

Pros และ Cons ในการวัดด้วยมาตราประเมินรวมค่า โดยให้ผู้ตอบรายงานตนเอง (Pros and Cons of Summated Rating Method with Self-Report)

ดวงเดือน พันธุมนาวิน*

บทคัดย่อ

การวิจัยทางสังคมพฤติกรรมศาสตร์ มุ่งศึกษามนุษย์ และสังคมมนุษย์ เครื่องมือวัดที่นิยมใช้กันมาก มาตลอดช่วง 40 ปี หลังนี้ คือ เครื่องมือวัดแบบมาตราประเมินรวมค่า (Summated ratings method) ในบทความนี้ได้กล่าวถึงลักษณะสำคัญของเครื่องมือวัดประเภทนี้ที่ทำให้เป็นที่นิยมใช้ และข้อจำกัดที่ผู้ใช้ควรระมัดระวังเป็นพิเศษ เมื่อนำเครื่องมือวัดชนิดนี้มาใช้ในการให้ผู้ตอบรายงานเกี่ยวกับตนเอง (Self-report) ทำให้ได้ข้อมูลที่ลึกซึ้งและกว้างขวาง ครอบคลุมปรากฏการณ์ในอดีต ปัจจุบัน และอนาคตของผู้ตอบได้เป็นอย่างดี ได้มีการศึกษาเปรียบเทียบวิธีวัดแบบมาตราประเมินตนเอง (Self ratings) นี้กับวิธีวัดแบบอื่นๆ โดยใช้ตัวอย่างจากงานวิจัยการอบรมเลี้ยงดูเด็กในประเทศไทยและต่างประเทศหลายเรื่อง นอกจากนี้ บทความนี้ยังได้กล่าวถึง วิธีการควบคุมที่สำคัญ คือ การตรวจจับการตอบตามความนิยมของสังคม (Social desirability) เพื่อป้องกันและลดข้อผิดพลาด เมื่อใช้วิธีการวัดดังกล่าว

Abstract

Empirical studies in social-behavioral sciences aim at studying human individuals and human societies. The most popular measure which has been used widely for more than 40 years, is the “Summated ratings method”. This article pinpoints many important characteristics which have encouraged the use of this method. However, this is not without the shortcomings which the users should try to control. When the summated ratings method is combined with self report, which is

* ศาสตราจารย์ อดิศักดิ์ อดิศักดิ์ประจักษ์พัฒนาศักดิ์และสิ่งแวดล้อม สถาบันบัณฑิตพัฒนบริหารศาสตร์ และผู้อำนวยการโครงการวิจัยแม่บท : การวิจัยและพัฒนาระบบพฤติกรรมไทย สำนักงานคณะกรรมการวิจัยแห่งชาติ

called “self ratings”, it can generate rich data which can encompass phenomena in the past as well as at present and in the future of the respondents. Comparisons were made between self ratings method and other techniques with emphasis on studies in child rearing from Thailand as well as studies from abroad. In addition, this article also examines the important control technique of “social desirability” in an attempt to improve this type of measurement.

มนุษย์เป็นสิ่งมีชีวิตประเภทเดียวที่มีความสามารถในการศึกษาวิจัยสิ่งรอบตัว รวมทั้งวิจัยมนุษย์ด้วยกันเอง หรือแม้แต่วิเคราะห์ วิจัยรณตนเอง นักจิตวิทยาพฤติกรรมศาสตร์ทั้งหลายมีภารกิจในการแสวงหาความจริงเกี่ยวกับสาเหตุต่างๆ ของจิตใจ และพฤติกรรมของมนุษย์ที่แปรเปลี่ยนอยู่เสมอ ตลอดจนศึกษาผลที่หลากหลายอันเกิดจากจิต และพฤติกรรมของบุคคลในสถานการณ์ต่างๆ ด้วย

สิ่งที่นักวิชาการอื่นๆ ศึกษาวิจัย มักเป็นสัตว์ ต้นไม้ หรือสิ่งที่ไม่มีชีวิต เช่น ดิน น้ำ ลม ไฟ ซึ่งพูดได้ตอบกับผู้นักวิจัยไม่ได้ แต่นักวิจัยทางจิตวิทยาพฤติกรรมศาสตร์ได้เปรียบทางด้านที่สิ่งที่ตนศึกษา สามารถสื่อสารได้ตอบกับนักวิจัยได้อย่างดี จนบางครั้งทำให้นักวิจัยในบางสาขาวิชาที่ศึกษามนุษย์คาดหวังจากผู้ถูกศึกษามากเกินไป จนกระทั่งไปถามถึงสาเหตุต่างๆ ที่ตนอยากรู้ จากผู้ที่ตนจะต้องเข้าไปวัด และวิจัย เพื่อพิสูจน์ถึงสาเหตุต่างๆ ของจิต และพฤติกรรมของคนเหล่านั้น โดยไม่ได้ตระหนักว่าผู้ที่กระทำพฤติกรรมต่างๆ มักไม่ทราบสาเหตุที่แท้จริงของการกระทำของตน เขาอาจให้คำตอบไปต่างๆ นานา จึงเป็นหน้าที่ของนักวิจัยที่จะต้องพิสูจน์สาเหตุกับผลด้วยการวัดและการวิจัยที่สามารถจะให้ความจริงที่น่าเชื่อถือได้ นั่นคือนักวิจัยจะต้องใช้หลักและวิธีทางวิทยาศาสตร์อย่างเคร่งครัด เพื่อให้ได้ข้อมูลและข้อสรุปที่เป็นความจริงมากกว่าความเท็จ

