำแนกรายข้อ ค่าความยากง่ายรายข้อ และอำนาจตัวลวง มีคุณลักษณะที่แตกต่างกันไป แต่คุณสมบัติของการวิเคราะห์รายข้อเหล่านี้มีความสัมพันธ์กัน ซึ่งมีรายละเอียดดังนี้

1. ความสัมพันธ์ระหว่างความยากง่ายและอำนาจจำแนก

ความยากง่ายมีผลต่ออำนาจจำแนกรายข้อโดยตรง เช่น ถ้าทุกคนเลือกคำตอบที่ถูก (ค่า $P = 1$) หรือเลือกคำตอบที่ผิด (ค่า $P = 0$) ค่าอำนาจจำแนกรายข้อจะเป็น 0 นั่นคือ ข้อที่ง่ายมากหรือยากมากไม่สามารถใช้จำแนกผู้ตอบที่ได้คะแนนสูงออกจากผู้ตอบที่ได้คะแนนต่ำ ดังนั้น ค่า P ที่สุดโต่ง คือเข้าใกล้ 1 หรือ 0 จะมีค่าอำนาจจำแนกรายข้อน้อยกว่าค่า P ที่อยู่ตรงกลาง เท่ากับ .50 ซึ่งจะมีค่าอำนาจจำแนกสูงสุด เท่ากับ 1

2. ความสัมพันธ์ระหว่างความยากง่ายและอำนาจตัวลวง

ความยากง่ายมีผลต่อคุณภาพของตัวลวงโดยตรง เช่น ถ้าคนส่วนใหญ่เลือกคำตอบที่ถูก (ค่า P เข้าใกล้ 1) แสดงว่า ผู้ตอบรู้ได้ทันทีว่าตัวลวงนั้นเป็นตัวเลือกที่ผิด ในขณะเดียวกัน ถ้าคนส่วนใหญ่เลือกคำตอบที่ผิด (ค่า P เข้าใกล้ 0) แสดงว่า ตัวลวงที่มีอยู่มีความใกล้เคียงกับตัวเลือกที่ถูกเป็นอย่างมาก ซึ่งเป็นสิ่งที่ทำให้ข้อนั้นมีความยาก

3. ความสัมพันธ์ระหว่างอำนาจตัวลงและอำนาจจำแนก

คุณภาพตัวลงมีผลต่ออำนาจจำแนกรายข้อได้ เนื่องจากคุณภาพตัวลงส่งผลต่อความยากง่าย และความยากง่ายมีผลต่ออำนาจจำแนกโดยตรง ดังนั้น ตัวลงจึงส่งผลทางอ้อมต่ออำนาจจำแนก เช่น ถ้าค่าอำนาจจำแนกเป็นลบ ต้องทำการตรวจสอบตัวลงอย่างรอบคอบ ซึ่งอาจเป็นไปได้ว่า ผู้ตอบที่ได้คะแนนสูงหรือผู้ที่มีความรู้ อาจถูกทำให้เข้าใจผิดด้วยภาษาหรือถ้อยคำของตัวลง

จากที่กล่าวมาทั้งหมด อำนาจจำแนกรายข้อ ความยากง่ายรายข้อ และอำนาจตัวลง เป็นการวิเคราะห์รายข้อ เพื่อตรวจสอบคุณภาพของเครื่องมือ ที่ใช้กระบวนการทางสถิติในการกระทำการวิเคราะห์ ซึ่งเป็นการวิเคราะห์รายข้อในเชิงปริมาณ นอกจากนี้ ยังมีการวิเคราะห์รายข้อในเชิงคุณภาพ ซึ่งเป็นการวิเคราะห์รายข้อโดยใช้เทคนิควิธีการทางด้านคุณภาพ ซึ่งกล่าวถึงในลำดับถัดไป

การวิเคราะห์รายข้อเชิงคุณภาพ (Qualitative Item Analysis)

สำหรับการวิเคราะห์รายข้อ มักจะมีการคำนวณทางสถิติ หรือใช้กระบวนการทางด้านปริมาณ ซึ่งเป็นแนวคิดวิธีการหนึ่งเพื่อทำความเข้าใจคุณภาพของเครื่องมือวัด นอกจากนี้ ยังมีวิธีการอีกแนวคิดหนึ่ง นั่นคือ วิธีการวิจัยทางด้านคุณภาพ (qualitative methods) ซึ่งเป็นเทคนิคของการสร้างและการวิเคราะห์ข้อมูลที่มุ่งเน้นถ้อยคำทางด้านภาษามากกว่ากระบวนการทางคณิตศาสตร์หรือสถิติ

Cohen et al. (2013) กล่าวว่า การวิเคราะห์รายข้อเชิงคุณภาพ เป็นกระบวนการที่ไม่ได้ใช้สถิติ เพื่อค้นหาว่าข้อคำถามที่สร้างขึ้นมาใช้ได้หรือไม่ ลักษณะของการวิเคราะห์รายข้อจะทำการเปรียบเทียบข้อคำถามแต่ละข้อกับข้อคำถามข้ออื่น ๆ และเปรียบเทียบกับภาพรวมของแบบทดสอบ/แบบสอบถาม ซึ่งวิธีการวิจัยเชิงคุณภาพเกี่ยวข้องกับการค้นหาประเด็นปัญหาที่เกิดขึ้นผ่านถ้อยคำทางภาษา เช่น การสัมภาษณ์ และการอภิปรายกลุ่ม กับผู้ทำแบบทดสอบ/แบบสอบถาม และบุคคลอื่น ๆ ที่เกี่ยวข้อง

โดยการวิเคราะห์รายข้อเชิงคุณภาพที่นักวิจัยให้ความสนใจ มีหัวข้อและรายละเอียดสรุปได้ดังตารางที่ 7-1

ตารางที่ 7-1 การวิเคราะห์รายข้อเชิงคุณภาพรูปแบบต่าง ๆ

หัวข้อ	ตัวอย่างคำถาม
ความไวทางวัฒนธรรม (Cultural Sensitivity)	คุณรู้สึกที่ข้อคำถามหรือมุมมองของการทดสอบเป็นการเลือกปฏิบัติกับบุคคลหรือไม่ ถ้าใช่ ทำไม?
ความเที่ยงตรงเชิงปรากฏ (Face Validity)	แบบทดสอบวัดในสิ่งที่คาดว่าจะต้องการวัดหรือไม่ ถ้าไม่ อะไรเป็นสิ่งที่ตรงข้ามกับความคาดหวังของคุณ?
ผู้จัดการทดสอบ (Test Administrator)	พฤติกรรมของผู้จัดการทดสอบมีผลต่อการทำแบบทดสอบหรือไม่ ถ้าใช่ อย่างไร?
สภาพแวดล้อมการทดสอบ (Test Environment)	เงื่อนไขในห้องสอบมีผลต่อการทำแบบทดสอบหรือไม่ ถ้าใช่ อย่างไร?

ตารางที่ 7-1 (ต่อ) การวิเคราะห์รายชื่อเชิงคุณภาพรูปแบบต่าง ๆ

หัวข้อ	ตัวอย่างคำถาม
ความยุติธรรมของแบบทดสอบ (Test Fairness)	คุณคิดว่าแบบทดสอบมีความยุติธรรมต่อสิ่งที่ต้องการจะวัดหรือไม่ ทำไม?
ภาษาของแบบทดสอบ (Test Language)	มีคำชี้แนะหรือคำสั่งของแบบทดสอบที่ยากต่อการทำความเข้าใจหรือไม่?
ความยาวของแบบทดสอบ (Test Length)	คุณรู้สึกว่าความยาวของแบบทดสอบสอดคล้องกับเวลาที่ใช้ในการทำแบบทดสอบ และจำนวนข้อคำถามเป็นอย่างไร?
การเดาของผู้ทำแบบทดสอบ (Testtaker's Guessing)	คุณเดาข้อคำถามหรือไม่ ร้อยละของข้อคำถามที่คุณเดาเป็นเท่าไร ใช้เทคนิคการเดาหรือไม่ หรือแค่สุ่ม?
ความซื่อสัตย์ของผู้ทำแบบทดสอบ (Testtaker's Integrity)	คุณคิดว่ามีการลอกในระหว่างการทดสอบหรือไม่ ถ้าใช่ จงบอกวิธีการที่คุณคิดว่าอาจมีการใช้?
สภาพจิตใจ/ร่างกายของผู้ทำแบบทดสอบ ในช่วงเริ่มต้น (Testtaker's Mental/Physical State Upon Entry)	ให้คุณบอกถึงสภาพจิตใจในช่วงเริ่มต้นทำแบบทดสอบว่าเป็นอย่างไร คุณคิดว่าสภาพจิตใจมีผลต่อการทำแบบทดสอบหรือไม่ ถ้าใช่ อย่างไร และให้คุณบอกถึงสภาพร่างกายในช่วงเริ่มต้นทำแบบทดสอบว่าเป็นอย่างไร คุณคิดว่าสภาพร่างกายมีผลต่อการทำแบบทดสอบหรือไม่ ถ้าใช่ อย่างไร?
สภาพจิตใจ/ร่างกายของผู้ทำแบบทดสอบ ในระหว่างการทำแบบทดสอบ (Testtaker's Mental/Physical State During the Test)	ให้คุณบอกถึงสภาพจิตใจในระหว่างทำแบบทดสอบว่าเป็นอย่างไร คุณคิดว่าสภาพจิตใจมีผลต่อการทำแบบทดสอบหรือไม่ ถ้าใช่ อย่างไร และให้คุณบอกถึงสภาพร่างกายในระหว่างทำแบบทดสอบว่าเป็นอย่างไร คุณคิดว่าสภาพร่างกายมีผลต่อการทำแบบทดสอบหรือไม่ ถ้าใช่ อย่างไร?
ความประทับใจในภาพรวมของผู้ทำแบบทดสอบ (Testtaker's Overall Impressions)	ความประทับใจในภาพรวมของแบบทดสอบนี้มีอะไรบ้าง มีข้อเสนอแนะอะไรบ้าง สำหรับการปรับปรุงแบบทดสอบให้ดีขึ้น?
ความชอบของผู้ทำแบบทดสอบ (Testtaker's Preferences)	คุณเจอส่วนที่เกี่ยวข้องกับการศึกษา ความบันเทิงหรือสิ่งอื่น ๆ ในแบบทดสอบหรือไม่ อะไรคือสิ่งที่คุณชอบหรือไม่ชอบ และมีส่วนไหนที่ทำให้เกิดความวิตกกังวลหรือโกรธ ทำไม?
การเตรียมตัวของผู้ทำแบบทดสอบ (Testtaker's Preparation)	คุณเตรียมตัวทำแบบทดสอบนี้อย่างไร ถ้าคุณไปให้คำแนะนำผู้อื่นถึงวิธีการเตรียมตัว คุณจะบอกอะไรพวกเขาเหล่านั้น?

นอกจากนี้ การวิเคราะห์รายชื่อเชิงคุณภาพ ยังมีวิธีการอื่น ๆ (Cohen et al., 2013; Miller et al., 2013) ได้แก่

1. การขอความเห็นจากผู้เชี่ยวชาญ (Expert Panels) โดยการทบทวนความไว (Sensitivity review) เป็นการศึกษาเกี่ยวกับข้อคำถามในกระบวนการของการพัฒนาแบบทดสอบ ซึ่งข้อคำถามแต่ละข้อจะถูกตรวจสอบถึงความยุติธรรมต่อผู้ทำแบบทดสอบทุกคน และการใช้ภาษาที่รุนแรง การเหมารวม หรือสภาพแวดล้อมต่าง ๆ ซึ่งวิธีการนี้จะขอความเห็นจากผู้เชี่ยวชาญเกี่ยวกับอคติที่อาจเกิดขึ้นจากการทำแบบทดสอบ

2. วิธีการคิดตั้ง ๆ (Think Aloud) เป็นวิธีการเชิงคุณภาพที่ออกแบบมาเพื่อสอบถามความคิดเห็นของผู้ทำแบบทดสอบในระหว่างการทำแบบทดสอบ โดยมีวิธีการ คือ การถามตอบระหว่างผู้ทำแบบทดสอบและผู้ให้ทำแบบทดสอบแบบ 1 ต่อ 1 นั่นคือ จะให้ผู้ทำแบบทดสอบเป็นคนทำแบบทดสอบ แล้วดูผลการตอบสนองของผู้ทำแบบทดสอบแต่ละข้อ เช่น ถ้าเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน การใช้คำอาจมีประโยชน์ในแง่การประเมินการตีความข้อคำถามผิดไป

ไม่ว่าจะใช้วิธีการใดก็ตาม ทั้งจากผู้ทำแบบทดสอบหรือผู้เชี่ยวชาญ ผู้พัฒนาแบบทดสอบอาจต้องการปรับเปลี่ยนหรือปรับปรุงแบบทดสอบ การปรับปรุงข้อคำถามมักเกี่ยวข้องกับการเปลี่ยนถ้อยคำ การตัดข้อคำถามทิ้ง หรือการสร้างข้อคำถามใหม่ ซึ่งกระบวนการนี้ถือเป็นขั้นตอนหนึ่งของการพัฒนาและตรวจสอบคุณภาพเครื่องมือวัด

และเมื่อพิจารณาทั้งการวิเคราะห์รายข้อเชิงปริมาณและเชิงคุณภาพแล้ว สุจิตรา เทียนสวัสดิ์ (2562) ได้ให้ความเห็นที่น่าสนใจว่า หากวิเคราะห์ทั้งเชิงปริมาณและเชิงคุณภาพได้ผลที่เหมือนกัน นักวิจัย/นักวิชาการไม่จำเป็นต้องใช้ทั้งสองวิธี ดังนั้นหากต้องการจะใช้ทั้ง 2 วิธี ผู้สร้างเครื่องมือน่าจะคาดหวังเพื่อให้ได้ผลที่แตกต่างกัน เพราะแต่ละวิธีมีข้อดีและข้อจำกัดแตกต่างกัน การใช้ร่วมกันทั้ง 2 วิธี จึงเสมือนการทำวิจัยแบบสามเส้า (Triangulation) เพื่อตรวจสอบผลที่ได้จากวิธีการเชิงปริมาณกับผลที่ได้จากวิธีการเชิงคุณภาพ โดยข้อดีคือ จะได้ข้อมูลที่มีความหลากหลาย มีความครอบคลุมสำหรับการปรับปรุงเครื่องมือ แต่ข้อเสียคือ ใช้เวลาและค่าใช้จ่ายมาก ดังนั้น ผู้สร้างเครื่องมือจึงควรทำการพิจารณาเลือกใช้วิธีที่เหมาะสมกับวัตถุประสงค์ของการพัฒนาเครื่องมือวัดนั้น ๆ

การปรับปรุงแบบทดสอบ (Test Revision)

Cohen et al. (2013) กล่าวถึง การปรับปรุงแบบทดสอบ (Test revision) สามารถแบ่งได้เป็น 2 ลักษณะ ซึ่งมีรายละเอียดดังนี้

1. การปรับปรุงแบบทดสอบเป็นขั้นตอนในการพัฒนาแบบทดสอบใหม่ (Test revision as a stage in new test development) การสร้างแบบทดสอบ/แบบสอบถามขึ้นมาใหม่ เริ่มต้นสร้างข้อคำถาม นำไปทดลองใช้ทำการวิเคราะห์รายข้อทั้งในเชิงปริมาณและเชิงคุณภาพ ทำให้ได้ข้อมูลที่สำคัญเพื่อนำไปสู่แบบทดสอบ/แบบสอบถามฉบับสมบูรณ์ เช่น บางข้อคำถามอาจถูกตัดทิ้ง หรืออาจเขียนข้อคำถามขึ้นมาใหม่ ซึ่งพิจารณาว่าข้อคำถามนั้นมีจุดอ่อนหรือจุดแข็งเป็นอย่างไร ถ้าหากมีจุดอ่อนหลายอย่าง ทั้งไม่มีอำนาจจำแนก มีความยากมาก อาจพิจารณาตัดข้อคำถามทิ้งมากกว่าที่จะปรับปรุงข้อคำถามใหม่

2. การปรับปรุงแบบทดสอบในการพัฒนาแบบทดสอบที่มีอยู่แล้ว (Test revision in the life cycle of an existing test) สำหรับแบบทดสอบ/แบบสอบถามที่พัฒนาขึ้นมาแล้ว การนำเครื่องมือวัดมาใช้ก็ต้องมีการปรับปรุงแบบทดสอบเช่นเดียวกัน โดยเครื่องมือวัดต่าง ๆ หากไปนำมาใช้แล้ว ควรทำการปรับปรุงเครื่องมือวัดก่อน โดยมีเงื่อนไขสำคัญที่ต้องทำการปรับปรุงเครื่องมือวัด ซึ่งมีรายละเอียดดังนี้

- 1) เครื่องมือวัดดูเหมือนจะล้าสมัย และผู้ทำแบบทดสอบในปัจจุบันไม่เกี่ยวข้องสัมพันธ์กับเครื่องมือวัดนั้น
- 2) เนื้อหาทางด้านภาษาของแบบทดสอบ รวมทั้งคำชี้แนะหรือคำสั่งในการทำแบบทดสอบ และข้อคำถามใช้คำศัพท์ที่ล้าสมัย ซึ่งยากต่อการทำความเข้าใจ
- 3) วัฒนธรรมมีการเปลี่ยนแปลงและคำที่ใช้ มีความหมายใหม่ มีคำหรือการแสดงออกในข้อคำถามหรือทิศทางของแบบทดสอบที่อาจรับรู้ได้ว่าไม่เหมาะสมหรือทำให้ไม่พอใจในกลุ่มบางกลุ่ม จึงต้องมีการเปลี่ยนแปลง
- 4) เกณฑ์ปกติของแบบทดสอบอาจไม่เพียงพอ อันเป็นผลมาจากการเปลี่ยนสมาชิกของกลุ่มในกลุ่มประชากรของผู้ทำแบบทดสอบ
- 5) เกณฑ์ปกติของแบบทดสอบอาจไม่เพียงพอ อันเป็นผลมาจากการเปลี่ยนแปลงของช่วงอายุในด้านความสามารถในการวัดข้ามช่วงเวลา ดังนั้น การขยายเกณฑ์ปกติของช่วงอายุจึงเป็นสิ่งจำเป็น
- 6) ความเชื่อมั่นและความเที่ยงตรงของแบบทดสอบ และประสิทธิผลของข้อคำถามแต่ละข้อ สามารถพัฒนาได้ดีขึ้นด้วยการปรับปรุงแบบทดสอบ
- 7) ทฤษฎีที่เกี่ยวข้องกับแบบทดสอบสามารถพัฒนาได้ และการเปลี่ยนแปลงนั้นสะท้อนถึงการออกแบบและเนื้อหาของแบบทดสอบ

ขั้นตอนในการปรับปรุงพัฒนาเครื่องมือวัดที่มีอยู่แล้ว มีลักษณะเช่นเดียวกันกับการสร้างเครื่องมือวัดขึ้นมาใหม่ นั่นคือ มีการนำไปทดลองใช้ การวิเคราะห์รายข้อ รวมทั้งการปรับปรุงเครื่องมือวัด เพื่อให้ได้เครื่องมือวัดฉบับสมบูรณ์

สรุป

การวิเคราะห์รายข้อ ถือเป็นกระบวนการหนึ่งที่มีความสำคัญในการพัฒนาเครื่องมือวิจัย เพราะเป็นตรวจสอบคุณภาพของเครื่องมือวัด หลังจากที่ได้สร้างเครื่องมือวัดตามวัตถุประสงค์ของการวัด และตามตัวแปรที่ศึกษา ว่าเครื่องมือวัดที่พัฒนาขึ้นมีคุณภาพเป็นอย่างไร ในแง่ของคุณลักษณะของข้อคำถามเป็นรายข้อ เช่น อำนาจจำแนกรายข้อ ความยากง่ายรายข้อ อำนาจตัวलग เป็นต้น นอกจากนี้ ยังมีการวิเคราะห์รายข้อในเชิงคุณภาพ ทั้งวิธีการที่ดำเนินการกับผู้ทำแบบทดสอบและผู้เชี่ยวชาญ ที่ช่วยทำให้เครื่องมือวัดมีความสมบูรณ์มากยิ่งขึ้น ซึ่งเหมาะสมกับตัวแปรทางด้านพฤติกรรมศาสตร์ ซึ่งเป็นตัวแปรที่มีความซับซ้อน การตรวจสอบคุณภาพเครื่องมือด้วยการวิเคราะห์รายข้อถือเป็นแนวทางในการสร้างและพัฒนาเครื่องมือทางด้านพฤติกรรมศาสตร์ให้มีความเหมาะสม รวมไปถึงการปรับปรุงเครื่องมือวัด ทั้งการสร้างเครื่องมือวัดขึ้นมาใหม่ และการนำเครื่องมือวัดที่มีการพัฒนาแล้วมาประยุกต์ใช้ด้วย

เอกสารอ้างอิง

- ณัฐภรณ์ หลาวทอง. (2561). *การสร้างเครื่องมือการวิจัยทางการศึกษา* (พิมพ์ครั้งที่ 2). กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ดุจเดือน พันธุนาวิน. (2551). หลักและเทคนิคการหาคุณภาพของแบบวัดชนิดมาตราประเมิณรวมค่า. *วารสารพัฒนาลังคม*, 10(1), 41-67.
- สุจิตรา เทียนสวัสดิ์. (2562). *การพัฒนาเครื่องมือสำหรับการวิจัยทางการแพทย์*. เชียงใหม่: สยามพิมพ์นานาชาติ.
- อรพินทร์ ชูชม. (2545). *เอกสารคำสอนวิชา วป 502 การสร้างและพัฒนาเครื่องมือวัดทางพฤติกรรมศาสตร์*. กรุงเทพฯ: สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- Aiken, L. R. (2003). *Psychological testing and assessment* (11th ed.). Boston: Allyn & Bacon.
- Browne, R. H. (1995). On the use of a pilot sample for sample size determination. *Statistics in Medicine*, 14(17), 1933-1940.
- Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2013). *Psychological testing and Assessment: An introduction to tests and measurement* (8th ed.). New York: McGraw-Hill.
- Fishman, J. A., & Galguera, T. (2003). *Introduction to test construction in the social and behavioral sciences: A practical guide*. Lanham, MD: Rowman & Littlefield Publisher.
- Gregory, R. J. (1996). *Psychological testing: History, principles, and applications* (2nd ed.). Boston: Allyn & Bacon.
- Johanson, G. A., & Brooks, G. P. (2010). Initial scale development: Sample size for pilot studies. *Educational and Psychological Measurement*, 70(3), 394-400.
- Kaplan, R. M., & Saccuzzo, D. P. (2013). *Psychological assessment and theory: Creating and using psychological tests* (8th ed.). Boston, MA: Cengage Learning.
- Miller, L. A., Lovler, R. L., & McIntire, S. A. (2013). *Foundations of psychological testing: A practical approach* (4th ed.). Los Angeles: Sage.
- Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing: Principles and applications* (6th ed.). Upper Saddle River, NJ: Prentice Hall.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use* (5th ed.). Oxford: Oxford University Press.
- Waltz, C., Strickland, O., & Lenz, E. (2005). *Measurement in nursing and health research* (3th ed.). New York: Springer Publishing Company.

บทที่ 8

ความเที่ยงตรง

บทที่ 8 เป็นการนำเสนอเกี่ยวกับความเที่ยงตรง โดยผู้เขียนต้องการให้ผู้อ่านได้ทราบและทำความเข้าใจเกี่ยวกับความเที่ยงตรง แนวคิดเบื้องต้นของความเที่ยงตรง รวมทั้งความเที่ยงตรงประเภทต่าง ๆ เนื่องจากความเที่ยงตรงแต่ละประเภท จะทำการวัดคุณสมบัติและคุณลักษณะของเครื่องมือวัดที่แตกต่างกัน ความเที่ยงตรงถือว่าเป็นวิธีการหนึ่งในการตรวจสอบคุณภาพของเครื่องมือวัด ว่าเครื่องมือวัดนั้นทำการวัดได้อย่างถูกต้องแม่นยำตรงตามสิ่งที่ต้องการวัดมากน้อยเพียงใด โดยบทนี้ จะช่วยให้ผู้วิจัยและนักวิชาการสามารถสร้างเครื่องมือวัดด้านต่างๆ ได้อย่างถูกต้องและแม่นยำ เพื่อช่วยให้กระบวนการพัฒนาเครื่องมือวิจัยทางพฤติกรรมศาสตร์มีความเหมาะสมและสมบูรณ์มากยิ่งขึ้น

แนวคิดเบื้องต้นของความเที่ยงตรง

ความเที่ยงตรง (Validity) หมายถึง ความถูกต้องแม่นยำของเครื่องมือวัดในการวัดสิ่งที่ต้องการวัด ดังนั้น เครื่องมือวัดที่มีความเที่ยงตรง หมายถึง เครื่องมือวัดนั้นมีคุณสมบัติที่วัดในสิ่งที่ต้องการวัดได้อย่างถูกต้องแม่นยำ และมีความเหมาะสมกับสิ่งที่ต้องการวัดให้มากที่สุดเท่าที่จะเป็นไปได้ (ไซติกา ภาชีผล, 2559; อรพินทร์ ชูชม, 2545) นอกจากนี้ Gregory (1996) กล่าวถึง ความเที่ยงตรง ว่าเป็นเครื่องมือที่มีความถูกต้องแม่นยำในแง่การอ้างอิงผลการวัด (inferences) ว่ามีความเหมาะสม มีความหมาย และมีประโยชน์ ดังนั้น เครื่องมือที่มีความเที่ยงตรง จะต้องอ้างอิงผลการวัดได้ถูกต้องกับสิ่งที่ต้องการวัด เช่น ถ้าแบบวัดใช้เพื่อวัดพฤติกรรมหนึ่ง ๆ คะแนนที่ได้ควรจะสะท้อนถึงพฤติกรรมนั้น ๆ หรือถ้าแบบทดสอบใช้วัดผลสัมฤทธิ์ทางการเรียนวิชาวิจัย คะแนนที่ได้จึงควรสะท้อนผลสัมฤทธิ์ทางการเรียนวิชาวิจัย

ความเที่ยงตรงเป็นคุณลักษณะพื้นฐานของเครื่องมือที่สำคัญที่สุด ถ้าเครื่องมือวัดขาดความเที่ยงตรงแล้ว แม้ว่าคุณสมบัติทางการวัดอื่น ๆ จะดีและเหมาะสม ข้อมูลที่ได้จะขาดความเหมาะสม และไม่มีประโยชน์

โดยปกติแล้ว ความเที่ยงตรงเกี่ยวข้องกับหลักฐานของการอ้างอิงผลการวัดที่มาจากคะแนนที่ได้จากการวัด ซึ่งสามารถจัดกลุ่มหลักฐานของการอ้างอิงผลการวัดได้เป็น 3 กลุ่ม ซึ่งมีรายละเอียดดังนี้ (ไซติกา ภาชีผล, 2559; อรพินทร์ ชูชม, 2545; Coaley, 2014; Cohen, Swerdlik, & Sturman, 2013; Gregory, 1996; Kaplan, & Saccuzzo, 2013; Miller, Lovler, & McIntire, 2013)

1. ผลที่ได้จากการวัดที่เป็นตัวแทนของเนื้อหาทั้งหมด นั่นคือ ความเที่ยงตรงเชิงเนื้อหา (Content Validity) เป็นการวัดความเที่ยงตรง โดยมุ่งประเมินหัวข้อ หรือเนื้อหาให้มีความครอบคลุมกับข้อคำถามของเครื่องมือวัดนั้น

2. ผลที่ได้จากการวัดที่ตรงตามเกณฑ์ที่เกี่ยวข้อง นั่นคือ ความเที่ยงตรงตามเกณฑ์ (Criterion-related Validity) เป็นการวัดความเที่ยงตรง โดยมุ่งประเมินความสัมพันธ์ของคะแนนที่ได้จากการวัด กับเครื่องมือวัดอื่น ๆ

3. ผลที่ได้จากการวัดที่เป็นตัวแทนโครงสร้าง นั่นคือ ความเที่ยงตรงเชิงโครงสร้าง (Construct Validity) เป็นการวัดความเที่ยงตรงที่ได้จากการทำความเข้าใจเกี่ยวกับคะแนนที่ได้จากการวัดว่ามีความสัมพันธ์กับเครื่องมือวัดอื่น ๆ รวมทั้งคะแนนที่ได้จากการวัดสามารถทำความเข้าใจภายใต้กรอบแนวคิดเชิงทฤษฎี ซึ่งช่วยให้เข้าใจโครงสร้างของเครื่องมือวัดนั้น

สำหรับความเที่ยงตรงทั้ง 3 รูปแบบ มักใช้กันอย่างกว้างขวางในการทำวิจัย เช่นเดียวกับการวิจัยทางด้านพฤติกรรมศาสตร์ การตรวจสอบเครื่องมือวัดด้วยความเที่ยงตรงนั้นไม่จำเป็นต้องกระทำทั้ง 3 รูปแบบทั้งหมด ขึ้นอยู่กับการใช้เครื่องมือวัดนั้น ว่าให้ความสำคัญกับการอ้างอิงผลการวัดในลักษณะใด โดยปกติแล้วความเที่ยงตรงเป็นแนวคิดเดียว (unitary) ที่มีแหล่งหลักฐานที่หลากหลาย ดังนั้น จึงไม่มีความเที่ยงตรงรูปแบบใดรูปแบบหนึ่งที่ดีกว่าความเที่ยงตรงรูปแบบอื่น ๆ แต่การใช้แหล่งหลักฐานของความเที่ยงตรงที่ผสมผสานกันจะยิ่งช่วยให้สามารถอ้างอิงผลการวัดหรือผลที่ได้จากการวัดว่ามีความเหมาะสม และมีประโยชน์มากยิ่งขึ้น (Cohen et al., 2013; Gregory, 1996; Miller et al., 2013)

ประเภทของความเที่ยงตรง

ความเที่ยงตรง เป็นคุณสมบัติที่สำคัญในการตรวจสอบคุณภาพเครื่องมือวัด และสำหรับเครื่องมือวัดทางด้านพฤติกรรมศาสตร์ ที่ทำการวัดตัวแปรที่มีความซับซ้อน ดังนั้น เพื่อให้เกิดความถูกต้องแม่นยำในการวัดเครื่องมือวัดทางด้านพฤติกรรมศาสตร์ จึงอาศัยการตรวจสอบความเที่ยงตรง ซึ่งสามารถทำการตรวจสอบความเที่ยงตรงได้หลายประเภท โดยมีรายละเอียดดังนี้ (โชติกา ภาชีผล, 2559; สุจิตรา เทียนสวัสดิ์, 2562; อรพินทร์ ชูชม, 2545; Coaley, 2014; Cohen et al., 2013; Gregory, 1996; Kaplan, & Saccuzzo, 2013; Miller et al., 2013)

1. ความเที่ยงตรงเชิงเนื้อหา

ความเที่ยงตรงเชิงเนื้อหา (Content Validity) หมายถึง คุณสมบัติของเครื่องมือวัดที่สามารถวัดได้ตรง และมีความครอบคลุมเนื้อหาตามที่ต้องการวัดได้อย่างครบถ้วน ยังใช้พิจารณาถึงเหมาะสมเพียงพอของแนวคิดหลักที่นำมาใช้ว่าออกแบบได้ครอบคลุมหรือไม่ โดยมีขั้นตอนที่สำคัญในการตรวจสอบความเที่ยงตรงเชิงเนื้อหาของเครื่องมือวัด ดังนี้

1.1 การวางแผนกำหนดพินช์เชียว Nunnally and Bernstein (1994) กล่าวว่า การสร้างเครื่องมือวัดใด ๆ ก็ตาม ให้มีความเที่ยงตรงเชิงเนื้อหาจะต้องเริ่มตั้งแต่ก่อนการสร้างแบบวัด ไม่ใช่ตรวจสอบหลังจากเครื่องมือวัดเสร็จเรียบร้อยแล้ว กล่าวคือ การสร้างเครื่องมือวัดให้มีความเที่ยงตรงเชิงเนื้อหา จะต้องเริ่มตั้งแต่

การวางแผนกำหนดพิมพ์เขียว (Blueprint) ของเนื้อหา รวมทั้งวิธีการวัดเนื้อหานั้นด้วย หรืออีกนัยหนึ่งคือ ต้องนำนิยามเชิงปฏิบัติการที่สร้างขึ้น และข้อคำถามที่สร้างขึ้น นำมาใส่ในแผนกำหนดพิมพ์เขียวของเนื้อหา

1.2 การพิจารณาตรวจสอบโดยผู้เชี่ยวชาญ ภายหลังจากสร้างแผนกำหนดพิมพ์เขียวเสร็จแล้ว จึงนำมาให้ผู้เชี่ยวชาญทำการพิจารณาความสัมพันธ์ระหว่างข้อความที่เขียนขึ้นมากับประเด็นเนื้อหา โดยทำการพิจารณาด้านต่าง ๆ ได้แก่ ความสอดคล้องกัน ความสามารถในการอ่านให้เข้าใจ ความถูกต้อง ภาษาที่ใช้ และการเป็นตัวแทนของสิ่งที่ต้องการวัด (Berk, 1990) ซึ่งเมื่อพิจารณาแล้ว ให้คะแนนตามเกณฑ์ที่กำหนดให้ อาจให้ความคิดเห็นอื่น ๆ เพิ่มเติมเกี่ยวกับการปรับปรุงเนื้อหาหรือข้อความต่าง ๆ นอกจากนี้ จำนวนผู้เชี่ยวชาญที่เหมาะสม แนะนำว่าอยู่ระหว่าง 3-5 คน (Lynn, 1986) และผู้เชี่ยวชาญควรมีความรู้หรือเชี่ยวชาญในเรื่องนั้น ๆ และควรมีอย่างน้อย 1 คนเชี่ยวชาญเรื่องการสร้างเครื่องมือด้วย (Davis, 1992)

สำหรับการวัดทางด้านพฤติกรรมศาสตร์ ค่าสถิติที่นิยมใช้ในการตรวจสอบความเที่ยงตรงเชิงเนื้อหา มากที่สุด คือ ค่าดัชนีความสอดคล้องระหว่างข้อคำถามกับวัตถุประสงค์หรือเนื้อหา (Index of Item Objective Congruence: IOC) โดยให้ผู้เชี่ยวชาญแต่ละคนประเมินข้อคำถามแต่ละข้อว่าสอดคล้องหรือตรงกับวัตถุประสงค์หรือเนื้อหาที่ต้องการวัดหรือไม่ โดยมีเกณฑ์การให้คะแนนดังนี้

- +1 คือ แน่ใจว่าข้อคำถามข้อนั้นวัดได้ตรงกับเนื้อหาที่ต้องการวัด
- 0 คือ ไม่แน่ใจว่าข้อคำถามข้อนั้นวัดได้ตรงกับเนื้อหาที่ต้องการวัด
- 1 คือ แน่ใจว่าข้อคำถามข้อนั้นวัดได้ไม่ตรงกับเนื้อหาที่ต้องการวัด

โดยมีสูตรคำนวณค่า IOC ดังนี้

$$IOC = \frac{\sum R}{N}$$

IOC	แทน ค่าดัชนีความสอดคล้องระหว่างข้อคำถามกับเนื้อหา
$\sum R$	แทน ผลรวมคะแนนที่ได้จากผู้เชี่ยวชาญ
N	แทน จำนวนผู้เชี่ยวชาญ

โดยเกณฑ์ที่กำหนดสำหรับค่าดัชนีความสอดคล้องระหว่างข้อคำถามกับเนื้อหา (IOC) ควรมีค่ามากกว่า .50 (Rovinelli & Hambleton, 1977) กล่าวคือ ถ้าค่า IOC ที่คำนวณได้มีค่ามากกว่าหรือเท่ากับ .50 แสดงว่า ข้อคำถามนั้นเป็นตัวแทนหรือมีความสอดคล้องระหว่างข้อคำถามกับเนื้อหา ส่วนข้อคำถามใดที่มีค่าต่ำกว่า .50 แสดงว่า ไม่สอดคล้องระหว่างข้อคำถามกับเนื้อหา ข้อคำถามนั้นจะถูกตัดทิ้งออกไป หรืออาจนำไปปรับปรุงแก้ไขใหม่ให้ดีขึ้น

ตัวอย่าง การตรวจสอบเครื่องมือวัดด้วยความเที่ยงตรงเชิงเนื้อหา

ก. การกำหนดพิมพ์เขียว (Blueprint)

การรับรู้ความสามารถของตน (Self-efficacy) มีเนื้อหาเกี่ยวกับการที่บุคคลตัดสินใจเกี่ยวกับความสามารถของตนเองในการจัดการและดำเนินการกระทำพฤติกรรมให้บรรลุเป้าหมายที่กำหนดไว้ได้

แบบวัดนี้พัฒนามาจากแบบวัดการรับรู้ความสามารถของตนเองโดยทั่วไปจากเซอร์เรอร์และคณะ (Sherer et al., 1982) จำนวนข้อคำถาม 5 ข้อ มีลักษณะเป็นมาตรวัดประมาณค่า 6 ระดับ ตั้งแต่ “ไม่จริงเลย” ถึง “จริงที่สุด” นักเรียนที่ได้คะแนนสูงกว่าแสดงว่าเป็นผู้มีการรับรู้ความสามารถของตนสูงกว่านักเรียนที่ได้คะแนนต่ำกว่า

ข้อความในแบบวัดการรับรู้ความสามารถของตนเอง มีด้วยกัน 2 ลักษณะ คือ ข้อความที่เป็นทางบวก และข้อความที่เป็นทางลบ โดยมีจำนวนข้อคำถามตามแต่ละองค์ประกอบ ดังนี้

องค์ประกอบ	ข้อความทางบวก	ข้อความทางลบ
การรับรู้ความสามารถของตนเอง	1, 4, 5	2, 3
รวม	3	2
เหลือ	5	

ข้อคำถาม	ความคิดเห็นของผู้เชี่ยวชาญ			ข้อเสนอแนะ
	เห็นด้วย (+1)	ไม่แน่ใจ (0)	ไม่เห็นด้วย (-1)	
(+) 1. เมื่อฉันวางแผนไว้ ฉันมั่นใจว่าสามารถทำตามแผนที่วางไว้ได้				
(-) 2. ปัญหาของฉันคือฉันไม่สามารถทำงานให้สำเร็จได้ ทั้ง ๆ ที่ควรจะต้องทำ				
(-) 3. ฉันมักจะล้มเลิกการทำงานก่อนที่จะทำงานนั้นจนสำเร็จ				
(+) 4. ถึงแม้ว่าฉันจะต้องทำในสิ่งที่ไม่ถูกใจ ฉันก็จะพยายามทำงานสำเร็จ				
(+) 5. หากฉันตัดสินใจจะทำอะไรแล้ว ฉันจะมุ่งมั่นทำต่อไปจนกว่าสิ่งนั้นจะสำเร็จ				