การที่นักวิจัยจะได้ความจริงจากการทำวิจัยของเขาได้อย่างมากนั้น ย่อมขึ้นอยู่กับเครื่องมือวัดตัวแปรอิสระ และตัวแปรตามต่างๆ ที่มีมาตรฐานสูงเป็นอันดับแรก

การวัดด้วยมาตรประเมินรวมค่า

เป็นเวลานานมากกว่า 50 ปีมาแล้ว ที่นักวิจัย และนักพัฒนาทางด้าน สังคมศาสตร์ และพฤติกรรมศาสตร์ โดยเฉพาะในสาขาจิตวิทยานิยมใช้เครื่องมือ วัดแบบมาตรประเมิน (Rating – scales method, Guilford, 1954 : 263) และมี ผู้ยืนยันว่ายังเป็นที่ยอมรับใช้มากขึ้นจนถึงปัจจุบัน (Kerlinger & Lee, 2000 : 736)

เครื่องมือวัดแบบมาตรประเมินนี้ อาจใช้ได้หลายกรณี ทั้งในกรณีที่ให้ ผู้ตอบประเมินคนอื่น หรือสิ่งอื่น และที่เป็นการรายงานเกี่ยวกับลักษณะของผู้ตอบเอง (เรียกว่า self rating) โดยเริ่มใช้เป็นต้นแบบครั้งแรกในการวัดทัศนคติด้วยวิธีของ Likert ที่เริ่มตั้งแต่ ค.ศ. 1932 (Likert, 1932 และ Guilford, 1954 : 459) โดยให้หลาย ข้อ แต่ละข้อมีมาตร 5 หน่วยประกอบ คะแนนรวมกันจากการตอบบนมาตรเหล่านี้ แสดงถึงทิศทาง (ชอบ – ไม่ชอบ) และปริมาณของทัศนคติของผู้ตอบต่อวัตถุทาง ทัศนคตินั้น

วิธีวัดลักษณะของผู้ตอบเองด้วยมาตรประเมินรวมค่านี้นี้ได้มีผู้นำมาใช้ใน ประเทศไทยเป็นครั้งแรก ตั้งแต่ พ.ศ. 2510 เมื่อเก็บข้อมูลในงานวิจัยเรื่อง “อิทธิพล ของสังคมต่อทัศนคติของวัยรุ่น” (ดวงเดือน พันธุมนาวิน และบุญยิ่ง เจริญยิ่ง, 2517)

งานวิจัยทางจิตวิทยา และพฤติกรรมศาสตร์ในปัจจุบัน ทั้งในประเทศไทย และต่างประเทศแทบทุกเรื่อง จะต้องใช้แบบวัดชนิดมาตรประเมินรวมค่า และ ให้ผู้ตอบรายงานเกี่ยวกับตนเองด้วย โดยใช้แบบวัดสองชนิดผสมดังกล่าวในการ วัดทั้งตัวแปรอิสระ และตัวแปรตามต่างๆ ในการวิจัยเดียวกัน ซึ่งถือว่าอาจทำให้ งานวิจัยนั้นมีข้อจำกัดได้ จำเป็นที่นักวิจัยจะต้องหาทางออกที่ดีในการดำเนินการ ป้องกันหลายๆ วิธีจึงจะทำให้งานนั้นเป็นเลิศได้

ความนิยมใช้มาตรประเมินรวมค่าในการวัดแบบ self report นี้ นับวันจะเข้ามาแทนที่การวัดแบบอื่นๆ มากขึ้น แต่นักวิจัยย่อมไม่ทำตามๆ กัน ไม่ม่งมาย ไม่ลอกเลียนแบบ แต่ต้องเป็นผู้มีเหตุมีผล รู้ข้อดี และข้อจำกัด ของวิธีการวัดทั้งที่ตนนำมาใช้และไม่เลือกมาใช้อย่างถ่องแท้ โดยรู้จักพิจารณาว่าแบบวัด และวิธีวัด ชนิดที่ตนตัดสินใจเลือกเอามาใช้ในงานวิจัยของตนนั้น จะต้องดี มีประโยชน์ มากกว่าที่จะมีข้อเสีย หรือมีโทษ โดยเฉพาะในบริบทและกลุ่มตัวอย่างประเภทที่ตนศึกษา นอกจากนี้ นักวิจัยจะต้องพยายามด้วยตนเองในการเพิ่มข้อดี และลดข้อเสียเหล่านั้นให้มากด้วย

บทความนี้ จึงมีจุดมุ่งหมายที่จะนำเสนอข้อดี และข้อจำกัด ของวิธีการวัดดังกล่าว ในกรณีและสถานการณ์ในการวิจัยต่างๆ ตลอดจนเสนอแนะแนวทางการจัดการเพิ่มข้อดี และป้องกันหรือลดข้อเสียให้มากที่สุดในงานวิจัยที่จะทำต่อไป

ลักษณะของมาตรประเมินรวมค่า

มาตรประเมินรวมค่า (Summated rating scales) อาจแยกอธิบายเป็น 3 คำ เริ่มด้วย “มาตร” (scale) ซึ่งหมายถึง ตัวเลข หรือสัญลักษณ์ชุดหนึ่ง ที่มีการสร้างขึ้นไว้เพื่อให้แก่ลักษณะหนึ่งของบุคคลต่างๆ หรือให้แก่พฤติกรรมของเขาเหล่านั้นที่ต่างกันไป โดยยึดตามกฎ หรือระเบียบที่กำหนดขึ้น (Kerlinger & Lee, 2000 : 708) ในการใช้มาตรหนึ่งๆ ผู้ใช้จะต้อง “ประเมิน” หรือทำการพิจารณาตัดสินสิ่งใดสิ่งหนึ่ง ดังนั้น “มาตรประเมิน” (rating scales) จึงเหมาะที่จะให้มนุษย์เป็นผู้ใช้ โดยจะให้เขาประเมินผู้อื่นทางด้านการกระทำ ความรู้สึกนึกคิด หรือ ประเมินวัตถุ สิ่งเร้า หรือสิ่งที่เป็นนามธรรมก็ได้ ตลอดจนการประเมินและรายงานเกี่ยวกับจิตลักษณะ พฤติกรรม และการรับรู้สภาพแวดล้อมของตนเอง (ที่เรียกว่า self report)

มาตรประเมิน มีหลายแบบตั้งแต่การประเมินตามประเภทคำตอบที่มีให้เลือก ซึ่งเป็นคำตอบด้านหนึ่งที่ต่างกันเชิงปริมาณ หรือแปลงปริมาณต่างๆ เป็นตัวเลข เช่น จาก 1 ถึง 6 มาตรประเมินที่มีลักษณะครบถ้วน จะต้องมีการใช้เส้น (หน่วย) ที่ยาวเท่ากัน ห่างเท่ากัน (เรียกว่า graphic rating scale) และมีคำอธิบายประกอบแต่ละเส้น (หน่วย) ที่มีความหมายเชิงปริมาณที่ต่างเป็นช่วงเท่าๆ กัน ดังตัวอย่าง

ข้อ บวก

คะแนน	6	5	4	3	2	1
-------	---	---	---	---	---	---

ข้อ ลบ

คะแนน	1	2	3	4	5	6
	<u>จริง</u>	<u>จริง</u>	<u>ค่อนข้าง</u>	<u>ค่อนข้าง</u>	<u>ไม่</u>	<u>ไม่จริง</u>
	ที่สุด		จริง	ไม่จริง	จริง	เลย

“มาตรประเมิน” นี้ถ้าสร้างอย่างมีมาตรฐานตามข้างบนนี้ จะมีลักษณะเป็นมาตรแบบปรวิสัย (objective scale) ซึ่งหมายถึงว่าถ้านักวิจัยต่างๆ สร้างตามกฎระเบียบที่กำหนดไว้นี้ (กำหนดโดยความเห็นชอบร่วมกันของผู้ชำนาญการ) จะสามารถให้คะแนนได้ตรงกันหมด นอกจากนี้การให้แต่ละหน่วยเป็นคะแนน มีการแบ่งเส้น ความห่างของเส้น และคำอธิบายใต้แต่ละหน่วย ดังในมาตรข้างบนนี้ จะทำให้มาตรประเมินมีลักษณะของการใช้หน่วยการวัดต่างๆ ซึ่งแต่ละหน่วยมีค่าเท่ากัน (equal – unit scale) และเป็นไปตามข้อตกลงเบื้องต้นว่าแบบวัดดังกล่าวอยู่ในระดับ Interval scale (Kerlinger & Lee, 2000 : 736-738) ซึ่งทำให้นักวิจัยเมื่อนำข้อมูลจากการวัดในลักษณะนี้ไปใช้เป็นตัวแปรตาม จะสามารถใช้สถิติแบบ Parametric ได้โดยไม่ผิดกติกา (ดูตาราง 1 Guilford ,1954 : 11-18)

อย่างไรก็ตาม นักสถิติวิจัยทางจิตวิทยารุ่นหลังๆ ยังมีความเห็นที่แตกต่างจากที่ Guilford นำเสนอตามตาราง 1 ดังกล่าว โดยแบ่งเป็น 3 ฝ่าย คือ ฝ่ายของ Guilford ที่เสนอแนะว่าเครื่องมือวัดที่ให้ข้อมูลอยู่ในมาตราต่างประเภทกัน เมื่อเป็นตัวแปรตามจะต้องใช้สถิติวิเคราะห์ที่เหมาะสมกับระดับการวัดดังที่กล่าวไปแล้ว ฝ่ายที่สองเห็นด้วยกับหลักการของฝ่ายแรก แต่แบ่งประเภทของมาตราต่างไปจากของ Guilford ส่วนฝ่ายที่สาม เห็นว่า มาตรวัด 4 ประเภท กับ