ข. การประเมินความสอดคล้องระหว่างข้อคำถามกับวัตถุประสงค์

ข้อที่	ความคิดเห็นของผู้เชี่ยวชาญ					Σ R	IOC
	คนที่ 1	คนที่ 2	คนที่ 3	คนที่ 4	คนที่ 5		
1	+1	+1	+1	+1	+1	+5	1.00
2	+1	0	+1	+1	+1	+4	0.80
3	+1	+1	0	-1	0	+1	0.20
4	+1	-1	-1	0	0	-1	-0.20
5	+1	+1	0	0	+1	+3	0.60

จากตารางข้างต้น จึงเห็นว่า ข้อที่ 1, 2 และ 5 มีค่า IOC มากกว่า 0.50 แสดงว่า มีความสอดคล้องกันระหว่างข้อคำถามกับเนื้อหา แสดงว่าเป็นตัวแทนของตัวแปรการรับรู้ความสามารถของตนเอง นั่นคือ ข้อคำถามวัดได้ตรงกับเนื้อหาการวัดของตัวแปรการรับรู้ความสามารถของตนเอง ส่วนข้อที่ 3 และ 4 มีค่า IOC ต่ำกว่า 0.50 แสดงว่าไม่เป็นตัวแทนของการวัดการรับรู้ความสามารถของตนเอง นั่นคือ ข้อคำถามวัดได้ไม่ตรงกับเนื้อหาการวัดของตัวแปรการรับรู้ความสามารถของตนเอง จึงพิจารณาให้ตัดข้อคำถามทิ้ง

2. ความเที่ยงตรงตามเกณฑ์

ความเที่ยงตรงตามเกณฑ์ (Criterion Validity) หมายถึง คุณสมบัติของเครื่องมือวัดที่สามารถวัดได้ตรงตามเกณฑ์ที่เกี่ยวข้อง เกณฑ์ (Criteria) เป็นผลลัพธ์จากเครื่องมือวัดที่ใช้สำหรับการทำนาย นั่นคือ เครื่องมือวัดที่มีความเที่ยงตรงตามเกณฑ์ ใช้วิธีการตรวจสอบด้วยค่าสหสัมพันธ์ระหว่างคะแนนที่ได้จากการวัดและเครื่องมือวัดอื่น ๆ (เกณฑ์ที่เกี่ยวข้อง) จะมีประโยชน์ในการประเมินว่าเครื่องมือวัดที่สร้างขึ้นสามารถทำนายและ/หรือตัดสินใจเกี่ยวกับบุคคล (Coaley, 2014; Cohen et al., 2013; Miller et al., 2013) เช่น คะแนนที่ได้จากความสามารถในการขาย อาจมีความสัมพันธ์กับเกณฑ์ที่เกี่ยวข้อง ได้แก่ จำนวนสินค้าที่ขายได้ มูลค่าเงินที่ขายได้ หรือจำนวนของลูกค้าที่มาใช้ซ้ำ เป็นต้น

คุณลักษณะของเกณฑ์ที่ดี (Characteristics of a Good Criterion)

เกณฑ์ (Criterion) เป็นการวัดผลลัพธ์ใด ๆ ก็ตาม ที่เครื่องมือวัดใช้ในการตรวจสอบเทียบเคียง (Validate) ดังนั้น เกณฑ์ที่ดีควรมีคุณลักษณะที่สำคัญ ดังนี้ (อรพินทร์ ชูชม, 2545; Cohen et al., 2013; Gregory, 1996)

1) ต้องเป็นเกณฑ์ที่เชื่อถือได้ (Reliable) เกณฑ์ที่เชื่อถือได้จะเป็นดัชนีที่มีประโยชน์ในการตรวจสอบเครื่องมือวัด ถ้าความหมายของความเชื่อมั่น คือ ความคงเส้นคงวาของคะแนนที่วัด ดังนั้น ความต้องการการวัดที่เชื่อถือได้จึงเป็นสิ่งที่ชัดเจน ถ้าการไม่มีความเชื่อมั่น หมายถึง การไม่สามารถทำนายได้ เกณฑ์ที่เชื่อถือไม่ได้ก็จะทำนายผลไม่ได้ แม้ว่าเครื่องมือวัดจะดีมาน้อยเพียงใดก็ตาม

จากขีดจำกัดทางทฤษฎีของสัมประสิทธิ์ความเที่ยงตรงถูกจำกัดโดยความเชื่อมั่นของเครื่องมือวัดและความเชื่อมั่นของเกณฑ์ จากสูตร

$$r_{xy} = \sqrt{(r_{xx})(r_{yy})}$$

จากสูตรดังกล่าว ค่าสัมประสิทธิ์ความเที่ยงตรง (r_{xy}) มีค่าน้อยกว่าหรือเท่ากับรากที่สองของผลคูณระหว่างความเชื่อมั่นของเครื่องมือวัด (r_{xx}) และความเชื่อมั่นของเกณฑ์ (r_{yy}) เสมอ หรืออีกนัยหนึ่งคือ ถ้าความเชื่อมั่นของเครื่องมือวัด หรือความเชื่อมั่นของเกณฑ์ (หรืออาจทั้งสอง) มีค่าต่ำแล้ว สัมประสิทธิ์ความเที่ยงตรงจะมีค่าต่ำลงไปด้วย นั่นคือ ความเชื่อมั่นทั้งเครื่องมือวัดและเกณฑ์ส่งผลต่อความเที่ยงตรง

2) ต้องเป็นเกณฑ์ที่มีความเหมาะสม (Appropriate) กับเครื่องมือที่ใช้วัด เกณฑ์ที่เลือกมาควรมีความถูกต้องและมีเหตุผลในการเลือกเกณฑ์นั้น ๆ จะต้องเป็นเกณฑ์ที่มีความเกี่ยวข้อง (relevant) ด้วย ยกตัวอย่างเช่น กรณีของแบบทดสอบความสนใจ ซึ่งไม่ชัดเจนว่า เกณฑ์ที่ใช้ของแบบทดสอบความสนใจ คือ ความพึงพอใจความสำเร็จ หรือความต่อเนื่องในการกระทำกิจกรรมนั้น หรือแม้แต่แบบทดสอบลักษณะอาการทางจิตเวช เกณฑ์ที่ใช้อาจเป็นการวินิจฉัยโรคทางจิต เป็นต้น ดังนั้น การเลือกตัวแปรที่เป็นเกณฑ์จึงควรระมัดระวังด้วย

3) ต้องเป็นเกณฑ์ที่เป็นอิสระจากการปนเปื้อน (Free of contamination) ตัวเครื่องมือวัดจะต้องเป็นอิสระจากการปนเปื้อน ยกตัวอย่างเช่น แบบวัดเหตุการณ์ในชีวิต เป็นแบบวัดที่วัดเหตุการณ์ความเครียดในชีวิตหลาย ๆ ด้าน (เช่น การหย่า การเลื่อนขั้น เป็นต้น) โดยแบบวัดเหตุการณ์ในชีวิตมีความสัมพันธ์กับตัวแปรเกณฑ์ เช่น ความเจ็บป่วยทางร่างกาย และความรบกวนทางจิตใจ อย่างไรก็ตาม มีตัวแปรเกณฑ์มากมายที่มีข้อคำถามที่คล้ายกันหรือเหมือนกันกับแบบวัดเหตุการณ์ในชีวิต เช่น แบบทดสอบคัดกรองอาการทางจิตที่ทำการตรวจสอบความเปลี่ยนแปลงด้านการกิน ด้านการนอน และด้านกิจกรรมทางสังคม ซึ่งแบบวัดเหตุการณ์ในชีวิตก็วัดความเปลี่ยนแปลงด้านการกิน นอน และกิจกรรมสังคมเช่นเดียวกัน ถ้าแบบทดสอบคัดกรองมีข้อคำถามเช่นเดียวกับแบบวัดเหตุการณ์ในชีวิต ค่าความสัมพันธ์ระหว่างเครื่องมือวัดทั้งสองนี้อาจมีค่าต่ำหรือสูงกว่าความเป็นจริง นอกจากนี้เกณฑ์ที่ปนเปื้อน (criterion contamination) มีความเป็นไปได้ว่าอาจเกิดจากการให้คะแนนของผู้เชี่ยวชาญ หากผู้เชี่ยวชาญทราบคะแนนความรู้ของผู้เข้ารับการทดสอบ ข้อมูลนี้อาจจะไปมีผลต่อการให้คะแนนทั้งอย่างตั้งใจและไม่ตั้งใจ

ดังนั้น เมื่อทราบคุณลักษณะของเกณฑ์ที่ดีทั้งสามแล้ว จึงควรทำความเข้าใจถึงการตรวจสอบความเที่ยงตามเกณฑ์ โดยให้ความสนใจอยู่ 2 ประเภท ซึ่งมีรายละเอียดดังนี้ (โชติกา ภาษีผล, 2559; สุจิตรา เทียนสวัสดิ์, 2562; Cohen et al., 2013; Gregory, 1996; Miller et al., 2013)

1. ความเที่ยงตรงตามสภาพ

ความเที่ยงตรงตามสภาพ (Concurrent Validity) หมายถึง คุณสมบัติของเครื่องมือวัดที่สามารถวัดได้ตรงตามสภาพที่เป็นจริงในสภาวะปัจจุบัน นั่นคือ เป็นความสามารถของเครื่องมือวัดในการเทียบกับเกณฑ์ในปัจจุบัน เป็นการเก็บข้อมูลด้วยเครื่องมือทั้ง 2 ชุด ในเวลาเดียวกัน ทั้งเครื่องมือที่วัดและเครื่องมือที่เป็นเกณฑ์ เช่น ถ้าต้องการตรวจสอบความเที่ยงตรงตามสภาพของแบบทดสอบผลสัมฤทธิ์ทางการเรียนที่สร้างขึ้นว่ามีความเที่ยงตรงตามสภาพหรือไม่ ก็ต้องทำการหาเครื่องมือที่เป็นมาตรฐานหรือเครื่องมือที่มีความน่าเชื่อถือที่วัดผลสัมฤทธิ์ทางการเรียนมาเป็นเกณฑ์เปรียบเทียบ ถ้าคะแนนจากแบบทดสอบมีค่าสัมประสิทธิ์สหสัมพันธ์สูง ก็แสดงให้เห็นว่าแบบทดสอบผลสัมฤทธิ์ที่สร้างขึ้นมีความเที่ยงตรงตามสภาพ

2. ความเที่ยงตรงเชิงทำนาย

ความเที่ยงตรงเชิงทำนาย (Predictive Validity) หมายถึง คุณสมบัติของเครื่องมือวัดที่สามารถวัดได้ตรงตามสภาพความเป็นจริงของสิ่งนั้นที่จะเกิดขึ้นในอนาคต หรือสามารถนำผลการวัดไปทำนายลักษณะหรือพฤติกรรมต่าง ๆ ได้ นั่นคือ เป็นความสามารถของเครื่องมือวัดในการเทียบกับเกณฑ์ในอนาคต โดยเริ่มต้นเก็บข้อมูลเครื่องมือวัดก่อน แล้วจึงทำการเก็บข้อมูลเครื่องมือมาตรฐานที่เป็นเกณฑ์ในภายหลัง เช่น ถ้าต้องการตรวจสอบความเที่ยงตรงเชิงทำนายของแบบทดสอบคัดเลือกเข้าทำงานที่สร้างขึ้นว่ามีความเที่ยงตรงเชิงทำนายหรือไม่ ถ้าบุคคลได้ทำแบบทดสอบคัดเลือกเข้าทำงานได้คะแนนสูง แล้วภายหลังมีคะแนนผลการปฏิบัติงานในระดับสูง ขณะที่บุคคลได้ทำแบบทดสอบคัดเลือกเข้าทำงานได้คะแนนต่ำกว่า แล้วภายหลังมีคะแนนผลการปฏิบัติงานในระดับต่ำกว่า แสดงว่า แบบทดสอบคัดเลือกเข้าทำงานที่สร้างขึ้นว่ามีความเที่ยงตรงเชิงทำนาย นั่นคือ สามารถทำนายผลการปฏิบัติงานของบุคคลได้

โดยทั้งความเที่ยงตรงเชิงสภาพและความเที่ยงตรงเชิงทำนาย มีวิธีการตรวจสอบความเที่ยงตรง ส่วนใหญ่ใช้ด้วยวิธีการวิเคราะห์สหสัมพันธ์ (โชติกา ภาชีผล, 2559; สุจิตรา เทียนสวัสดิ์, 2562; Kaplan & Saccuzzo, 2013) ซึ่งวิธีการตรวจสอบด้วยการวิเคราะห์ค่าสหสัมพันธ์ของเครื่องมือวัดกับตัวแปรเกณฑ์ เรียกว่า “สัมประสิทธิ์ความเที่ยงตรง” (Validity coefficient) ซึ่งมีรายละเอียดดังนี้

- ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson Product Moment Correlation Coefficient) สำหรับเครื่องมือวัด ทั้ง 2 แบบ เป็นข้อมูลการวัดระดับ Interval

(1) หากเป็นความสัมพันธ์ระหว่างคะแนนจากเครื่องมือวัดกับคะแนนจากเครื่องมือที่เป็นเกณฑ์ในปัจจุบัน เป็นการหาความเที่ยงตรงตามสภาพ

(2) หากเป็นความสัมพันธ์ระหว่างคะแนนจากเครื่องมือวัดกับคะแนนจากเครื่องมือที่เป็นเกณฑ์ในอนาคต เป็นการหาความเที่ยงตรงเชิงทำนาย

- หากข้อมูลการวัดค่าใดค่าหนึ่งหรือทั้งสองค่าเป็นข้อมูลที่ระดับต่ำกว่า Interval ให้เลือกใช้วิธีวิเคราะห์ทางสถิติอื่นที่เหมาะสม เช่น Spearman เป็นต้น

- สำหรับ ในกรณีที่มีตัวแปรอื่นที่อาจจะร่วมทำนายตัวแปรผลลัพธ์ด้วย สถิติที่เหมาะสม คือ การวิเคราะห์การถดถอย (regression analysis)

ตัวอย่าง การตรวจสอบความเที่ยงตรงตามสภาพ

แบบประเมินสติ (อรรวรรณ ศิลปกิจ, 2558)

เครื่องมือวัด: แบบประเมินสติ (Mindfulness Assessment Scale: MAS) โดยแบบประเมินสติสร้างตามหลักพุทธศาสนา

เครื่องมือวัดที่เป็นเกณฑ์: แบบประเมิน Philadelphia Mindfulness Scale (PHLMS)

สถิติวิเคราะห์: ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

ผลการวิจัยพบว่า แบบประเมินสติ (MAS) มีความสัมพันธ์ทางบวกกับแบบประเมิน Philadelphia Mindfulness Scale ด้านตระหนักรู้ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($r = .65$)

สรุปว่า แบบประเมินสติมีความเที่ยงตรงตามสภาพ

มาตรวัดการเห็นคุณค่าในตนเองจากความสัมพันธ์ (นรุตม์ พรประสิทธิ์, 2562)

เครื่องมือวัด: มาตรวัดการเห็นคุณค่าในตนเองจากความสัมพันธ์ (Relational Self-esteem Scale)

เครื่องมือวัดที่เป็นเกณฑ์: มาตรวัดการเห็นคุณค่าในตนเองจากอัตลักษณ์แห่งตน

สถิติวิเคราะห์: ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

ผลการวิจัยพบว่า มาตรวัดการเห็นคุณค่าในตนเองจากความสัมพันธ์ มีความสัมพันธ์ทางบวกกับมาตรวัดการเห็นคุณค่าในตนเองจากอัตลักษณ์แห่งตน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($r = .63$)

สรุปว่า มาตรวัดการเห็นคุณค่าในตนเองจากความสัมพันธ์มีความเที่ยงตรงตามสภาพ

ตัวอย่าง การตรวจสอบความเที่ยงตรงเชิงทำนาย

มาตรวัดความรุนแรงในการฆ่าตัวตาย (Lindh et al., 2018)

เครื่องมือวัด: มาตรวัดความรุนแรงในการฆ่าตัวตาย (Columbia-Suicide Severity Rating Scale)

เครื่องมือวัดที่เป็นเกณฑ์: แบบวัดพฤติกรรมกรรมการฆ่าตัวตาย (วัดใน 6 เดือนต่อมา)

สถิติวิเคราะห์: การวิเคราะห์การถดถอยแบบโลจิสติกส์

ผลการวิจัยพบว่า เมื่อควบคุมประวัติความพยายามในการฆ่าตัวตาย คะแนนจากมาตรวัดความรุนแรงในการฆ่าตัวตาย สามารถทำนายความพยายามในการฆ่าตัวตาย ได้เป็น 1.08 เท่า อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (OR = 1.08)

สรุปว่า มาตรวัดความรุนแรงในการฆ่าตัวตายมีความเที่ยงตรงเชิงทำนาย โดยสามารถทำนายความพยายามในการฆ่าตัวตายได้

แบบวัดการรับรู้ความสามารถของตนเองในการเลิกสูบบุหรี่ (Spek et al., 2013)

เครื่องมือวัด: แบบวัดการรับรู้ความสามารถของตนเองในการเลิกสูบบุหรี่ (Smoking Abstinence Self-Efficacy Questionnaire)

เครื่องมือวัดที่เป็นเกณฑ์: สถานะการสูบบุหรี่ (วัดใน 52 สัปดาห์ต่อมา)

สถิติวิเคราะห์: การวิเคราะห์การถดถอยแบบโลจิสติกส์

ผลการวิจัยพบว่า ผู้ที่ได้คะแนนจากแบบวัดการรับรู้ความสามารถของตนเองในการเลิกสูบบุหรี่สูง สามารถทำนายการเลิกสูบบุหรี่ได้เป็น 1.85 เท่า อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (OR = 1.85)

สรุปว่า แบบวัดการรับรู้ความสามารถของตนเองในการเลิกสูบบุหรี่มีความเที่ยงตรงเชิงทำนาย โดยสามารถทำนายการเลิกสูบบุหรี่ได้

3. ความเที่ยงตรงเชิงโครงสร้าง

ความเที่ยงตรงเชิงโครงสร้าง (Construct Validity) หมายถึง คุณสมบัติของเครื่องมือวัดที่สามารถวัดได้ตรงตามคุณลักษณะ หรือแนวคิดทฤษฎีต่าง ๆ ตามโครงสร้างของตัวแปรที่ศึกษา โดยโครงสร้าง (Construct) มีลักษณะที่เป็นนามธรรม หรือเป็นความคิดที่พัฒนาขึ้นเพื่อบรรยายหรืออธิบายพฤติกรรมหนึ่ง ๆ รวมทั้งเป็นคุณลักษณะ (traits) ที่ไม่สามารถสังเกตเห็นได้ (Cohen et al., 2013) ดังนั้น จึงต้องอธิบายให้ชัดเจนว่า โครงสร้างมาจากแนวคิดทฤษฎีที่ใช้อธิบายถึงสิ่งที่ต้องการจะวัด ว่ามีองค์ประกอบใดบ้าง เช่น ตามแนวคิดทฤษฎีความรู้

ด้านสุขภาพ กล่าวหาว่า ความรอบรู้ด้านสุขภาพ ควรประกอบไปด้วย การเข้าถึงข้อมูลสุขภาพ การเข้าใจข้อมูลสุขภาพ การวิเคราะห์ข้อมูลสุขภาพ และการประยุกต์ใช้ข้อมูลสุขภาพ กล่าวคือ เมื่อสร้างเครื่องมือวัดขึ้นโดยให้มีความสัมพันธ์สอดคล้องกับกรอบแนวคิดหรือตามทฤษฎีที่กำหนด แล้วนำเครื่องมือวัดไปทดสอบกับกลุ่มตัวอย่าง แล้วยังพบว่า เป็นจริงตามทฤษฎี แสดงว่า เครื่องมือวัดความรอบรู้ด้านสุขภาพนั้นมีความเที่ยงตรงเชิงโครงสร้าง (โชติกา ภาชีผล, 2559; Kaplan & Saccuzzo, 2013) หรืออีกนัยหนึ่ง เช่น ตัวแปรการรับรู้ความสามารถของตนเอง ความหุนตัว ความหวัง การมองโลกในแง่ดี เป็นคุณลักษณะที่ไม่สามารถสังเกตเห็นได้ และวัดได้โดยตรง จึงต้องกำหนดนิยามเชิงปฏิบัติการให้เข้าใจง่าย ดังนั้น เครื่องมือวัดต้องกำหนดนิยามให้ชัดเจนว่าครอบคลุมคุณลักษณะใดบ้าง แล้วจึงทำการสร้างข้อคำถามให้ครอบคลุมคุณลักษณะนั้น ๆ หากสร้างเครื่องมือวัดได้ครอบคลุมคุณลักษณะเหล่านี้ ก็แสดงว่า เครื่องมือวัดนั้นมีความเที่ยงตรงเชิงโครงสร้าง โดยการตรวจสอบความเที่ยงตรงเชิงโครงสร้างมีหลายวิธี ซึ่งมีรายละเอียดดังนี้ (โชติกา ภาชีผล, 2559; สุจิตรา เทียนสวัสดิ์, 2562; อรพินทร์ ชูชม, 2545; Cohen et al., 2013; Gregory, 1996)

3.1 วิธีการหาความเป็นเอกพันธ์ของเครื่องมือวัด (Test Homogeneity) ความเป็นเอกพันธ์ของเครื่องมือวัด คือ การที่เครื่องมือวัดมีคุณลักษณะแบบเดียวกัน วัดเพียงคุณลักษณะเดียว สำหรับวิธีการตรวจสอบคือการหาความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับ ข้อคำถามใดที่แทบไม่มีความสัมพันธ์ หรือมีความสัมพันธ์กับคะแนนรวมน้อย ให้ปรับปรุงข้อคำถามใหม่หรือตัดทิ้งไปเลย เนื่องจากข้อคำถามนั้นไม่ได้วัดโครงสร้างตัวแปรนั้น ถึงแม้ว่าการหาความเป็นเอกพันธ์นี้เป็นการตรวจสอบว่าข้อคำถามทั้งหมดวัดเรื่องเดียวกันก็ตาม แต่ก็ยังไม่เพียงพอสำหรับการตรวจสอบความเที่ยงตรงเชิงโครงสร้าง ดังนั้น จึงเป็นสิ่งสำคัญที่ควรพิจารณาการตรวจสอบความเที่ยงตรงเชิงโครงสร้างอื่น ๆ ประกอบกันด้วย

3.2 วิธีการวิเคราะห์องค์ประกอบ (Factor Analysis) การวิเคราะห์องค์ประกอบถือเป็นเทคนิคการวิเคราะห์ทางสถิติที่ช่วยรวมกลุ่มตัวแปรหรือข้อคำถามที่มีความสัมพันธ์กันไว้ในกลุ่มที่ชัดเจน ทำให้เข้าใจลักษณะของข้อมูล แบบแผน โครงสร้าง รวมทั้งความสัมพันธ์ได้ง่ายยิ่งขึ้น สำหรับเครื่องมือวัดทางด้านพฤติกรรมศาสตร์ที่มีความซับซ้อน สังเกตและวัดได้ยาก นักวิจัย/นักวิชาการส่วนใหญ่ใช้การวิเคราะห์องค์ประกอบในการตรวจสอบความเที่ยงตรงเชิงโครงสร้าง การวิเคราะห์องค์ประกอบ ประกอบด้วย 2 วิธีการ ดังนี้

(1) การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis: EFA) เป็นวิธีการตรวจสอบความเที่ยงตรงเชิงโครงสร้าง เพื่อสำรวจดูว่าเครื่องมือวัดนั้นมีองค์ประกอบใดบ้าง มีลักษณะเป็นเช่นไร มีข้อคำถามใดบ้างที่มีความสัมพันธ์เกี่ยวข้องกัน จัดกลุ่มรวมกันเป็นองค์ประกอบเดียวกัน โดยแต่ละองค์ประกอบคือ โครงสร้างของเครื่องมือวัดนั้น ๆ นอกจากนี้ ยังเป็นการค้นหาลักษณะที่นักวิจัย/นักวิชาการยังมีแนวคิดทฤษฎีที่ไม่เพียงพอที่จะสรุปหรือยืนยันเกี่ยวกับโครงสร้างองค์ประกอบได้ (Finch, 2013; Flora & Flake, 2017; Haig, 2005; Watkins, 2018)

(2) การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) เป็นวิธีการตรวจสอบความเที่ยงตรงเชิงโครงสร้างว่าเป็นไปตามโครงสร้างตามที่แนวคิดทฤษฎีกำหนดไว้หรือไม่ เพื่อเป็นการยืนยันว่าเครื่องมือวัดที่สร้างขึ้นตามแนวคิดทฤษฎี โดยกำหนดสมมติฐานได้ว่า เครื่องมือวัดมีองค์ประกอบและแต่ละองค์ประกอบมีข้อคำถามใดบ้าง การวิเคราะห์เป็นการตรวจสอบว่าข้อมูลที่เก็บรวบรวมข้อมูลมา มีความสอดคล้องกับโครงสร้างองค์ประกอบที่กำหนดไว้แล้วหรือไม่ อย่างไร หรืออีกนัยหนึ่งคือ เป็นการยืนยันว่าเครื่องมือวัดเป็นไปตามแนวคิดทฤษฎีที่กำหนดไว้แล้วนั่นเอง (Flora & Flake, 2017; Hoyle, 2004)

โดยวิธีการตรวจสอบความเที่ยงตรงเชิงโครงสร้าง หากมีแนวคิดทฤษฎีอยู่แล้ว ให้ใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) ได้เลย ส่วนถ้าหากแนวคิดทฤษฎียังไม่ชัดเจนหรือไม่เพียงพอที่จะสรุปได้ว่ามีองค์ประกอบใดบ้าง ให้ใช้การวิเคราะห์องค์ประกอบเชิงสำรวจ (EFA) ก่อน แล้วจึงตามด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) เพื่อทำการยืนยันองค์ประกอบที่ได้ (Burns & Grove, 2009) โดยเนื้อหารายละเอียดสำหรับการวิเคราะห์องค์ประกอบจะอธิบายในบทที่ 10

ตัวอย่าง การตรวจสอบความเที่ยงตรงเชิงโครงสร้างด้วยวิธีการวิเคราะห์องค์ประกอบ แบบวัดพฤติกรรมป้องกันการสูบบุหรี่ (ชาริน สุวรรณวงศ์ และศรีณย์ พิมพ์ทอง, 2563)

เครื่องมือวัด: แบบวัดพฤติกรรมป้องกันการสูบบุหรี่

ความเป็นมา: ยังขาดแนวคิดและทฤษฎีเกี่ยวกับพฤติกรรมป้องกันการสูบบุหรี่ในกลุ่มวัยรุ่น รวมทั้งงานวิจัยที่เกี่ยวข้องยังขาดความชัดเจน

วัตถุประสงค์การวิจัย: เพื่อสร้างและตรวจสอบแบบวัดพฤติกรรมป้องกันการสูบบุหรี่ด้วยการวิเคราะห์องค์ประกอบเชิงสำรวจ และการวิเคราะห์องค์ประกอบเชิงยืนยัน

ผลการวิจัย: เมื่อทำการวิเคราะห์ด้วยการวิเคราะห์องค์ประกอบเชิงสำรวจ (EFA) พบว่า ได้สกัดองค์ประกอบด้วยวิธีการวิเคราะห์องค์ประกอบหลัก (Principal component) และหมุนแกนองค์ประกอบด้วยวิธี Varimax ทำให้ได้ 2 องค์ประกอบ คือ

- 1) การหลีกเลี่ยงสถานการณ์ช่วยต่อการสูบบุหรี่
- 2) การจัดการตนเองต่อการสูบบุหรี่

จากนั้นจึงทำการวิเคราะห์องค์ประกอบเชิงยืนยัน พบว่า เครื่องมือวัดพฤติกรรมป้องกันการสูบบุหรี่ 2 องค์ประกอบ มีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ดังนี้

องค์ประกอบที่ 1 การหลีกเลี่ยงสถานการณ์ช่วยต่อการสูบบุหรี่ มีจำนวน 3 ข้อ

องค์ประกอบที่ 2 การจัดการตนเองต่อการสูบบุหรี่ มีจำนวน 5 ข้อ

สรุปว่า แบบวัดพฤติกรรมป้องกันการสูบบุหรี่ในกลุ่มวัยรุ่นที่สร้างขึ้น มีความเที่ยงตรงเชิงโครงสร้าง

แบบวัดความเป็นมืออาชีพทางการแพทย์ (ธนยศ สุมาลย์โรจน์, นำชัย ศุภฤกษ์ชัยสกุล, และเจียรชัย งามทิพย์วัฒนา, 2561)

<p>เครื่องมือวัด: แบบวัดความเป็นมืออาชีพทางการแพทย์</p> <p>ความเป็นมา: มีแนวคิดความเป็นมืออาชีพทางการแพทย์ของ American Board of Internal Medicine (ABIM) จำนวน 6 องค์ประกอบ</p> <p>วัตถุประสงค์การวิจัย: เพื่อพัฒนาและตรวจสอบคุณภาพแบบวัดความเป็นมืออาชีพทางการแพทย์สำหรับนักศึกษาแพทย์</p> <p>ผลการวิจัย: แบบวัดความเป็นมืออาชีพทางการแพทย์ จำนวน 32 ข้อ ทำการวิเคราะห์องค์ประกอบเชิงยืนยัน พบว่า เครื่องมือวัดความเป็นมืออาชีพทางการแพทย์ 6 องค์ประกอบ มีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ดังนี้</p> <p>องค์ประกอบที่ 1 การเห็นแก่ประโยชน์ของผู้อื่นมากกว่าประโยชน์ส่วนตน จำนวน 6 ข้อ</p> <p>องค์ประกอบที่ 2 การมีความรับผิดชอบทั้งต่อตนเอง ผู้ป่วย สังคม และวิชาชีพ จำนวน 5 ข้อ</p> <p>องค์ประกอบที่ 3 การหมั่นพัฒนาตนเองอย่างสม่ำเสมอ จำนวน 5 ข้อ</p> <p>องค์ประกอบที่ 4 การให้ความสำคัญต่อบทบาทของวิชาชีพ จำนวน 5 ข้อ</p> <p>องค์ประกอบที่ 5 การมีความซื่อสัตย์และยุติธรรม จำนวน 5 ข้อ</p> <p>องค์ประกอบที่ 6 การเคารพในสิทธิของผู้อื่น จำนวน 6 ข้อ</p> <p>สรุปว่า แบบวัดความเป็นมืออาชีพทางการแพทย์ที่สร้างขึ้นตามแนวคิดความเป็นมืออาชีพทางการแพทย์ของ ABIM มีความเที่ยงตรงเชิงโครงสร้าง</p>
--

3.3 วิธีการเทียบกับกลุ่มที่รู้จักหรือกลุ่มตรงข้าม (Method of Known Group or Contrasted Group) วิธีการนี้มีหลักการว่า เครื่องมือวัดที่ออกแบบมาใช้วัดคุณลักษณะใดลักษณะหนึ่ง ถ้าเครื่องมือมีความเที่ยงตรงเชิงโครงสร้างต้องสามารถจำแนกหรือบอกความแตกต่างระหว่างกลุ่มที่รู้ว่ามีความรู้ว่ามีคุณลักษณะนั้นกับกลุ่มที่รู้ว่าไม่มีคุณลักษณะนั้น ดังนั้น จึงใช้วิธีการตรวจสอบด้วยวิธีการเปรียบเทียบความแตกต่างอย่างน้อย 2 กลุ่มในสิ่งที่ต้องการวัด แล้วนำคะแนนจากทั้งสองกลุ่มมาวิเคราะห์เปรียบเทียบความแตกต่างด้วยการทดสอบค่าที (t-test) หรือกรณีมากกว่า 2 กลุ่ม ให้ใช้การวิเคราะห์แปรปรวน (ANOVA) ถ้าพบว่ามีความแตกต่างกัน แสดงว่า แบบวัดที่สร้างขึ้นมีความเที่ยงตรงเชิงโครงสร้าง (โชติกา ภาษิมผล, 2559; สุจิตรา เทียนสวัสดิ์, 2562)

ตัวอย่าง การตรวจสอบความเที่ยงตรงเชิงโครงสร้างด้วยวิธีการเทียบกับกลุ่มรู้ชัดหรือกลุ่มตรงข้าม
แบบวัดทักษะการทำอาหารและการรับประทานอาหารเพื่อสุขภาพ (Jomori et al., 2017)

เครื่องมือวัด: แบบวัดทักษะการทำอาหารและการรับประทานอาหารเพื่อสุขภาพ
ความเป็นมา: กลุ่มเพศหญิงและกลุ่มที่มีความรู้การทำอาหารระดับสูง มีทักษะการทำอาหารและการรับประทานอาหารเพื่อสุขภาพสูงกว่า
วัตถุประสงค์การวิจัย: เพื่อตรวจสอบความเที่ยงตรงเชิงโครงสร้างด้วยวิธีการเทียบกับกลุ่มรู้ชัดของแบบวัดทักษะการทำอาหารและการรับประทานอาหารเพื่อสุขภาพ
สถิติวิเคราะห์: การทดสอบค่าที (t-test)
ผลการวิจัย: เพศและความรู้การทำอาหารที่แตกต่างกัน มีผลต่อทักษะการทำอาหารและการรับประทานอาหารเพื่อสุขภาพแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05
สรุปว่า แบบวัดทักษะการทำอาหารและการรับประทานอาหารเพื่อสุขภาพที่สร้างขึ้น มีความเที่ยงตรงเชิงโครงสร้าง

3.4 วิธีการวิเคราะห์พหุคุณลักษณะพหุวิธี (Multitrait-multimethod) วิธีการนี้เป็นแนวคิดของ Campbell and Fiske (1959) เป็นการวิเคราะห์ความสัมพันธ์ระหว่างการวัดอย่างน้อยสองตัวแปร แต่ละตัวแปรใช้วิธีการวัดที่ต่างกันอย่างน้อยสองวิธี แล้วนำข้อมูลที่ได้นำมาวิเคราะห์ความสัมพันธ์ ทำให้ได้ค่าสัมประสิทธิ์สหสัมพันธ์ทั้งหมด โดยแสดงเป็นตารางเมทริกซ์ (Matrix) เรียกว่า “เมทริกซ์พหุคุณลักษณะพหุวิธี” (multitrait-multimethod matrix: MTMM) เช่น การวัดคุณลักษณะที่แตกต่างกัน 2 ลักษณะ ได้แก่ การมองโลกในแง่ดี และความหวัง โดยใช้วิธีการวัดต่างกัน 2 วิธี คือ แบบวัดรายงานตนเอง และแบบสังเกต ดังตารางที่ 8-1

ตารางที่ 8-1 ลักษณะเมทริกซ์พหุคุณลักษณะพหุวิธี เมื่อวัด 2 ลักษณะ ด้วยการวัด 2 วิธี

คุณลักษณะ	คุณลักษณะ	การมองโลกในแง่ดี		ความหวัง	
	วิธีการวัด	แบบวัด	แบบสังเกต	แบบวัด	แบบสังเกต
การมองโลกในแง่ดี	แบบวัด	1			
	แบบสังเกต	2	1		
ความหวัง	แบบวัด	3	4	1	
	แบบสังเกต	4	3	2	1

ค่าสัมประสิทธิ์สหสัมพันธ์ที่ปรากฏใน MTMM เมทริกซ์ สามารถจัดได้เป็น 4 กลุ่ม (Miller et al., 2013; Nunnally & Bernstein, 1994) ซึ่งมีรายละเอียดดังนี้

1) สัมประสิทธิ์ความเชื่อมั่น (reliability coefficient) คือ ความสอดคล้องภายใน (internal consistency) ของแบบวัดหรือการวัดแต่ละวิธี นั่นคือ สัมประสิทธิ์สหสัมพันธ์ระหว่างคุณลักษณะเดียวกัน โดยใช้วิธีการวัดเดียวกัน เปรียบเหมือนการวัดซ้ำ ซึ่งควรจะมีค่าเท่ากับ 1.00 แต่อย่างไรก็ตาม ในช่องนี้ได้ใส่ค่าสัมประสิทธิ์ค่าความเชื่อมั่นในเมทริกซ์ นั่นคือ ตัวเลขหมายเลข 1 จากตารางที่ 8-1

2) สัมประสิทธิ์สหสัมพันธ์ระหว่างคุณลักษณะเดียวกันที่วัดด้วยวิธีการวัดต่างกัน (monotrait-heteromethod correlation coefficient) จะมีความแปรปรวนร่วมระหว่างคุณลักษณะที่วัด (shared trait variance) เรียกว่า สัมประสิทธิ์ความเที่ยงตรงเชิงเสมือน (convergent validity coefficient) นั่นคือ ตัวเลขหมายเลข 2 จากตารางที่ 8-1

3) สัมประสิทธิ์สหสัมพันธ์ระหว่างคุณลักษณะต่างกันที่วัดด้วยวิธีการวัดเดียวกัน (heterotrait-monomethod correlation coefficient) จะมีความแปรปรวนร่วมระหว่างวิธีการวัด (shared method variance) เรียกว่า สัมประสิทธิ์ความเที่ยงตรงเชิงจำแนก (discriminant validity coefficient) นั่นคือ ตัวเลขหมายเลข 3 จากตารางที่ 8-1

4) สัมประสิทธิ์สหสัมพันธ์ระหว่างคุณลักษณะต่างกันที่วัดด้วยวิธีการวัดต่างกัน (heterotrait-heteromethod correlation coefficient) โดยจะไม่มีค่าแปรปรวนร่วมทั้งคุณลักษณะและวิธีการวัด นั่นคือ ตัวเลขหมายเลข 4 จากตารางที่ 8-1

สำหรับการแปลผลค่าสัมประสิทธิ์สหสัมพันธ์ในเมทริกซ์พหุลักษณะพหุวิธี พิจารณาได้ดังนี้ (Nunnally & Bernstein, 1994)

(1) ค่าสัมประสิทธิ์ความเชื่อมั่น (หมายเลข 1) ควรจะมีค่าสูงสุด และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคุณลักษณะต่างกันวัดด้วยวิธีการวัดที่ต่างกันด้วย (หมายเลข 4) ควรจะมีค่าต่ำสุด

(2) ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคุณลักษณะเดียวกัน แต่วัดด้วยวิธีการวัดที่ต่างกัน (หมายเลข 2) ควรมีค่าสูง เพื่อแสดงว่า เครื่องมือวัดนั้นมีความเที่ยงตรงเชิงเสมือน (convergent validity)

(3) ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคุณลักษณะที่แตกต่างกัน แต่วัดด้วยวิธีการวัดเดียวกัน (หมายเลข 3) ควรมีค่าต่ำ เพื่อแสดงว่า เครื่องมือวัดนั้นมีความเที่ยงตรงเชิงจำแนก (discriminant validity)

(4) ถ้าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคุณลักษณะที่แตกต่างกัน แต่วัดด้วยวิธีเดียวกัน (หมายเลข 3) มีค่าสูงกว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคุณลักษณะเดียวกัน แต่วัดด้วยวิธีการวัดที่ต่างกัน (หมายเลข 2) แสดงให้เห็นว่า วิธีการวัดมีความสำคัญมากกว่าสิ่งที่ต้องการวัด หรือเรียกว่า ความแปรปรวนของวิธีการวัดเดียวกัน (common method variance) (Streiner, Norman, & Cairney, 2015)