ตาราง 1 สถิติวิเคราะห์ที่ใช้การวัดเหมาะสมกับลักษณะตัวแปรตามที่อยู่ในมาตราต่างกัน

สถิติวิเคราะห์	การวัดในมาตร 4 ประเภท			
	Nominal Scale	Ordinal Scale	Interval Scale	Ratio Scale
Frequencies	✓	✓	✓	✓
Modes (ฐานนิยม)	✓	✓	✓	✓
Contingency correlation coeff.	✓	✓	✓	✓
Medians (มัธยฐาน)	-	✓	✓	✓
Centiles	-	✓	✓	✓
Rank – order coeff. of correl.	-	✓	✓	✓
Means (ค่าเฉลี่ย)	-	-	✓	✓
Standard deviation	-	-	✓	✓
Pearson's r	-	-	✓	✓
Coeff. of variation	-	-	-	✓
r^2	-	-	-	✓

(ตาม Guilford, 1954 : 11-18)

สถิติวิเคราะห์ที่ใช้ได้ไม่จำเป็น ต้องขึ้นต่อกันดังของสองฝ่ายแรก (Michell, 1986) ข้อขัดแย้งข้างบนนี้เกี่ยวข้องกับมาก โดยเฉพาะกับตัวแปรตามที่อยู่ในมาตร ordinal scale ในบางกรณี อาจใช้สถิติวิเคราะห์แบบ parametric ก็ได้ (Davison & Sharma, 1990) แต่เหตุการณ์ที่พบมากโดยเฉพาะในประเทศไทย คือ การใช้สถิติวิเคราะห์แบบ nonparametric กับข้อมูลที่จะอยู่ในระดับของ Interval scale ทำให้การวิเคราะห์นั้น พิจารณาลักษณะของข้อมูลตามการเรียงลำดับ ความมากน้อยแบบหน่วยที่มีช่วงไม่เท่า หรือพิจารณาเปรียบเทียบปริมาณว่า **มากกว่า เท่ากัน หรือน้อยกว่า** เท่านั้น มิได้ใช้ค่าที่เป็นค่าต่อเนื่อง เช่น จาก 10 ถึง 60 คะแนน เป็นต้น สถิติแบบ nonparametric จึงใช้แต่ลักษณะของข้อมูลของค่าจากตัวแปรตามที่หายา และไม่นำมาใช้ให้ครบถ้วนในหลายลักษณะเท่าสถิติแบบ parametric ซึ่งยังให้ความแตกต่าง และการกระจายของคะแนนที่มีค่าต่อเนื่อง และมีการพิจารณาค่าต่ำสุด ค่าเฉลี่ย และค่าสูงสุด ด้วยสถิติแบบ parametric จึงให้ข้อสรุปผลวิจัยที่ละเอียดและน่าเชื่อถือมากกว่าได้

คำว่า Interval Scale มักมีผู้เข้าใจว่า “เป็นช่วงเท่า” โดยตกคำว่า “หน่วย” บนมาตราในการวัดนั้นๆ ไป หน่วยในการวัด ถ้าเป็นการวัดเวลา อาจเป็นนาที หรือ ชั่วโมง หรือวัน หรือปี ทุกนาที่เท่ากัน คือมี 60 วินาที ทุกวันเท่ากันคือมี 24 ชั่วโมง ทุกคะแนนบนมาตราเท่ากันคือ แต่ละช่วงต่างกัน 1 คะแนน ไม่ว่า 3 กับ 2 หรือ 6 กับ 5 ความแตกต่างที่เท่ากับ 1 ในทั้ง 2 ช่วงนี้ เป็น 1 ที่มีค่าเท่ากัน ความเข้าใจเรื่อง Interval scale ถ้าต่างไปจากที่เขียนนี้ จะผิดทั้งหมด เช่น ที่เข้าใจผิดว่านักวิจัย แบ่งคะแนนเป็น 2 ช่วง คือ 1 ถึง 5 กับ 6 ถึง 10 เป็นการแสดงความเป็นช่วงเท่าของคะแนนที่นักวิจัยใช้ การที่นักวิจัยคนนี้ทำได้ดังกล่าว เพราะอาศัยความมีหน่วยต่างๆ ที่เท่ากัน บนมาตราในการวัดของเขาเป็นเบื้องต้นนั่นเอง

ความสำคัญในประเด็นนี้ คือ การสร้างมาตร การกำหนดคะแนนแต่ละ หน่วย การขีดเส้นยาวเท่ากัน ห่างเท่ากัน และมีคำอธิบายประกอบหน่วยใน ความหมายต่างเชิงปริมาณเท่าๆ กัน เท่านั้น ที่ทำให้มาตรวัดของท่านอยู่ในระดับ Interval scale ได้

ที่กล่าวมานี้คือ ข้อจำกัดของมาตรประเมินที่อาจเกิดได้ ถ้านักวิจัยไม่ จัดการให้มีลักษณะที่ใกล้เคียง Interval scale ให้มากที่สุด