ในปัจจุบัน พบว่า วิธีการตรวจสอบความเที่ยงตรงเชิงโครงสร้างด้วยการวิเคราะห์พหุคุณลักษณะ พหุวิธีมีข้อจำกัดต่าง ๆ มากมาย เนื่องจากการขาดการเก็บรวบรวมข้อมูลที่ใช้หลายวิธีการวัดสำหรับตัวแปรหนึ่ง ๆ (Brown, 2015) รวมทั้งมีความเป็นไปได้ที่จะเกิดค่าสัมประสิทธิ์สหสัมพันธ์ที่อาจสูงกว่าความเป็นจริง (Conway, Lievens, Scullen, & Lance, 2004) นอกจากนี้ การเก็บข้อมูลหลายวิธีการวัดยังใช้ค่าใช้จ่ายสูง และเสียเวลาค่อนข้างมาก รวมทั้งอาจส่งผลกระทบต่อคำตอบของกลุ่มตัวอย่างได้ (Waltz, Strickland, & Lenz, 2005)

ที่ผ่านมา ได้มีการคิดหาวิธีการวิเคราะห์ข้อมูลอื่นอีกหลายวิธีที่มักใช้กันมากคือ การวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) นอกจากนี้ วิธีการ MTMM ที่มุ่งเน้นวิธีการวัดหลายวิธีอาจทำให้เกิดอคติด้านวิธีการวัด จึงให้ใช้วิธีการวัดเดียวในหลายคุณลักษณะ (สุจิตรา เทียนสวัสดิ์, 2562; Bagozzi, Yi, & Phillips, 1991) โดยมีรายละเอียดเกี่ยวกับความเที่ยงตรงเชิงเสมือนและเชิงจำแนก (Trochim & Donnelly, 2008) ดังนี้

(1) ความเที่ยงตรงเชิงเสมือน (convergent validity) เป็นการวัดในเครื่องมือวัดเดียวกัน โดยมุ่งเน้นความสอดคล้องกันของข้อคำถามต่าง ๆ ในตัวแปรนั้น ๆ ใช้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อคำถามภายในตัวแปรเดียวกัน (within-construct correlation)

(2) ความเที่ยงตรงเชิงจำแนก (discriminant validity) เป็นการวัดในเครื่องมือวัดที่ต่างกัน ใช้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อคำถามของตัวแปรที่ต่างกัน (cross-construct correlations)

ปัจจัยที่ส่งผลต่อความเที่ยงตรง

ปัจจัยที่ส่งผลต่อความเที่ยงตรงของเครื่องมือวัด สามารถสรุปแหล่งที่สำคัญที่ส่งผลต่อความเที่ยงตรงได้ 4 แหล่ง (โชติกา ภาชีผล, 2559; อรพินทร์ ชูชม, 2545) ดังนี้

1. ปัจจัยจากตัวแบบทดสอบ/แบบสอบถามเอง

กระบวนการสร้างเครื่องมือวัดและส่วนประกอบของข้อคำถามที่มีผลต่อความเที่ยงตรง ดังนี้

- คำชี้แจง/คำสั่งที่ไม่ชัดเจน ทำให้เกิดการตีความหมายผิดไป ซึ่งมีส่วนทำให้ความเที่ยงตรงลดน้อยลง
- ใช้โครงสร้างทางด้านภาษาที่ซับซ้อนเกินไป ทำให้เกิดความสับสนได้ และวัดได้ไม่ตรงกับสิ่งที่ต้องการวัด นั่นคือ คะแนนที่ได้ไม่สามารถเป็นตัวแทนที่ดีได้
- ระดับความยากง่ายที่ไม่เหมาะสม เช่น มีข้อคำถามที่ง่ายหรือยากเกินไป ไม่สามารถจำแนกผู้สอบได้อย่างน่าเชื่อถือ ส่งผลต่อความเที่ยงตรงที่ต่ำ
- ข้อคำถามที่ชี้นำคำตอบ ทำให้ผู้ตอบค้นหาคำตอบจากคำที่ชี้นำ ส่งผลทำให้ไม่ได้วัดสิ่งที่ต้องการอย่างแท้จริง
- ข้อคำถามที่มีจำนวนข้อน้อยเกินไป หรือข้อคำถามที่ไม่ได้วัดเนื้อหาของสิ่งที่ต้องการวัด จะทำให้ความเที่ยงตรงลดลง

ดังนั้น การสร้างและพัฒนาเครื่องมือวัดจึงเป็นสิ่งสำคัญที่ผู้สร้างต้องพยายามลดหรือขจัดปัญหาเหล่านี้ ออกไปให้ได้มากที่สุด เพื่อให้เครื่องมือวัดมีความเที่ยงตรงมากยิ่งขึ้น

2. ปัจจัยจากการบริหารจัดการและการให้คะแนน

การบริหารจัดการ รวมทั้งการให้คะแนนของเครื่องมือ ส่งผลต่อความเที่ยงตรง ดังนี้

- เวลาที่ใช้ในการทำแบบทดสอบ/แบบสอบถามไม่เหมาะสม อาจให้เวลามากเกินไป หรือน้อยเกินไป มีผลทำให้ความเที่ยงตรงลดลง

- สภาพแวดล้อมในการทำแบบทดสอบ/แบบสอบถามที่ไม่เหมาะสม ทั้งทางด้านกายภาพ เช่น อุณหภูมิห้อง แสง เสียงรบกวน เป็นต้น และทางด้านจิตวิทยา เช่น ความตึงเครียด เป็นต้น สิ่งเหล่านี้ส่งผลให้ความเที่ยงตรงลดลง

- การตรวจให้คะแนนขาดความเป็นปรนัย ขาดเกณฑ์การให้คะแนนที่เหมาะสม ผู้ตรวจให้คะแนนลำเอียง ส่งผลให้ความเที่ยงตรงลดลง

3. ปัจจัยจากตัวผู้ตอบเอง

สำหรับตัวผู้ตอบเอง ก็มีผลต่อความเที่ยงตรง ดังนี้

- ความเป็นเอกพันธ์ของกลุ่มผู้ตอบ เช่น กลุ่มผู้ตอบอาจมีความใกล้เคียงกันในบางตัวแปรที่อาจมีผลต่อความเที่ยงตรงได้ เช่น ความสามารถ ทักษะ สติปัญญา ภูมิหลัง เป็นต้น ซึ่งมีผลทำให้เกิดการกระจายของคะแนนลดลง ส่งผลต่อความเที่ยงตรงได้

- นิสัยในการทำแบบทดสอบ/แบบสอบถาม เช่น การชอบเดาคำตอบ ทำให้เกิดความคลาดเคลื่อนของการตอบมาก ส่งผลให้ความเที่ยงตรงลดลง

- สภาพความพร้อมทางร่างกายและจิตใจ เช่น มีอาการเจ็บป่วย เหนื่อย ตาล้า เหนื่อยเพลีย วิตกกังวล ย่อมทำให้มีประสิทธิภาพในการทำแบบทดสอบ/แบบสอบถามลดลง ส่งผลต่อความเที่ยงตรงได้

4. ปัจจัยจากเกณฑ์ที่ใช้อ้างอิง

สำหรับเกณฑ์ที่ใช้อ้างอิงนั้นมีความสำคัญ สามารถส่งผลต่อความเที่ยงตรง ดังนี้

- ความเหมาะสมในการนำเกณฑ์มาใช้เปรียบเทียบ ถ้านำเครื่องมือวัดที่มีคุณภาพต่ำมาทำการเปรียบเทียบ ก็ย่อมส่งผลต่อความเที่ยงตรงลดลง

- ความเหมาะสมของแนวคิดทฤษฎีที่นำมาใช้ เช่น แนวคิดทฤษฎียังไม่ได้เป็นที่ยอมรับมาใช้ดำเนินโครงสร้าง ซึ่งมีผลทำให้สิ่งที่วัดหรือศึกษาอาจไม่ใช่คุณลักษณะที่ต้องการในโครงสร้างของการวัด ย่อมส่งผลให้ความเที่ยงตรงลดลง

สรุป

ความเที่ยงตรง ถือเป็นกระบวนการหนึ่งที่มีความสำคัญในการพัฒนาเครื่องมือวิจัย เพราะเป็นการตรวจสอบคุณภาพของเครื่องมือวัด หากเครื่องมือวัดขาดความเที่ยงตรง แม้จะมีคุณสมบัติทางการวัดอื่น ๆ ที่ดีและเหมาะสมก็ตาม เครื่องมือวัดนั้นจะขาดประโยชน์ทันที โดยเครื่องมือวัดสามารถดำเนินการตรวจสอบความเที่ยงตรงได้ 3 ประเภทหลัก ๆ นั่นคือ ความเที่ยงตรงเชิงเนื้อหา ความเที่ยงตรงตามเกณฑ์ และความเที่ยงตรงเชิงโครงสร้าง ความเที่ยงตรงช่วยทำให้เครื่องมือวัดมีความสมบูรณ์มากยิ่งขึ้น ซึ่งเหมาะสมกับตัวแปรทางด้านพฤติกรรมศาสตร์ ซึ่งเป็นตัวแปรที่มีความซับซ้อน การตรวจสอบคุณภาพเครื่องมือด้วยความเที่ยงตรงถือเป็นแนวทางในการสร้างและพัฒนาเครื่องมือทางด้านพฤติกรรมศาสตร์ให้มีความเหมาะสม และสามารถวัดได้ตรงจุดมากขึ้น

เอกสารอ้างอิง

- ชาริน สุวรรณวงศ์, และศรัณย์ พิมพ์ทอง. (2563). การสร้างและตรวจสอบคุณภาพแบบวัดพฤติกรรมป้องกัน การสูบบุหรี่ในกลุ่มวัยรุ่น. *วารสารเกื้อการุณย์*, 27(1), 46-58.
- โชติกา ภาชีผล. (2559). *การวัดและประเมินผลการเรียนรู้*. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ธนยศ สุมาลัยโรจน์, นำชัย ศุภฤกษ์ชัยสกุล, และเอียรชัย งามทิพย์วัฒนา. (2561). ความเป็นมืออาชีพทางการแพทย์: การพัฒนาและตรวจสอบคุณภาพของแบบวัด. *วารสารพฤติกรรมศาสตร์เพื่อการพัฒนา*, 10(1), 126-145.
- นรุตม์ พรประสิทธิ์. (2562). การพัฒนามาตรวัดการเห็นคุณค่าในตนเองจากความสัมพันธ์. *วารสารสมาคมนักวิจัย*, 24(2), 200-211.
- อรพินทร์ ชูชม. (2545). *เอกสารคำสอนวิชา วป 502 การสร้างและพัฒนาเครื่องมือวัดทางพฤติกรรมศาสตร์*. กรุงเทพฯ: สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- อรรวรรณ ศิลปกิจ. (2558). การสร้างแบบประเมินสติ. *วารสารสุขภาพจิตแห่งประเทศไทย*, 23(2), 72-90.
- สุจิตรา เทียนสวัสดิ์. (2562). *การพัฒนาเครื่องมือสำหรับการวิจัยทางการพยาบาล*. เชียงใหม่: สยามพิมพ์นานา.
- Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly*, 36(3), 421-458.
- Berk, R. A. (1990). Importance of expert judgment in content-related validity evidence. *Western Journal of Nursing Research*, 12(5), 659-671.

- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York: The Guilford Press.
- Burns, N., & Grove, S. (2009). *The practice of nursing research: Appraisal, synthesis, and generation of evidence* (6th ed.). St. Louise, MO: Saunders/Elsevier.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81-105.
- Coaley, K. (2014). *An introduction to psychological assessment & psychometrics* (2nd ed.). Los Angeles: Sage.
- Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2013). *Psychological testing and Assessment: An introduction to tests and measurement* (8th ed.). New York: McGraw-Hill.
- Conway, J. M., Lievens, F., Scullen, F., & Lance, C. E. (2004). Bias in the correlated uniqueness model for MTMM data. *Structural Equation Modeling*, *11*(4), 535-559.
- Davis, L. (1992). Instrument review: Getting the most from your panel of experts. *Applied Nursing Research*, *5*(4), 104-107.
- Finch, W. H. (2013). Exploratory factor analysis. In T. Teo (Ed.), *Handbook of quantitative methods for educational research* (pp. 167-186). Rotterdam, Netherlands: Sense Publishers.
- Flora, D. B., & Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale development and validation. *Canadian Journal of Behavioural Science*, *49*(2), 78-88.
- Gregory, R. J. (1996). *Psychological testing: History, principles, and applications* (2nd ed.). Boston: Allyn & Bacon.
- Haig, B. D. (2005). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research*, *40*(3), 303-329.
- Hoyle, R. H. (2004). Confirmatory factor analysis. In M. Lewis-Beck, A. Bryman, & T. Liao (Eds.), *Encyclopedia of social science research methods* (Vol. 1, pp. 169-175). Thousand Oaks, CA: Sage.
- Jomori, M. M., Proença, R. P. C., Echevarria-Guanilo, M. E., Bernardo, G. L., Uggioni, P. L., & Fernandes, A. C. (2017). Construct validity of Brazilian cooking skills and healthy eating questionnaire by the known-groups method. *British Food Journal*, *119*(5), 1003-1016.

- Kaplan, R. M., & Saccuzzo, D. P. (2013). *Psychological assessment and theory: Creating and using psychological tests* (8th ed.). Boston, MA: Cengage Learning.
- Lindh, A. U., Waern, M., Beckman, K., Renberg, E. S., Dahlin, M., & Runeson, B. (2018). Short term risk of non-fatal and fatal suicidal behaviours: The predictive validity of the Columbia-Suicide Severity Rating Scale in a Swedish adult psychiatric population with a recent episode of self-harm. *BMC Psychiatry, 18*, 319.
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382-385.
- Miller, L. A., Lovler, R. L., & McIntire, S. A. (2013). *Foundations of psychological testing: A practical approach* (4th ed.). Los Angeles: Sage.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Rovinelli, R. J., & Hambleton, R. K. (1977). On the use of content specialists in the assessment of criterion-referenced test item validity. *Dutch Journal for Educational Research, 2*, 49-60.
- Spek, V., Lemmens, F., Chatrou, M., van Kempen, S., Pouwer, F., & Pop, V. (2013). Development of a smoking abstinence self-efficacy questionnaire. *International Journal of Behavioral Medicine, 20*, 444-449.
- Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use* (5th ed.). Oxford: Oxford University Press.
- Trochim, W. M. K., & Donnelly, J. P. (2008). *Research methods knowledge base* (3rd ed.). Mason, OH: Atomic Dog/Cengage Learning.
- Waltz, C., Strickland, O., & Lenz, E. (2005). *Measurement in nursing and health research* (3rd ed.). New York: Springer Publishing Company.
- Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology, 44*(3), 219-246.

บทที่ 9

ความเชื่อมั่น

บทที่ 9 เป็นการนำเสนอเกี่ยวกับความเชื่อมั่น โดยผู้เขียนต้องการให้ผู้อ่านได้ทราบและทำความเข้าใจเกี่ยวกับความเชื่อมั่น แนวคิดเบื้องต้นของความเชื่อมั่น รวมทั้งความเชื่อมั่นประเภทต่าง ๆ เนื่องจากความเชื่อมั่นแต่ละประเภท จะทำให้การตรวจสอบคุณภาพเครื่องมือวัดทั้งฉบับที่ไม่เหมือนกัน ทั้งนี้ขึ้นอยู่กับลักษณะของข้อมูล ด้วย ความเชื่อมั่นถือว่าเป็นวิธีการหนึ่งในการตรวจสอบคุณภาพของเครื่องมือวัด ว่าเครื่องมือวัดนั้นมีคุณสมบัติของการวัดที่มีความคงเส้นคงวาหรือไม่ ไม่ว่าจะวัดกี่ครั้ง ผลที่ได้ควรจะเหมือนเดิมหรือใกล้เคียงกับของเดิม โดยบทนี้จะช่วยให้ผู้วิจัยและนักวิชาการสามารถสร้างเครื่องมือวัดในสาขาพฤติกรรมศาสตร์ได้อย่างเชื่อถือ เพื่อช่วย กระบวนการพัฒนาเครื่องมือวิจัยทางพฤติกรรมศาสตร์มีความเหมาะสมและสมบูรณ์มากยิ่งขึ้น

แนวคิดเบื้องต้นของความเชื่อมั่น

คุณลักษณะที่สำคัญของเครื่องมือวัด นั่นคือ ความเชื่อมั่น (Reliability) เป็นมาตรฐานที่จะกำหนดว่า ข้อมูลมีความน่าเชื่อถือมากน้อยเพียงใด (Miller, Lovler, & McIntire, 2013) เมื่อใดก็ตามที่มีการนำเครื่องมือวัดไปใช้ เครื่องมือวัดที่มีความเชื่อมั่น ไม่ว่าจะวัดกี่ครั้งก็ได้ผลเหมือนเดิมหรือใกล้เคียงของเดิม ดังนั้น ความเชื่อมั่น (Reliability) จึงหมายความว่า ความคงที่หรือความคงเส้นคงวาของผลที่ได้ไม่ว่าจะวัดกี่ครั้งก็ตาม และเครื่องมือวัดที่มีความเชื่อมั่น หมายถึง เครื่องมือวัดนั้นมีคุณสมบัติที่ให้ผลที่คงเส้นคงวา (consistency) มีความคงที่ไม่ว่าจะวัดกี่ครั้งก็ตาม (ศิริชัย กาญจนวาสี, 2548; อรพินทร์ ชูชม, 2545; Cohen, Swerdlik, & Sturman, 2013; Gregory, 1996; Reynolds, Livingston, & Willson, 2010; Sax & Newton, 1997) กล่าวคือ กลุ่มตัวอย่างสถานการณ์เดียวกันและไม่ได้มีการเรียนรู้เพิ่มเติม ผลที่ได้จากการวัดควรจะเหมือนเดิมหรือใกล้เคียงของเดิม เช่น ให้กลุ่มตัวอย่างทำแบบวัดบุคลิกภาพห้าองค์ประกอบในครั้งแรก ได้เว้นระยะเวลาไป 1 เดือน นำแบบวัดบุคลิกภาพห้าองค์ประกอบชุดเดิมหรือชุดที่คล้ายคลึงกันมาทำอีกครั้ง ได้ผลใกล้เคียงกับครั้งแรก แสดงว่า แบบวัดบุคลิกภาพห้าองค์ประกอบดังกล่าวมีความน่าเชื่อถือ หรือกล่าวได้ว่า เครื่องมือวัดบุคลิกภาพห้าองค์ประกอบมีความเชื่อมั่น

จากแนวคิดทฤษฎีการทดสอบดั้งเดิม (Classical test theory) หรือเรียกว่าทฤษฎีคะแนนจริงและความคลาดเคลื่อน (the theory of true and error scores) มีพื้นฐานความคิดที่ว่า คะแนนการทดสอบเป็นผลมาจากสองปัจจัย (Gregory, 1996) คือ 1) ปัจจัยที่ส่งเสริมความคงเส้นคงวา ประกอบด้วย คุณลักษณะคงที่ของบุคคลที่ทำการวัด และ 2) ปัจจัยที่ส่งเสริมความไม่คงเส้นคงวา ได้แก่ คุณลักษณะของบุคคล แบบทดสอบหรือสถานการณ์ซึ่งไม่เกี่ยวข้องกับคุณลักษณะที่ทำการวัด แต่อย่างไรก็ตามมีผลต่อคะแนนการทดสอบ โดยปัจจัยแรก

เป็นสิ่งที่ปรารถนา เนื่องจากเป็นค่าคะแนนคุณลักษณะที่แท้จริง ในขณะที่ปัจจัยที่สองแสดงถึงปัจจัยความคลาดเคลื่อนที่หลีกเลี่ยงไม่ได้ของการวัด ดังนั้น จึงสามารถเขียนเป็นสมการได้ดังนี้

$$X = T + e$$

เมื่อ X คือ คะแนนที่วัดได้
 T คือ คะแนนจริง
 e คือ ความคลาดเคลื่อนของการวัด

ดังนั้น ความคลาดเคลื่อนทางการวัด (error of measurement) จึงเป็นผลต่างของคะแนนที่วัดได้กับคะแนนจริง ยังเป็นสิ่งที่ลดประโยชน์ของการวัดด้วย (Reynolds et al., 2010) ดังสมการ

$$e = X - T$$

จากสมการดังกล่าวเกี่ยวกับความคลาดเคลื่อนทางการวัด สามารถเป็นได้ทั้งค่าบวกและค่าลบ ถ้าความคลาดเคลื่อนทางการวัดเป็นบวก แสดงว่า คะแนนที่วัดได้มากกว่าคะแนนจริง หรือในทางกลับกันถ้าความคลาดเคลื่อนทางการวัดเป็นลบ แสดงว่า คะแนนที่วัดได้น้อยกว่าคะแนนจริง ถึงแม้ว่าจะเป็นไปไม่ได้ที่จะกำจัดความคลาดเคลื่อนทางการวัดทั้งหมด แต่ผู้สร้างสามารถลดความคลาดเคลื่อนของการวัดจากแหล่งต่าง ๆ ให้เหลือน้อยที่สุด

แหล่งความคลาดเคลื่อนของการวัด (Sources of measurement error) มีรายละเอียดที่สำคัญดังนี้ (Cohen et al., 2013; Gregory, 1996; Reynolds et al., 2010)

1. การสร้างเครื่องมือวัด (Test construction) โดยแหล่งความคลาดเคลื่อนของการวัดที่เกิดขึ้นในช่วงของการสร้างเครื่องมือวัด คือ การสุ่มข้อคำถาม (item sampling) หรือการสุ่มเนื้อหา (content sampling) หมายถึง ความแตกต่างของข้อคำถามภายในเครื่องมือวัดนั้น รวมทั้งความแตกต่างของข้อคำถามระหว่างเครื่องมือวัด เช่น เครื่องมือวัดทางด้านพฤติกรรมศาสตร์ อาจมีการวัดทักษะ ความรู้ บุคลิกภาพ หรือพฤติกรรมของบุคคล ความแตกต่างอาจพบได้จากคำที่ใช้ รวมทั้งเนื้อหาที่นำมาใช้ ยกตัวอย่างเช่น เนื้อหาที่นำมาใช้ในแบบทดสอบผลสัมฤทธิ์ทางการเรียน นักเรียนอาจจะคุ้นเคยกับเนื้อหาพอสมควร ซึ่งทำแบบทดสอบได้ดีกว่านักเรียนคนอื่น อาจเกิดความคลาดเคลื่อนทางการวัดได้ ดังนั้น การออกแบบควรลดความคลาดเคลื่อนทางการวัดที่เกิดจากการสุ่มข้อคำถามให้เหลือน้อยที่สุด

2. การบริหารจัดการเครื่องมือวัด (Test administration) โดยแหล่งความคลาดเคลื่อนของการวัดที่เกิดขึ้นในช่วงของการบริหารจัดการเครื่องมือวัด อาจมีอิทธิพลต่อความตั้งใจและแรงจูงใจในการทำแบบทดสอบ/แบบสอบถาม ปฏิบัติการของผู้ทำแบบทดสอบถือเป็นแหล่งความคลาดเคลื่อนของการวัดรูปแบบหนึ่ง เช่น อุณหภูมิห้อง ระดับของแสง อากาศถ่ายเท และเสียง นอกจากสภาพแวดล้อมแล้ว ยังมีปัจจัยจากตัวผู้ทำแบบทดสอบ เช่น ความวิตกกังวล แรงจูงใจ ความตั้งใจ และความเหนื่อยล้าของตัวผู้ทดสอบ อาจเป็นไปได้ว่านอนหลับไม่เพียงพอ จึงทำให้ขาดสมาธิ มีผลต่อการอ่านข้อความผิดพลาดได้ รวมทั้งยังมีปัจจัยจากผู้ทดสอบ (examiner) ด้วย เช่น ในการสอบปากเปล่า ผู้ทดสอบอาจบอกเป็นนัย ๆ เกี่ยวกับคำสำคัญนั้น ๆ ด้วยลักษณะของการส่ายหัว การเคลื่อนไหวของตา หรือท่าทางอื่น ๆ ที่บ่งบอกถึงคำตอบที่ถูกต้องหรือไม่ถูกต้อง

3. การให้คะแนนเครื่องมือวัด (Test Scoring) หากการให้คะแนนด้วยคอมพิวเตอร์ ก็จะขจัดความคลาดเคลื่อนที่อาจเกิดขึ้นได้ แต่บางเครื่องมือวัดต้องให้คะแนนโดยบุคคลที่ได้รับการฝึกฝนมา ซึ่งสำหรับการให้คะแนนแบบอัตวิสัย (Subjective) ผู้ให้คะแนนจึงเป็นแหล่งความคลาดเคลื่อนของการวัดที่สำคัญ

สำหรับแหล่งความคลาดเคลื่อนของการวัดดังกล่าว มีผลต่อความเชื่อมั่นของเครื่องมือวัด ดังนั้นการประมาณค่าความเชื่อมั่น จึงต้องนำแนวคิดความแปรปรวนเข้ามาในสมการด้วย จึงมีสมการดังนี้ (Gregory, 1996; Reynolds et al., 2010)

$$\sigma_x^2 = \sigma_T^2 + \sigma_e^2$$

เมื่อ σ_x^2 คือ ความแปรปรวนของคะแนนที่วัดได้

σ_T^2 คือ ความแปรปรวนของคะแนนจริง

σ_e^2 คือ ความแปรปรวนของความคลาดเคลื่อนของการวัด

สัมประสิทธิ์ความเชื่อมั่น (Reliability coefficient) หรือ r_{xx} เป็นอิทธิพลเชิงสัมพันธ์ของคะแนนจริงและคะแนนความคลาดเคลื่อนของการวัดที่มีต่อคะแนนที่วัดได้ โดยสำหรับด้านคณิตศาสตร์ สัมประสิทธิ์ความเชื่อมั่นเป็นส่วนหนึ่งของความแปรปรวนของคะแนนจริงและความแปรปรวนของคะแนนที่วัดได้ ดังนี้

$$r_{xx} = \frac{\sigma_T^2}{\sigma_x^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_e^2}$$

จากสมการจะเห็นได้ว่า ถ้าความแปรปรวนของความคลาดเคลื่อนของการวัดเข้าใกล้ 0 มาก หรือไม่มี ความคลาดเคลื่อนของการวัด สิ่งที่วัดได้คือ คะแนนจริง แสดงว่า การวัดมีความน่าเชื่อถือได้อย่างสมบูรณ์ คือ

สัมประสิทธิ์ความเชื่อมั่นมีค่าเท่ากับ 1 หรือถ้าความแปรปรวนของความคลาดเคลื่อนของการวัดมีขนาดใหญ่มาก ค่าสัมประสิทธิ์ความเชื่อมั่นจะมีค่าน้อย ซึ่งในทางทฤษฎีจำกัดไว้ที่ 0 ดังนั้น ช่วงสัมประสิทธิ์ความเชื่อมั่นที่เป็นไปได้คือ ระหว่าง 0-1 และในทางปฏิบัติ เครื่องมือวัดมีสัมประสิทธิ์ความเชื่อมั่นในช่วงดังกล่าว แต่ถ้าเข้าใกล้ 1 ถือว่าเป็นเครื่องมือที่มีความน่าเชื่อถือ

วิธีการประมาณค่าความเชื่อมั่น

ความเชื่อมั่นเป็นคุณสมบัติที่สำคัญในการตรวจสอบคุณภาพเครื่องมือวัด ว่าเครื่องมือวัดนั้น ๆ มีความคงที่หรือคงเส้นคงวามากน้อยเพียงใด การประมาณค่าความเชื่อมั่นสามารถทำได้หลายวิธี ในการประมาณค่าความเชื่อมั่น หรือเรียกอีกอย่างว่า “สัมประสิทธิ์ความเชื่อมั่น” (Reliability Coefficient) โดยวิธีการประมาณค่าความเชื่อมั่น (Reliability estimate) สามารถสรุปได้เป็น 4 คุณลักษณะที่สำคัญ (โชติกา ภาชีผล, 2558; อรพินทร์ ชูชม, 2545; Cohen et al., 2013; Gregory, 1996; Kaplan & Saccuzzo, 2013; Reynolds et al., 2010; Sax & Newton, 1997) ดังตารางที่ 9-1

ตารางที่ 9-1 คุณลักษณะของวิธีการประมาณค่าความเชื่อมั่น

คุณลักษณะ	คำอธิบาย
ความคงที่ (Stability)	ความคงเส้นคงวาของผลการวัด โดยใช้เครื่องมือวัดในการทดสอบซ้ำ เช่น 4 ความเชื่อมั่นแบบวัดซ้ำ
ความเท่าเทียมกัน (Equivalence)	ความคงเส้นคงวาของผลการวัดของเครื่องมือวัด 2 อันในช่วงเวลาเดียวกัน 4 และเนื้อหาเดียวกัน เช่น ความเชื่อมั่นแบบคู่ขนาน
ความสอดคล้องภายใน (Internal Consistency)	ความคงเส้นคงวาของผลการวัดของข้อคำถามในเครื่องมือวัดนั้น ๆ เช่น ความเชื่อมั่นแบบวิธีแบ่งครึ่ง ความเชื่อมั่นแบบวิธีของคูเดอร์-ริชาร์ดสัน และความเชื่อมั่นแบบสัมประสิทธิ์แอลฟาของครอนบาค
ความเห็นระหว่างผู้ประเมิน (Inter-rater)	ความคงเส้นคงวาของผลการวัดที่ทำการประเมินโดยผู้ประเมินมากกว่า 1 คน หรือระดับความสอดคล้อง (agreement) ของผู้ประเมินต่อเครื่องมือวัดหนึ่ง ๆ เช่น ความเชื่อมั่นระหว่างผู้ประเมิน

สำหรับวิธีการประมาณค่าความเชื่อมั่น สามารถสรุปรายละเอียดได้ดังนี้

1. วิธีการประมาณค่าความเชื่อมั่นแบบวัดซ้ำ

การประมาณค่าความเชื่อมั่นแบบวัดซ้ำ (Test-Retest Reliability) หรือเรียกอีกอย่างว่าสัมประสิทธิ์ความคงที่ (Coefficient of stability) หมายถึง การประมาณค่าความเชื่อมั่นที่ได้มาจากการจับคู่ความสัมพันธ์ของ

คะแนนจากบุคคลเดียวกันต่อแบบทดสอบเดียวกัน แต่เก็บรวบรวมข้อมูลคนละช่วงเวลากัน หรืออีกนัยหนึ่งคือ การนำเครื่องมือวัดชุดเดียวกันไปทดสอบกับกลุ่มตัวอย่างเดิมสองครั้งในช่วงเวลาที่ต่างกัน เป็นดัชนีความคงที่ของ เครื่องมือวัดข้ามช่วงเวลา เนื่องจากมีเครื่องมือวัดจำนวนมากเป็นการวัดคุณลักษณะที่คงที่ จึงคาดหวังว่าเมื่อเวลา ผ่านไป เครื่องมือนั้นยังคงมีความคงที่เท่าเดิม

สิ่งสำคัญที่ควรพิจารณาเมื่อทำการคำนวณและประเมินความเชื่อมั่นแบบวัดซ้ำ นั่นคือ ช่วงเวลาระหว่าง การทดสอบทั้ง 2 แบบวัด ถ้าช่วงเวลาในการวัดซ้ำมีระยะเวลาที่สั้นมาก (เช่น รายชั่วโมง หรือรายวัน) การประมาณ ค่าความเชื่อมั่นแบบวัดซ้ำอาจจะให้ค่าสูงเกินกว่าความเป็นจริง อันเป็นผลมาจากความทรงจำและการฝึกฝนมา จากการทดสอบในครั้งแรก หรือถ้าช่วงเวลาในการวัดซ้ำมีระยะเวลาที่ยากมาก การประมาณค่าความเชื่อมั่น แบบวัดซ้ำอาจมีค่าต่ำลง มีความเป็นไปได้ว่าเกิดความไม่คงที่ของคะแนน และอาจมีการเปลี่ยนแปลงเกิดขึ้นใน ช่วงเวลาดังกล่าว (Streiner, Norman, Cairney, 2015) สำหรับในทางปฏิบัติ ไม่มีช่วงเวลาใดเป็นช่วงเวลาที่ดี ที่สุดในการประมาณค่าความเชื่อมั่นแบบวัดซ้ำ ดังนั้น สิ่งที่ควรพิจารณาคือ ความยาวของช่วงเวลาทำการวัดซ้ำ ซึ่งขึ้นอยู่กับบริบทของคุณลักษณะนั้นด้วย (Reynolds et al., 2010)

เช่น งานวิจัยของ Marx et al. (2003) ที่ได้ศึกษาความเชื่อมั่นแบบวัดซ้ำของเครื่องมือวัดสถานะทาง สุขภาพในกลุ่มผู้ป่วยที่เชื่อว่าตนเองอยู่ในสภาวะคงที่ พบว่า กรอบช่วงเวลาที่ใช้วัดตัวแปรสถานะทางสุขภาพ ตั้งแต่ 2 วัน ถึง 2 สัปดาห์ มีค่าความเชื่อมั่นแบบวัดซ้ำที่ไม่แตกต่างกัน

นอกจากนี้ การใช้วิธีการประมาณค่าความเชื่อมั่นแบบวัดซ้ำมีความเหมาะสมมากที่สุดกับตัวแปร บุคลิกภาพ ซึ่งมีความคงที่ของคะแนนในช่วงเวลาสั้น ๆ (McCrae, Kurtz, Yamagata, & Terracciano, 2011)

สำหรับวิธีการประมาณค่าความเชื่อมั่นแบบวัดซ้ำ นำคะแนนที่วัดได้จากการทดสอบครั้งแรก และ การทดสอบครั้งหลัง มาคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson Product Moment Correlation Coefficient) นอกจากนี้ ได้หาค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น (Intraclass Correlation Coefficient: ICC) ซึ่งเป็นการปรับแก้ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันที่ให้เพียงการวัดค่าสหสัมพันธ์ โดยค่า ICC เป็นการวัดค่าความเชื่อมั่นที่มีความเหมาะสมมากกว่า ซึ่งค่า ICC นั้นสะท้อนทั้งระดับค่าสหสัมพันธ์ และความสอดคล้อง (agreement) ของเครื่องมือวัด (Koo & Li, 2016) โดยค่า ICC ได้ถูกนำมาใช้กัน อย่างแพร่หลายในการประเมินความเชื่อมั่นแบบวัดซ้ำ (Gravesande, Richardson, Griffith, & Scott, 2019; Merkulova et al., 2018; Qin et al., 2019; Schuck, 2004)

ตัวอย่าง การคำนวณค่าความเชื่อมั่นแบบวัดซ้ำ

แบบวัดบุคลิกภาพเชิงรุก ทดสอบกับกลุ่มตัวอย่าง 30 คน มีวิธีการวิเคราะห์ตามโปรแกรมทางสถิติ ดังนี้

ก. ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

- เตรียมข้อมูล แล้วใช้คำสั่ง: Analyze > Correlate > Bivariate นำตัวแปรที่วัดครั้งแรกและครั้งที่สองใส่เข้าไปในช่อง Variables แล้วกดเลือก Pearson จากนั้น จึงกด OK

Correlations

		b1	b2
b1	Pearson Correlation	1	.738**
	Sig. (2-tailed)		.000
	N	30	30
b2	Pearson Correlation	.738**	1
	Sig. (2-tailed)	.000	
	N	30	30

** . Correlation is significant at the 0.01 level (2-tailed).

จากผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน พบว่า แบบวัดบุคลิกภาพเชิงรุกที่วัดครั้งที่ 1 และวัดครั้งที่ 2 มีค่า r เท่ากับ .738 นั่นคือ ค่าสัมประสิทธิ์ความเชื่อมั่นแบบวัดซ้ำของแบบวัดบุคลิกภาพเชิงรุก มีค่าเท่ากับ .738 แสดงว่า แบบวัดบุคลิกภาพเชิงรุกมีความน่าเชื่อถือได้สูงโดยให้ผลการวัดที่คงที่

ข. ค่าสัมประสิทธิ์สหสัมพันธ์แบบภายในชั้น

- เตรียมข้อมูล แล้วใช้คำสั่ง: Analyze > Scale > Reliability Analysis นำตัวแปรที่วัดครั้งแรกและครั้งที่สองใส่เข้าไปในช่อง Items แล้วกดเลือก Statistics จากนั้น จึงเลือก Intraclass correlation coefficient ในส่วน Model ให้เลือก Two-Way Mixed และ Type ให้เลือก Absolute Agreement แล้วจึงกด Continue และ OK

Intraclass Correlation Coefficient

	Intraclass Correlation ^b	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.753 ^a	.544	.874	6.966	29	29	.000
Average Measures	.859 ^c	.704	.933	6.966	29	29	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

- a. The estimator is the same, whether the interaction effect is present or not.
- b. Type A intraclass correlation coefficients using an absolute agreement definition.
- c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

สำหรับสาเหตุที่เลือก Model: Two-Way Mixed Effect โมเดลนี้มีความเหมาะสมกับการประมาณค่าความเชื่อมั่นแบบวัดซ้ำ เพราะการวัดซ้ำไม่เกี่ยวข้องกับกลุ่มตัวอย่างแบบสุ่ม และสำหรับการเลือก Type: Absolutely Agreement เหมาะสมกับความเชื่อมั่นแบบวัดซ้ำ เพราะการวัดจะไร้ความหมาย ถ้าหากไม่มีความสอดคล้องระหว่างการวัดซ้ำ (Shrout & Fleiss, 1979)

จากผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น พบว่า แบบวัดบุคลิกภาพเชิงรุกที่วัดครั้งที่ 1 และวัดครั้งที่ 2 มีค่า ICC เท่ากับ .859 (ให้พิจารณาที่ Average Measure เนื่องจากเป็นการวัด 2 ครั้ง จึงนำมาเฉลี่ยกัน) นั่นคือ ค่าสัมประสิทธิ์ความเชื่อมั่นแบบวัดซ้ำของแบบวัดบุคลิกภาพเชิงรุก มีค่าเท่ากับ .859 แสดงว่า แบบวัดบุคลิกภาพเชิงรุกมีความน่าเชื่อถือได้สูงโดยให้ผลการวัดที่คงที่

2. วิธีการประมาณค่าความเชื่อมั่นแบบคู่ขนาน

การประมาณค่าความเชื่อมั่นแบบคู่ขนาน (Parallel-Forms Reliability) หรือเรียกอีกอย่างว่าสัมประสิทธิ์ความเท่าเทียมกัน (Coefficient of Equivalence) หมายถึง การประมาณค่าความเชื่อมั่นของเครื่องมือวัดที่มีความคล้ายหรือเท่าเทียมกันในด้านของเนื้อหา รูปแบบข้อคำถาม ค่าเฉลี่ยและความแปรปรวน ไปทำการทดสอบกับกลุ่มตัวอย่างเดิมในเวลาเดียวกัน แล้วนำข้อมูลที่ได้จากการวัดของเครื่องมือวัดทั้ง 2 ชุด มาคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

นอกจากนี้ การประมาณค่าความเชื่อมั่นของเครื่องมือวัดที่สามารถทดแทนกันได้ (Alternative-Forms Reliability) หมายถึง การประมาณค่าความเชื่อมั่นจากเครื่องมือวัด 2 ชุดที่สร้างขึ้นที่แตกต่างกัน แต่สร้างมาเป็นคู่ขนานกัน ดังนั้น การสร้างเครื่องมือวัดที่สามารถทดแทนกันได้ จึงถูกออกแบบมาเพื่อให้ความเท่าเทียมกันในแง่ของเนื้อหา และระดับความยาก หรือมีโครงสร้างการวัดอันเดียวกันที่มีค่าเฉลี่ยและความแปรปรวนของคะแนนที่วัดทั้ง 2 ชุด มีความเท่ากันหรือใกล้เคียงกัน โดยในการนำเครื่องมือวัดทั้ง 2 ชุดไปใช้ทดสอบกับกลุ่มตัวอย่างนั้น ควรสลับลำดับของการทำให้สมดุลกัน อาจทำการสุ่มกลุ่มตัวอย่างร้อยละ 50 ให้ทำเครื่องมือวัด A ก่อนเครื่องมือวัด B ส่วนที่เหลือทำเครื่องมือวัด B ก่อน A (ศิริชัย กาญจนวาสิ, 2548)

การประมาณค่าความเชื่อมั่นแบบคู่ขนานมีปัญหาในเรื่องของความยากในทางปฏิบัติที่จะสร้างเครื่องมือวัดทั้งสองชุดให้มีลักษณะที่เหมือนกันทุกประการได้ ซึ่งการพัฒนาเครื่องมือวัดดังกล่าวเป็นสิ่งที่ยาก ใช้เวลานาน และมีราคาแพง

ตัวอย่าง การคำนวณค่าความเชื่อมั่นแบบคู่ขนาน

แบบวัดบุคลิกภาพเชิงรุก ทดสอบกับกลุ่มตัวอย่าง 30 คน มีวิธีการวิเคราะห์ตามโปรแกรมทางสถิติ ดังนี้
- เตรียมข้อมูล แล้วใช้คำสั่ง: Analyze > Scale > Reliability Analysis นำตัวแปรที่วัด 2 ชุด ใส่เข้า

ไปในช่อง Items ในส่วน Model ให้เลือก Parallel หรือ Strict Parallel แล้วจึงกด OK

Test for Model Goodness of Fit

Chi-Square	Value	.598
	df	1
	Sig	.439
Log of Determinant of	Unconstrained Matrix	-2.244
	Constrained Matrix	-2.223

Under the parallel model assumption

Reliability Statistics

Common Variance	.485
True Variance	.356
Error Variance	.129
Common Inter-Item Correlation	.735
Reliability of Scale	.847
Reliability of Scale (Unbiased)	.858

สำหรับ Model: Parallel: ข้อคำถามทุกข้อมีความแปรปรวนและคลาดเคลื่อนของความแปรปรวนเท่ากัน ส่วน Strict Parallel ข้อคำถามทุกข้อมีลักษณะเดียวกับ Parallel แต่มีความเท่ากันของค่าเฉลี่ย (Hayes & Pritchard, 2013)

จากผลการวิเคราะห์ความเชื่อมั่นแบบคู่ขนาน พบว่า ถ้าค่านัยสำคัญทางสถิติเกินกว่า .05 แสดงว่ายอมรับสมมติฐาน Parallel model นั่นคือ ยอมรับว่าแบบวัดบุคลิกภาพเชิงรุกมีความเท่าเทียมกันหรือมีความคู่ขนานกัน และเมื่อพิจารณาค่าความเชื่อมั่นแบบคู่ขนาน พบว่า มีค่าความเชื่อมั่นแบบคู่ขนาน เท่ากับ .847 แสดงว่าแบบวัดบุคลิกภาพเชิงรุกทั้ง 2 แบบวัดมีความน่าเชื่อถือได้สูงโดยให้ผลการวัดที่คู่ขนานกัน หรือมีความเท่าเทียมกัน

3. วิธีการประมาณค่าความเชื่อมั่นแบบสอดคล้องภายใน

การประมาณค่าความเชื่อมั่นแบบสอดคล้องภายใน (Internal Consistency Reliability) เป็นวิธีการประมาณค่าที่นิยมใช้กันมากสำหรับเครื่องมือวัดทางด้านพฤติกรรมศาสตร์ หรือเรียกอีกอย่างหนึ่งว่า สัมประสิทธิ์ความสอดคล้องภายใน (Coefficient of Internal Consistency) หมายถึง การประมาณค่าความเชื่อมั่นของเครื่องมือวัดเพื่อตรวจสอบความสอดคล้องของข้อคำถามในเครื่องมือวัดว่าวัดสิ่งเดียวกัน ถ้าข้อคำถามวัดสิ่งเดียวกัน เครื่องมือวัดก็น่าจะมีความสอดคล้องภายในสูง ในทางกลับกัน หากข้อคำถามวัดเนื้อหาที่ไม่เกี่ยวข้องสัมพันธ์กัน คะแนนจะไม่สอดคล้องกัน ส่งผลให้เครื่องมือวัดมีความสอดคล้องภายในต่ำ (ศิริชัย กาญจนวาสี, 2548)

วิธีการประมาณค่าความเชื่อมั่นแบบสอดคล้องภายใน มีหลายวิธีดังนี้

3.1 วิธีการแบบแบ่งครึ่ง (Split-Half Method) โดยการประมาณค่าความเชื่อมั่นแบบแบ่งครึ่ง หมายถึง การหาค่าสหสัมพันธ์ของคะแนน 2 ส่วน ที่ได้จากการแบ่งครึ่งของเครื่องมือวัดที่มีความเท่าเทียมกันในการวัดครั้งเดียว การวัดความเชื่อมั่นวิธีการนี้มีประโยชน์ตรงที่ไม่สามารถวัดความเชื่อมั่นของเครื่องมือวัด 2 ชุด หรือทำการทดสอบ 2 ครั้ง อาจเนื่องจากการใช้เวลาหรือมีค่าใช้จ่ายมาก สำหรับการแบ่งเป็น 2 ส่วนให้มีความเท่าเทียมกัน อาจมีการแบ่งเป็นชุดข้อคู่กับข้อคี่ หรือแบ่งเป็นครึ่งแรกกับครึ่งหลัง ซึ่งต้องวางแผนให้สองส่วนคู่ขนานกันก่อน โดยมีวิธีการคำนวณค่าความเชื่อมั่นแบบแบ่งครึ่ง (Cohen et al., 2013) ดังนี้

- (1) แบ่งเครื่องมือวัดออกเป็น 2 ส่วนที่เท่าเทียมกัน
- (2) คำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันระหว่างคะแนนของเครื่องมือวัดในแต่ละส่วน (r_{xy})
- (3) ปรับแก้ค่าความเชื่อมั่นแบบแบ่งครึ่งโดยใช้สูตรคำนวณ The Spearman-Brown formula (r_{sb})

ซึ่งมีสูตรดังนี้

$$r_{sb} = \frac{2r_{xy}}{1+r_{xy}}$$

สมมติว่า ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันของเครื่องมือวัดส่วนที่ 1 และส่วนที่ 2 มีค่าเท่ากับ .80 เราสามารถนำค่า $r_{xy} = .80$ ใส่ในสูตรดังกล่าว โดยคำนวณออกมาได้ค่าความเชื่อมั่นแบบแบ่งครึ่งเท่ากับ .889

ตัวอย่าง การคำนวณค่าความเชื่อมั่นแบบแบ่งครึ่ง

แบบวัดบุคลิกภาพเชิงรุก ทดสอบกับกลุ่มตัวอย่าง 30 คน มีวิธีการวิเคราะห์ตามโปรแกรมทางสถิติ ดังนี้

- เตรียมข้อมูล แล้วใช้คำสั่ง: Analyze > Scale > Reliability Analysis นำตัวแปรที่วัด 2 ชุด ใส่เข้าไปในช่อง Items ในส่วน Model ให้เลือก Split-half แล้วจึงกด OK

Reliability Statistics

Cronbach's Alpha	Part 1	Value	.848
		N of Items	4 ^a
	Part 2	Value	.838
		N of Items	4 ^b
	Total N of Items		8
Correlation Between Forms			.563
Spearman-Brown Coefficient	Equal Length		.721
	Unequal Length		.721
Guttman Split-Half Coefficient			.720

a. The items are: b1, b2, b3, b4.

b. The items are: b5, b6, b7, b8.

จากผลการวิเคราะห์ความเชื่อมั่นแบบแบ่งครึ่ง จำนวน 8 ข้อ โดยแบ่งเป็นส่วนแรก 4 ข้อ และส่วนหลัง 4 ข้อ พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันของเครื่องมือวัด 2 ส่วน มีค่าเท่ากับ .563 และเมื่อปรับแก้ด้วยสูตร spearman-Brown formula พบว่า มีค่าความเชื่อมั่นแบบแบ่งครึ่ง เท่ากับ .721 แสดงว่า แบบวัดบุคลิกภาพเชิงรุกทั้ง 2 ส่วนมีความน่าเชื่อถือได้

3.2 วิธีการของคูเดอร์-ริชาร์ดสัน (Kuder-Richardson Method) โดยการประมาณค่าความเชื่อมั่นด้วยวิธีการของคูเดอร์-ริชาร์ดสัน เป็นวิธีการที่แก้ไขจุดอ่อนของวิธีการแบบแบ่งครึ่ง โดยได้คิดค้นหาวิธีการประมาณค่าความเชื่อมั่นที่ไม่จำเป็นต้องแบ่งครึ่งเครื่องมือวัด และค่าความเชื่อมั่นสามารถคำนวณได้จากการทดสอบกลุ่มตัวอย่างเพียงครั้งเดียว วิธีการนี้ได้พัฒนาสูตรการคำนวณค่าความเชื่อมั่นที่รู้จักกันดีว่า Kuder-Richardson Formula 20 (หรือ KR20) และ Kuder-Richardson Formula 21 (หรือ KR21) โดยสูตรดังกล่าว ใช้ในกรณีที่มีการให้คะแนนเป็น 0 กับ 1 นั่นคือ ทำถูกต้อง 1 คะแนน และทำผิดได้ 0 คะแนน โดยส่วนใหญ่เป็นการวัดทางด้านการรู้คิด (cognitive) อย่างเช่นแบบทดสอบผลสัมฤทธิ์ทางการเรียน เป็นต้น โดยสำหรับ KR20 มีสูตรคำนวณ ดังนี้

$$KR20 = \frac{k}{k-1} \left[1 - \frac{\sum pq}{S_x^2} \right]$$

เมื่อ	KR20	คือ ความเชื่อมั่นของเครื่องมือวัด
	k	คือ จำนวนข้อของเครื่องมือวัด
	p	คือ สัดส่วนของคนที่ทำถูกในแต่ละข้อ
		= $\frac{\text{จำนวนคนที่ทำถูก}}{\text{จำนวนคนทั้งหมด}}$
	q	คือ สัดส่วนของคนที่ทำผิดในแต่ละข้อ = 1 - p
	S_x^2	คือ ความแปรปรวนของคะแนนรวม

นอกจากนี้ ยังได้พัฒนาสูตร KR21 ซึ่งกล่าวว่า สูตรการคำนวณค่าความเชื่อมั่นไม่ได้ต้องการค่า p และ q ในทุกข้อคำถาม ดังนั้น จึงได้ประมาณค่าคะแนนรวมของ pq ให้เป็นค่าคะแนนเฉลี่ย และวิธีการ KR21 มีข้อตกลงเบื้องต้นว่า ข้อคำถามทุกข้อมีค่าความยากง่ายรายข้อเท่ากัน หรือทุกข้อมีค่าเฉลี่ยความยากง่ายอยู่ที่ร้อยละ 50 ซึ่งในทางปฏิบัติเป็นการยากมากในการทำให้ค่าความยากง่ายรายข้อเท่ากันทุกข้อคำถาม (Kaplan & Saccuzzo, 2013) โดยสำหรับ KR21 มีสูตรคำนวณดังนี้

$$KR21 = \frac{k}{k-1} \left[1 - \frac{\bar{x}(k - \bar{x})}{kS_x^2} \right]$$

เมื่อ	KR21	คือ ความเชื่อมั่นของเครื่องมือวัด
	k	คือ จำนวนข้อของเครื่องมือวัด
	\bar{x}	คือ ค่าเฉลี่ยของเครื่องมือวัด
	S_x^2	คือ ความแปรปรวนของคะแนนรวม

3.3 วิธีการหาค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Coefficient Alpha) เป็นวิธีการประมาณค่าที่ Cronbach ได้ดัดแปลงมาจากสูตร KR20 ซึ่งให้ความสำคัญกับระบบการให้คะแนนไม่จำเป็นต้องเป็น 0 และ 1 คะแนนเท่านั้น แต่สามารถเป็นแบบใดก็ได้ ส่วนใหญ่วิธีการหาค่าสัมประสิทธิ์แอลฟาของครอนบาคมักใช้กันเป็นจำนวนมากสำหรับเครื่องมือวัดทางด้านพฤติกรรมศาสตร์ เพราะเป็นเครื่องมือที่มักใช้ระบบการให้คะแนนแบบมาตราประมาณค่า (Rating Scale) สูตรในการคำนวณมีความคล้ายคลึงกับสูตร KR20 โดยเปลี่ยนจาก $\sum pq$ ให้เป็น $\sum S_i^2$ ซึ่งแสดงถึงผลรวมความแปรปรวนของคะแนนของแต่ละบุคคล ซึ่งค่าสัมประสิทธิ์แอลฟาของครอนบาคจะให้ค่าสัมประสิทธิ์ความเชื่อมั่นที่มากกว่าค่า KR20 เนื่องจากความแปรปรวนของคะแนนมีแตกต่างมากกว่าการให้คะแนนแบบ 0 และ 1 (Kaplan & Saccuzzo, 2013) โดยมีสูตรในการคำนวณดังนี้

$$\alpha = \frac{k}{k-1} \left[1 - \frac{\sum S_i^2}{S_x^2} \right]$$

เมื่อ	α	คือ ค่าสัมประสิทธิ์แอลฟาของครอนบาค
	k	คือ จำนวนข้อของเครื่องมือวัด
	S_i^2	คือ ความแปรปรวนของคะแนนในแต่ละข้อคำถาม
	S_x^2	คือ ความแปรปรวนของคะแนนรวมของเครื่องมือวัดทั้งฉบับ

วิธีการหาค่าสัมประสิทธิ์แอลฟาของครอนบาค เป็นวิธีหนึ่งที่ยอมรับกันอย่างแพร่หลายสำหรับการประมาณค่าความเชื่อมั่นแบบสอดคล้องภายใน โดยเฉพาะอย่างยิ่งการวัดทางด้านพฤติกรรมศาสตร์ เพราะสะดวกและง่ายต่อการนำไปใช้ เนื่องจากทำการทดสอบกับกลุ่มตัวอย่างเพียงครั้งเดียว รวมทั้งสามารถใช้ได้อย่างกว้างขวางกับเครื่องมือวัดที่ให้คะแนนเป็น 0 และ 1 ได้ หรือกำหนดการให้คะแนนเป็นมาตราประมาณค่าหลายระดับได้ (ศิริชัย กาญจนวาสี, 2548)

ตัวอย่าง การคำนวณค่าความเชื่อมั่นแบบสัมประสิทธิ์แอลฟาของครอนบาค

แบบวัดบุคลิกภาพเชิงรุก จำนวน 14 ข้อ ทดสอบกับกลุ่มตัวอย่าง 30 คน มีวิธีการวิเคราะห์ตามโปรแกรมทางสถิติ ดังนี้

- เตรียมข้อมูล แล้วใช้คำสั่ง: Analyze > Scale > Reliability Analysis นำตัวแปรที่วัดทั้งหมดใส่เข้าไปในช่อง Items ในส่วน Model ให้เลือก Alpha แล้วจึงกด Statistics สำหรับในช่อง Descriptives for ให้เลือกทั้งหมด Item, Scale และ Scale if item deleted แล้วจึงกด Continue และ OK

Reliability Statistics

Cronbach's Alpha	N of Items
.926	14

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
b1	55.57	59.564	.349	.929
b2	55.70	58.907	.462	.926
b3	55.47	59.499	.480	.925
b4	55.63	55.964	.754	.918
b5	55.87	56.051	.664	.920
b6	55.73	59.375	.469	.926
b7	55.53	56.189	.876	.916
b8	55.90	52.852	.835	.914
b9	55.77	55.909	.739	.918
b10	55.87	50.740	.878	.912
b11	56.10	50.645	.698	.922
b12	55.90	55.817	.582	.923
b13	55.73	53.030	.801	.915
b14	55.67	52.851	.765	.917

จากผลการวิเคราะห์ความเชื่อมั่นแบบสัมประสิทธิ์แอลฟาของครอนบาค จำนวน 14 ข้อ พบว่า ค่าความเชื่อมั่นแบบสัมประสิทธิ์สหสัมพันธ์แอลฟาของครอนบาคของเครื่องมือวัด มีค่าเท่ากับ .926 แสดงว่า มีความเชื่อมั่นแบบสอดคล้องภายในสูง แบบวัดบุคลิกภาพเชิงรุกมีความน่าเชื่อถือได้สูง

โดยปกติแล้ว ค่าความเชื่อมั่นแบบสัมประสิทธิ์แอลฟาของครอนบาคควรมีขั้นต่ำที่ .70 สำหรับใช้ในการวิจัย และหากใช้ประยุกต์ในคลินิก ควรมีค่าความเชื่อมั่นขั้นต่ำที่ .90 (Bland & Altman, 1997; De Vellis, 2003; Nunnally & Bernstein, 1994; Streiner et al., 2015) หากคำนวณค่าความเชื่อมั่นดังกล่าวด้วยโปรแกรมทางสถิติ แล้วพบว่า มีค่าความเชื่อมั่นต่ำกว่า 0.70 แสดงว่ามีข้อคำถามที่ไม่สอดคล้องกันภายใน หรือมีความสอดคล้องกันน้อย ให้พิจารณาที่ตาราง Item Statistics ตรง Cronbach's Alpha If Item Deleted นั่นคือ หากตัดข้อคำถามข้อใด ๆ หนึ่ง ค่าความเชื่อมั่นก็จะเพิ่มขึ้น

นอกจากนี้ ระบบการให้คะแนนแบบ 0 และ 1 ที่ให้คะแนนถูกผิด ซึ่งเป็นการหาความเชื่อมั่นแบบสอดคล้องภายในด้วยวิธีการ Kuder-Richardson (KR20, KR21) ก็สามารถคำนวณด้วยโปรแกรมทางสถิติ โดยใช้วิธีการและคำสั่งเดียวกันกับการหาค่าสัมประสิทธิ์แอลฟาของครอนบาคเช่นกัน

4. วิธีการประมาณค่าความเชื่อมั่นระหว่างผู้ประเมิน

การประมาณค่าความเชื่อมั่นระหว่างผู้ประเมิน (Inter-Rater Reliability) หรือเรียกอีกอย่างว่า ความสอดคล้องระหว่างผู้ประเมิน (Inter-rater agreement) หรือสัมประสิทธิ์ความเชื่อมั่นระหว่างผู้ประเมิน (Coefficient of Inter-Rater Reliability) หมายถึง ระดับความเห็นสอดคล้องหรือความคงเส้นคงวาของผู้ประเมิน เครื่องมือวัดมากกว่า 2 คนขึ้นไป นั่นคือ การเปรียบเทียบการประเมินของผู้ประเมินว่าแต่ละคำตอบมีความสอดคล้องหรือเห็นตรงกันมากแค่ไหน โดยการประมาณค่าความเชื่อมั่นระหว่างผู้ประเมินเป็นกระบวนการที่ตรงไปตรงมา เป็นลักษณะของการทดสอบครั้งเดียว และมีผู้ประเมินให้คะแนนแต่ละเครื่องมือวัดอย่างเป็นอิสระต่อกัน ถ้าค่าสัมประสิทธิ์ความเชื่อมั่นสูง แสดงว่า ผู้ประเมินได้รับการอบรมฝึกฝนที่เพียงพอและมีความเป็นระบบ และถ้าหากพบความสอดคล้องกันต่ำ อาจจะต้องย้อนกลับไปดูปัญหาว่าเกิดจากอะไร เช่น อาจเกิดจากความไม่กระจำงัดของเกณฑ์การให้คะแนน

การประมาณค่าความเชื่อมั่นระหว่างผู้ประเมิน แบ่งตามระดับข้อมูลการวัด คือ Kappa สำหรับข้อมูลการวัด Nominal ส่วนสัมประสิทธิ์สหสัมพันธ์ภายในชั้น (Intra-class correlation coefficient: ICC) สำหรับข้อมูลการวัด Ordinal, Interval และ Ratio ซึ่งมีรายละเอียดดังนี้ (Hallgren, 2012; Koo & Li, 2016; Zapf, Castell, Morawietz, & Karch, 2016)

4.1 Cohen's Kappa เป็นการหาความเชื่อมั่นระหว่างผู้ประเมิน โดยใช้ข้อมูลการวัดแบบ Nominal โดยใช้สถิติ Kappa ในการวัดระดับความสอดคล้องระหว่างผู้ประเมินที่ให้คะแนนแบบ Nominal และเป็นการแก้ไขความสอดคล้องที่เกิดขึ้นโดยบังเอิญ โดยสูตรในการประมาณค่า Kappa สามารถคำนวณได้ดังนี้

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

เมื่อ K คือ Cohen's Kappa
 P_a คือ ร้อยละความสอดคล้องที่สังเกตได้

$$= \frac{\text{จำนวนการสังเกตที่เห็นสอดคล้อง}}{\text{จำนวนรวมทั้งหมด}}$$

 P_e คือ ความน่าจะเป็นของความสอดคล้องโดยบังเอิญ

ตัวอย่าง การคำนวณค่าเชื่อมั่นระหว่างผู้ประเมินด้วยสถิติ Cohen's Kappa

โดยมีผู้ประเมิน 2 คน ประเมินว่ามีพฤติกรรมก้าวร้าวในห้องเรียน โดยประเมินว่ามีหรือไม่มี

ดังตาราง

		ผู้ประเมินคนที่ 1		
		ไม่มี	มี	รวม
ผู้ประเมินคนที่ 2	ไม่มี	40	15	55
	มี	10	35	45
	รวม	50	50	100

ค่า P_a คือ ร้อยละของความสอดคล้องที่สังเกตได้ นั่นคือ $(40+35)/100 = .75$

ค่า P_e คือ ความน่าจะเป็นของความสอดคล้องโดยบังเอิญ ซึ่งได้มาจากโอกาสความน่าจะเป็นที่สอดคล้องกันที่แสดงว่ามีพฤติกรรมก้าวร้าว ระหว่างผู้ประเมินที่ 1 ($50/100 = .50$) และจากผู้ประเมินที่ 2 ($45/100 = .45$) จึงมีค่าเท่ากับ $.50 \times .45 = .225$ และโอกาสความน่าจะเป็นที่เกิดความสอดคล้องโดยบังเอิญที่แสดงว่าไม่มีพฤติกรรมก้าวร้าว จะเป็น $(1 - .50) \times (1 - .45) = .275$ ดังนั้น โอกาสความน่าจะเป็นของความสอดคล้องโดยบังเอิญ คือ $.225 + .275 = .50$

$$\text{ค่า } K = (.75 - .50) / (1 - .50) = .50$$

ค่าสถิติ Kappa ควรมีค่าอยู่ระหว่าง -1 ถึง 1 นั่นคือ 1 หมายถึง สอดคล้องกันอย่างสมบูรณ์ และ -1 หมายถึง ไม่สอดคล้องกันอย่างสมบูรณ์ โดย Cohen ได้แนะนำการแปลผลค่า Kappa ดังนี้ (McHugh, 2012)

น้อยกว่าหรือเท่ากับ 0	คือ ไม่สอดคล้องเลย
0.01 – 0.20	คือ สอดคล้องกันเล็กน้อย
0.21 – 0.40	คือ สอดคล้องกันพอใช้
0.41 – 0.60	คือ สอดคล้องกันปานกลาง
0.61 – 0.80	คือ สอดคล้องกันมาก
0.81 – 1.00	คือ สอดคล้องกันอย่างสมบูรณ์

ตัวอย่าง การคำนวณค่าเชื่อมั่นระหว่างผู้ประเมินด้วยสถิติ Cohen's Kappa

แบบสังเกตพฤติกรรมก้าวร้าว ประเมินโดยผู้ประเมิน 2 คน ให้คะแนนแบบ 0 และ 1 ทดสอบกับกลุ่มตัวอย่าง 30 คน มีวิธีการวิเคราะห์ตามโปรแกรมทางสถิติ ดังนี้

- เตรียมข้อมูล แล้วใช้คำสั่ง: Analyze > Descriptive Statistics > Crosstabs นำตัวแปรที่ประเมินโดยผู้ประเมินคนที่ 1 ใส่เข้าไปในช่อง Row ส่วนผู้ประเมินคนที่ 2 ใส่เข้าไปในช่อง Column แล้วจึงกด Statistics ให้เลือกทั้งหมด Kappa แล้วจึงกด Continue และ OK

Rater1 * Rater2 Crosstabulation

Count		Rater2		Total
		.00	1.00	
Rater1	.00	12	1	13
	1.00	3	14	17
Total		15	15	30

Symmetric Measures

		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Measure of Agreement	Kappa	.733	.123	4.053	.000
N of Valid Cases		30			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

จากผลการวิเคราะห์ความเชื่อมั่นระหว่างผู้ประเมินด้วยสถิติ Cohen's Kappa โดยผู้ประเมินจำนวน 2 คน พบว่า ค่าความเชื่อมั่นแบบระหว่างผู้ประเมิน มีค่าเท่ากับ .733 แสดงว่า มีความเชื่อมั่นระหว่างผู้ประเมินสูง แบบสังเกตพฤติกรรมก้าวร้าวที่ประเมินโดยผู้ประเมิน 2 คน มีความสอดคล้องกันมาก

4.2 ค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น (Intraclass Correlation Coefficient: ICC) นอกจากค่า ICC จะใช้ในการประมาณค่าความเชื่อมั่นแบบวัดซ้ำแล้ว ยังประมาณค่าความเชื่อมั่นระหว่างผู้ประเมินด้วย โดย McGraw and Wong (1996) กล่าวถึง ค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น สามารถพิจารณาได้จากรูปแบบ (Model) ประเภท (Type) และนิยาม (Definition)

ตัวอย่างการคำนวณค่าความเชื่อมั่นระหว่างผู้ประเมินด้วยค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น

แบบสังเกตการมีส่วนร่วมในชั้นเรียน ประเมินโดยผู้ประเมิน 3 คน ให้คะแนนเป็นมาตรฐานประมาณค่า 5 ระดับ ทดสอบกับกลุ่มตัวอย่าง 30 คน มีวิธีการวิเคราะห์ตามโปรแกรมทางสถิติ ดังนี้

- เตรียมข้อมูล แล้วใช้คำสั่ง: Analyze > Scale > Reliability Analysis นำตัวแปรที่ประเมินโดยผู้ประเมิน (ในที่นี้ มีผู้ประเมิน 3 คน) ใส่เข้าไปในช่อง Items แล้วกดเลือก Statistics จากนั้น จึงเลือก Intraclass correlation coefficient ในส่วน Model ให้เลือก Two-Way Mixed และ Type ให้เลือก Absolute Agreement แล้วจึงกด Continue และ OK

Intraclass Correlation Coefficient

	Intraclass Correlation ^b	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.679 ^a	.503	.817	7.502	29	58	.000
Average Measures	.864 ^c	.752	.930	7.502	29	58	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

- The estimator is the same, whether the interaction effect is present or not.
- Type A intraclass correlation coefficients using an absolute agreement definition.
- This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

สำหรับสาเหตุที่เลือก Model: Two-Way Mixed Effect เนื่องจากเรากำหนดผู้ประเมินไว้แล้ว จึงมีลักษณะเป็น Fixed และกลุ่มตัวอย่างที่ประเมิน มีลักษณะเป็นแบบสุ่ม โมเดลนี้เกี่ยวข้องกับผู้ประเมินที่กำหนดไว้แล้ว ถ้าเลือกผู้ประเมินจากการสุ่มตัวอย่างเข้ามาจะใช้ Model: Two-Way Random แทน และสำหรับการเลือก Type: Absolutely Agreement เนื่องจากเราต้องการทราบความสอดคล้องกันของผู้ประเมิน ส่วน Consistency เหมาะกับระดับความสัมพันธ์ของคะแนนมากกว่า

จากผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น พบว่า แบบสังเกตการมีส่วนร่วมในชั้นเรียนที่ประเมินโดยผู้ประเมิน จำนวน 3 คน มีค่า ICC เท่ากับ .864 (ให้พิจารณาที่ Average Measure เนื่องจากเป็นความสอดคล้องของผู้ประเมิน 3 คน จึงนำมาเฉลี่ยกัน) นั่นคือ ค่าสัมประสิทธิ์ความเชื่อมั่นระหว่างผู้ประเมินของแบบสังเกตการมีส่วนร่วมในชั้นเรียน มีค่าเท่ากับ .864 แสดงว่า มีความเชื่อมั่นระหว่างผู้ประเมินสูง แบบสังเกตการมีส่วนร่วมในชั้นเรียนที่ประเมินโดยผู้ประเมิน 3 คน มีความสอดคล้องกันมาก

สำหรับเกณฑ์การแปลผลค่าความเชื่อมั่นระหว่างผู้ประเมินด้วยค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น (ICC) สามารถอธิบายได้ดังนี้ (Koo & Li, 2016; Portney & Watkins, 2000)

- ค่า ICC ต่ำกว่า .50 แสดงว่า มีค่าความเชื่อมั่นต่ำ
- ค่า ICC ระหว่าง .50 - .74 แสดงว่า มีค่าความเชื่อมั่นปานกลาง
- ค่า ICC ระหว่าง .75 - .90 แสดงว่า มีค่าความเชื่อมั่นที่ดี
- ค่า ICC มากกว่า .90 แสดงว่า มีค่าความเชื่อมั่นที่ดีเยี่ยม

ปัจจัยที่ส่งผลต่อความเชื่อมั่น

สำหรับปัจจัยที่ส่งผลต่อความเชื่อมั่นของเครื่องมือวัดนั้น สามารถสรุปได้ดังรายละเอียดต่อไปนี้ (โชติกา ภาชีผล, 2559; ศิริชัย กาญจนวาศี, 2548; อรพินทร์ ชูชม, 2545; Kaplan & Saccuzzo, 2013)

1. ความยาวของเครื่องมือวัด (Test Length)

โดยปกติแล้ว เครื่องมือวัดที่มีจำนวนข้อมาก ยิ่งทำให้ค่าความเชื่อมั่นเพิ่มสูงขึ้น ทั้งนี้ เนื่องจากการเพิ่มจำนวนข้อที่มากทำให้เครื่องมือวัดนั้นสามารถวัดเนื้อหาได้มากกว่าจำนวนข้อที่น้อยหรือสั้น เครื่องมือวัดที่ข้อสั้นไม่สามารถสุ่มตัวอย่างเนื้อหาพฤติกรรมที่จะทำการวัดได้เพียงพอ เครื่องมือวัดที่ยาวขึ้นจะมีความน่าเชื่อถือมากขึ้น หากเครื่องมือวัดนั้นมีคุณภาพที่ดีแบบเดียวกัน

2. ความเป็นเอกพันธ์ของกลุ่มตัวอย่าง (Group Homogeneity)

เครื่องมือวัดออกแบบมาเพื่อวัดความแตกต่างระหว่างบุคคล ถ้ากลุ่มตัวอย่างมีความเป็นเอกพันธ์มากเท่าใด (นั่นคือ แทบไม่แตกต่างกันเลย หรือมีความสามารถบางอย่างใกล้เคียงกัน) เครื่องมือวัดนั้นจะมีความเชื่อมั่นลดลงตามไปด้วย เครื่องมือวัดที่เชื่อถือได้ควรสะท้อนให้เห็นถึงความแตกต่างระหว่างบุคคลได้ ดังนั้น กลุ่มตัวอย่างที่มีคุณลักษณะที่แตกต่างกันมาก (Group Heterogeneity) หรือมีการกระจายของคะแนนมาก ก็จะได้ค่าความเชื่อมั่นที่สูงขึ้น

3. ความยากง่ายของเครื่องมือวัด (Test Difficulty)

ความยากง่ายของเครื่องมือวัดมีผลต่อความเชื่อมั่นได้ เนื่องจากค่าความเชื่อมั่นของเครื่องมือวัดขึ้นอยู่กับความแปรปรวนของคะแนน ถ้าความแปรปรวนมีน้อยจะได้ค่าความเชื่อมั่นที่ต่ำกว่าความแปรปรวนมาก ถ้าแบบทดสอบยากเกินไป หรือง่ายเกินไป เกือบทุกคนทำผิดหมด หรือเกือบทุกคนทำถูกหมด การกระจายของ

คะแนนมีน้อย คะแนนไม่ห่างกันมาก ซึ่งสะท้อนถึงความเป็นเอกพันธ์ในข้อก่อนหน้า ค่าความเชื่อมั่นก็จะน้อยลงตามไปด้วย

4. ความจำกัดของเวลา (Time Limit)

ถ้าเวลาในการทำเครื่องมือวัดถูกจำกัดมาก เช่น การให้เวลาน้อยเกินไป ทำข้อสอบไม่ทัน ก็อาจมีการเดาคำตอบเกิดขึ้นมากกว่าปกติ หรือการให้เวลามากเกินไปก็เช่นกัน เป็นไปได้ว่า กลุ่มตัวอย่างที่มีความสามารถแตกต่างกัน อาจมีโอกาสดตอบได้คะแนนที่ไม่ต่างกันมาก ส่งผลให้เกิดการกระจายของคะแนนน้อยกว่าที่ควรจะเป็น ทำให้ส่งผลต่อความเชื่อมั่นของเครื่องมือวัด สำหรับเครื่องมือวัดที่ออกแบบและสร้างมาอย่างดีแล้ว กำหนดเวลาในการทำอย่างเหมาะสม ก็จะส่งผลต่อความเชื่อมั่นที่สูง

5. วิธีการที่ใช้ประมาณค่าความเชื่อมั่น (Methods of Estimating Reliabilities)

การประมาณค่าความเชื่อมั่นของเครื่องมือวัดนั้นมีหลายวิธีการและให้ผลลัพธ์ที่แตกต่างกัน วิธีการแต่ละวิธีการจึงเหมาะสมกับเครื่องมือวัดที่มีคุณลักษณะและวัตถุประสงค์ในการใช้ที่แตกต่างกันไป เช่น หากเป็นเครื่องมือวัดที่เป็นแบบทดสอบประเภทความเร็ว (speed test) ไม่ควรใช้วิธีการแบบแบ่งครึ่งหรือวิธีการสอดคล้องภายใน เพราะจะได้ค่าความเชื่อมั่นที่สูงกว่าความเป็นจริง เป็นต้น

สรุป

ความเชื่อมั่น ถือเป็นกระบวนการหนึ่งที่มีความสำคัญในการพัฒนาเครื่องมือวิจัย เพราะเป็นการตรวจสอบคุณภาพของเครื่องมือวัด ความเชื่อมั่นเป็นคุณสมบัติที่สะท้อนถึงเครื่องมือวัดที่มีความคงที่หรือความคงเส้นคงวา ไม่ว่าจะวัดกี่ครั้งก็ได้เท่าเดิมหรือใกล้เคียงของเดิมมากที่สุด โดยมีวิธีการประมาณค่าความเชื่อมั่นหลัก ๆ อยู่ 4 วิธี ได้แก่ วิธีการประมาณค่าความเชื่อมั่นแบบวัดซ้ำ วิธีการประมาณค่าความเชื่อมั่นแบบคู่ขนาน วิธีการประมาณค่าความเชื่อมั่นแบบสอดคล้องภายใน และวิธีการประมาณค่าความเชื่อมั่นระหว่างผู้ประเมิน ความเชื่อมั่นช่วยทำให้เครื่องมือวัดมีความสมบูรณ์มากยิ่งขึ้น ซึ่งเหมาะสมกับตัวแปรทางด้านพฤติกรรมศาสตร์ ซึ่งเป็นตัวแปรที่มีความซับซ้อน การตรวจสอบคุณภาพเครื่องมือด้วยความเชื่อมั่นถือเป็นแนวทางในการสร้างและพัฒนาเครื่องมือทางด้านพฤติกรรมศาสตร์ให้มีความเหมาะสมมากยิ่งขึ้น

เอกสารอ้างอิง

- โชติกา ภาชีผล. (2559). *การวัดและประเมินผลการเรียนรู้*. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2548). *ทฤษฎีการทดสอบแบบดั้งเดิม (พิมพ์ครั้งที่ 5)*. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- อรพินทร์ ชูชม. (2545). *เอกสารคำสอนวิชา วป 502 การสร้างและพัฒนาเครื่องมือวัดทางพฤติกรรมศาสตร์*. กรุงเทพฯ: สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- Bland, J. M., & Altman, D. G. (1997). Cronbach's alpha. *BMJ*, 314(7080), 572.
- Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2013). *Psychological testing and Assessment: An introduction to tests and measurement* (8th ed.). New York: McGraw-Hill.
- De Vellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.
- Gravesande, J., Richardson, J., Griffith, L., & Scott, F. (2019). Test-retest reliability, internal consistency, construct validity and factor structure of a falls risk perception questionnaire in older adults with type 2 diabetes mellitus: A prospective cohort study. *Archives of Physiotherapy*, 9, 14.
- Gregory, R. J. (1996). *Psychological testing: History, principles, and applications* (2nd ed.). Boston: Allyn & Bacon.
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23-34.
- Hayes, M., & Pritchard, J. (2013). *Estimation of internal reliability*. Coventry: Office of Qualification and Examinations Regulation.
- Kaplan, R. M., & Saccuzzo, D. P. (2013). *Psychological assessment and theory: Creating and using psychological tests* (8th ed.). Boston, MA: Cengage Learning.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163.
- Marx, R. G., Menezes, A., Horovitz, L., Jones, E. C., & Warren, R. F. (2003). A comparison of two time intervals for test-retest reliability of health status instruments. *Journal of Clinical Epidemiology*, 56, 730-735.

- McCrae, R. R., Kourtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review, 15*(1), 28-50.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods, 1*(1), 30-46.
- McHugh, M. L. (2012). Interrater reliability: The kappa statistics. *Biochemia Medica, 22*(3), 276-282.
- Merkulova, E., Savostyanov, A., Bocharov, A., Proshina, E., & Knyazev, G. (2018). Test-retest reliability of a social interaction task. *Behavioral Sciences, 8*(10), 97.
- Miller, L. A., Lovler, R. L., & McIntire, S. A. (2013). *Foundations of psychological testing: A practical approach* (4th ed.). Los Angeles: Sage.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: Applications to practice* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.
- Qin, S., Nelson, L., McLeod, L., Eremenco, S., & Coons, S. J. (2019). Assessing test-retest reliability of patient-reported outcome measures using intraclass correlation coefficients: Recommendations for selecting and documenting the analytical formula. *Quality of Life Research, 28*, 1029-1033.
- Reynolds, C. R., Livingston, R. B., & Willson, V. (2010). *Measurement and assessment in education* (2nd ed.). Upper Saddle River, NJ: Pearson Education.
- Sax, G., & Newton, J. W. (1997). *Principles of educational and psychological measurement and evaluation* (4th ed.). Belmont, CA: Wadsworth Publishing Company.
- Schuck, P. (2004). Assessing reproducibility for interval data in health-related quality of life questionnaires: Which coefficient should be used? *Quality of Life Research, 13*(3), 571-586.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420-428.
- Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use* (5th ed.). Oxford: Oxford University Press.
- Zapf, A., Castell, S., Morawietz, L., & Karch, A. (2016). Measuring inter-rater reliability for nominal data – which coefficients and confidence intervals are appropriate? *BMC Medical Research Methodology, 16*, 93.