คำว่า “รวมค่า” หมายถึง การใช้มาตรวัดหลายมาตรจากหลายข้อ แล้วนำ คะแนนของทุกมาตรของผู้ตอบคนหนึ่งๆ มารวมกันเป็นคะแนนรวมที่แสดงปริมาณ การมีลักษณะนั้นของผู้ตอบคนนั้น ขึ้นอยู่กับลักษณะที่ต้องการวัด มาตรประเมิน อาจใช้เพียงมาตรเดียว เช่น วัดคุณภาพชีวิตโดยรวมของบุคคล ซึ่งเป็นการรับรู้ สภาวะต่างๆ ที่น่าพอใจในชีวิตของตน หรือความรู้สึกพอใจในปัจจุบันโดยรวม (Leelakulthanit & Day, 1993) แต่การวัดดังกล่าว อาจใช้หลายมาตรหลายข้อวัด คุณภาพชีวิตด้านต่างๆ ให้ละเอียดครอบคลุมมากขึ้นเช่น ด้านกายภาพ ด้านจิต ด้านเศรษฐกิจ ด้านสังคม (Spilker, 1990 และภมรินทร์ เชาวน์จินดา, 2542) แล้ว นำมารวมกัน ก็ต้องใช้หลายข้อหลายมาตรในแต่ละด้าน เพื่อให้ครอบคลุม ความหมายของแต่ละด้านได้ครบถ้วน จึงอาจใช้วัดแบบที่มีมากถึง 20-30 ข้อ แล้วนำ คะแนนบนมาตรทั้งหมดที่ผู้ตอบเลือกตอบมารวมกัน เรียกมาตรประเมินรวมค่า (Summated ratings method) ซึ่งกำลังเป็นที่นิยมใช้กันมาก แต่การวัดด้วยมาตร เดียวมักมีค่าความน่าเชื่อถือ (reliability) ต่ำมาก ไม่เกิน (.25) ซึ่งทำให้ผลวิจัย บิดเบือนได้มากจนถึงขั้นไม่พบค่าทางสถิติที่สูงในการวิเคราะห์ข้อมูล เพราะเหตุนี้ ได้ (Schmidt & Hunter, 1996) ดังนั้นการวัดโดยใช้มาตรประเมินเพียงข้อเดียว จึงไม่เหมาะสม แต่ถ้าใช้ 15 ข้อรวมกัน ค่าความน่าเชื่อถืออาจสูงถึง .84 ได้

มาตรประเมินรวมค่า มีลักษณะที่สำคัญ 5 ประการ คือ

ประการแรก มาตรประเมินรวมค่า ต้องมีข้อและมาตรประกอบเป็นคู่ๆ อย่างน้อย 10 ข้อขึ้นไป ซึ่งจะเป็นตัวแทนที่ดีของข้อต่างๆ มากมายที่รวมอยู่ในแดน (domain) หรือในจักรวาล (universe) ของข้อความทั้งหลาย แดน หรือจักรวาล กำหนดจากความหมายของลักษณะที่จะวัด เช่น วัดทัศนคติ มี 3 องค์ประกอบ รวมเป็นจักรวาล มี 3 แดนย่อย การนำประโยคหรือข้อความมาจากแต่ละแดนเป็นจำนวนมากเท่าไรก็จะเป็นตัวแทนที่ดีของการวัดแดนนั้นได้มากขึ้นทุกข้อที่นำมาใช้มีค่าเท่ากันหมดตามวิธีของลิเชิต (แต่ตามวิธีของเทอร์สโตน แต่ละประโยคมีค่าต่างกันตามที่นักวิจัยได้หาไว้ก่อนแล้ว) การใช้ข้อหรือประโยคต่างๆ ที่แตกต่างกันแต่มีค่าเท่าเทียมกัน จึงทำให้การแบ่งจำนวนข้อที่มีมากเป็นแบบวัดคูขนาน 2 แบบวัด ทำได้โดยไม่มีข้อจำกัด

ประการที่สอง มาตรประเมินรวมค่า เหมาะที่จะให้ผู้ตอบแสดงปริมาณการยอมรับแต่ละข้อในปริมาณต่างๆ แต่ละข้อถ้ามีมากหน่วยบนมาตรประกอบ เช่น 6 หรือ 8 หน่วย ย่อมทำให้เกิดค่าที่ตอบแตกต่างกันได้กว้างกว่าการใช้มาตรที่มี 4 หรือ 5 หน่วย เมื่อใช้หลายข้อหลายมาตร แล้วนำคะแนนคำตอบมารวมกัน จะยิ่งวัดได้ละเอียดแตกต่างกันได้มากยิ่งขึ้น ทำให้เป็นแบบวัดที่ไวต่อการพบความแตกต่าง หรือพบความเปลี่ยนแปลงของคะแนนแม้เพียงปริมาณเล็กน้อย ซึ่งเป็นข้อดีของแบบวัดชนิดมาตรประเมินรวมค่านี้ (Kerlinger & Lee, 2000 หน้า 739)