บทที่ 10

การวิเคราะห์องค์ประกอบ

บทที่ 10 เป็นการนำเสนอเกี่ยวกับการวิเคราะห์องค์ประกอบ โดยผู้เขียนต้องการให้ผู้อ่านได้ทราบและทำความเข้าใจเกี่ยวกับการวิเคราะห์องค์ประกอบ แนวคิดเบื้องต้นของการวิเคราะห์องค์ประกอบ รวมทั้งการวิเคราะห์องค์ประกอบประเภทต่าง ๆ ซึ่งการวิเคราะห์องค์ประกอบเป็นวิธีการวิเคราะห์ทางสถิติที่สามารถตรวจสอบ 14 ความเที่ยงตรงเชิงโครงสร้าง ความเที่ยงตรงเชิงเสมือน และความเที่ยงตรงเชิงจำแนกได้ สำหรับเครื่องมือวัดในสาขาพฤติกรรมศาสตร์ ผู้เขียนเห็นว่า การวิเคราะห์องค์ประกอบถือเป็นวิธีการที่สำคัญที่ช่วยให้เครื่องมือวัดพฤติกรรมศาสตร์มีคุณภาพที่ดีและเหมาะสม รวมทั้งสามารถวัดได้ถูกต้องตามโครงสร้างของการวัด โดยบทนี้ 14 จะช่วยให้ผู้วิจัยและนักวิชาการสามารถสร้างเครื่องมือวัดพฤติกรรมศาสตร์ได้อย่างมีคุณภาพ เพื่อช่วยทุกระบวนการพัฒนาเครื่องมือวิจัยทางพฤติกรรมศาสตร์ที่มีความเหมาะสมและสมบูรณ์มากยิ่งขึ้น

แนวคิดเบื้องต้นของการวิเคราะห์องค์ประกอบ

การวิเคราะห์องค์ประกอบ (Factor Analysis) เป็นเทคนิควิธีการวิเคราะห์ทางสถิติวิธีการหนึ่งที่มีประโยชน์ในการตรวจสอบความเที่ยงตรงเชิงโครงสร้าง (Construct Validity) รวมทั้งความเที่ยงตรงเชิงเสมือน (Convergent Validity) และความเที่ยงตรงเชิงจำแนก (Discriminant Validity) เป็นต้น (Cole, 1987; Gregory, 1996; Miller, Lovler, & McIntire, 2013)

การวิเคราะห์องค์ประกอบ มีวัตถุประสงค์เพื่อลดปริมาณข้อมูลหรือลดจำนวนตัวแปรในการวิจัย โดยอาศัยการรวมตัวแปรที่มีความสัมพันธ์กันมาสร้างเป็นองค์ประกอบ (factor) ที่อยู่เบื้องหลังของตัวแปรนั้น ๆ ช่วยให้สามารถเข้าใจโครงสร้างของตัวแปรนั้นเป็นอย่างดี รวมทั้งยังช่วยลดความคลาดเคลื่อนในการวัดได้อีกด้วย (measurement error) (ไชยันต์ สกกุลศรีประเสริฐ, 2556; Miller et al., 2013)

Hair et al. (2006) ได้กล่าวถึง การวิเคราะห์องค์ประกอบ ว่าเป็นเครื่องมือสำหรับการวิเคราะห์โครงสร้างของความสัมพันธ์ระหว่างตัวแปรต่าง ๆ (เช่น คะแนนเครื่องมือวัด ข้อคำถาม รูปแบบการตอบ) โดยกำหนดชุดของตัวแปรที่มีความสัมพันธ์กันสูง หรือนั่นคือ องค์ประกอบ (factor)

ดังนั้น โดยสรุปการวิเคราะห์องค์ประกอบเป็นเทคนิคการวิเคราะห์ทางสถิติที่ช่วยรวมกลุ่มตัวแปรหรือข้อคำถามที่มีความสัมพันธ์กันไว้ในกลุ่มที่ชัดเจน ทำให้เข้าใจลักษณะของข้อมูล แบบแผน โครงสร้าง รวมทั้งความสัมพันธ์ได้ง่ายยิ่งขึ้น โดยการวิเคราะห์องค์ประกอบ ประกอบด้วย 2 วิธีการ ดังนี้

(1) การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis: EFA) เป็นวิธีการตรวจสอบความเที่ยงตรงเชิงโครงสร้าง เพื่อสำรวจว่าเครื่องมือวัดนั้นมีองค์ประกอบใดบ้าง มีลักษณะเป็นเช่นไร มีข้อคำถามใดบ้างที่มีความสัมพันธ์เกี่ยวข้องกัน จัดกลุ่มรวมกันเป็นองค์ประกอบเดียวกัน โดยแต่ละองค์ประกอบคือ โครงสร้างของเครื่องมือวัดนั้น ๆ นอกจากนี้ ยังเป็นการค้นหาองค์ประกอบที่นักวิจัย/นักวิชาการยังมีแนวคิดทฤษฎีที่ไม่เพียงพอที่จะสรุปหรือยืนยันเกี่ยวกับโครงสร้างองค์ประกอบได้ (Finch, 2013; Flora & Flake, 2017; Haig, 2005; Watkins, 2018)

(2) การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) เป็นวิธีการตรวจสอบความเที่ยงตรงเชิงโครงสร้างว่าเป็นไปตามโครงสร้างตามที่แนวคิดทฤษฎีกำหนดไว้หรือไม่ เพื่อเป็นการยืนยันว่าเครื่องมือวัดที่สร้างขึ้นตามแนวคิดทฤษฎี โดยกำหนดสมมติฐานได้ว่า เครื่องมือวัดมีกี่องค์ประกอบ และแต่ละองค์ประกอบมีข้อคำถามใดบ้าง การวิเคราะห์เป็นการตรวจสอบว่าข้อมูลที่เก็บรวบรวมข้อมูลมา มีความสอดคล้องกับโครงสร้างองค์ประกอบที่กำหนดไว้แล้วหรือไม่ อย่างไร หรืออีกนัยหนึ่งคือ เป็นการยืนยันว่าเครื่องมือวัดเป็นไปตามแนวคิดทฤษฎีที่กำหนดไว้แล้วนั่นเอง (Flora & Flake, 2017; Hoyle, 2004)

สำหรับการวิเคราะห์องค์ประกอบ มี 2 รูปแบบที่สำคัญ คือ การวิเคราะห์องค์ประกอบเชิงสำรวจ และการวิเคราะห์องค์ประกอบเชิงยืนยัน ซึ่งเป็นเครื่องมือที่ใช้ในการพัฒนาและตรวจสอบโครงสร้างและการวัดทางทฤษฎี ในบทนี้ ผู้เขียนนำเสนอวิธีการวิเคราะห์องค์ประกอบทั้ง 2 รูปแบบ พร้อมยกตัวอย่างวิธีการวิเคราะห์ซึ่งนำเสนอในลำดับถัดไป

การวิเคราะห์องค์ประกอบเชิงสำรวจ

การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis: EFA) เป็นวิธีการสำรวจและระบุองค์ประกอบ โดยที่ไม่มีสมมติฐานกำหนดไว้ เป็นเครื่องมือที่ใช้ระบุความสัมพันธ์ระหว่างตัวแปรกับองค์ประกอบ ดังนั้น จึงไม่สามารถทราบได้เลยว่าองค์ประกอบที่จะได้มีกี่องค์ประกอบ และองค์ประกอบแต่ละด้านมีชื่อว่าอะไร เนื่องจากข้อมูลความสัมพันธ์ดังกล่าวมาจากการวิเคราะห์ทางสถิติ ไม่ได้มาจากแนวคิดทฤษฎี (ไซยันต์ สุกศรีประเสริฐ, 2556)

สำหรับกระบวนการวิเคราะห์องค์ประกอบเชิงสำรวจ มี 7 ขั้นตอนที่สำคัญ (Hair et al., 2006) ซึ่งมีรายละเอียดดังนี้

ขั้นตอนที่ 1 วัตถุประสงค์ของการวิเคราะห์องค์ประกอบ (Objectives of Factor Analysis)

วัตถุประสงค์ของการวิเคราะห์องค์ประกอบ คือ การหาวิธีการเพื่อสรุปข้อมูลซึ่งอยู่ในรูปของจำนวนตัวแปร ไปสู่การสร้างองค์ประกอบใหม่ให้น้อยลง นั่นคือ การลดทอนข้อมูล (Data reduction) เช่น มีความต้องการลดคุณลักษณะของข้อคำถาม 30 ข้อ โดยรวมกันให้เหลือเป็นองค์ประกอบจำนวนน้อยลง

ขั้นตอนที่ 2 การออกแบบการวิเคราะห์องค์ประกอบ (Designing a Factor Analysis)

การออกแบบการวิเคราะห์องค์ประกอบ ประกอบด้วยการตัดสินใจพื้นฐานที่สำคัญ 3 อย่าง ได้แก่

(1) การคำนวณข้อมูล (เมทริกซ์สหสัมพันธ์) เพื่อให้ตรงกับวัตถุประสงค์ของตัวแปรที่จัดกลุ่ม โดยควรมีการคำนวณค่าสหสัมพันธ์ระหว่างตัวแปรต่าง ๆ ซึ่งตัวแปรส่วนใหญ่จะเป็นตัวแปรที่เป็นข้อมูลการวัด Interval ขึ้นไป สำหรับข้อมูลการวัด Nominal และ Ordinal จะต้องถูกแปลงให้เป็นตัวแปรหุ่น (dummy variables) ก่อนจึงจะวิเคราะห์ได้

(2) การออกแบบการศึกษาจำนวนตัวแปร คุณสมบัติทางการวัดของตัวแปร และรูปแบบของตัวแปร โดยนักวิจัยควรพยายามลดจำนวนตัวแปรให้เหลือน้อยที่สุด โดยวิธีการนี้ช่วยให้เห็นแบบแผนระหว่างกลุ่มตัวแปร

(3) ความจำเป็นของขนาดตัวอย่าง (sample size) สำหรับขนาดกลุ่มตัวอย่างควรมากกว่า 100 คน หรือสำหรับกฎทั่วไป จำนวนขั้นต่ำของกลุ่มตัวอย่างต่อตัวแปร ควรเป็น 5:1 ส่วนขนาดกลุ่มตัวอย่างที่เป็นที่ยอมรับ อยู่ที่ 10:1 นั่นคือ หากมีจำนวนข้อคำถาม 30 ข้อ จำนวนกลุ่มตัวอย่างควรมีจำนวนอย่างน้อย 150 คน (5:1) และ 300 คน (10:1)

ขั้นตอนที่ 3 ข้อตกลงเบื้องต้นในการวิเคราะห์องค์ประกอบ (Assumptions in Factor Analysis)

นักวิจัยต้องแน่ใจว่ากลุ่มตัวอย่างมีลักษณะที่เหมือนกัน (Homogenous) ยกตัวอย่างการวิเคราะห์องค์ประกอบจะไม่เหมาะสมหากชุดข้อคำถามมีความแตกต่างกันของเพศ เมื่อเอาข้อมูลทั้งเพศชายและเพศหญิงมารวมกัน ผลการวิเคราะห์สหสัมพันธ์และโครงสร้างองค์ประกอบจะเป็นตัวแทนที่ไม่ดีสำหรับโครงสร้างตัวแปรนั้น

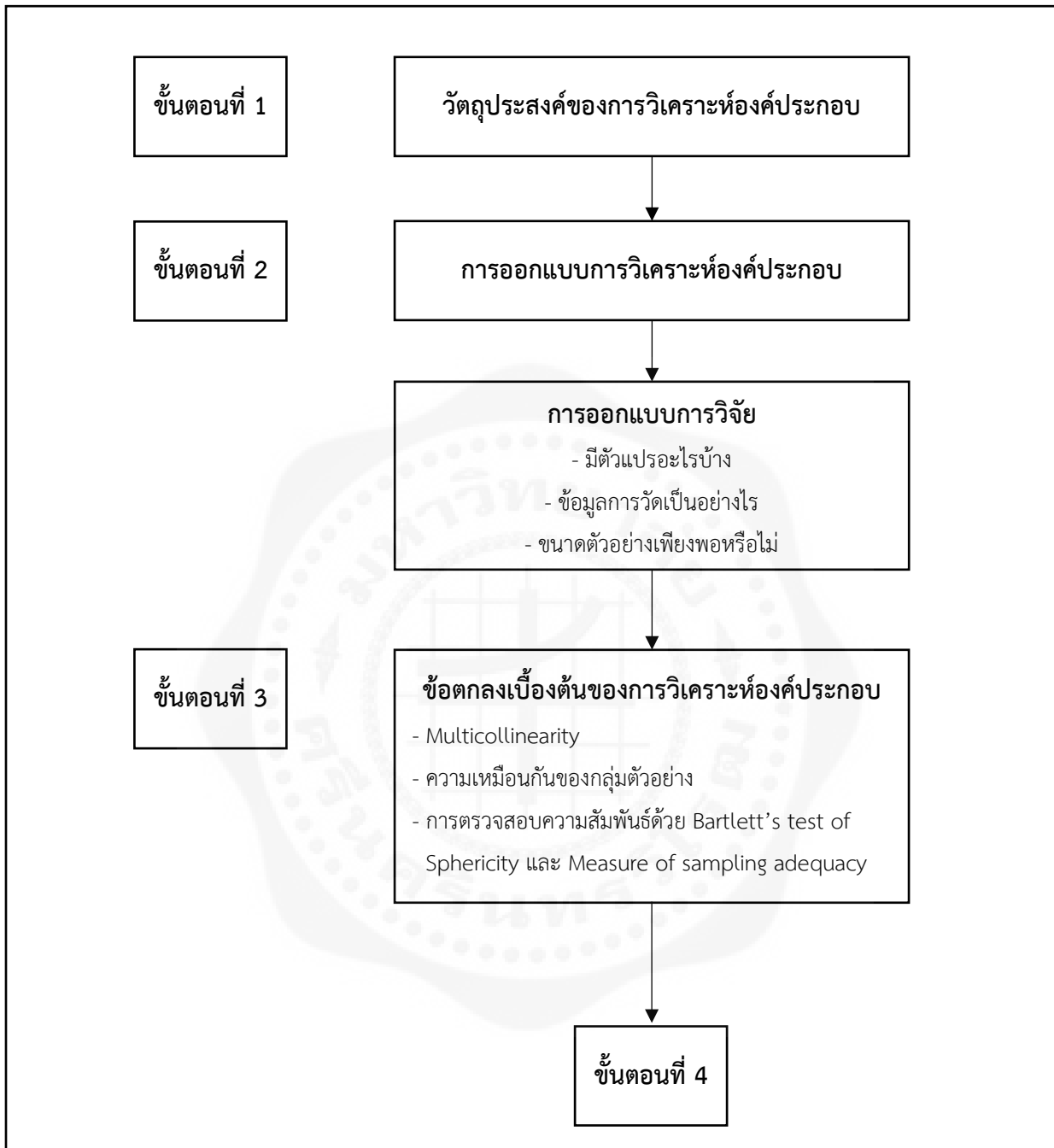
สำหรับการตรวจสอบข้อตกลงเบื้องต้น ควรพิจารณาปัญหา Multicollinearity เพื่อตรวจสอบระดับความสัมพันธ์ระหว่างตัวแปรที่ไม่สูงกันจนเกินไป Kline (2005) กล่าวว่า ควรมีค่าสหสัมพันธ์สูงไม่เกิน .85

การพิจารณาว่าตัวแปรมีความสัมพันธ์กันเพียงพอที่จะสร้างองค์ประกอบขึ้นมาได้นั้น (Williams, Onsmann, & Brown, 2010) ควรพิจารณาจาก

(1) ค่า Bartlett test of sphericity เป็นสถิติทดสอบที่เหมาะสมสำหรับการวิเคราะห์องค์ประกอบ ในการตรวจสอบว่าเมทริกซ์สหสัมพันธ์มีนัยสำคัญหรือไม่ ซึ่งค่า p-value ต้องน้อยกว่า 0.05 จึงจะชี้ได้ว่ามีค่าสหสัมพันธ์ระหว่างตัวแปรที่เพียงพอสำหรับการวิเคราะห์องค์ประกอบ

(2) Measure of sampling adequacy (MSA) เป็นการวัดปริมาณระดับค่าความสัมพันธ์ระหว่างตัวแปร และความเหมาะสมในการวิเคราะห์องค์ประกอบ โดยมีค่าอยู่ระหว่าง 0 ถึง 1 โดยควรมีค่ามากกว่า 0.50 ถึงจะเป็นค่าที่เหมาะสมสำหรับการวิเคราะห์องค์ประกอบต่อไป หากค่าต่ำกว่า 0.50 ให้ทำการตัดตัวแปรบางตัวทิ้ง เพื่อให้ค่า MSA เป็นที่ยอมรับได้ เมื่อผ่านเกณฑ์ดังกล่าวแล้ว จึงสามารถทำการประเมินและตัดสินใจที่จะวิเคราะห์องค์ประกอบต่อไป

โดยในขั้นตอนที่ 1-3 แสดงได้ดังภาพ 10-1



ภาพที่ 10-1 ขั้นตอนที่ 1-3 วิธีการวิเคราะห์องค์ประกอบเชิงสำรวจ
ที่มา: Hair et al. (2006)

ขั้นตอนที่ 4 การได้องค์ประกอบและการประเมินความสอดคล้องภาพรวม (Deriving Factors and Assessing Overall Fit)

เมื่อผ่านขั้นตอนที่ 1-3 แล้ว ขั้นตอนต่อไป คือ การเตรียมวิเคราะห์องค์ประกอบเพื่อระบุโครงสร้างของความสัมพันธ์ โดยการตัดสินใจดังกล่าว ขึ้นอยู่กับ

(1) วิธีการสกัดองค์ประกอบ ด้วยการวิเคราะห์องค์ประกอบร่วม (common factor analysis) หรือการวิเคราะห์ส่วนประกอบ (Component analysis) โดยวิธีการสกัดองค์ประกอบไม่ว่าวิธีใดก็ตาม ต้องพิจารณาวัตถุประสงค์ของการวิเคราะห์องค์ประกอบร่วมกับความรู้เกี่ยวกับคุณลักษณะพื้นฐานของความสัมพันธ์ระหว่างตัวแปร นั่นคือ วิธีการวิเคราะห์ส่วนประกอบ จะสนใจที่การลดจำนวนตัวแปรทั้งหมดให้อยู่ในองค์ประกอบที่มีจำนวนน้อยเพื่อให้องค์ประกอบเป็นตัวเก็บค่าข้อมูลเดิมให้มากที่สุด ส่วนการวิเคราะห์องค์ประกอบร่วม จะสนใจที่การหาองค์ประกอบที่เป็นตัวแทนของกลุ่มตัวแปรบางกลุ่มที่มีค่าร่วมกัน (common) กับองค์ประกอบนั้น ๆ นั่นคือ ตัวแปรทั้งหมดไม่ได้เป็นตัวก่อให้เกิดองค์ประกอบใดองค์ประกอบหนึ่ง โดยส่วนใหญ่ นักวิจัยมักใช้วิธีการวิเคราะห์ส่วนประกอบหลัก (Principal Component Analysis)

(2) จำนวนองค์ประกอบที่เลือกมาเป็นตัวแทนโครงสร้างในข้อมูล โดยเทคนิคที่มักใช้กันคือใช้ค่าไอเกน (eigenvalues) นั่นคือ ถ้าค่าไอเกน มีค่ามากกว่า 1 แสดงว่า องค์ประกอบนั้นมีความสำคัญ หากค่าไอเกนต่ำกว่า 1 แสดงว่าองค์ประกอบนั้นไม่มีความสำคัญ และต้องถูกตัดทิ้งออกไป โดยการนำค่าไอเกนเพื่อเป็นจุดตัด (cutoff) มีความน่าเชื่อถือมาก เมื่อจำนวนตัวแปรอยู่ระหว่าง 20-50 ตัวแปร หากมีจำนวนตัวแปรน้อยกว่า 20 ตัวแปร มีแนวโน้มสกัดองค์ประกอบได้น้อย และหากมีจำนวนตัวแปรมากกว่า 50 ตัวแปร มีแนวโน้มสกัดองค์ประกอบได้มาก รวมทั้งวิธีหนึ่งที่ทำให้ได้จำนวนองค์ประกอบ คือ เกณฑ์ร้อยละความแปรปรวน โดยองค์ประกอบจะต้องอธิบายร้อยละของความแปรปรวนได้ร้อยละ 60 ขึ้นไป

ขั้นตอนที่ 5 การแปลความหมายขององค์ประกอบ (Interpreting the Factors)

เมื่อผ่านขั้นตอนที่ 4 แล้ว สิ่งสำคัญในขั้นตอนต่อไป คือ กระบวนการของการแปลความหมายขององค์ประกอบ ซึ่งประกอบด้วย

(1) การหมุนแกนองค์ประกอบ (Factor rotation) เป็นเครื่องมือที่สำคัญที่สุดในขั้นตอนนี้ โดยการหมุนแกนช่วยจัดสรรความแปรปรวนใหม่ให้ง่ายขึ้น ซึ่งในทางทฤษฎี คือ ช่วยให้เกิดแบบแผนองค์ประกอบมากขึ้น

การหมุนแกนองค์ประกอบ มี 2 วิธี คือ วิธีการหมุนแกนแบบมุมฉาก (Orthogonal rotation) เป็นวิธีการที่มีเป้าหมายเพื่อลดทอนข้อมูลให้มีจำนวนตัวแปรน้อยลง ได้แก่ Varimax, Equimax และ Quartimax โดยวิธี Varimax เป็นวิธีที่ใช้กันมาก ได้รับการพิสูจน์แล้วว่าเป็นวิธีการที่ประสบความสำเร็จในการวิเคราะห์ และวิธีการหมุนแกนแบบมุมแหลม (Oblique rotation) เป็นวิธีการที่เหมาะสมกับการได้องค์ประกอบหรือตัวแปรที่มีความหมายจำนวนหนึ่ง ได้แก่ Oblimin และ Promax

(2) ค่านัยสำคัญของน้ำหนักองค์ประกอบ (Factor loading significance) โดยพิจารณาที่ค่าน้ำหนักองค์ประกอบ โดยค่าที่ยอมรับได้มีค่าระหว่าง 0.30-0.40 เป็นระดับขั้นต่ำของการตีความโครงสร้าง ส่วนค่าน้ำหนักองค์ประกอบมากกว่า 0.50 ได้รับการยอมรับว่ามีนัยสำคัญในทางปฏิบัติ และค่าน้ำหนักองค์ประกอบที่มีค่ามากกว่า 0.70 ถึงจะเป็นตัวแทนที่ดีของโครงสร้าง

(3) การแปลความหมายองค์ประกอบ (Factor interpretation) สำหรับการตีความเมทริกซ์น้ำหนักองค์ประกอบต้องใช้เกณฑ์ในการตัดสิน โดยดำเนินการตามวิธีการ 5 ขั้นตอน ดังนี้

(3.1) การตรวจสอบเมทริกซ์น้ำหนักองค์ประกอบ ประกอบด้วย ค่าน้ำหนักองค์ประกอบของแต่ละตัวแปรแต่ละตัวแปรของแต่ละองค์ประกอบ

(3.2) การระบุค่าน้ำหนักองค์ประกอบที่มีนัยสำคัญในแต่ละตัวแปร โดยเริ่มต้นจากตัวแปรแรกขององค์ประกอบแรก เพื่อหาค่าน้ำหนักองค์ประกอบค่าสูงสุดในแต่ละองค์ประกอบ โดยดำเนินการในทุกองค์ประกอบ แต่สิ่งที่น่าสนใจที่นักวิจัยอาจพบเจอคือ ตัวแปรหนึ่งมีค่าน้ำหนักองค์ประกอบที่มีนัยสำคัญในหลายองค์ประกอบ เรียกว่า ค่าน้ำหนักไขว้ (Cross-loading) หรือมีน้ำหนักองค์ประกอบร่วมกันมากกว่า 1 องค์ประกอบ ซึ่งอาจพิจารณาให้ตัดตัวแปรนั้นทิ้ง

(3.3) ประเมินค่าความร่วมมือกันของตัวแปร ถ้าค่าความร่วมมือกันน้อยกว่า 0.50 แสดงว่า แต่ละตัวแปรอธิบายปริมาณความแปรปรวนที่ยังไม่เพียงพอ

(3.4) กำหนดโมเดลองค์ประกอบใหม่ถ้าต้องการ โดยหากพบปัญหาอย่างหนึ่งอย่างใด เช่น ตัวแปรที่มีค่าน้ำหนักองค์ประกอบที่ไม่มีนัยสำคัญ ค่าความร่วมมือกันของตัวแปรมีค่าต่ำเกินไป หรือตัวแปรมีค่าน้ำหนักไขว้ เป็นต้น หากเจอสถานการณ์ดังกล่าว สามารถผสมผสานวิธีการแก้ปัญหาดังกล่าวนี้ได้ดังนี้

- เพิกเฉยตัวแปรที่เป็นปัญหา และตีความผลดังกล่าว แต่นักวิจัยพึงตระหนักด้วยว่า ตัวแปรนั้นเป็นตัวแทนที่ไม่ดี

- ประเมินตัวแปรที่มีความเป็นไปได้ที่จะต้องตัดทิ้ง ขึ้นอยู่กับค่าความร่วมมือกัน (communality) เมื่อตัดทิ้งแล้ว ให้กำหนดโมเดลใหม่ หรือวิเคราะห์องค์ประกอบใหม่

- ใช้วิธีการหมุนแกนอื่น ๆ

- เพิ่ม/ลดจำนวนองค์ประกอบ เพื่อดูว่าหากโครงสร้างองค์ประกอบใหญ่ขึ้นหรือเล็กลง จะเห็นตัวแปรที่เป็นปัญหา

- การปรับวิธีการสกัดองค์ประกอบใหม่ ใช้วิธีส่วนประกอบหลักหรือวิธีวิเคราะห์องค์ประกอบรวม ไม่ว่าจะใช้วิธีใดก็ตาม เป้าหมายสำคัญคือ เพื่อให้ได้โครงสร้างองค์ประกอบที่สนับสนุนทั้งในเชิงประจักษ์และเชิงแนวคิด สามารถใช้วิธีการหลากหลายเพื่อพัฒนาโครงสร้างดังกล่าวให้ดีขึ้นได้

(3.5) การตั้งชื่อองค์ประกอบ เมื่อผ่านขั้นตอนข้างต้นมาแล้ว นักวิจัยจะต้องตั้งชื่อองค์ประกอบเพื่อเป็นตัวแทนกลุ่มตัวแปรนั้น ๆ ต้องเป็นชื่อที่เหมาะสมที่สะท้อนถึงความเป็นองค์ประกอบ

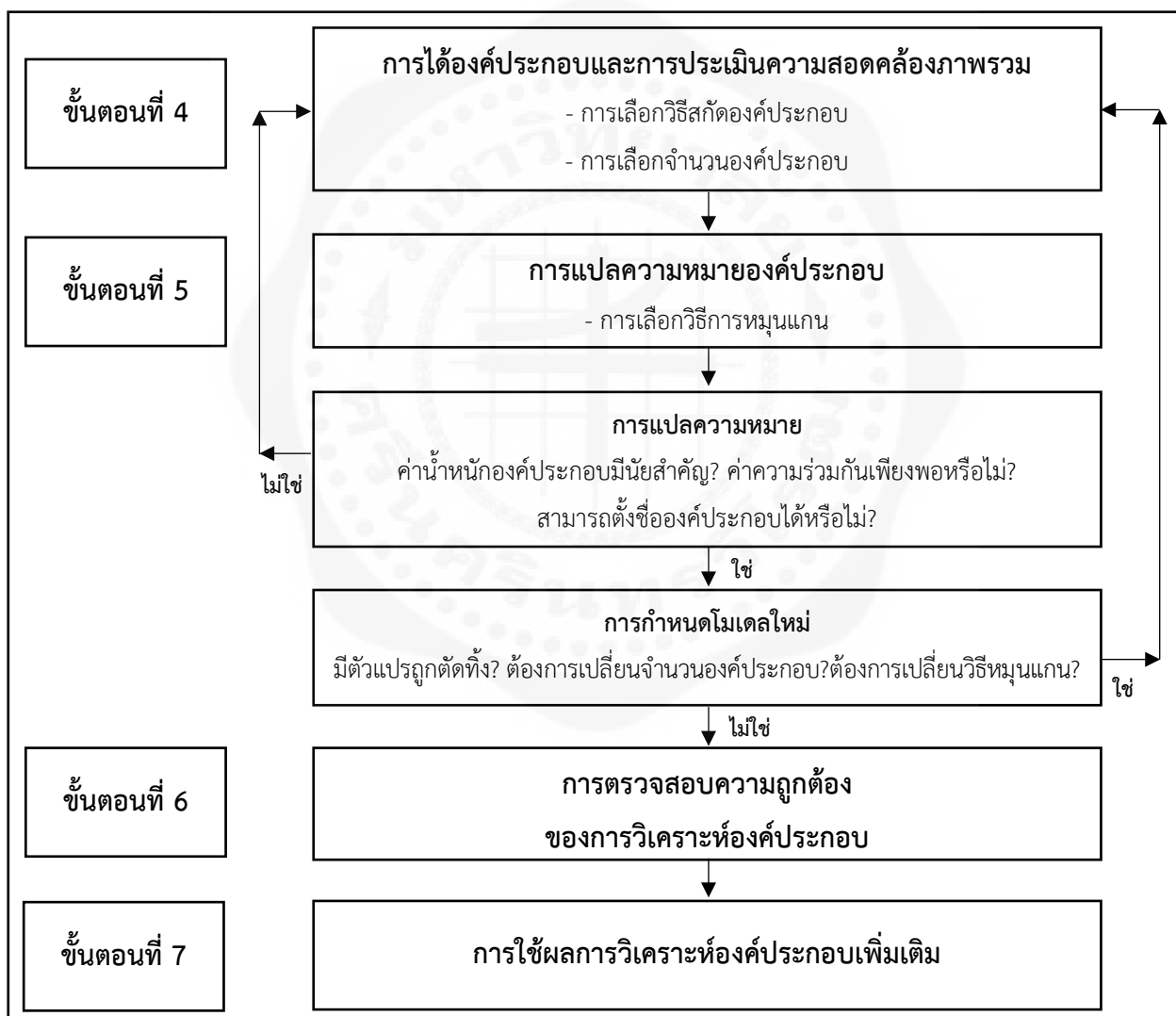
ขั้นตอนที่ 6 การตรวจสอบความถูกต้องของการวิเคราะห์องค์ประกอบ (Validation of Factor Analysis)

ขั้นตอนนี้เกี่ยวข้องกับการประเมินระดับการนำไปใช้อ้างอิงผลไปยังกลุ่มประชากร โดยมีวิธีการตรวจสอบความถูกต้องของผลการวิเคราะห์องค์ประกอบ ดังนี้

(1) การใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) เพื่อยืนยันผลการวิเคราะห์องค์ประกอบดังกล่าว ด้วยการวิเคราะห์โมเดลสมการเชิงโครงสร้าง (Structural Equation Modeling: SEM)

(2) การประเมินความคงที่ของโครงสร้างองค์ประกอบ อาจทำการแบ่งกลุ่มตัวอย่าง เพื่อทำการวิเคราะห์องค์ประกอบ แล้วมาทำการเปรียบเทียบโครงสร้างองค์ประกอบของทั้งสองกลุ่มตัวอย่าง

โดยในขั้นตอนที่ 4-7 แสดงได้ดังภาพที่ 10-2



ภาพที่ 10-2 ขั้นตอนที่ 4-7 วิธีการวิเคราะห์องค์ประกอบเชิงสำรวจ

ที่มา: Hair et al. (2006)

ขั้นตอนที่ 7 การใช้ผลการวิเคราะห์องค์ประกอบเพิ่มเติม (Additional Uses of Factor Analysis Results)

เช่น อาจทำการคัดเลือกตัวแปรที่มีค่าน้ำหนักองค์ประกอบสูงสุดในแต่ละองค์ประกอบ มาใช้เป็นตัวแทนขององค์ประกอบนั้น หรือการสร้างคะแนนรวม (Summated Scale) โดยนำตัวแปรที่มีค่าน้ำหนักองค์ประกอบสูงในองค์ประกอบนั้น มาทำการรวมคะแนนกัน เป็นการนำตัวชี้วัดหลายตัว มาทำการวัดเหลือเพียงการวัดเดียว (single measure)

ตัวอย่าง การวิเคราะห์องค์ประกอบเชิงสำรวจ

ผู้วิจัยได้สร้างข้อคำถาม 14 ข้อ ทำการวัดพฤติกรรมป้องกันความก้าวร้าว ซึ่งต้องการค้นหาองค์ประกอบของพฤติกรรมป้องกันความก้าวร้าวในกลุ่มวัยรุ่นนอคาชีวิตศึกษา มีวิธีการวิเคราะห์ตามโปรแกรมทางสถิติ ดังนี้

- เตรียมข้อมูล แล้วใช้คำสั่ง: Analyze > Dimension Reduction > Factor นำตัวแปรทั้งหมด 14 ข้อ ใส่เข้าไปในช่อง Variables แล้วเลือก Descriptive กดเลือก Coefficients และ KMO and Bartlett's test of sphericity จึงกด Continue จากนั้น จึงเลือก Extraction แล้วเลือก Method: Principal components จึงกด Continue ต่อมา เลือก Rotation กดเลือก Method: Varimax จึงกด Continue และ OK

Correlation Matrix

	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	b11	b12	b13	b14
Correlation b1	1.000	.746	.716	.298	.127	.235	.315	.189	.485	.314	.070	-.021	.094	.092
b2	.746	1.000	.623	.440	.370	.099	.335	.402	.459	.377	.097	.056	.169	.231
b3	.716	.623	1.000	.657	.293	.343	.605	.359	.487	.438	.147	.093	.282	.136
b4	.298	.440	.657	1.000	.748	.393	.808	.678	.509	.724	.383	.393	.667	.535
b5	.127	.370	.293	.748	1.000	.276	.615	.683	.383	.557	.408	.483	.599	.649
b6	.235	.099	.343	.393	.276	1.000	.571	.560	.386	.475	.300	.136	.411	.417
b7	.315	.335	.605	.808	.615	.571	1.000	.770	.614	.814	.651	.505	.787	.684
b8	.189	.402	.359	.678	.683	.560	.770	1.000	.631	.732	.603	.560	.715	.709
b9	.485	.459	.487	.509	.383	.386	.614	.631	1.000	.792	.594	.350	.606	.546
b10	.314	.377	.438	.724	.557	.475	.814	.732	.792	1.000	.715	.455	.778	.723
b11	.070	.097	.147	.383	.408	.300	.651	.603	.594	.715	1.000	.789	.692	.690
b12	-.021	.056	.093	.393	.483	.136	.505	.560	.350	.455	.789	1.000	.627	.667
b13	.094	.169	.282	.667	.599	.411	.787	.715	.606	.778	.692	.627	1.000	.752
b14	.092	.231	.136	.535	.649	.417	.684	.709	.546	.723	.690	.667	.752	1.000

เมื่อพิจารณาแล้วพบว่า ไม่พบปัญหา Multicollinearity จึงพิจารณาค่าอื่น ๆ ต่อไป

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.769
Bartlett's Test of Sphericity	Approx. Chi-Square	387.045
	df	91
	Sig.	.000

โดยค่า KMO มากกว่า 0.50 และค่า Bartlett's test of sphericity มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่า ข้อคำถามทั้ง 14 ข้อ มีความสัมพันธ์กัน สามารถวิเคราะห์องค์ประกอบเชิงสำรวจได้ โดยได้ทำการสกัดองค์ประกอบด้วยวิธี Principal Component Analysis และหมุนแกนแบบมุมฉาก วิธี Varimax

Communalities

	Initial	Extraction
b1	1.000	.788
b2	1.000	.705
b3	1.000	.796
b4	1.000	.703
b5	1.000	.542
b6	1.000	.306
b7	1.000	.834
b8	1.000	.766
b9	1.000	.628
b10	1.000	.811
b11	1.000	.717
b12	1.000	.640
b13	1.000	.803
b14	1.000	.772

Extraction Method: Principal
Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.508	53.627	53.627	7.508	53.627	53.627	6.436	45.973	45.973
2	2.305	16.462	70.089	2.305	16.462	70.089	3.376	24.116	70.089
3	.962	6.870	76.959						
4	.921	6.578	83.537						
5	.581	4.148	87.685						
6	.510	3.641	91.325						
7	.292	2.088	93.413						
8	.230	1.645	95.058						
9	.212	1.511	96.569						
10	.163	1.166	97.735						
11	.127	.908	98.643						
12	.102	.728	99.372						
13	.052	.370	99.742						
14	.036	.258	100.000						

Extraction Method: Principal Component Analysis.