ประการที่สาม มาตรประเมินรวมค่านี้ นักวิจัยด้วยการเรียนรู้และฝึกทักษะมาก่อน จะสามารถสร้างแบบวัดชนิดนี้ได้อย่างมีมาตรฐานสูง มีลักษณะของเครื่องมือวัดที่สร้างตามหลักวิทยาศาสตร์หลายหลัก เช่น หลักการมีนิยามปฏิบัติการ หลักการควบคุมขณะวัดและการตีความคะแนนจากการวัด นอกจากนี้ แบบวัดที่ใช้วัดตัวแปรหนึ่งๆ ยังสามารถสร้างให้มีคุณสมบัติสำคัญของเครื่องมือวัด คือ มีค่าจำแนกรายข้อ มีความเที่ยงตรง และความคงที่ได้โดย

นักวิจัยอาจใช้เวลา และงบประมาณที่ไม่มากเท่าการสร้างเครื่องมือชนิดอื่นๆ (Spector, 1992 หน้า 2) และเป็นแบบวัดที่สามารถสร้างเพื่อใช้วัดลักษณะต่างๆ ได้มากมาย (Kerlinger & Lee, 2000 หน้า 739) ตั้งแต่ทัศนคติ บุคลิกภาพ พฤติกรรม และการรับรู้เกี่ยวกับสภาวะแวดล้อมของผู้ตอบ เป็นต้น

ประการที่ดี เครื่องมือวัดแบบมาตรประเมินรวมค่า มีลักษณะที่น่าสนใจ ใช้ง่าย และผู้ตอบตอบสนองได้ง่าย โดยการขีดตอบที่หน่วยใดหน่วยหนึ่งบนมาตราที่ใช้ประกอบแต่ละประโยค ทำให้ใช้ได้จำนวนมากข้อในคราวเดียวกัน ในการวัดตัวแปรอิสระและตัวแปรตามได้หลายตัวในงานวิจัยหนึ่งๆ

ประการที่ห้า เครื่องมือวัดแบบมาตรประเมินรวมค่า ยังเหมาะสมที่จะใช้กับผู้ตอบที่มีความสามารถอ่านข้อเขียนได้คล่องแคล่ว และเข้าใจได้ง่ายด้วยตนเอง โดยไม่ต้องให้มีผู้มาอธิบายข้อความมากนัก ผู้ใช้มาตร 6 หน่วยขึ้นไป ต้องเป็นผู้มีความสามารถคิดแยกแยะในเชิงปริมาณที่ละเอียด มิใช่ตอบได้อย่างหยาบๆ แต่เพียงว่า “จริง” กับ “ไม่จริง” แต่สามารถให้คำตอบว่า “จริง” ในเชิงปริมาณ ในระดับต่างๆ ได้ด้วย เช่น “ค่อนข้างจริง” “จริง” และ “จริงที่สุด” จะสังเกตได้ว่า ในโลกยุคปัจจุบัน ประชาชนมีการศึกษามากขึ้น ส่วนใหญ่สามารถคิดได้ในเชิงปริมาณในหลายระดับ แบบวัดชนิดมาตรประเมินรวมค่า จึงเหมาะสมกับคนรุ่นใหม่ในยุคปัจจุบัน และจะเหมาะสมมากขึ้นเรื่อยๆ ในการใช้วัดคนในอนาคต

ส่วนข้อเสียที่มาพร้อมกัน ก็คือ ผู้ตอบอาจมีลักษณะการตอบสนองพิเศษของตน (response set) เช่น ชอบตอบบนหน่วยที่อยู่กลางมาตร บางคนชอบตอบว่า “เห็นด้วย” เสมอ บางคนชอบตอบว่า “ไม่เห็นด้วย” นอกจากนี้ ยังมีผู้ระบุถึงข้อผิดพลาดอีกหลายประการของผู้ตอบในการใช้มาตรประเมินลักษณะของผู้อื่น โดยผู้ตอบอาจจะประเมินลักษณะอื่นๆ ตามลักษณะที่เด่นของคนนั้น (halo effect) เช่น ถ้ารู้สึกว่าเขาเป็นคนใจบุญ ก็จะมีแนวโน้มว่าเขาบริจาจาคเงิน

มากกว่าคนอื่น หรือประเมินลักษณะหนึ่งให้สอดคล้องกับลักษณะอื่นๆ ของคนๆ นั้น โดยไม่ได้พิจารณาแยก จากกันโดยเด็ดขาด เช่น บุคคลประเมินเพื่อนของเขาในข้อต้นๆ ว่า เพื่อนของเขาชอบไปเที่ยวสถาบันเทิงมาก ดังนั้นข้อต่อมา เขาจะรายงานว่ เพื่อนคนนี้ มักมีรายได้ไม่พอกับค่าใช้จ่าย ตามไปด้วย (logical error)

การใช้มาตรประเมิน ไม่ว่าผู้ตอบจะประเมินผู้อื่น หรือประเมินตนเอง ก็อาจจะเกิดข้อผิดพลาดในรูปแบบต่างๆ ได้อีกมาก นอกจากที่กล่าวไปแล้ว เช่น ประเมินคนอื่นว่ามีลักษณะที่ตรงข้ามกับตน (contrast error) หรือ ขณะตอบมีการประเมินบนมาตรที่อยู่ใกล้เคียงกันไปในทำนองเดียวกันได้มากด้วย (proximity error) (Guildford, 1954 หน้า 278-280)

ข้อผิดพลาดทั้งหลายนี้อาจป้องกันได้อยู่บ้าง โดยการจัดการควบคุมวิธีการตอบในลักษณะต่างๆ และหรือฝึกอบรมผู้ตอบให้ระมัดระวังการเกิดข้อผิดพลาดเหล่านี้ ก่อนที่จะให้เขาตอบแบบวัด นอกจากนั้น ยังมีผู้แนะนำวิธีการวิเคราะห์ทางสถิติ ที่จะสามารถปรับแก้ข้อผิดพลาดที่เกิดขึ้นได้ แต่ก็มีผู้ที่นำวิธีการเหล่านี้ไปปฏิบัติไม่มากนัก จึงยังปรากฏข้อผิดพลาดให้ได้พบเห็นอยู่ตลอดมา (Saal และคณะ, 1980) ดังนั้น ผู้สร้างและใช้แบบวัดชนิดมาตรประเมินรวมค่า จึงควรให้ความสำคัญในการป้องกันการเกิดข้อผิดพลาดเหล่านี้ให้มากขึ้นต่อไปด้วย

หลักฐานที่แสดงว่า เครื่องมือวัดแบบมาตรประเมินรวมค่า เป็นที่นิยมในหมู่นักวิชาการสาขามากกว่าวิธีวัดแบบอื่นๆ นั้น สังเกตได้จากการเปลี่ยนวิธีวัดจิตลักษณะที่สำคัญต่างๆ ตัวอย่างเช่น การวัดความสามารถในการใช้เหตุผลเชิงจริยธรรม ตามทฤษฎีของโคลเบอร์ก (Kohlberg, 1964, 1976) ซึ่งเริ่มต้นโดยใช้แบบวัดชนิดปลายเปิด (open – ended questions) ให้ผู้ตอบเขียนตอบให้มากที่สุด และใช้ 5 ถึง 7 เรื่อง ซึ่งไม่สะดวกในการใช้วัด เพราะเสียเวลามากในการตอบ และต้องนำ

คำตอบมาวิเคราะห์เนื้อหา (content analysis) เพื่อให้คะแนนว่าผู้ตอบอยู่ชั้นใดใน 6 ชั้น ต่อมาผู้สร้างแบบวัดแบบมีตัวเลือกให้เลือกตอบตามชั้นของโคลเบอร์ก 6 ชั้น ในแต่ละเรื่อง (multiple – choice questions) ซึ่งอาจใช้ถึง 10 เรื่อง นำชั้นของคำตอบที่ผู้ตอบเลือกใน 10 ข้อ มารวมแล้วหาค่าเฉลี่ยเป็นชั้นเหตุผลเชิงจริยธรรมของผู้ตอบ (ดวงเดือน พันธุมนาวิน และเพ็ญแข ประจวบปัจฉิก, 2520) ซึ่งใช้ตอบและคิดคะแนนได้สะดวกมากขึ้น ส่วนในต่างประเทศมีแบบวัดทำนองเดียวกันนี้ที่ใช้ถึง 12 ตัวเลือก ต่อมา มีผู้ใช้วิธีให้ผู้ตอบเรียงลำดับเหตุผล (Ranking) ที่ตนยอมรับได้ทั้งหมดที่มีให้ แล้วหาเปอร์เซ็นต์ของจำนวนเหตุผลชั้น 5 และ 6 จากข้อทั้งหมดที่ผู้ตอบคนนั้นเลือกไว้ เพื่อวัดปริมาณการใช้เหตุผลเชิงจริยธรรมในชั้น 5 และ 6 (%P) (Rest และคณะ, 1997) แต่ก็ยังเป็นวิธีที่ยุ่งยากและเรื่องที่ใช้ไม่เหมาะกับคนไทย ต่อมาโกศล มีคุณ (Meekun, 1999) จึงใช้เรื่องแบบไทยๆ แล้วให้เหตุผลชั้น 5 และชั้น 6 อย่างละเจ็ด เพื่อให้ผู้ตอบประเมินบนมาตรา 6 หน่วยประกอบ จาก “เห็นด้วยอย่างยิ่ง” ถึง “ไม่เห็นด้วยเลย” ให้ผู้ตอบเลือก แล้วเอาคะแนนที่ตอบคุณด้วยชั้นของประโยคนั้น แล้วบวกกับคะแนนที่คุณแล้วของอีกประโยคหนึ่งในเรื่องเดียวกัน มีเรื่องขัดแย้งระหว่างผลประโยชน์ของฝ่ายต่างๆ ที่ใช้ 10 เรื่อง ใช้คะแนนรวมทั้งหมด เป็นแบบวัดชนิดมาตราประเมินรวมค่าที่ใช้ได้ดีในวิจัยต่างๆ มาจนถึงปัจจุบัน (ดุจเดือน พันธุมนาวิน, 2550)