Rotated Component Matrix^a

	Component	
	1	2
b1	-.016	.888
b2	.113	.832
b3	.188	.872
b4	.660	.518
b5	.695	.242
b6	.465	.298
b7	.809	.424
b8	.822	.300
b9	.607	.509
b10	.814	.385
b11	.846	-.038
b12	.785	-.154
b13	.888	.121
b14	.877	.052

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

จากผลการวิเคราะห์องค์ประกอบเชิงสำรวจ พบว่า มีค่าไอเกน เกิน 1 อยู่ 2 องค์ประกอบ และสามารถอธิบายความแปรปรวนของตัวแปรได้ร้อยละ 70.089 ส่วนข้อคำถามข้อที่ 6 มีค่าความร่วมกัน (Communality) ต่ำกว่า 0.50 รวมทั้งข้อคำถามข้อที่ 4 และ 9 มีลักษณะของค่าน้ำหนักไขว้ คือ มีค่าน้ำหนักองค์ประกอบสูงมากกว่า 1 องค์ประกอบ จึงตัดทิ้งข้อคำถาม 4, 6 และ 9 ออกไป แล้วทำการคำนวณใหม่อีกครั้ง

Rotated Component Matrix^a

	Component	
	1	2
b1	.001	.904
b2	.140	.861
b3	.192	.863
b5	.701	.242
b7	.807	.408
b8	.820	.286
b10	.809	.357
b11	.856	-.032
b12	.803	-.129
b13	.888	.106
b14	.884	.056

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

จากการคำนวณใหม่อีกครั้ง พบว่า มีค่าไอเกน เกิน 1 อยู่ 2 องค์ประกอบ และสามารถอธิบายความแปรปรวนของตัวแปรได้ร้อยละ 74.928 และเมื่อพิจารณาแล้ว ไม่พบปัญหา โดยองค์ประกอบที่ 1 ประกอบด้วยข้อคำถาม 1, 2 และ 3 และองค์ประกอบที่ 2 ประกอบด้วยข้อคำถาม 5, 7, 8, 10-14 ต่อมา จึงต้องทำการตั้งชื่อองค์ประกอบที่สะท้อนข้อคำถามดังกล่าว ได้แก่ การหลีกเลี่ยงสถานการณ์ที่ทำให้เกิดความก้าวร้าว และการจัดการควบคุมตนเอง เป็นต้น

สำหรับการตรวจสอบความถูกต้องของการวิเคราะห์องค์ประกอบ (Validation) อาจทำได้ด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน เพื่อยืนยันว่า องค์ประกอบทั้งสองเป็นตัวแทนที่ดีของโครงสร้างพฤติกรรมป้องกันความก้าวร้าวในกลุ่มวัยรุ่นนออาชีวศึกษา

การวิเคราะห์องค์ประกอบเชิงยืนยัน

การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) เป็นวิธีการยืนยันองค์ประกอบโดยที่มีโมเดลการวัดอยู่แล้ว ทราบว่ามีความจำนวนองค์ประกอบอยู่เท่าไร แต่ละองค์ประกอบชื่ออะไร และประกอบด้วยตัวแปรอะไรบ้าง เป็นเครื่องมือที่ใช้พิสูจน์ความถูกต้องของโมเดลการวัด นั่นคือ มีแนวคิดทฤษฎีอยู่ก่อนแล้ว ซึ่งเป็นการตรวจสอบยืนยันองค์ประกอบตามแนวคิดทฤษฎีดังกล่าว (ไซยันต์ สกฤษรีประเสริฐ, 2556) โดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน ใช้เทคนิคการวิเคราะห์โมเดลสมการเชิงโครงสร้าง (SEM) ด้วยโปรแกรมต่าง ๆ เช่น LISREL, AMOS, Mplus เป็นต้น

สำหรับกระบวนการวิเคราะห์องค์ประกอบเชิงยืนยันด้วยเทคนิคการวิเคราะห์โมเดลสมการเชิงโครงสร้างมี 4 ขั้นตอนที่สำคัญ (Hair et al., 2006) ซึ่งมีรายละเอียดดังนี้

ขั้นตอนที่ 1 การกำหนดนิยามตัวแปร (Defining Individual Constructs)

สำหรับการสร้างตัวแปร จำเป็นต้องกำหนดนิยามเชิงปฏิบัติการให้ชัดเจน และทำการพัฒนาข้อคำถามจากนิยามดังกล่าว โดยมักใช้การวิจัยเชิงคุณภาพในการพัฒนาข้อคำถามด้วย จากนั้นกำหนดมาตรวัดที่ใช้ นำไปสู่การประเมินความเที่ยงตรงเชิงเนื้อหา และนำไปทดลองใช้ เพื่อวิเคราะห์ข้อคำถามเป็นรายข้อ เป็นการกล่าเครื่องมือวัดให้สมบูรณ์มากยิ่งขึ้น นอกจากนี้ ถ้าหากไม่มีแนวคิดทฤษฎีที่ชัดเจน ให้ใช้การวิเคราะห์องค์ประกอบเชิงสำรวจเพื่อตรวจสอบจำนวนองค์ประกอบและแบบแผนของน้ำหนักองค์ประกอบ โดยก่อนการวิเคราะห์องค์ประกอบเชิงยืนยัน นักวิจัยควรพิจารณาถึงจำนวนข้อคำถามที่เพียงพอในแต่ละองค์ประกอบ โดยตัวแปรหนึ่ง ๆ ควรมีข้อคำถามขั้นต่ำ 3 ข้อ

ขั้นตอนที่ 2 การพัฒนาโมเดลการวัดในภาพรวม (Developing the Overall Measurement Model)

ในขั้นตอนนี้ นักวิจัยต้องพิจารณาอย่างระมัดระวังเกี่ยวกับตัวแปรแต่ละตัวรวมกันเป็นร่างรูปแบบการวัดโดยภาพรวม ซึ่งมีสิ่งที่จะต้องพิจารณา ดังนี้

- ตัวแปรสังเกตได้แต่ละตัวต้องถูกอธิบายด้วยตัวแปรแฝงเพียงตัวเดียว (unidimensional measures) หรือค่า cross-loading มีค่าเท่ากับ 0 หากเกิด cross-loading ขึ้น แสดงว่า มีตัวแปรสังเกตได้ 1 ตัวที่มีความสัมพันธ์กับตัวแปรแฝงมากกว่า 1 ตัว แสดงว่า เครื่องมือวัดขาดความเที่ยงตรงเชิงโครงสร้าง

- ตัวแปรแฝงควรวัดด้วยตัวแปรสังเกตได้อย่างน้อย 3 ตัวแปร จึงจะสามารถระบุความเป็นไปได้ของพารามิเตอร์เพียงค่าเดียว (Identification) ซึ่งคำนวณได้จาก $[(p)(p + 1)]/2$ โดย p เป็นจำนวนตัวแปรที่สังเกตได้ของโมเดล

(1) หากพารามิเตอร์ที่ประมาณได้มีค่ามากกว่าจำนวนสมาชิกในเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของตัวแปรสังเกตได้ โมเดลจะไม่สามารถประมาณค่าพารามิเตอร์ได้ เรียกว่า โมเดลระบุไม่พอดี (underidentified model)

(2) หากพารามิเตอร์ที่ประมาณได้มีค่าเท่ากับจำนวนสมาชิกในเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของตัวแปรสังเกตได้ เรียกว่า โมเดลที่ระบุพอดี (just-identified model) ซึ่งลักษณะนี้จะมีตัวแปรสังเกตได้ 3 ตัว ต่อตัวแปรแฝง 1 ตัว

(3) หากพารามิเตอร์ที่ประมาณได้มีค่าน้อยกว่าจำนวนสมาชิกในเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของตัวแปรสังเกตได้ เรียกว่า โมเดลระบุเกินพอดี (overidentified model)

ดังนั้น ตัวแปรสังเกตได้ขั้นต่ำควรมีน้อยกว่า 3 ตัวขึ้นไป

ขั้นตอนที่ 3 การออกแบบการศึกษาเพื่อสร้างผลการวิจัยเชิงประจักษ์ (Designing a Study to Produce Empirical Results)

ขั้นตอนนี้ เริ่มต้นวิเคราะห์องค์ประกอบเชิงยืนยัน ต้องมีตัวแปรสังเกตได้ 3 ตัว จึงจะเป็นเงื่อนไขที่เพียงพอสำหรับการระบุค่าพารามิเตอร์ที่เป็นไปได้ รวมทั้งการประมาณค่าพารามิเตอร์ที่เหมาะสม เช่น Maximum Likelihood (ML) และ Generalized Least Square (GLS) เหมาะกับข้อมูลที่มีการแจกแจงปกติ และมีกลุ่มตัวอย่างขนาดใหญ่ ส่วน Weight Least Square (WLS) เหมาะกับข้อมูลที่ไม่ได้แจกแจงปกติ และมีกลุ่มตัวอย่างขนาดใหญ่ (Schumacker & Lomaz, 2010)

ขั้นตอนที่ 4 การประเมินความเที่ยงตรงโมเดลการวัด (Assessing Measurement Model Validity)

เมื่อโมเดลการวัดมีการระบุอย่างถูกต้องแล้ว เทคนิคโมเดลสมการเชิงโครงสร้างจะช่วยประมาณค่าเครื่องมือวัดเกี่ยวกับความสัมพันธ์ระหว่างตัวแปรต่าง ๆ โดยพิจารณา ดังนี้

- การประเมินความสอดคล้องกลมกลืน โดยดัชนีวัดความสอดคล้องกลมกลืนที่มักใช้กัน ได้แก่ GFI, CFI, AGFI, SRMR, RMSEA รวมทั้งค่าสถิติไคสแควร์ หรือค่าสถิติไคสแควร์สัมพัทธ์ สำหรับการเลือกใช้ค่าดัชนีดังกล่าว ควรเลือกใช้ประมาณ 3-4 ค่าดัชนี ก็ถือว่าเป็นหลักฐานความสอดคล้องกลมกลืนของโมเดลที่เพียงพอ (Hair et al., 2006) โมเดลที่รายงานค่าไคสแควร์และระดับองศาอิสระ (degree of freedom: df) ค่า CFI และค่า RMSEA แสดงถึง ข้อมูลที่เพียงพอสำหรับการประเมินโมเดล และสำหรับ Schumacker and Lomax (2010) กล่าวว่า ค่า GFI, AGFI และ CFI (ต้องมีค่ามากกว่า 0.90) ค่า SRMR (ต้องมีค่าต่ำกว่า 0.05) และค่า RMSEA (ต้องมีค่าต่ำกว่า 0.08) แสดงถึง โมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์ โดยสำหรับกฎในการใช้ค่าดัชนีความสอดคล้องกลมกลืน ให้ใช้ค่าดัชนีหลากหลายรูปแบบ และต้องมีค่าดัชนีที่ผ่านเกณฑ์ที่พิจารณา

- เมื่อค่าดัชนีต่าง ๆ ยังไม่ผ่านเกณฑ์ สามารถทำการปรับโมเดลได้หลายรูปแบบ ได้แก่

- (1) ค่าอิทธิพลเส้นทาง (Path Estimates) ค่าที่ยอมรับได้อยู่ที่ 0.50 หากมีค่าต่ำกว่า ให้พิจารณาตัดตัวแปรสังเกตทิ้ง เนื่องจากไม่เป็นตัวแทนโครงสร้างในการวัด หรือหากเจอค่าที่มากกว่า +1 หรือต่ำกว่า -1 เป็นสัญญาณที่บอกว่าข้อมูลมีปัญหา อาจต้องทำการตรวจสอบถึงข้อคำถามที่อาจจะไม่ได้ recode จากลบเป็นบวกก็ได้ เป็นต้น

(2) ค่าส่วนที่เหลือมาตรฐาน (Standardized Residuals) สามารถเป็นได้ทั้งค่าบวกและค่าลบ โดยอาจพิจารณาตัดทิ้งถ้าหากมีค่ามากกว่า +4.0 หรือต่ำกว่า -4.0

(3) ค่าดัชนีปรับแก้ (Modification Indices) เป็นค่าที่คำนวณความสัมพันธ์ที่อาจเกิดขึ้นได้ ซึ่งเมื่อทำการจับคู่หรือ free ความสัมพันธ์ จะช่วยให้ค่าไคสแควร์มีค่าลดลงได้ โดยหากมีค่าตั้งแต่ 4 ขึ้นไป ควรทำการ free เส้นความสัมพันธ์ของความคลาดเคลื่อนดังกล่าว ซึ่งเป็นเงื่อนไขของ CFA ที่ยอมให้เส้นความคลาดเคลื่อนมีความสัมพันธ์กันได้ ในขณะที่ EFA ไม่ยอมให้เกิดความสัมพันธ์ดังกล่าว

ตัวอย่าง การวิเคราะห์องค์ประกอบเชิงยืนยัน

ผู้วิจัยได้สร้างข้อคำถาม 14 ข้อ ทำการวัดพฤติกรรมป้องกันตนเองจากฝุ่นละออง ซึ่งต้องการค้นหาองค์ประกอบของพฤติกรรมป้องกันตนเองจากฝุ่นละออง มีวิธีการวิเคราะห์ตามโปรแกรมทางสถิติ LISREL ดังนี้

```
Observed variables
b1-b14

Latent variables
BH1 BH2

Covariance Matrix from file: bh2
Sample size = 390

Relationships
b1-b5 = BH1
b6-b14 = BH2

LISREL outputs: ME=ML MI SC RS
Path diagram
End of problem
```

Syntax

เมื่อทำการวิเคราะห์แล้ว สิ่งแรกที่ควรพิจารณา คือ ค่าดัชนีความสอดคล้องต่าง ๆ

```
Goodness of Fit Statistics

Degrees of Freedom = 76
Minimum Fit Function Chi-Square = 392.59 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 451.61 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 375.61
90 Percent Confidence Interval for NCP = (312.41 ; 446.31)

Minimum Fit Function Value = 1.01
Population Discrepancy Function Value (F0) = 0.97
90 Percent Confidence Interval for F0 = (0.80 ; 1.15)
Root Mean Square Error of Approximation (RMSEA) = 0.11
90 Percent Confidence Interval for RMSEA = (0.10 ; 0.12)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.00
```

Normed Fit Index (NFI) = 0.97
 Non-Normed Fit Index (NNFI) = 0.97
 Parsimony Normed Fit Index (PNFI) = 0.81
Comparative Fit Index (CFI) = 0.98
 Incremental Fit Index (IFI) = 0.98
 Relative Fit Index (RFI) = 0.97

Critical N (CN) = 107.60

Root Mean Square Residual (RMR) = 0.032

Standardized RMR = 0.041

Goodness of Fit Index (GFI) = 0.86

Adjusted Goodness of Fit Index (AGFI) = 0.80

Parsimony Goodness of Fit Index (PGFI) = 0.62

จากค่าดัชนีความสอดคล้องดังกล่าว (พิจารณาตัวทึบ) พบว่า ค่าไคสแควร์ยังไม่ผ่านเกณฑ์ นั่นคือ มีนัยสำคัญทางสถิติที่ระดับ .05 ค่า RMSEA = 0.11 (ต้องมีค่าน้อยกว่า 0.08) และค่า GFI = 0.86 (ต้องมีค่ามากกว่า 0.90) ส่วนค่า CFI = 0.98 (มีค่ามากกว่า 0.90) และค่า SRMR = 0.041 (มีค่าน้อยกว่า 0.05) โดยเมื่อโมเดลยังไม่สอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ดังนั้น จึงต้องพิจารณาค่าน้ำหนักอิทธิพล ค่าส่วนที่เหลือมาตรฐาน หรือค่าดัชนีปรับแก้

Largest Negative Standardized Residuals			
Residual for	b4 and	b2	-3.17
Residual for	b5 and	b3	-2.81
Residual for	b9 and	b1	-2.69
Residual for	b10 and	b2	-3.04
Residual for	b11 and	b6	-3.20
Residual for	b11 and	b7	-4.26
Residual for	b12 and	b1	-2.69
Residual for	b12 and	b6	-3.20
Residual for	b12 and	b7	-5.40
Residual for	b12 and	b9	-3.44
Residual for	b13 and	b8	-4.61
Residual for	b13 and	b9	-4.05
Residual for	b13 and	b10	-2.97
Residual for	b14 and	b8	-4.06
Largest Positive Standardized Residuals			
Residual for	b2 and	b1	5.10
Residual for	b4 and	b3	4.15
Residual for	b7 and	b6	5.24
Residual for	b8 and	b4	2.64
Residual for	b9 and	b8	3.90
Residual for	b10 and	b9	3.62
Residual for	b12 and	b5	2.70
Residual for	b12 and	b11	7.49
Residual for	b13 and	b5	3.05
Residual for	b13 and	b11	5.49
Residual for	b13 and	b12	7.66
Residual for	b14 and	b12	3.23
Residual for	b14 and	b13	4.37

Modification Indices and Expected Change

Modification Indices for LAMBDA-X

	BH1	BH2
	-----	-----
b1	- -	5.95
b2	- -	5.66
b3	- -	0.00
b4	- -	3.17
b5	- -	11.94
b6	3.16	- -
b7	1.39	- -
b8	5.55	- -
b9	0.03	- -
b10	0.96	- -
b11	2.28	- -
b12	1.99	- -
b13	0.50	- -
b14	1.21	- -

Modification Indices for THETA-DELTA

	b1	b2	b3	b4	b5	b6
	-----	-----	-----	-----	-----	-----
b1	- -					
b2	25.97	- -				
b3	6.32	0.17	- -			
b4	3.80	10.04	17.19	- -		
b5	2.62	0.00	7.92	0.61	- -	
b6	3.65	0.66	0.61	0.02	0.14	- -
b7	0.12	0.29	1.55	1.48	3.36	27.51
b8	0.31	5.18	0.19	0.56	2.08	0.15
b9	5.68	0.07	0.07	2.02	0.30	4.38
b10	0.63	6.91	0.04	6.10	0.20	1.89
b11	0.01	0.75	0.70	10.70	7.56	10.27
b12	1.77	0.08	0.01	2.53	7.22	10.26
b13	0.16	0.65	0.94	0.17	7.12	3.58
b14	1.35	0.76	0.28	0.20	0.00	0.58

Modification Indices for THETA-DELTA

	b7	b8	b9	b10	b11	b12
	-----	-----	-----	-----	-----	-----
b7	- -					
b8	3.13	- -				
b9	0.07	15.24	- -			
b10	1.67	2.94	13.08	- -		
b11	18.12	0.43	2.34	6.60	- -	
b12	29.15	3.87	11.81	0.36	56.15	- -
b13	0.48	21.28	16.41	8.81	30.16	58.68
b14	0.01	16.46	1.54	0.06	0.29	10.45

Modification Indices for THETA-DELTA

	b13	b14
	-----	-----
b13	- -	
b14	19.06	- -

โดยเมื่อพิจารณาค่า Path Estimate ไม่พบปัญหา พบว่า มีค่าเกินกว่า 0.70 ดังนั้น จึงพิจารณาค่า residual ของตัวแปรสังเกตที่ 12 และ 13 มีค่าสูงสุด เช่นเดียวกับกับค่าดัชนีปรับแก้ ของตัวแปรสังเกตที่ 12 และ 13 ดังนั้น จึงเพิ่มเส้นความสัมพันธ์ของความคลาดเคลื่อนดังกล่าว โดยใส่คำสั่ง “set error between ... and ...” โดยเพิ่มคำสั่งทีละคู่ แล้ววิเคราะห์ใหม่ แล้วทำไปเรื่อย ๆ จนกระทั่งค่าดัชนีความสอดคล้องกลมกลืนจะผ่านเกณฑ์ และเมื่อทำการปรับแก้โมเดลจนเสร็จสิ้น ได้ผลดังนี้

Relationships

b1-b5 = BH1

b6-b14 = BH2

```
set error between b12 and b13
set error between b11 and b12
set error between b11 and b13
set error between b1 and b2
set error between b13 and b14
set error between b12 and b14
set error between b6 and b7
```

Syntax

Goodness of Fit Statistics

Degrees of Freedom = 60

Minimum Fit Function Chi-Square = 79.14 (P = 0.050)

Normal Theory Weighted Least Squares Chi-Square = 76.99 (P = 0.069)

Estimated Non-centrality Parameter (NCP) = 16.99

90 Percent Confidence Interval for NCP = (0.0 ; 43.67)

Minimum Fit Function Value = 0.20

Population Discrepancy Function Value (F0) = 0.044

90 Percent Confidence Interval for F0 = (0.0 ; 0.11)

Root Mean Square Error of Approximation (RMSEA) = 0.027

90 Percent Confidence Interval for RMSEA = (0.0 ; 0.043)

P-Value for Test of Close Fit (RMSEA < 0.05) = 0.99

Normed Fit Index (NFI) = 0.99

Non-Normed Fit Index (NNFI) = 1.00

Parsimony Normed Fit Index (PNFI) = 0.66

Comparative Fit Index (CFI) = 1.00

Incremental Fit Index (IFI) = 1.00

Relative Fit Index (RFI) = 0.99

Critical N (CN) = 435.44

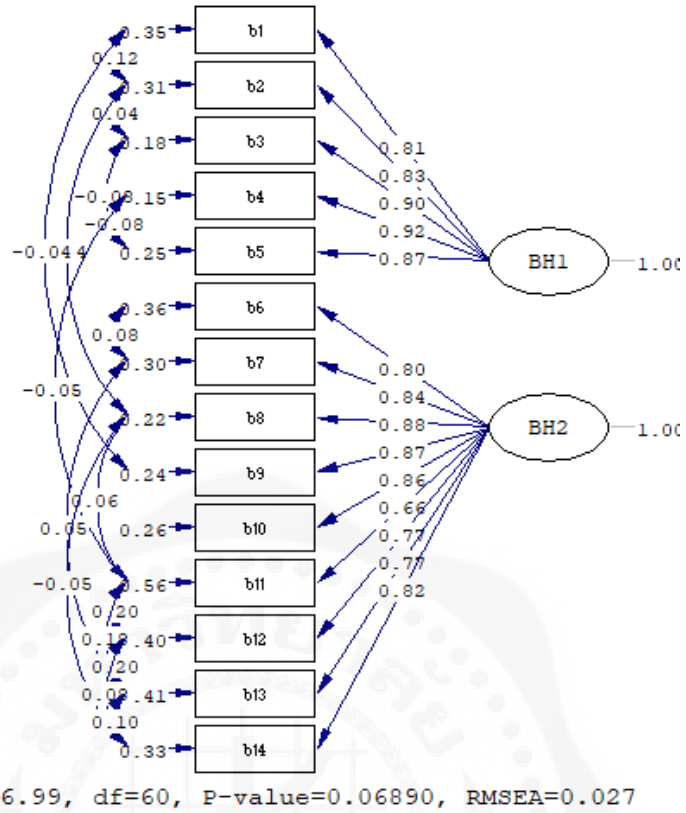
Root Mean Square Residual (RMR) = 0.015

Standardized RMR = 0.019

Goodness of Fit Index (GFI) = 0.97

Adjusted Goodness of Fit Index (AGFI) = 0.95

Parsimony Goodness of Fit Index (PGFI) = 0.56



เมื่อทำการปรับแก้โมเดลแล้ว พบว่า โมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ พิจารณาจากค่าไคสแควร์ซึ่งไม่มีนัยสำคัญทางสถิติ ($\chi^2 = 76.99$, p-value = 0.069) ค่า GFI และ CFI มีค่ามากกว่า 0.90 ส่วนค่า SRMR = 0.019 (น้อยกว่า 0.05) และค่า RMSEA = 0.027 (น้อยกว่า 0.08) จากนั้นจึงพิจารณาค่าน้ำหนักองค์ประกอบ ค่าความคลาดเคลื่อนมาตรฐาน สถิติที รวมทั้งสัมประสิทธิ์ความเชื่อมั่น (R^2)

Completely Standardized Solution		
LAMBDA-X	LAMBDA-X	
	BH1	BH2
b1	0.81	--
b2	0.83	--
b3	0.90	--
b4	0.92	--
b5	0.87	--
b6	--	0.80
b7	--	0.84
b8	--	0.88
b9	--	0.87
b10	--	0.86
b11	--	0.66
b12	--	0.77
b13	--	0.77
b14	--	0.82

LISREL Estimates (Maximum Likelihood)		
LAMBDA-X	LAMBDA-X	
	BH1	BH2
b1	0.70 (0.04) 19.15	--
b2	0.70 (0.04) 19.84	--
b3	0.78 (0.03) 22.62	--
b4	0.76 (0.03) 23.45	--
b5	0.77 (0.04) 20.15	--

ค่า SE	
b1	0.04
b2	0.04
b3	0.03
b4	0.03
b5	0.04

สถิติที	
b1	19.15
b2	19.84
b3	22.62
b4	23.45
b5	20.15

Squared Multiple Correlations for X - Variables

b1	b2	b3	b4	b5	b6
0.65	0.69	0.82	0.85	0.75	0.64

Squared Multiple Correlations for X - Variables

b7	b8	b9	b10	b11	b12
0.70	0.78	0.76	0.74	0.44	0.60

Squared Multiple Correlations for X - Variables

b13	b14
0.59	0.67

ค่า R²

THETA-DELTA

	b1	b2	b3	b4	b5	b6
b1	0.35					
b2	0.12	0.31				
b3	- -	0.04	0.18			
b4	- -	- -	- -	0.15		
b5	- -	- -	-0.08	-0.08	0.25	
b6	- -	- -	- -	- -	- -	0.36

THETA-DELTA

	b7	b8	b9	b10	b11	b12
b7	0.30					
b8	- -	0.22				
b9	- -	- -	0.24			
b10	- -	- -	- -	0.26		
b11	- -	0.06	- -	- -	0.56	
b12	- -	- -	- -	- -	0.20	0.40

THETA-DELTA

	b13	b14
b13	0.41	
b14	0.10	0.33

ค่า Measurement Error

จากผลดังกล่าว นำมาเขียนลงตารางนำเสนอผลการวิเคราะห์ที่ได้ดังนี้

ตัวแปรแฝง	ตัวแปรสังเกต	น้ำหนักองค์ประกอบ			R ²
		Factor Loading	SE	t-value	
BH1					
	b1	0.81	0.04	19.15**	0.65
	b2	0.83	0.04	19.84**	0.69
	b3	0.90	0.03	22.62**	0.82
	b4	0.92	0.03	23.45**	0.85
	b5	0.87	0.04	20.15**	0.75
BH2					
	b6	0.80	0.03	18.87**	0.64
	b7	0.84	0.03	20.10**	0.70
	b8	0.88	0.03	21.87**	0.78
	b9	0.87	0.03	21.63**	0.76
	b10	0.86	0.03	21.09**	0.74
	b11	0.66	0.05	14.38**	0.44
	b12	0.77	0.04	17.91**	0.60
	b13	0.77	0.04	17.74**	0.59
	b14	0.82	0.04	19.42**	0.67

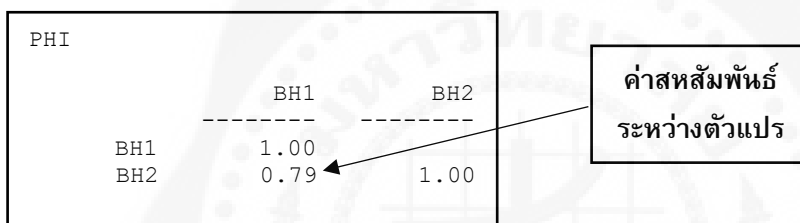
จากตารางแสดงผลการวิเคราะห์องค์ประกอบเชิงยืนยัน พบว่า เครื่องมือวัดประกอบด้วย 2 องค์ประกอบ นั่นคือ ตัวแปรแฝงที่ 1 มีตัวแปรสังเกตทั้ง 5 ตัว มีค่าน้ำหนักองค์ประกอบ อยู่ระหว่าง 0.83-0.96 โดยตัวแปรสังเกตทั้ง 5 ตัว มีนัยสำคัญทางสถิติที่ระดับ .01 และตัวแปรสังเกตทั้ง 5 ตัว มีค่าสัมประสิทธิ์ความเชื่อมั่น (R²) อยู่ระหว่าง 0.65-0.85

และตัวแปรแฝงที่ 2 ตัวแปรสังเกตทั้ง 9 ตัว มีค่าน้ำหนักองค์ประกอบ อยู่ระหว่าง 0.66-0.88 โดยตัวแปรสังเกตทั้ง 9 ตัว มีนัยสำคัญทางสถิติที่ระดับ .01 และตัวแปรสังเกตทั้ง 9 ตัว มีค่าสัมประสิทธิ์ความเชื่อมั่น (R²) อยู่ระหว่าง 0.44-0.85

นอกจากนี้ ยังได้ทำการตรวจสอบความเที่ยงตรงเชิงเสมือน (Convergent Validity) และความเที่ยงตรงเชิงจำแนก (Discriminant Validity)

ความเที่ยงตรงเชิงเสมือน เป็นการตรวจสอบว่าตัวแปรสังเกตได้ทั้งหลายเป็นตัวแปรเดียวกัน จึงคำนวณหาค่าความเชื่อมั่นของตัวแปรแฝง (Construct reliability: CR) ควรมีค่ามากกว่า 0.70 และค่าเฉลี่ยความแปรปรวนที่ถูกสกัดด้วยองค์ประกอบ (Average variance extracted: AVE) ควรมีค่ามากกว่า 0.50 (Fornell & Larcker, 1981; Hair et al., 2006) โดยค่า CR และ AVE ที่ผ่านเกณฑ์ดังกล่าว แสดงถึง หลักฐานที่เพียงพอว่าเครื่องมือวัดมีความเที่ยงตรงเชิงเสมือน

ส่วนความเที่ยงตรงเชิงจำแนก เป็นการตรวจสอบว่าตัวแปรแฝงแต่ละตัวจำแนกออกจากกันชัดเจน โดยพิจารณาจากค่า AVE ต้องมีค่ามากกว่าค่าสหสัมพันธ์ระหว่างตัวแปรแฝงยกกำลังสอง (Fornell & Larcker, 1981; Hair et al., 2006) เป็นหลักฐานที่เพียงพอว่าเครื่องมือวัดมีความเที่ยงตรงเชิงจำแนก ถ้าหากค่า AVE มีค่าน้อยกว่า แสดงว่า ตัวแปรแฝงมีค่าน้ำหนักไขว้ (cross-loading) ทำให้เกิดปัญหาความเที่ยงตรงเชิงจำแนก คือไม่สามารถจำแนกตัวแปรแฝงออกจากกันได้



	BH1	BH2
b1	0.81	
b2	0.83	
b3	0.90	
b4	0.92	
b5	0.87	
b6		0.80
b7		0.84
b8		0.88
b9		0.87
b10		0.86
b11		0.66
b12		0.77
b13		0.77
b14		0.82
CR	0.94	0.97
AVE	0.75	0.72

คำนวณจาก $[\text{ผลรวมของ loading}]^2$ หารด้วย $[(\text{ผลรวมของ loading})^2 + \text{ผลรวมของ Measurement Error}]$

คำนวณจาก $[\text{ผลรวมของ loading}]^2$ หารด้วย $[(\text{ผลรวมของ loading})^2 + \text{ผลรวมของ Measurement Error}]$

จากตาราง พบว่า ความเชื่อมั่นของตัวแปรแฝง (CR) มีค่าอยู่ระหว่าง 0.94-0.97 (ซึ่งมีค่ามากกว่า 0.70) และค่าเฉลี่ยความแปรปรวนที่ถูกสกัดด้วยองค์ประกอบ (AVE) มีค่าอยู่ระหว่าง 0.72-0.75 (ซึ่งมีค่ามากกว่า 0.50) ดังนั้น จึงเป็นหลักฐานที่เพียงพอว่า เครื่องมือวัดนี้มีความเที่ยงตรงเชิงเสมือน

จากการวิเคราะห์พบว่า ค่าสหสัมพันธ์ระหว่างตัวแปรแฝง เท่ากับ 0.79 ดังนั้น ค่าสหสัมพันธ์ระหว่างตัวแปรแฝงยกกำลังสอง จึงเท่ากับ 0.6241 โดยเมื่อพิจารณากับค่า AVE พบว่า ค่า AVE มีค่ามากกว่าค่าสหสัมพันธ์ระหว่างตัวแปรแฝงยกกำลังสอง ($0.72-0.75 > 0.62$) ดังนั้น จึงเป็นหลักฐานเพียงพอว่า เครื่องมือวัดนี้มีความเที่ยงตรงเชิงจำแนก

ดังนั้น จากหลักฐานทั้งหมด จึงสามารถกล่าวได้ว่า เครื่องมือวัดมีความเที่ยงตรงเชิงโครงสร้าง แสดงให้เห็นว่า เครื่องมือวัดนี้มีคุณสมบัติทางการวัดที่ดีและเหมาะสม

สรุป

การวิเคราะห์องค์ประกอบถือเป็นเทคนิควิธีการวิเคราะห์ทางสถิติหนึ่งที่มีความสำคัญในการพัฒนาเครื่องมือวิจัย โดยเฉพาะเครื่องมือวัดทางด้านพฤติกรรมศาสตร์ ซึ่งเป็นเครื่องมือวัดที่ต้องการตรวจสอบคุณภาพในด้านความเที่ยงตรงเชิงโครงสร้าง ความเที่ยงตรงเชิงเสมือน และความเที่ยงตรงเชิงจำแนก โดยการวิเคราะห์องค์ประกอบ ทั้งการวิเคราะห์องค์ประกอบเชิงสำรวจ และการวิเคราะห์องค์ประกอบเชิงยืนยัน โดยทั้งสองวิธีมีวิธีการวิเคราะห์ที่แตกต่างกันไปตามวัตถุประสงค์ของการพัฒนาและตรวจสอบเครื่องมือวัด การวิเคราะห์องค์ประกอบเป็นวิธีการหนึ่งซึ่งช่วยให้เครื่องมือวัดมีความสมบูรณ์มากยิ่งขึ้น ซึ่งเหมาะสมกับตัวแปรทางด้านพฤติกรรมศาสตร์ ซึ่งเป็นตัวแปรที่มีความซับซ้อน การตรวจสอบคุณภาพเครื่องมือด้วยการวิเคราะห์องค์ประกอบ จึงถือเป็นแนวทางในการสร้างและพัฒนาเครื่องมือทางด้านพฤติกรรมศาสตร์ให้มีความเหมาะสมมากยิ่งขึ้น

เอกสารอ้างอิง

- ไชยันต์ สกกุลศรีประเสริฐ. (2556). การวิเคราะห์องค์ประกอบเชิงยืนยัน. *วารสารจิตวิทยาคลินิก*, 44(1), 1-16.
- Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology*, 55, 584-594.
- Finch, W. H. (2013). Exploratory factor analysis. In T. Teo (Ed.), *Handbook of quantitative methods for educational research* (pp. 167-186). Rotterdam, Netherlands: Sense Publishers.

- Flora, D. B., & Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale development and validation. *Canadian Journal of Behavioural Science, 49*(2), 78-88.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50.
- Gregory, R. J. (1996). *Psychological testing: History, principles, and applications* (2nd ed.). Boston: Allyn & Bacon.
- Haig, B. D. (2005). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research, 40*(3), 303-329.
- Hair, Jr., J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Pearson Education.
- Hoyle, R. H. (2004). Confirmatory factor analysis. In M. Lewis-Beck, A. Bryman, & T. Liao (Eds.), *Encyclopedia of social science research methods* (Vol. 1, pp. 169-175). Thousand Oaks, CA: Sage.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: The Guilford Press.
- Miller, L. A., Lovler, R. L., & McIntire, S. A. (2013). *Foundations of psychological testing: A practical approach* (4th ed.). Los Angeles: Sage.
- Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.). New York: Routledge.
- Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology, 44*(3), 219-246.
- Williams, B., Onsmann, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Journal of Emergency Primary Health Care, 8*(3), Article 990399.

บทที่ 11

ทฤษฎีการตอบสนองรายข้อ

บทที่ 11 เป็นการนำเสนอเกี่ยวกับทฤษฎีการตอบสนองรายข้อ เนื่องจากในบทที่ผ่านมา เป็นทฤษฎีการวัดแบบดั้งเดิม ผู้เขียนจึงมีความต้องการที่จะนำเสนอทฤษฎีการตอบสนองรายข้อ ซึ่งได้รับความสนใจเพิ่มมากขึ้นในปัจจุบัน โดยผู้เขียนต้องการให้ผู้อ่านได้ทำความเข้าใจเกี่ยวกับภาพรวมของทฤษฎีการตอบสนองรายข้อ แนวคิดเบื้องต้นของทฤษฎีการตอบสนองรายข้อ โมเดลการตอบสนองรายข้อ สำหรับเครื่องมือวัดพฤติกรรมศาสตร์ ผู้เขียนเห็นว่า โมเดลการตอบสนองรายข้อเป็นวิธีการหนึ่งและเป็นวิธีการวัดแนวใหม่ที่จะช่วยให้เครื่องมือวัดพฤติกรรมศาสตร์มีคุณภาพที่ดีและเหมาะสม เนื่องจากทฤษฎีการวัดแนวใหม่ช่วยอธิบายความสัมพันธ์ระหว่างคุณลักษณะที่มีอยู่ในตัวบุคคลกับข้อคำถามได้เป็นอย่างดี โดยบทนี้ จะช่วยให้ผู้วิจัยและนักวิชาการสามารถสร้างและพัฒนาเครื่องมือวัดพฤติกรรมศาสตร์ได้อย่างมีคุณภาพ ช่วยให้กระบวนการพัฒนาเครื่องมือวิจัยทางพฤติกรรมศาสตร์มีความสมบูรณ์มากยิ่งขึ้น

ภาพรวมของทฤษฎีการตอบสนองรายข้อ

โดยปกติแล้ว ในการวัดทางพฤติกรรมศาสตร์ มักใช้ทฤษฎีการวัดแบบดั้งเดิมหรือที่สามารถเรียกได้อีกชื่อหนึ่งว่า ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) โดยที่ผ่านมา ในระยะหลัง ๆ ทฤษฎีการตอบสนองรายข้อ (Item Response Theory) ได้รับความสนใจเพิ่มมากขึ้น และมักจะมองว่าเป็นทฤษฎีที่มีความทันสมัยกว่าทฤษฎีการทดสอบแบบดั้งเดิม (De Boeck & Wilson, 2004; Embretson & Reise, 2000; Nering & Ostini, 2010) และเป็นวิธีการที่ก้าวข้ามข้อจำกัดบางอย่างที่เกิดจากการทฤษฎีการทดสอบแบบดั้งเดิม (Hambleton, 1994) แต่อย่างไรก็ตาม มีการนำทฤษฎีการตอบสนองรายข้อไปใช้ในการวัดทางด้านจิตวิทยา การศึกษา และพฤติกรรม และเกิดการเปลี่ยนแปลงในด้านบวกเกี่ยวกับการพัฒนาเครื่องมือวัดทางด้านจิตวิทยา (Embretson & Yang, 2006; Hambleton & Jodoin, 2003; Al Nima et al., 2020; Zampetakis, Lerakis, Kafetsios, & Moustakis, 2015; Zanon, Hutz, Yoo, & Hambleton, 2016)

ข้อจำกัดของทฤษฎีการทดสอบดั้งเดิม

ทฤษฎีการตอบสนองรายข้อเกิดขึ้นหรือได้รับการพัฒนาเนื่องมาจากข้อจำกัดของทฤษฎีการทดสอบดั้งเดิม มีรายละเอียด ดังนี้ (Streiner, Norman, & Cairney, 2015)

1. การขึ้นกับกลุ่มตัวอย่าง (Sample dependency) ได้แก่ 1) ถ้าเครื่องมือวัดบริหารจัดการให้กับบุคคลที่มีความแตกต่างกันบางอย่าง เช่น มีการวินิจฉัยแตกต่างกัน หรือเคยเข้าชั้นเรียนพิเศษ จึงมีความจำเป็นต้องตรวจสอบคุณสมบัติทางการวัดใหม่อีกครั้ง และบางทีอาจพัฒนาเกณฑ์ปกติใหม่ ในกระบวนการสร้างเกณฑ์ปกติใหม่ ถ้ามีข้อคำถามเปลี่ยนแปลงไป หรือข้อคำถามนั้นถูกตัดทิ้ง เพื่อพัฒนาเครื่องมือวัดฉบับสั้น หรือสกัดองค์ประกอบย่อยจากเครื่องมือวัด ดังนั้น เมื่อพิจารณาข้อจำกัดถึงคะแนน ค่าความเชื่อมั่นและความเที่ยงตรงจึงขึ้นอยู่กับความเป็นเอกพันธ์ของกลุ่มตัวอย่าง และ 2) มีความเป็นไปได้ที่จะแยกคุณสมบัติของการวัดออกจากคุณลักษณะของบุคคลที่ทำแบบวัด คือ คะแนนจากเครื่องมือวัดขึ้นกับคุณลักษณะที่กลุ่มตัวอย่างมี ดังนั้น คุณลักษณะของเครื่องมือวัดเปลี่ยน เมื่อทำการวัดในกลุ่มที่แตกต่างกัน และคุณลักษณะของกลุ่มเปลี่ยนแปลงไป เมื่อใช้เครื่องมือวัดที่แตกต่างกัน

2. ความเท่าเทียมกันของข้อคำถาม (Assumption of item equivalence) โดยทฤษฎีการทดสอบดั้งเดิม แต่ละข้อคำถามมีส่วนทำให้เกิดคะแนนสุดท้ายที่เท่าเทียมกัน นั่นคือ แม้ว่าจะถ่วงน้ำหนักแต่ละข้อคำถามแตกต่างกัน คะแนนรวมก็มาจากคะแนนของข้อคำถามมารวมกัน โดยไม่ได้คำนึงถึงว่าแต่ละข้อคำถามมีความสัมพันธ์กับตัวแปรได้ดีแค่ไหน อย่างไรก็ตาม บางข้อคำถามมีความสำคัญในการเป็นตัวแทนคุณลักษณะได้ดีกว่าข้ออื่น แต่ก็ไม่มีทางที่จะสร้างเครื่องมือวัดได้อย่างมีประสิทธิภาพ

3. ความคลาดเคลื่อนมาตรฐานของการวัด (Assumptions of the standard error of measurement) โดยทฤษฎีการทดสอบดั้งเดิม มีความคลาดเคลื่อนของการวัดของเครื่องมือวัด ถ้าเครื่องมือเป็นโค้งปกติมากหรือน้อย แสดงว่ามีบุคคลอยู่ตรงกลางจำนวนมาก เนื่องจากกลุ่มตัวอย่างมีขนาดใหญ่ ความคลาดเคลื่อนมาตรฐานของการวัดที่มีขนาดเล็กจะอยู่ตรงกลาง และเพิ่มขึ้นเมื่อคะแนนสูงขึ้นและลดลง อย่างไรก็ตาม เป็นการยากที่จะคำนวณความคลาดเคลื่อนมาตรฐานของการวัดที่แท้จริงได้ และไม่มีสถิติที่อธิบายในจุดนี้ ดังนั้น เมื่อมีการคำนวณความคลาดเคลื่อนมาตรฐานของการวัด จึงคาดการณ์ว่ามีค่าเท่ากันในแต่ละบุคคล เนื่องจากการถดถอยของค่าเฉลี่ย บุคคลจะมีค่าสูงในค่าสุดโต่งของเครื่องมือวัด มากกว่าบุคคลที่อยู่ตรงกลาง

4. ปัญหาการทดสอบที่เท่ากัน (Problems equating tests) โดยทฤษฎีการทดสอบดั้งเดิม เป็นการยากที่จะให้คะแนนแต่ละคนเท่ากันได้ในการทดสอบที่แตกต่างกัน ซึ่งเป็นปัญหาสำหรับการวิจัยระยะยาวด้วย 2 เหตุผลสำคัญ คือ 1) ข้ามช่วงเวลา เครื่องมือวัดมีการปรับปรุง มีเกณฑ์ปกติใหม่ และ 2) ความมีวุฒิภาวะของเด็ก เด็กอาจได้รับการทดสอบที่แตกต่างกัน จึงยากที่จะดูแนวโน้มการเจริญเติบโตข้ามช่วงเวลา คะแนนอาจได้รับอิทธิพลจากการเปลี่ยนแปลงดังกล่าว

ความแตกต่างระหว่างทฤษฎีการทดสอบดั้งเดิมและทฤษฎีการตอบสนองรายข้อ

ความแตกต่างระหว่างทฤษฎีการทดสอบดั้งเดิมและทฤษฎีการตอบสนองรายข้อสามารถอธิบายในภาพรวม (Hambleton, Swaminathan, & Roger, 1991) มี 3 ประการที่สำคัญ ซึ่งมีรายละเอียดดังนี้

1. การวัดเป็นรายข้อกับการวัดแบบภาพรวมทั้งหมด ความแตกต่างที่สำคัญสนใจที่คุณสมบัติของข้อต่าง ๆ ในแบบรายข้อที่ประกอบกันขึ้นมาเป็นมาตรวัด ข้อแตกต่างคือ วิธีการแบบดั้งเดิมนั้นมีแนวโน้มที่จะมุ่งเน้นไปที่คุณลักษณะของมาตรวัดแบบเป็นภาพรวม ตัวอย่างพิจารณาจากค่าความเชื่อมั่น เช่น เมื่อก้าวถึงค่าสัมประสิทธิ์แอลฟาของครอนบาค สังเกตเห็นว่าแอลฟานั้นสามารถเพิ่มขึ้นได้โดยการเพิ่มจำนวนข้อ หรือการเพิ่มค่าความสัมพันธ์ระหว่างข้อโดยเฉลี่ยอย่างใดอย่างหนึ่ง ดังนั้น ไม่ว่าจะเป็นการเพิ่มจำนวนข้อ หรือการสร้างข้อคำถามที่ดีขึ้น จะเป็นการเพิ่มค่าความเชื่อมั่นให้สูงขึ้น โดยวิธีการตอบสนองรายข้อ มองว่าค่าความเชื่อมั่นเป็นพื้นฐานของข้อคำถามแต่ละข้อแบบแยกข้อกันไป ในขณะที่วิธีการแบบดั้งเดิมมีแนวโน้มว่ามาตรวัดนั้นเป็นการวัดแบบภาพรวมเป็นหน่วยเดียว เช่น การวิเคราะห์ความสัมพันธ์ของข้อทั้งหมดสามารถวิเคราะห์และตัดสินได้ว่าข้อคำถามแต่ละข้อสามารถทำให้เกิดความเชื่อมั่นของมาตรวัดในภาพรวมได้อย่างไร

2. การระบุรายข้อโดยใช้ระดับที่มีความเฉพาะสำหรับคุณลักษณะที่วัด ความแตกต่างคือ วิธีการตอบสนองรายข้อ วิเคราะห์ว่าระดับการวัดของคุณลักษณะหนึ่ง ๆ อยู่ระดับใดอย่างชัดเจน ซึ่งมีอิทธิพลอย่างมากต่อข้อคำถามในการวัด ข้อคำถามในการวัดที่แตกต่างกันอาจจะถูกทำให้ “กลายเป็น” ข้อที่อยู่ในระดับคุณลักษณะที่แตกต่างกัน และในท้ายที่สุด อาจส่งผลต่อการเปลี่ยนแปลงในปริมาณที่ต่างกันและมีผลต่อความต่อเนื่องทั้งหมดทุกระดับของคุณลักษณะนั้น ตัวอย่าง ข้อคำถาม “บางครั้งฉันรู้สึกแย่” อาจเป็นไปได้ว่าวัดระดับความเศร้าที่ต่ำกว่าภาวะเศร้าหรือซึมเศร้า ซึ่งอยู่ในระดับที่ต่ำกว่าข้อการวัดที่บอกว่า “ฉันรู้สึกว่าชีวิตไม่มีค่าพอที่จะอยู่ต่อไป” ข้อแรกอาจจะวัดเพื่อจำแนกความแตกต่างของใครบางคนผู้ซึ่งไม่ค่อยมีประสบการณ์ในการประสบกับความรู้สึกเศร้าหรือภาวะซึมเศร้าออกจากผู้ที่ประสบกับความรู้สึกเช่นนั้นบ่อยกว่า ซึ่งอาจไม่ใช่เครื่องมือที่ดีพอที่จะวัดและจำแนกผู้ตอบที่ประสบกับความเศร้าเพียงครั้งคราวออกจากผู้ที่ประสบกับความเศร้าบ่อยครั้ง ในอีกด้านหนึ่งข้อที่สองอาจจะสามารถวัดและจำแนกผู้ตอบเฉพาะคนที่อยู่ในช่วงคำตอบที่สุดโต่งได้เท่านั้น ผู้ที่อยู่ในช่วงรอยต่อที่เป็นช่วงต่อเนื่องที่สำคัญอาจจะตกไปจากข้อนี้ เนื่องด้วยมีผู้ตอบที่มีภาวะเศร้าหรือซึมเศร้าที่รุนแรงกว่ามากตอบข้อนี้อย่างชัดเจนกว่า วิธีการตอบสนองรายข้อจะระบุและจำแนกข้อทั้งสองข้อนี้ว่าเป็นการแสดงให้เห็นถึงจุดที่แตกต่างกันบนความต่อเนื่องหนึ่ง ๆ ของภาวะซึมเศร้าได้

การจำแนกว่าข้อใดเกี่ยวข้องกับช่วงใดในช่วงความต่อเนื่องของตัวแปรที่วัดนั้น วิธีการตอบสนองรายข้อช่วยให้ผู้พัฒนาเครื่องมือวัดแสดงช่วงค่าคะแนนทั้งหมดในช่วงความต่อเนื่องของคุณลักษณะที่ถูกวัดได้ด้วยข้อคำถามต่าง ๆ ที่มีความเหมาะสม นั่นคือ ข้อคำถามสามารถทำให้เกิดมาตรวัดที่แสดงระดับต่ำ ปานกลาง และสูง

ของคุณลักษณะที่ศึกษาได้อย่างรวดเร็วกว่า ซึ่งวิธีการนี้มีความเกี่ยวข้องกับข้อคำถามเป็นรายข้ออย่างชัดเจนซึ่งสัมพันธ์กับช่วงความต่อเนื่องของการวัดคุณลักษณะหนึ่ง ๆ ที่เกี่ยวข้อง ในขณะที่วิธีแบบดั้งเดิมสามารถให้ผลที่คล้ายคลึงกัน (เช่นการตรวจสอบสัดส่วนของผู้ตอบที่เห็นด้วยกับข้อคำถามแต่ละข้อกับคะแนนรวมที่เป็นแบบฉบับของผู้ตอบที่เห็นด้วยและไม่เห็นด้วย) กล่าวโดยสรุปคือ หากพิจารณาแล้วมุมมองการเน้นที่รายข้อในวิธีการตอบสนองรายข้อ สามารถส่งเสริมให้เกิดความเชื่อมั่นของมาตรวัดได้

3. การนำเสนอที่เห็นภาพชัดเจนของคุณลักษณะรายข้อและคุณลักษณะมาตรวัด วิธีการตอบสนองรายข้อทำให้ใช้ภาพกราฟฟิคประกอบเกิดขึ้นอย่างกว้างขวาง ซึ่งเป็นการนำเสนอคุณสมบัติรายข้อรวมถึงมาตรวัดทั้งหมด เนื่องจากวิธีการนี้เกิดมาจากบริบทของการทดสอบความสามารถ คำศัพท์และศัพท์เฉพาะต่าง ๆ จึงมักเกี่ยวข้องกับเนื้อหาสาระในแนวนั้น และวิธีการนี้เป็นชุดของรูปแบบการปฏิบัติมากกว่าที่จะเป็นทฤษฎีที่เฉพาะเจาะจงที่มีกระบวนการเพียงชุดเดียว

แนวคิดเบื้องต้นของทฤษฎีการตอบสนองรายข้อ

ทฤษฎีการตอบสนองรายข้อ เป็นทฤษฎีทางด้านสถิติที่ประกอบด้วย โมเดลทางคณิตศาสตร์ที่หลากหลายซึ่งมีคุณลักษณะเพื่อทำนายคะแนนของบุคคลจากความสามารถหรือคุณลักษณะแฝง (Latent variables) และเพื่อสร้างความสัมพันธ์ระหว่างผลการปฏิบัติการของข้อคำถามของบุคคลและชุดของคุณลักษณะภายใต้ผลการปฏิบัติการของข้อคำถาม ผ่านการทำหน้าที่ที่เรียกว่า โค้งแสดงคุณลักษณะรายข้อ (Item characteristic curve: ICC) (Hambleton et al., 1991) หรืออธิบายอีกนัยหนึ่งว่า

1. คะแนนของบุคคล หรือการตอบข้อคำถามของบุคคล สามารถทำนายหรืออธิบายได้ด้วยความสามารถหรือคุณลักษณะแฝงของบุคคล
2. ความสัมพันธ์ระหว่างคะแนนของบุคคล หรือการตอบข้อคำถามของบุคคล กับความสามารถหรือคุณลักษณะแฝงในตัวบุคคล สามารถอธิบายได้ด้วยฟังก์ชันคณิตศาสตร์ที่เรียกว่า โค้งแสดงคุณลักษณะรายข้อ

การนำทฤษฎีการตอบสนองรายข้อมาใช้ทำการศึกษาทางพฤติกรรมศาสตร์ จำเป็นต้องมีการตรวจสอบข้อตกลงเบื้องต้น 2 ประการ ได้แก่ (ศิริชัย กาญจนวาสี, 2555; Hambleton et al., 1991)

1. ความเป็นมิติเดียว (Unidimensionality) การที่ชุดของข้อคำถามในมาตรวัดหรือเครื่องมือวัด มุ่งวัดคุณลักษณะแฝงเพียงคุณลักษณะเดียว (หรือความสามารถของบุคคลเพียงความสามารถเดียว) ซึ่งในระยะต่อมา มีการพัฒนาทฤษฎีการตอบสนองรายข้อแบบพหุมิติ (Multidimensional Item Response Theory: MIRT) เนื่องจากข้อจำกัดเกี่ยวกับความเป็นมิติเดียวที่ละเลยความสัมพันธ์ระหว่างคุณลักษณะแฝง อาจทำให้ผลการวัดขาดความถูกต้อง (Reckase, 2009) จากการวิจัยของ Immekus, Snyder and Ralston (2019) เมื่อเปรียบเทียบ

ผลการวิเคราะห์ด้วยทฤษฎีการตอบสนองรายข้อแบบพหุมิติ (MIRT) กับการวิเคราะห์องค์ประกอบเชิงยืนยันพบว่า MIRT มีความใกล้เคียงกันในเรื่องความยืดหยุ่นในการทดสอบโมเดลเพื่อประเมินโครงสร้างองค์ประกอบของเครื่องมือวัด

2. ความเป็นอิสระ (Local Independence) หมายถึง เมื่อทำการควบคุมคุณลักษณะแฝงหรือความสามารถของบุคคลให้คงที่แล้ว ผลของการตอบข้อคำถามในข้อคำถามแต่ละคู่ จะต้องมีความเป็นอิสระจากกันทางสถิติ หรือกล่าวอีกนัยหนึ่งคือ คุณลักษณะแฝงภายในบุคคลต้องเป็นปัจจัยเดียวที่ทำให้การตอบข้อคำถามของผู้ทดสอบมีความแตกต่างกัน โดยความแตกต่างกันของการตอบข้อคำถามต้องไม่ได้เกิดขึ้นจากความสัมพันธ์ในรูปแบบอื่น

3. โค้งแสดงคุณลักษณะรายข้อ (ICC) เป็นฟังก์ชันทางคณิตศาสตร์ ใช้อธิบายความสัมพันธ์ระหว่างความน่าจะเป็นที่ผู้ตอบจะตอบข้อคำถามได้ถูกต้อง กับระดับความสามารถที่วัดได้ด้วยชุดของข้อคำถามหรือเครื่องมือวัดนั้น โดยความน่าจะเป็นในการตอบข้อคำถามได้ถูกต้องขึ้นอยู่กับโค้งแสดงคุณลักษณะรายข้อในแต่ละโมเดลที่เลือกใช้ ซึ่งรูปร่าง (Shape) ของโค้งแสดงคุณลักษณะรายข้อในแต่ละข้อมีคุณสมบัติไม่แปรเปลี่ยน (Invariant) ไปตามกลุ่มตัวอย่างที่ใช้ ดังนั้น ทำให้ความน่าจะเป็นในการตอบข้อคำถามได้ถูกต้องไม่แปรเปลี่ยนไปด้วย

4. เครื่องมือวัดที่ใช้ ต้องไม่เป็นเครื่องมือวัดประเภทความเร็ว (Speed test) โดยการจัดการทดสอบจะต้องไม่เป็นไปในลักษณะที่จำกัดทางด้านเวลา หรือสถานการณ์ที่แข่งขันกันด้วยเวลา เนื่องจากความสามารถเป็นปัจจัยสำคัญต่อผลการสอบ ทฤษฎีการตอบสนองรายข้อจึงไม่ควรเป็นแบบทดสอบประเภทความเร็ว (Speed test)

พารามิเตอร์ในทฤษฎีการตอบสนองรายข้อ

พารามิเตอร์ในทฤษฎีการตอบสนองรายข้อ จะต้องไม่มีความแปรเปลี่ยนไปทั้งข้อคำถามและผู้ตอบ พารามิเตอร์จะมีความเท่าเทียมกันในระหว่างกลุ่มหรือเงื่อนไขของการวัด ดังนั้น ถ้ามีความแปรปรวนของพารามิเตอร์เมื่อเก็บรวบรวมในกลุ่มประชากรที่แตกต่างกัน จะทำให้เกิดการอ้างอิงผลการวิจัยที่ผิด (Rupp & Zumbo, 2004)

สำหรับพารามิเตอร์ (Parameter) ในทฤษฎีการตอบสนองรายข้อ สามารถแบ่งได้เป็น 2 ชนิด คือ

- 1) พารามิเตอร์ของผู้ตอบ (Person parameter) ได้แก่ ระดับความสามารถหรือคุณลักษณะแฝงของผู้ตอบ และ
- 2) พารามิเตอร์ของข้อคำถาม (Item parameter) ได้แก่ ค่าความยาก อำนาจจำแนก สัมประสิทธิ์การเดา นอกจากนี้ค่าพารามิเตอร์แล้ว ยังมีค่าคงที่ด้วย (Baker, 2001; Hambleton & Swaminathan, 1985) ซึ่งมีรายละเอียดดังนี้

พารามิเตอร์ของผู้ตอบ (Person parameter)

พารามิเตอร์ของผู้ตอบ ประกอบด้วย

θ คือ ระดับความสามารถหรือคุณลักษณะแฝงของผู้ตอบ โดยประมาณจากโมเดลทฤษฎีการตอบสนองรายข้อ

$P_i(\theta)$ คือ ความน่าจะเป็นที่ผู้ตอบที่มีความสามารถ จะตอบข้อคำถามข้อที่ i ได้ถูกต้อง ส่วนมาก ค่า P จะมีค่าอยู่ระหว่าง -3.00 ถึง $+3.00$ โดยค่าที่เป็นลบ แสดงว่ามีความสามารถหรือคุณลักษณะนั้นต่ำ และค่าที่เป็นบวก แสดงว่ามีความสามารถหรือคุณลักษณะนั้นสูง

พารามิเตอร์ของข้อคำถาม (Item parameter)

พารามิเตอร์ของข้อคำถาม ประกอบด้วย

b_i คือ ค่าพารามิเตอร์ความยากของข้อคำถามข้อที่ i (Difficulty parameter) เป็นค่าที่แสดงตำแหน่งของโค้งแสดงคุณลักษณะรายข้อ ณ จุดความสามารถของผู้ตอบ ที่มีโอกาสตอบข้อคำถามได้ถูกต้อง ในทางทฤษฎี b_i จะมีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัติ ข้อคำถามที่มีค่า b_i อยู่ระหว่าง -3.00 ถึง $+3.00$ โดยค่าที่เป็นลบ แสดงว่าข้อคำถามข้อนั้นง่าย และค่าที่เป็นบวก แสดงว่าข้อคำถามข้อนั้นยาก

เป้าหมายในการวิเคราะห์ค่าความยากของข้อคำถาม นั้นคือ เพื่อสร้างการรับรู้เกี่ยวกับความจริงที่ว่าบุคคลจะต้องใช้คุณสมบัติมากเพียงใดจึงจะผ่านข้อคำถามนั้น เช่น หากบุคคลตอบข้อคำถามเกี่ยวภาวะความซึมเศร้าได้ จะหมายความว่าบุคคลผู้นั้นเกี่ยวข้องกับระดับความซึมเศร้าทันที โดยไม่คำนึงถึงว่าบุคคลนั้นเป็นใคร หรือระดับเฉลี่ยของภาวะความซึมเศร้าจะมากเพียงใดในกลุ่มตัวอย่างที่ทำการศึกษา

a_i คือ ค่าพารามิเตอร์อำนาจจำแนกของข้อคำถามข้อที่ i (Discrimination parameter) เป็นค่าความชันของโค้งแสดงคุณลักษณะรายข้อ ณ ตำแหน่ง b_i ในทางทฤษฎี a_i จะมีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัติ ข้อคำถามมีค่า a_i อยู่ระหว่าง 0.50 ถึง 2.50 เพราะข้อคำถามที่มีค่า a_i ที่เป็นลบ แสดงว่าเป็นข้อคำถามที่ไม่ดีต้องนำออกจากชุดของข้อคำถามนั้น ส่วนค่า $a_i = 0$ หรือเข้าใกล้ 0 มาก แสดงว่าข้อคำถามนั้นไม่มีอำนาจในการจำแนก หรือมีอำนาจจำแนกต่ำ และค่า a_i ที่เป็นบวกและมีค่าสูง แสดงว่าข้อคำถามข้อนั้นสามารถจำแนกได้ดี

ดังนั้น เครื่องมือหรือข้อคำถามที่สามารถจำแนกได้ดีจะมีช่วงคะแนนของเหตุการณ์ที่จะมีผลลัพธ์ออกมาว่ากำกวมน้อย ยังมีอำนาจในการจำแนกน้อยลงเพียงใด เครื่องมือหรือข้อคำถามข้อนั้นก็จะมีค่าความกำกวมมากและไม่ชัดเจนมากขึ้นเท่านั้น

c_i คือ สัมประสิทธิ์การเดาของข้อคำถามข้อที่ i (Guessing parameter) เป็นค่าแสดงความน่าจะเป็นในการเดาข้อคำถามข้อที่ i ซึ่งโอกาสในการตอบข้อคำถามได้ถูกต้องของผู้ตอบที่มีความสามารถต่ำ จะเป็นค่ากำกับต่ำสุด (Lower asymptote) ของโค้งแสดงคุณลักษณะรายข้อ (ICC) ในทางทฤษฎี c_i จะมีค่าอยู่ระหว่าง 0 ถึง 1 โดยทั่วไปนิยมใช้ข้อคำถามที่มีค่า c_i ไม่เกิน 0.30

ผู้ตอบที่ไม่ได้รู้คำตอบที่ถูกต้องอย่างแท้จริงก็สามารถได้คะแนนในข้อคำถามนั้นโดยการเดา ซึ่งการเดาและได้คะแนนเช่นนี้ เรียกว่า ผลบวกปลอม แสดงให้เห็นถึงการบังชี้ที่เป็นผลบวกว่าผู้ตอบมีความรู้ความสามารถในระดับหนึ่งแม้ว่าบุคคลนั้นจะไม่ได้รู้คำตอบที่ถูกต้องของข้อคำถามนั้น ๆ อย่างแท้จริง

ค่าคงที่ (Constant)

ค่าคงที่ ประกอบด้วย

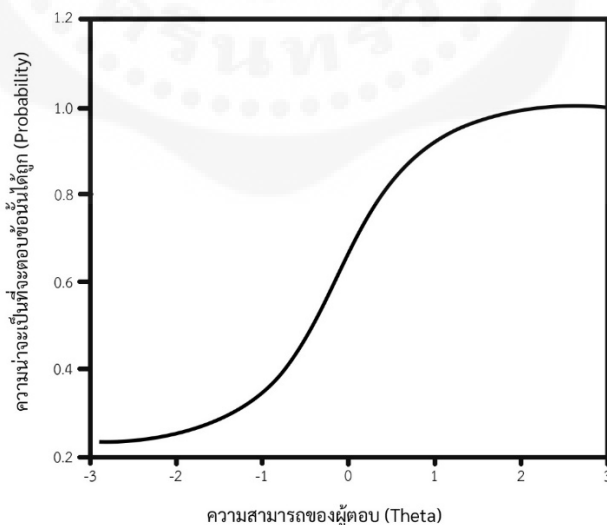
e คือ ค่าคงที่ของ Natural Log มีค่าเท่ากับ 2.71828

D คือ ค่าองค์ประกอบของการปรับสเกล (Scaling factor) กำหนดให้เป็น 1 หรือ 1.70 เพื่อให้ฟังก์ชันโลจิสติก (Logistic function) มีค่าใกล้เคียงกับฟังก์ชันปกติสะสม (Normal ogive function) หรือมีค่าต่างเกินไม่เกิน 0.01

โดยพารามิเตอร์ของข้อคำถามทั้งสาม ได้แก่ ค่าความยากง่าย ค่าอำนาจจำแนก และสัมประสิทธิ์การเดามีความสัมพันธ์กับความคลาดเคลื่อนของการวัด หากค่าความยากง่ายของข้อคำถามไม่เหมาะสม มีช่วงความคลุมเครือมากระหว่างคะแนนผ่านกับคะแนนตก หรือข้อคำถามบ่งชี้คุณลักษณะหนึ่ง ๆ แม้ว่าคุณลักษณะนั้นจะไม่มีก็ตาม จะสามารถสรุปได้ว่าข้อคำถามนั้นมีแนวโน้มที่จะทำให้เกิดความคลาดเคลื่อน โดยทฤษฎีการตอบสนองรายข้อประเมินมุมมองประสิทธิภาพของข้อคำถาม และวิธีการเลือกข้อคำถามที่มีแนวโน้มที่ดีในบริบทหนึ่ง ๆ

โค้งแสดงคุณลักษณะรายข้อ

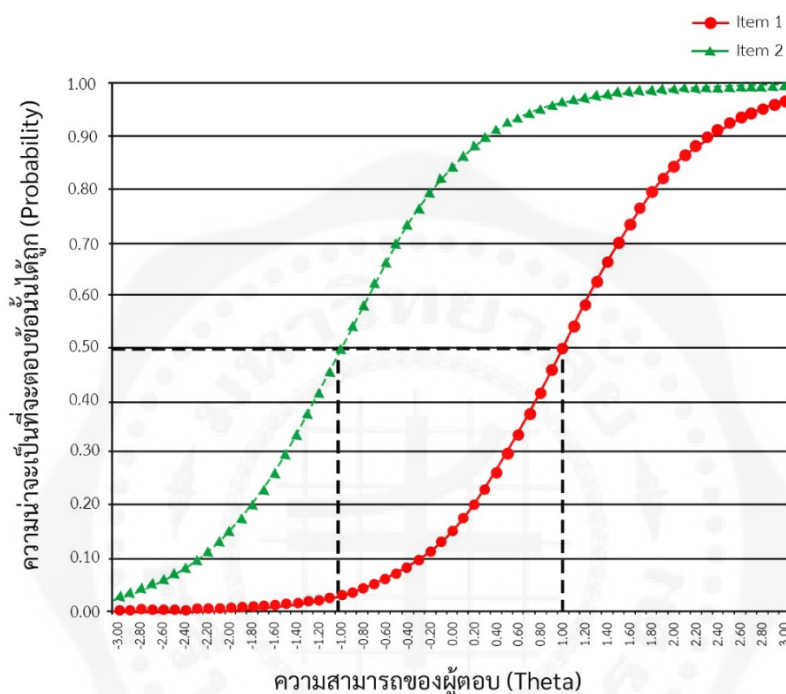
ค่าความยากง่าย ค่าอำนาจจำแนก และความน่าจะเป็นในการเดา สามารถสรุปรูปแบบ ของโค้งแสดงคุณลักษณะรายข้อ (item-characteristic curve: ICC) ซึ่งได้นำเสนอประสิทธิภาพของข้อคำถามในรูปแบบกราฟฟิก โดยปกติแล้ว โค้งแสดงคุณลักษณะรายข้อ จะมีลักษณะเป็นแบบรูปตัว S และส่วนต่าง ๆ ของโค้งจะแสดงให้เห็นข้อมูลเกี่ยวกับพารามิเตอร์ทั้งสาม ดังภาพที่ 11-1



ภาพที่ 11-1 ตัวอย่างโค้งแสดงคุณลักษณะรายข้อ (ICC)

ที่มา: ปรับมาจาก De Beer (2004)

จากภาพที่ 1 แสดงให้เห็นว่าโค้งแสดงคุณลักษณะรายข้อ (ICC) จะมีลักษณะอย่างไร แกน X แสดงจุดเด่นของคุณลักษณะแฝงหรือความสามารถที่ถูกวัด (เช่น ความรู้ จุดแข็ง ความแม่นยำ ภาวะซึมเศร้า หรือเหตุการณ์ที่สามารถวัดได้) ส่วนแกน Y แสดงความน่าจะเป็นที่ผู้ตอบที่มีความสามารถ จะตอบข้อคำถามข้อนั้นได้ ดังนั้น โค้งแสดงคุณลักษณะรายข้อสามารถใช้เพื่อประเมินคุณภาพของข้อคำถามได้อย่างไรนั้น โดยแท้จริงแล้วสามารถทำได้ง่ายกว่าหากมองไปที่แผนภูมิที่แสดงให้เห็นข้อคำถามสองข้อที่สามารถนำมาเปรียบเทียบกันได้ ดังภาพที่ 11-2



ภาพที่ 11-2 ตัวอย่างโค้งแสดงคุณลักษณะรายข้อ (ICC) เปรียบเทียบข้อคำถาม 2 ข้อ

ที่มา: ปรับมาจาก Sideridis, Tsaousis and Al Harbi (2016)

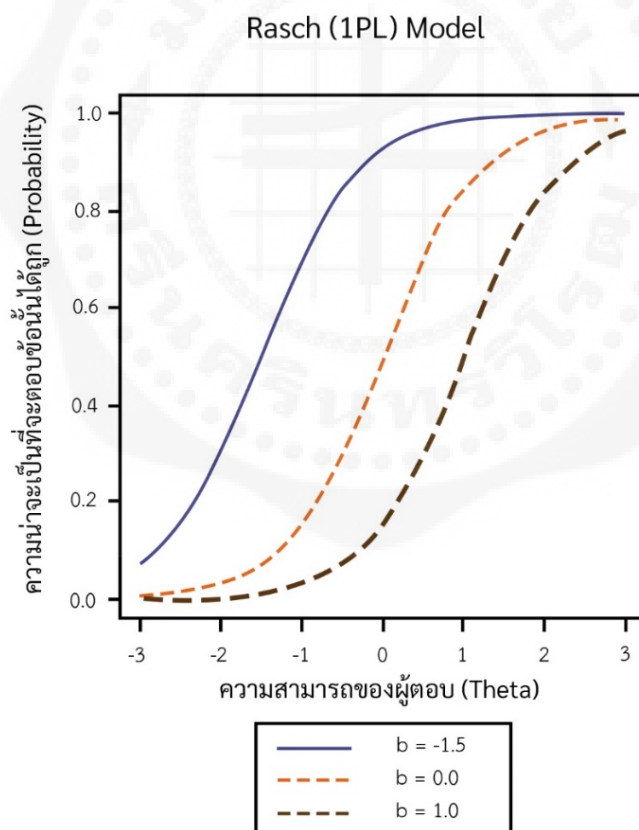
จากภาพที่ 2 แสดงให้เห็นว่า ข้อคำถามข้อที่ 1 และข้อที่ 2 มีความน่าจะเป็นที่จะตอบข้อนั้นได้ ถึงร้อยละ 50 เท่ากัน แต่ข้อที่ 2 แสดงระดับความสามารถที่ถูกวัดน้อยกว่า แสดงว่า ข้อคำถามข้อที่ 2 เป็นข้อที่ง่าย คนที่มีความสามารถน้อย มีโอกาสที่จะตอบข้อนั้นถูกถึงร้อยละ 50

โมเดลการตอบสนองรายข้อ

ทฤษฎีการตอบสนองรายข้อมีคุณลักษณะเป็นโมเดลทางคณิตศาสตร์ แต่ละโมเดลมีรายละเอียดที่แตกต่างกัน การเลือกใช้โมเดลการตอบสนองรายข้อจึงต้องพิจารณาความเหมาะสมของโมเดล แต่ละโมเดลมีความเหมาะสมกับข้อมูลที่แตกต่างกันออกไป โมเดลการตอบสนองรายข้อสามารถแบ่งตามชนิดของข้อมูลได้ 2 ประเภท (ศิริชัย กาญจนวาสี, 2555; Hambleton et al., 1991; Zanon et al., 2016) ซึ่งมีรายละเอียดดังนี้

1. โมเดลการตอบสนองรายข้อแบบตรวจให้คะแนน 2 ค่า (Dichotomous IRT models) โดยโมเดลนี้ ส่วนใหญ่จะเป็นเครื่องมือวัดทางด้านความรู้คิด (Cognitive test) มีการตรวจให้คะแนนแบบ 0 หรือ 1 นั่นคือ “ถูก” หรือ “ผิด” หรืออาจจะเป็นเครื่องมือวัดทางด้านอารมณ์ความรู้สึก (Affective test) มีตัวเลือก “เห็นด้วย” หรือ “ไม่เห็นด้วย” รวมทั้งเครื่องมือวัดทางด้านพฤติกรรม (Behavioral test) มีตัวเลือก “ใช่” หรือ “ไม่ใช่” โดยการตอบข้อคำถามดังกล่าว ล้วนมีการตรวจให้คะแนนแบบสองค่า โดยโมเดลที่นิยมใช้กัน ได้แก่

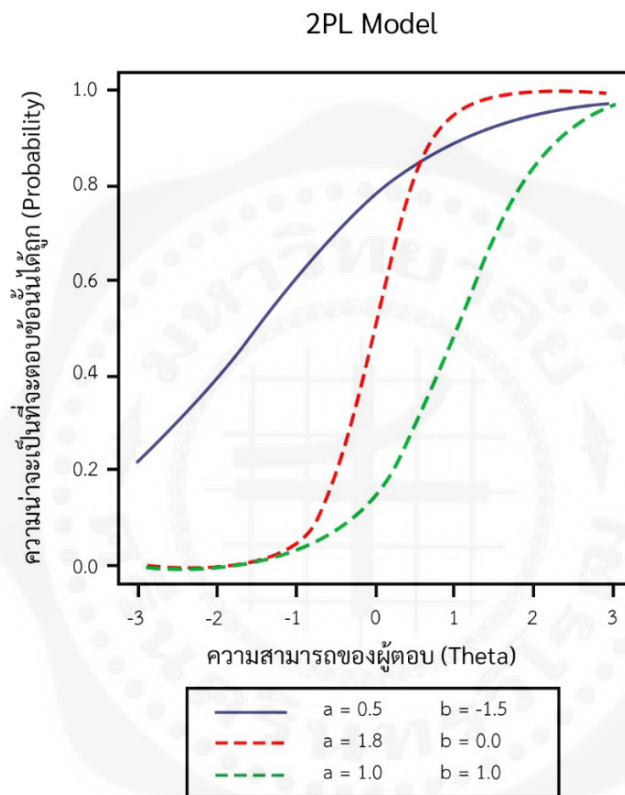
1.1 โมเดลโลจิสติกแบบ 1 พารามิเตอร์ (One-parameter logistic model: 1PLM) เป็นโมเดลที่มีความเรียบง่ายที่สุด เป็นการประมาณค่าความน่าจะเป็นที่บุคคลจะตอบข้อคำถามข้อหนึ่ง ๆ มีข้อตกลงเบื้องต้นว่า ข้อคำถามแต่ละข้อมีพารามิเตอร์ $c = 0$ และมีพารามิเตอร์ a เท่ากัน แต่มีความแตกต่างกันเฉพาะพารามิเตอร์ b เท่านั้น นั่นคือ ข้อคำถามแต่ละข้อไม่มีสัมประสิทธิ์การเดา มีอำนาจจำแนกเท่ากัน แต่มีความยากง่ายต่างกัน ดังนั้น พารามิเตอร์ความยากของข้อคำถามที่เพิ่มมากขึ้น ผู้ตอบที่มีความสามารถมากขึ้น จึงจะตอบข้อคำถามนั้น ได้ถูก ดังภาพที่ 11-3



ภาพที่ 11-3 โมเดลโลจิสติกแบบ 1 พารามิเตอร์

ที่มา: ปรับมาจาก Salkind (2010)

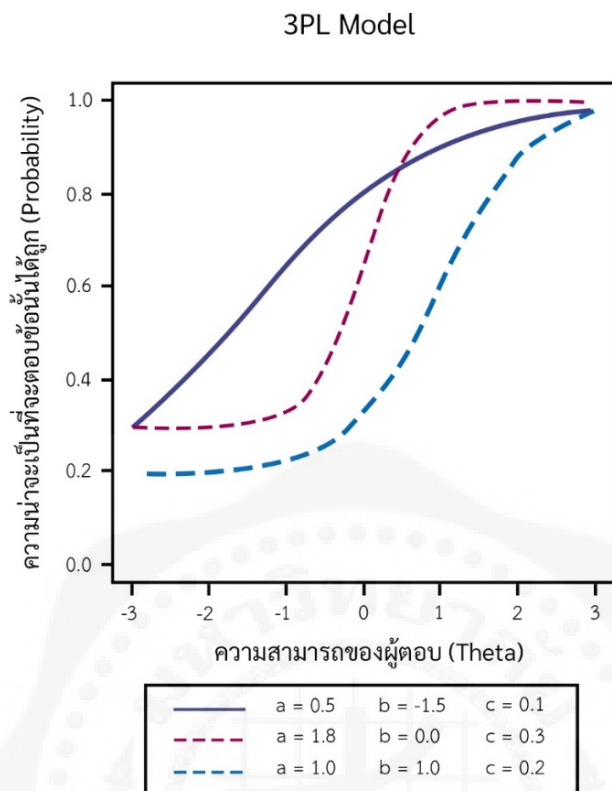
1.2 โมเดลโลจิสติกแบบ 2 พารามิเตอร์ (Two-parameter logistic model: 2PLM) เป็นการประมาณค่าความน่าจะเป็นที่บุคคลจะตอบข้อคำถามข้อหนึ่ง ๆ ได้ถูกต้อง มีข้อตกลงเบื้องต้นที่แตกต่างจาก 1PLM คือ มีความแตกต่างกันทั้งพารามิเตอร์ a และ b นั่นคือ ข้อคำถามแต่ละข้อไม่มีสัมประสิทธิ์การเดา มีอำนาจจำแนกและความยากง่ายต่างกัน ดังนั้น จึงเหมาะกับเครื่องมือวัดที่ไม่สามารถอธิบายความแตกต่างระหว่างบุคคลได้ด้วยระดับความสามารถของบุคคลเพียงอย่างเดียว แต่มีมุมมองอื่น ๆ ในการอธิบายความแตกต่างของบุคคลเพิ่มเติมด้วย ดังภาพที่ 11-4



ภาพที่ 11-4 โมเดลโลจิสติกแบบ 2 พารามิเตอร์

ที่มา: ปรับมาจาก Salkind (2010)

1.3 โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (Three-parameter logistic model: 3PLM) เป็นการประมาณพารามิเตอร์ทั้งค่าความยากง่ายและอำนาจจำแนก และค่าความน่าจะเป็นที่บุคคลจะเดาคำตอบโดยผลของการเดามีผลต่อความยากง่ายของข้อคำถาม และคะแนนโดยรวมของผู้ตอบ มีข้อตกลงเบื้องต้นที่แตกต่างจาก 1PLM และ 2PLM คือ มีความแตกต่างกันทั้งพารามิเตอร์ a , b และ c นั่นคือ ข้อคำถามแต่ละข้อมีสัมประสิทธิ์การเดา มีอำนาจจำแนกและความยากง่ายต่างกัน ดังนั้น จึงเหมาะสำหรับเครื่องมือวัดที่เป็นแบบหลายตัวเลือก เนื่องจากผู้ตอบสามารถเดาคำตอบได้ ดังภาพที่ 11-5



ภาพที่ 11-5 โมเดลโลจิสติกแบบ 3 พารามิเตอร์
ที่มา: ปรับมาจาก Salkind (2010)

2. โมเดลการตอบสนองรายข้อแบบตรวจให้คะแนนมากกว่า 2 ค่า (Polytomous IRT models) โดยโมเดลนี้ส่วนใหญ่จะเป็นเครื่องมือวัดทางด้านพฤติกรรมศาสตร์ มีการตรวจให้คะแนนมากกว่า 2 ค่า เช่น เครื่องมือวัดที่ใช้มาตรวัดลิเคิร์ต (Likert scale) หรือมาตราประมาณค่า (Rating scale) ซึ่งเป็นเครื่องมือที่มีลักษณะเป็นรายการคำตอบแบบเรียงลำดับหลายตัวเลือก (Ordered categorical response) หากเลือก “ไม่จริงเลย” ให้ 1 คะแนน “ไม่จริง” ให้ 2 คะแนน “ปานกลาง” ให้ 3 คะแนน “จริง” ให้ 4 คะแนน และ “จริงที่สุด” ให้ 5 คะแนน เป็นต้น โดยโมเดลที่นิยมใช้กัน ได้แก่

2.1 Graded Response Model (GRM) เป็นโมเดลที่ใช้กับเครื่องมือวัดที่แต่ละข้อคำถามมีรายการคำตอบแบบเรียงลำดับ โดยแต่ละข้ออาจมีจำนวนรายการคำตอบที่เท่ากันหรือไม่เท่ากันก็ได้ มีลักษณะเป็นโมเดลทั่วไปของโมเดลการตอบสนองรายข้อที่มี 2 พารามิเตอร์ นั่นคือ พารามิเตอร์ความยากง่าย และพารามิเตอร์อำนาจจำแนก จะนำมาใช้ในการคำนวณความน่าจะเป็นในการตอบข้อคำถามที่ถูกต้อง เนื่องจากโอกาสที่ผู้ตอบจะตอบข้อคำถามข้อใดข้อหนึ่งถูกต้องหรือไม่นั้น ไม่ได้ขึ้นกับระดับความสามารถหรือผู้ตอบเพียงอย่างเดียว

แต่ยังเกี่ยวข้องกับพารามิเตอร์ของข้อคำถามแต่ละข้อด้วย โดยในการวิเคราะห์โมเดล GRM นั้น ขนาดกลุ่มตัวอย่างที่ใช้ในการประมาณค่าพารามิเตอร์ ควรมีไม่ต่ำกว่า 500 คน (De Ayala, 2009) ซึ่งจะช่วยวิเคราะห์ความสอดคล้องกลมกลืนของโมเดล ประเมินความเป็นมิติเดียวของคุณลักษณะแฝงที่ต้องการวัดได้ สามารถประมาณค่าพารามิเตอร์ความชันได้ไม่ยากเกินไป และเกิดความมั่นใจได้ว่ามีข้อมูลการตอบในทุกตัวเลือกการรายการคำตอบ

2.2 Modified Graded Response Model (M-GRM) เป็นโมเดลที่ปรับปรุงมาจากโมเดล GRM เหมาะสำหรับการวิเคราะห์เครื่องมือวัดที่มีลักษณะเป็นมาตรฐานค่า โดยข้อคำถามจะต้องมีจำนวนรายการคำตอบที่เท่ากัน แต่สามารถมีค่าพารามิเตอร์ความชันแตกต่างกันได้ โมเดลนี้ใช้วิธีการคำนวณความน่าจะเป็นในการเลือกการคำตอบเหมือนโมเดล GRM โดยโมเดลนี้มีการประมาณค่าพารามิเตอร์ที่น้อยกว่าโมเดล GRM เนื่องจากข้อคำถามแต่ละข้อมีจำนวนรายการคำตอบที่เท่ากันตลอดทั้งเครื่องมือวัด

2.3 Partial Credit Model (PCM) เป็นโมเดลที่ออกแบบมาสำหรับวิเคราะห์ข้อคำถามที่มีหลายขั้นตอน โดยมีการตรวจและให้คะแนนสำหรับแต่ละขั้นตอนที่ผู้ตอบทำได้ พบได้บ่อยในแบบทดสอบผลสัมฤทธิ์ทางการเรียน เช่น โจทย์ปัญหาทางคณิตศาสตร์ นอกจากนี้ยังนำมาใช้วิเคราะห์เครื่องมือวัดเจตคติ หรือบุคลิกภาพที่มีการประเมินความเชื่อ หรือการตอบสนองต่อข้อความ โดยมีการให้คะแนนแบบหลายค่าด้วย โมเดลนี้แตกต่างจากโมเดล GRM และ M-GRM เนื่องจากเป็นโมเดลทางตรง (Direct model) นั่นคือ คำนวณค่าความน่าจะเป็นในการเลือกการคำตอบของแต่ละตัวเลือกได้โดยตรง โมเดลนี้เป็นโมเดลที่ขยายต่อมาจากโมเดล 1PLM ดังนั้นข้อคำถามแต่ละข้อจะมีความชันเท่ากัน คงเหลือเพียงพารามิเตอร์ความยากแบบลำดับขั้น (Step difficulty) ที่ใช้อธิบายระดับความยากของการได้คะแนนจากระดับหนึ่งไปยังคะแนนอีกระดับหนึ่ง หรือจากตัวเลือกการคำตอบหนึ่งไปสู่อีกตัวเลือกหนึ่ง นอกจากนี้ ค่านี้อิงบ่งบอกถึงระดับคุณลักษณะแฝงบริเวณที่ไค้ังการเลือกการคำตอบ 2 รายการตัดกัน หรือเรียกว่า พารามิเตอร์จุดตัดรายการคำตอบ (Category intersections parameter) โดยข้อคำถามที่มีตัวเลือกการคำตอบ 5 รายการ จะมีพารามิเตอร์ความยากแบบลำดับขั้น 4 ค่า จุดเด่นของโมเดลนี้ คือ สามารถใช้คะแนนดิบรวมของเครื่องมือวัด เพื่อประมาณค่าระดับคุณลักษณะแฝงได้ ดังนั้น บุคคลที่มีคะแนนดิบรวมของเครื่องมือวัดเท่ากัน จึงมีระดับคุณลักษณะแฝงเท่ากัน อย่างไรก็ตาม โมเดลนี้นำไปใช้ในทางปฏิบัติได้ยาก เพราะการประมาณค่าพารามิเตอร์อาจพบปัญหาว่าบริเวณจุดตัดไค้ังการเลือกการคำตอบไม่ได้เรียงลำดับจากง่ายไปยาก หรือจากระดับคุณลักษณะแฝงต่ำไปสูง ทำให้เกิดปรากฏการณ์ผกผัน (Reversal) แสดงให้เห็นว่า มีตัวเอกรายการคำตอบอย่างน้อย 1 รายการที่เป็นตัวเลือกที่ไม่สอดคล้องกับคุณลักษณะแฝงที่ต้องการวัด หรือข้อมูลที่เกิดขึ้นรวมมาไม่สอดคล้องกับโมเดล

2.4 Generalized Partial Credit Model (G-PCM) เป็นโมเดลที่ปรับปรุงจากโมเดล PCM โดยกำหนดให้ข้อคำถามแต่ละข้อในเครื่องมือวัด มีพารามิเตอร์ความชันแตกต่างกันได้ โดยพารามิเตอร์ความชันในโมเดลนี้ เป็นค่าที่บ่งบอกถึงระดับการตอบสนองต่อตัวเลือกรายการคำตอบที่แตกต่างกันไปในแต่ละข้อคำถาม ขึ้นอยู่กับระดับคุณลักษณะแฝงที่เปลี่ยนแปลงไป ในโมเดลนี้ หากข้อคำถามใดมีค่าพารามิเตอร์ความชันน้อยกว่า 1.0 โค้งการเลือกรายการคำตอบที่ได้จะมีลักษณะแบนราบ (คล้ายกับโมเดล PCM) แต่หากค่าพารามิเตอร์ความชันมากกว่า 1.0 โค้งการเลือกรายการคำตอบจะมีลักษณะโด่งมากกว่าโมเดล PCM

2.5 Rating Scale Model (RSM) เป็นกลุ่มของโมเดลที่มีหลากหลายรูปแบบ คล้ายกับโมเดล PCM แต่จุดแตกต่างกันคือ โมเดล PCM ข้อคำถามแต่ละข้อมีระดับความยากง่ายของแต่ละขั้นตอนที่แตกต่างกันได้ ส่วนโมเดลนี้จะมีชุดของพารามิเตอร์จุดตัดรายการคำตอบเพียงชุดเดียวที่ใช้ตลอดทั้งเครื่องมือวัด นั่นคือ ข้อคำถามแต่ละข้อในเครื่องมือวัดหนึ่ง ๆ มีระดับความยากง่ายหรือช่วงของแต่ละรายการคำตอบไม่แตกต่างกัน โดยโมเดลนี้เหมาะกับข้อคำถามที่มีการกำหนดช่วงรายการคำตอบแบบคงที่ ซึ่งใช้กับข้อคำถามทุกข้อเหมือนกันตลอดทั้งเครื่องมือวัด แต่ไม่เหมาะกับเครื่องมือวัดที่มีลักษณะการตอบข้อคำถามที่หลากหลาย

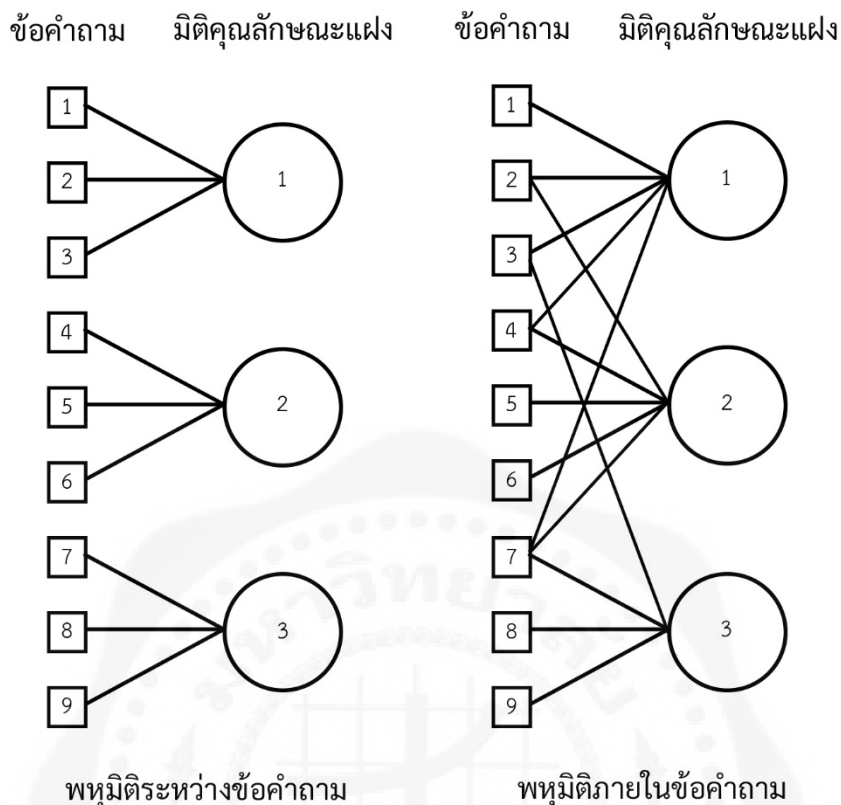
2.6 Nominal Response Model (NRM) เป็นโมเดลใช้วิเคราะห์ข้อคำถามที่รายการคำตอบไม่ได้จัดเรียงลำดับ ได้แก่ ข้อคำถามแบบหลายตัวเลือก (Multiple choices) ให้ความสำคัญกับคุณลักษณะตัวลวงซึ่งเป็นตัวเลือกหนึ่งในข้อคำถามด้วยเส้นร่องรอบ (Trace line)

ในปัจจุบันมีการพัฒนาโมเดลการตอบสนองรายข้อแบบพหุมิติ (Multidimensional IRT model: MIRT) มากขึ้น ซึ่งมุ่งวัดความแตกต่างของบุคคลที่กำหนดโดยคุณลักษณะแฝงมากกว่า 1 องค์ประกอบ ซึ่งสอดคล้องกับสภาพความเป็นจริงของการตอบข้อคำถามที่ต้องใช้คุณลักษณะที่หลากหลาย เพื่อที่จะตอบข้อคำถามข้อนั้นได้ จะช่วยสะท้อนถึงความซับซ้อนของปฏิสัมพันธ์ระหว่างผู้ตอบและข้อคำถามได้มากขึ้น (Reckase, 2009)

โมเดลการตอบสนองรายข้อแบบพหุมิติ

ข้อตกลงเบื้องต้นของโมเดลการตอบสนองรายข้อแบบพหุมิติ มีดังนี้ (Reckase, 2009)

1. ความเป็นพหุมิติระหว่างข้อคำถาม (Between-Items Multidimensionality) มีลักษณะของแต่ละข้อวัดคุณลักษณะแฝงเดียว แต่มีหลายคุณลักษณะแฝง และแต่ละคุณลักษณะแฝงมีความสัมพันธ์กัน
2. ความเป็นพหุมิติภายในข้อคำถาม (Within-Items Multidimensionality) มีลักษณะของข้อคำถามแต่ละข้อวัดหลายคุณลักษณะแฝง มีหลายคุณลักษณะแฝง และแต่ละคุณลักษณะแฝงมีความสัมพันธ์กัน



ภาพที่ 11-6 ข้อตกลงเบื้องต้นของโมเดลการตอบสนองรายข้อแบบพหุมิติ
ที่มา: ปรับมาจาก Wilson and Hoskens (2005)

โมเดลการตอบสนองรายข้อแบบพหุมิติ ประกอบด้วย พารามิเตอร์ที่บ่งบอกคุณลักษณะของบุคคลตั้งแต่ 2 พารามิเตอร์ขึ้นไป มิติการวัดที่หลากหลายทำให้ได้ข้อมูลการตอบข้อคำถามที่มีความสอดคล้องมากขึ้นเมื่อบุคคลตอบข้อคำถามที่มีความยากง่ายแตกต่างกันไป โมเดลนี้สามารถจัดกลุ่มได้เป็น 2 กลุ่ม คือ โมเดลเชิงสำรวจ (Exploratory models) และโมเดลเชิงยืนยัน (Confirmatory models) โดยเมื่อเปรียบเทียบกับ การวิเคราะห์องค์ประกอบ พบว่า โมเดลการตอบสนองรายข้อแบบพหุมิติเชิงสำรวจ (Exploratory multidimensional models) เกี่ยวข้องกับการประมาณค่าพารามิเตอร์ข้อคำถามและพารามิเตอร์บุคคลมากกว่า 1 มิติ เพื่อให้โมเดลมีความสอดคล้องกับข้อมูล โดยไม่มีการกำหนดจำนวนขององค์ประกอบอย่างชัดเจน ซึ่งมีลักษณะตรงกันข้ามกับโมเดลการตอบสนองรายข้อแบบพหุมิติเชิงยืนยัน (Confirmatory multidimensional models) เกี่ยวข้องกับการประมาณค่าพารามิเตอร์ โดยกำหนดมิติขององค์ประกอบไว้อย่างชัดเจน

การประเมินความสอดคล้องของโมเดล

เมื่อทำการเลือกโมเดลที่มีความเหมาะสมแล้ว จึงต้องทำการประเมินความสอดคล้องของโมเดล (Model fit) (Chernyshenko et al., 2001; Zanon et al., 2016) โดยมีรายละเอียดที่สำคัญ ดังนี้

1. การตรวจสอบข้อตั้งของเบื้องต้นของโมเดล (Checking Model Assumptions) โดยโมเดลการตอบสนองรายข้อส่วนใหญ่ทำการตรวจสอบข้อตั้งเบื้องต้นที่สำคัญ คือ ความเป็นมิติเดียว (Unidimensionality) และความเป็นอิสระ (Local independence) โดยที่ข้อตั้งเบื้องต้นความเป็นมิติเดียว แสดงถึง ความน่าจะเป็นหรือโอกาสของการตอบของบุคคลเป็นคุณลักษณะแฝงเดียวของบุคคล เมื่อคุณลักษณะแฝงภายในบุคคลมีความเป็นมิติเดียว จึงทำให้การตอบข้อคำถามของผู้ตอบมีความแตกต่างกัน นั้นแสดงถึง ผลของการตอบคำถามในข้อคำถามแต่ละข้อ มีความเป็นอิสระจากการกัน กล่าวโดยสรุป เมื่อโมเดลการตอบสนองรายข้อผ่านข้อตั้งเบื้องต้นในเรื่องของความเป็นมิติเดียว จึงทำให้ผ่านข้อตั้งเบื้องต้นในเรื่องความเป็นอิสระด้วย

สำหรับการตรวจสอบข้อตั้งเบื้องต้นความเป็นมิติเดียว มีการพัฒนาวิธีการตรวจสอบอย่างหลากหลาย โดยพบว่า วิธีการวิเคราะห์คู่ขนานแบบปรับแก้ (Modified parallel analysis: MPA) เป็นวิธีการหนึ่งที่มีประสิทธิภาพในการทดสอบว่าชุดของข้อคำถามมีความเป็นมิติเดียวที่มีความเพียงพอ (Drasgow & Lissak, 1983) นอกจากนี้ยังมีอีกวิธีการหนึ่งที่ใช้กันบ่อย คือ วิธีการ DIMTEST (Stout, 1987) แต่วิธีการนี้มีข้อจำกัดเรื่องเครื่องมือวัดฉบับสั้น โดยต้องมีอย่างน้อย 20 ข้อคำถาม จึงจะได้ผลที่มีความแม่นยำ ซึ่งเครื่องมือวัดบุคลิกภาพมักจะไม่เกิน 15 ข้อคำถาม

2. การตรวจสอบความสอดคล้องระหว่างโมเดลกับข้อมูล (Checking Model-Data Fit) โดย Drasgow et al. (1995) ได้แนะนำให้วิธีการภาพกราฟฟิกและวิธีการทางสถิติร่วมกัน เพื่อประเมินความเหมาะสมของการทำนายโมเดล

สำหรับวิธีการ Graphical fit plots เป็นวิธีการที่ใช้กันมากที่สุดสำหรับการอธิบายความสอดคล้องระหว่างโมเดลกับข้อมูล ซึ่งเป็นการพล็อตกราฟข้อคำถามและตัวเลือกการตอบ โดยประมาณจากกลุ่มตัวอย่างเปรียบเทียบ และสัดส่วนของการตอบได้ถูกต้องจากกลุ่มตัวอย่างตรวจสอบไขว้ ปัญหาของวิธีการนี้คือการประมาณความสามารถของบุคคลยากที่จะเท่ากับความสามารถแท้จริง เนื่องมาจากความคลาดเคลื่อนในการประมาณค่า ความคลาดเคลื่อนสามารถเปลี่ยนแปลงพล็อตได้ในกลุ่มตัวอย่างขนาดใหญ่

สำหรับวิธีการทดสอบความสอดคล้องกลมกลืนทางสถิติ (เช่น ค่าไค-สแควร์) เป็นวิธีการประเมินความสอดคล้องระหว่างโมเดลกับข้อมูลที่ใช้กันอย่างกว้างขวาง แต่ก็มีปัญหาความไม่สอดคล้องกันระหว่างโมเดลกับข้อมูล เนื่องจากขนาดกลุ่มตัวอย่าง ดังนั้น จึงพัฒนาวิธีการคำนวณทางสถิติ คือ ค่าไค-สแควร์ที่ปรับแก้ โดยเอาค่าไค-สแควร์ หารด้วยองศาอิสระ ซึ่งไม่ควรมีความมากกว่า 3.0

ยกตัวอย่าง การวิเคราะห์โมเดลการตอบสนองรายข้อแบบพหุมิติ

งานวิจัย: การพัฒนาแบบวัดจิตแห่งความเคารพและจิตแห่งจริยธรรมตามแนวคิดของการ์ดเนอร์สำหรับนักเรียนมัธยมศึกษาตอนต้น (น้ำทิพย์ งามอาภาวิชัย, 2561)

เครื่องมือวัด คือ เครื่องมือวัดจิตแห่งความเคารพและจิตแห่งจริยธรรม ประกอบด้วย 2 องค์ประกอบหลัก 7 องค์ประกอบย่อย ได้แก่ 1) จิตแห่งความเคารพ ประกอบด้วย การทำงานร่วมกัน การเคารพสิทธิผู้อื่น เมตตา และการคิดเชิงบวก และ 2) จิตแห่งจริยธรรม ประกอบด้วย จิตสาธารณะ วินัยในตนเอง และความรับผิดชอบ มีรายการคำตอบในรูปแบบมาตราประมาณค่า 5 ระดับ จำนวน 40 ข้อ

การวิเคราะห์ข้อมูล ได้แก่

1. ความยาก พิจารณาจากค่าพารามิเตอร์ของแต่ละรายการคำตอบ (b) ที่ได้จากการวิเคราะห์ด้วย Graded-Response Model (GRM) เนื่องจากแบบวัดโดยใช้ข้อคำถามที่มีลักษณะเชิงสถานการณ์ 5 ระดับ จึงต้องใช้การวิเคราะห์ความยากและอำนาจจำแนกด้วยโมเดลการตอบสนองรายข้อแบบตรวจให้คะแนนมากกว่า 2 ค่า (Polytomous IRT model) โดยการวิเคราะห์ด้วย GRM เป็นวิธีการที่เหมาะสมสำหรับแบบวัดที่แต่ละข้อคำถามมีรายการคำตอบแบบเรียงลำดับ เพื่อศึกษาบุคคลที่มีคุณลักษณะแฝงหรือความสามารถแตกต่างกัน มีความน่าจะเป็นในการตอบแต่ละรายการคำตอบแตกต่างกันอย่างไร

2. อำนาจจำแนก พิจารณาจากค่าพารามิเตอร์ความชัน (a) ที่ได้จากการวิเคราะห์ด้วย GRM เพื่อศึกษาว่าผู้ตอบที่มีความสามารถว่าสามารถจำแนกผู้ตอบที่มีความสามารถออกจากกันได้ดีเพียงใด

3. การวิเคราะห์ตามทฤษฎีการตอบสนองรายข้อแบบพหุมิติ เป็นการแสดงหลักฐานความเที่ยงตรงเชิงโครงสร้างของการวัดแบบพหุมิติ ซึ่งใช้การเปรียบเทียบความสอดคล้องของโมเดลการวัดกับข้อมูล (Model-data fit) ว่ามีความเหมาะสมหรือไม่ที่จะใช้การวิเคราะห์แบบพหุมิติ และทำการเปรียบเทียบโมเดลแบบมิติเดียวกับโมเดลพหุมิติ ซึ่งเป็นโมเดลที่ซ้อนสัมพันธ์กัน (Nested model) โดยพิจารณาจากการทดสอบความแตกต่างของไค-สแควร์ (Chi-square difference test) เพื่อทดสอบว่า โมเดลแบบพหุมิติมีความสอดคล้องกับข้อมูลมากกว่าโมเดลแบบมิติเดียวหรือไม่ หากการทดสอบมีนัยสำคัญทางสถิติแสดงว่า โมเดลแบบพหุมิติมีความสอดคล้องกับข้อมูล มากกว่าโมเดลแบบมิติเดียว

ผลการวิจัยพบว่า 1) การตรวจสอบคุณภาพของแบบวัดจิตแห่งความเคารพและจิตแห่งจริยธรรม การวิเคราะห์ค่าพารามิเตอร์ความยาก และอำนาจจำแนกตามทฤษฎีการตอบสนองรายข้อ (IRT) พบว่าค่าพารามิเตอร์อำนาจจำแนกแบบพหุมิติ มีค่าอยู่ระหว่าง 0.38-0.76 ส่วนค่าพารามิเตอร์ความยากแบบพหุมิติของแต่ละรายการคำตอบ (b) จากแบบวัดจิตแห่งความเคารพและจิตแห่งจริยธรรม พบว่า ข้อคำถามทุกข้อมีค่า $b_1 < b_2 < b_3 < b_4$ อาจกล่าวได้ว่า ผู้ที่มีคุณลักษณะแฝงหรือความสามารถสูง มีโอกาสเลือกรายการคำตอบที่มีคะแนนสูง และ 2) การตรวจสอบความเป็นพหุมิติ ซึ่งผลการพิจารณาค่าสถิติความเหมาะสมรายข้อบ่งชี้ ความเป็น

พหุมิติของโมเดลการวัดจิตแห่งความเคารพและจิตแห่งจริยธรรม จากการทดสอบรายคู่ (Bivariate Fit) ที่มีข้อคำถามเกินร้อยละ 18 ไม่สอดคล้องกับการวัดแบบมิติเดียว จึงจำเป็นต้องใช้การวิเคราะห์แบบพหุมิติ ส่วนการตรวจสอบความเป็นพหุมิติ โดยการตรวจสอบความสอดคล้องของโมเดลการวัดกับข้อมูลจากการเปรียบเทียบโมเดลซ้อนสัมพันธ์กัน (Nested Model) ได้แก่ โมเดลแบบมิติเดียวกับโมเดลแบบพหุมิติ โดยใช้วิธีทดสอบความแตกต่างของค่าไค-สแควร์ ผลการทดสอบแสดงให้เห็นว่า โมเดลแบบพหุมิติมีความสอดคล้องกับข้อมูลมากกว่าโมเดลแบบมิติเดียวอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า โมเดลการวัดจิตแห่งความเคารพและจิตแห่งจริยธรรมเหมาะสมกับการวัดแบบพหุมิติ

สรุป

ทฤษฎีการตอบสนองรายข้อ (IRT) ทำให้ทราบข้อมูลพื้นฐานสำคัญของทฤษฎีการตอบสนองรายข้อ ทั้งภาพรวมของทฤษฎีการตอบสนองรายข้อ ความแตกต่างระหว่างทฤษฎีการทดสอบดั้งเดิมและทฤษฎีการตอบสนองรายข้อ แนวคิดเบื้องต้นของทฤษฎีการตอบสนองรายข้อ พารามิเตอร์ในทฤษฎีการตอบสนองรายข้อ โค้งแสดงคุณลักษณะรายข้อ โมเดลการตอบสนองรายข้อ ซึ่งการเลือกโมเดลการตอบสนองรายข้อที่เหมาะสม และการพิจารณาว่าโมเดลดังกล่าวมีความสอดคล้องกับข้อมูล จึงเป็นการยืนยันว่าผลการวิเคราะห์มีความถูกต้องมากขึ้น ซึ่งช่วยให้แบบวัดทางด้านพฤติกรรมศาสตร์มีคุณภาพและได้มาตรฐานมากยิ่งขึ้น

เอกสารอ้างอิง

- น้ำทิพย์ งามอาภาณิชย์. (2561). การพัฒนาแบบวัดจิตแห่งความเคารพและจิตแห่งจริยธรรมตามแนวคิดของการวัดเนอร์สำหรับนักเรียนมัธยมศึกษาตอนต้น. *วารสารศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น*, 41(4), 23-36.
- ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่* (พิมพ์ครั้งที่ 4). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- Al Nima, A., Cloninger, K. M., Persson, B. N., Sikström, S., & Garcia, D. (2020). Validation of subjective well-being measures using item response theory. *Frontiers in Psychology*, 10, 3036.
- Baker, F. B. (2001). *The basics of item response theory*. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

- Chernyshenko, O. S., Stark, S., Chan, K., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research, 36*(4), 523-562.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press.
- De Beer, M. (2004). Use of differential item functioning (DIF) analysis for bias analysis in test construction. *SA Journal of Industrial Psychology, 30*(4), 52-58.
- De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer-Verlag.
- Drasgow, F., Levine, M. V., Tsien, S., Williams, B. A., & Mead, A. D. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement, 19*, 143-165.
- Drasgow, F., & Lissak, R. I. (1983). Modified parallel analysis: A procedure for examining the latent dimensionality of dichotomously scored item response. *Journal of Applied Psychology, 68*, 363-373.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Embretson, S. E., & Yang, X. (2006). Item response theory. In J. L. Green, G. Camilli, & P. B. Elmore (eds.), *Handbook of complementary methods in education research* (p. 385-409). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hambleton, R. K. (1994). Item response theory: A broad psychometric framework for measurement advances. *Psicothema, 6*(3), 535-556.
- Hambleton, R. K., & Jodoin, M. (2003). Item response theory: Models and features. In R. Fernández-Ballesteros (Ed.), *Encyclopedia of psychological assessment* (pp. 509-514). London: Sage.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer Academic Publishers.

- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Immekus, J. C., Snyder, K. E., & Ralston, P. A. (2019). Multidimensional item response theory for factor structure assessment in educational psychology research. *Frontiers in Education, 4*, 45.
- Nering, M. L., & Ostini, R. (2010). *Handbook of polytomous item response theory models*. New York: Routledge.
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer.
- Rupp, A. A., & Zumbo, B. D. (2004). A note on how to quantify and report whether IRT parameter invariance holds: When Pearson correlations are not enough. *Educational and Psychological Measurement, 64*(4), 588-599.
- Salkind, N. J. (2010). *Encyclopedia of research design*. Thousand Oaks, CA: Sage.
- Sideridis, G. D., Tsaousis, I., & Al Harbi, K. (2016). The impact of non-attempted and dually-attempted items on person abilities using item response theory. *Frontiers in Psychology, 7*, 1572.
- Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52*, 589-617.
- Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use* (5th ed.). Oxford: Oxford University Press.
- Wilson, M., & Hoskens, M. (2005). Multidimensional item responses: Multimethod-multitrait perspectives. In S. Alagumalai, D. D. Curtis, & N. Hungi (eds.), *Applied Rasch measurement: A book of exemplars* (pp. 287-307). Dordrecht, Netherlands: Springer.
- Zampetakis, L. A., Lerakis, M., Kafetsios, K., & Moustakis, V. (2015). Using item response theory to investigate the structure of anticipated affect: Do self-reports about future affective reactions conform to typical or maximal models? *Frontiers in Psychology, 6*, 1438.
- Zanon, C., Hutz, C. S., Yoo, H., & Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica, 29*, <https://doi.org/10.1186/s41155-016-0040-x>

ดัชนีคำศัพท์

ก

การนิยามตัวแปร	18
การนิยามเชิงทฤษฎี	18
การนิยามเชิงปฏิบัติการ	19
การปรับปรุงแบบทดสอบ	117
การประมาณค่าความเชื่อมั่นแบบคู่ขนาน	145
การประมาณค่าความเชื่อมั่นแบบวัดซ้ำ	142
ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน	143
ค่าสัมประสิทธิ์สหสัมพันธ์แบบภายในชั้น	144
การประมาณค่าความเชื่อมั่นแบบสอดคล้องภายใน	146
Split-Half Method	147
Kuder-Richardson Method	148
Cronbach's Coefficient Alpha	149
การประมาณค่าความเชื่อมั่นระหว่างผู้ประเมิน	151
Cohen's Kappa	151
Intraclass Correlation Coefficient: ICC	154
การประเมินการปฏิบัติ	98
การประเมินความสอดคล้องของโมเดล	196
การพัฒนาข้อคำถาม	4, 170
การรายงานตนเอง	97
การวัดทางจิตสรีรวิทยา	97
การวัดทางด้านการรู้คิด	46, 67, 148,
การวัดทางด้านความรู้สึก	70
การวัดทางด้านพฤติกรรม	11, 88, 89, 90, 105
การวิเคราะห์ตามทฤษฎีการตอบสนองรายข้อแบบพหุมิติ	197
การวิเคราะห์รายข้อ	104
การวิเคราะห์รายข้อเชิงคุณภาพ	115
การขอความเห็นจากผู้เชี่ยวชาญ	116
วิธีการคิดต่าง ๆ	117

ดัชนีคำศัพท์ (ต่อ)

การวิเคราะห์องค์ประกอบ	129, 159
การวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA)	130, 160, 170
การประเมินความสอดคล้องกลมกลืน	171, 196
ความเชื่อมั่นของตัวแปรแฝง	121
ความแปรปรวนที่ถูกลบด้วยองค์ประกอบ	179
ค่าน้ำหนักองค์ประกอบ	178
การวิเคราะห์องค์ประกอบเชิงสำรวจ (EFA)	129, 160
การสกัดองค์ประกอบ	163, 167
ค่าไอเกน	163, 169
การหมุนแกนองค์ประกอบ	163, 167
ค่าน้ำหนักองค์ประกอบ	164, 169
ค่าความร่วมกัน	164, 169
การสังเกตโดยตรง	93
การบันทึกการเล่าเรื่อง	93
การบันทึกเหตุการณ์	93
การบันทึกเป็นช่วง	94
การสัมภาษณ์เชิงพฤติกรรม	95
การสัมภาษณ์แบบมีโครงสร้าง	96
การสัมภาษณ์แบบกึ่งโครงสร้าง	96
การอ้างอิงผลการวัด	120
ค	
ความคลาดเคลื่อนในการใช้มาตรวัด	42
ความคลาดเคลื่อนทางการวัด	140
ความเชื่อมั่น	22, 139, 155
ความเที่ยงตรง	21, 120
ความเที่ยงตรงเชิงจำแนก	134, 179
ความเที่ยงตรงเชิงโครงสร้าง	121, 128, 159
วิธีการหาความเป็นเอกพันธ์ของเครื่องมือวัด	129, 155
วิธีการวิเคราะห์องค์ประกอบ	129, 159
วิธีการเทียบกับกลุ่มที่รู้ชัดหรือกลุ่มตรงข้าม	131
วิธีการวิเคราะห์หาคูณลักษณะพหุวิธี	132

ดัชนีคำศัพท์ (ต่อ)

ความเที่ยงตรงเชิงเนื้อหา	120, 121
การวางแผนกำหนดพิมพ์เขียว	121
ค่าดัชนีความสอดคล้องระหว่างข้อความกับเนื้อหา	122
ความเที่ยงตรงเชิงเสมือน	134, 179
ความเที่ยงตรงตามเกณฑ์	121, 124
ค่าสัมประสิทธิ์ความเที่ยงตรง	125
ความเที่ยงตรงเชิงสภาพ	126
ความเที่ยงตรงเชิงทำนาย	126
ความเป็นปรนัย	22
ความเป็นมิติเดียว	185, 196
ความเป็นอิสระ	186, 196
ความยากง่ายรายข้อ	110
ความสัมพันธ์ระหว่างการวิเคราะห์รายข้อรูปแบบต่าง ๆ	114
โค้งแสดงคุณลักษณะรายข้อ	186, 188
ท	
ทฤษฎีการตอบสนองรายข้อ	182, 185
ทฤษฎีการทดสอบดั้งเดิม	139, 182, 184
ทฤษฎีคะแนนจริง	139
บ	
แบบทดสอบความถนัด	53
แบบทดสอบความถนัดทั่วไป	54
แบบทดสอบความถนัดเฉพาะทาง	55
แบบทดสอบผลสัมฤทธิ์	58
แบบทดสอบปรนัย	63
- ข้อสอบแบบถูกผิด	63
- ข้อสอบแบบตัวเลือก	64
- ข้อสอบแบบจับคู่	65
แบบทดสอบอัตนัย	66
- ข้อสอบแบบไม่จำกัดคำตอบ	66
- ข้อสอบแบบจำกัดคำตอบ	66

ดัชนีคำศัพท์ (ต่อ)

แบบทดสอบสติปัญญา	47
แบบทดสอบสติปัญญารายบุคคล	48
แบบทดสอบสติปัญญารายกลุ่ม	52
แบบวัดค่านิยม	74
แบบวัดความสนใจ	71
แบบวัดบุคลิกภาพ	77
วิธีการวัดแบบไม่ให้รู้ตัว	82
เทคนิคการสังเกตแบบควบคุมได้	82
การรายงานตนเอง	83
เทคนิคการฉายออก	83
ป	
ปัจจัยที่ส่งผลต่อความเชื่อมั่น	155
พ	
พารามิเตอร์ในทฤษฎีการตอบสนองรายข้อ	186
พารามิเตอร์ของผู้ตอบ	186
พารามิเตอร์ของข้อคำถาม	187
ค่าคงที่	188
ม	
มาตรฐานค่าเชิงพฤติกรรม	97
มาตรวัด	26
มาตรวัดแบบเธอร์สโตน	29
มาตรวัดลิเคิร์ต	33
มาตรวัดออสกูด	38
มาตรประเมินค่า	41
โมเดลการตอบสนองรายข้อ	189
โมเดลการตอบสนองรายข้อแบบตรวจให้คะแนน 2 ค่า	190
โมเดลการตอบสนองรายข้อแบบตรวจให้คะแนนมากกว่า 2 ค่า	192
โมเดลการตอบสนองรายข้อแบบพหุมิติ	194

ดัชนีคำศัพท์ (ต่อ)

ร

ระดับของการวัด	15
มาตรการวัดนามบัญญัติ	15
มาตรการวัดเรียงอันดับ	16
มาตรการวัดช่วง/อันตรภาค	17
มาตรการวัดอัตราส่วน	17

ส

สัมประสิทธิ์ความเชื่อมั่น	141
---------------------------	-----

ห

หลักการวัด	10
หลักการสร้างมาตรวัด	27

อ

อำนาจจำแนก	22, 105
ค่าอำนาจจำแนกรายข้อ ด้วยดัชนีค่าอำนาจจำแนกรายข้อ	105
ค่าอำนาจจำแนกรายข้อ ด้วยการใช้การทดสอบค่าที	37, 107
ค่าอำนาจจำแนกรายข้อ ด้วยการทำค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างรายข้อกับคะแนนรวม	37, 109
อำนาจตัวลง	113

ประวัติผู้เขียน

ผู้ช่วยศาสตราจารย์ ดร. ศรัณย์ พิมพ์ทอง

ประวัติการศึกษา

- บัณฑิต (เกียรตินิยมอันดับ 2) มหาวิทยาลัยเกษตรศาสตร์
- เศรษฐศาสตรมหาบัณฑิต สถาบันบัณฑิตพัฒนบริหารศาสตร์
- ปรัชญาดุษฎีบัณฑิต พัฒนาสังคมและการจัดการสิ่งแวดล้อม สถาบันบัณฑิตพัฒนบริหารศาสตร์

การทำงาน

- อาจารย์ประจำ สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ

ความเชี่ยวชาญ

- การวิจัยพฤติกรรมศาสตร์
- การพัฒนาองค์การ
- การสร้างและพัฒนาเครื่องมือวัดทางพฤติกรรมศาสตร์
- Internet-Based Intervention