อาจกล่าวได้ว่า ความนิยมใช้แบบวัดชนิดมาตราประเมินรวมค่าทางจิตวิทยา และพฤติกรรมศาสตร์สาขาอื่นๆ นั้น ทำให้วิชาการทางด้านนี้เจริญขึ้นมากอย่างรวดเร็ว ที่เป็นดังนี้เพราะเครื่องมือวัดที่ใช้ในการวิจัยในแนวนี้นี้คือ กระดาษ และดินสอ เท่านั้น ไม่ต้องใช้เครื่องมือทางวิทยาศาสตร์ที่มีราคาแพง ซึ่งจะต้องเพิ่มทุนในการทำวิจัยขึ้นอีกมาก และอาจทำให้นักวิจัยในประเทศที่มีทุนน้อย เกิดความล้าหลัง และทิ้งห่างจากประเทศที่ร่ำรวยและทุ่มทุนให้แก่การวิจัย

ได้มาก จึงต้องขอบคุณการวัดชนิดมาตราประเมินที่ทำให้งานวิจัยที่ใช้การวัดประเภทนี้ในประเทศไทย มีความเจริญก้าวหน้ากว่าวิชาการสายอื่นๆ ด้วย

อย่างไรก็ตาม การใช้วิธีวัดด้วยกระดาษ ดินสอ แบบมาตราประเมินรวมค่านี้ จะมีความเคร่งครัดเหมาะสม ตามหลักและวิธีการทางวิทยาศาสตร์มากเพียงใด เพื่อป้องกันข้อผิดพลาดต่างๆ ให้มาก ย่อมขึ้นอยู่กับความพยายามในการสร้าง และการใช้แบบวัดชนิดนี้ของผู้วิจัย ที่จะทำให้ผลวิจัยของตนเป็นที่น่าเชื่อถือว่าเป็นความจริงมากตามไปด้วย

การวัดแบบให้ผู้ตอบรายงานตนเอง (Self report)

การให้ผู้ตอบรายงานเกี่ยวกับตนเองในเรื่องใดเรื่องหนึ่งนั้น (self report) ในวิชาการทางจิตวิทยามีการใช้ควบคู่กับการใช้มาตราประเมิน (rating scale) มาตั้งแต่เริ่มต้นของการวัดและการวิจัยเชิงปริมาณจึงเรียกว่าวิธี self rating แต่เดิมก่อน ค.ศ. 1950-1960 นั้น การศึกษาวิจัยมนุษย์ในวิชาการต่างๆ ใช้วิธีการเชิงคุณภาพทั้งสิ้น เช่น ให้เขียนบรรยายเกี่ยวกับตนเอง การตอบข้อสัมภาษณ์ หรือการใช้ข้อชี้วัดประวัติของบุคคล เป็นต้น ต่อมานักวิจัยเห็นว่า การวัดเชิงคุณภาพ ไม่สามารถให้รายละเอียดของลักษณะต่างๆ ในเชิงปริมาณในคุณภาพนั้นๆ ได้ เช่น การวัดเชิงคุณภาพ ให้ข้อมูลว่า คนนี้ฉลาด คนนั้นโง่ แต่การจะให้ทราบว่าใครฉลาดมากหรือน้อยเพียงใด หรือใครฉลาดมากกว่าใคร หรือใครฉลาดมากขึ้นกว่าเดิมเพียงใด ต้องอาศัยการวัดออกมาในเชิงปริมาณ เป็นคะแนน เช่น เครื่องมือวัดสติปัญญาเด็ก IQ เท่ากับ 100 แสดงว่า ฉลาดเป็นปกติตามอายุ แต่ถ้า IQ เกิน 100 แสดงว่า ฉลาดมากขึ้น เป็นลำดับ ถ้า IQ ต่ำกว่า 100 แสดงว่า ฉลาดน้อยลงเป็นลำดับ ทำให้สามารถเปรียบเทียบปริมาณความฉลาดกับผู้อื่น หรือกับตนเองในอดีตได้ การวัด IQ จึงเป็นจุดเริ่มต้นของการยอมรับการวัดทางจิตในเชิงปริมาณจากวิชาจิตวิทยาอย่างกว้างขวาง ตั้งแต่ ค.ศ. 1900 เป็นต้นมา

ส่วนการวัดทัศนคติของบุคคลด้วยมาตราประเมินรวมค่า เกิดตามมาภายหลัง และให้ผลการวัดในเชิงปริมาณ สร้างง่าย และใช้สะดวก มีการแทรกแซงให้เกิดข้อผิดพลาดโดยผู้ทำการวัดและสถานการณ์ในการวัดได้น้อยกว่าวิธีวัดแบบอื่นๆ เช่น การสังเกต การสัมภาษณ์ เป็นต้น ในขณะที่เดียวกัน แบบวัดทัศนคติของบุคคล ก็เป็นการให้บุคคลตอบเกี่ยวกับตนเอง ว่าตนเห็นประโยชน์เห็นโทษ ชอบ ไม่ชอบ พร้อมที่จะกระทำ หรือพร้อมที่จะหลีกเลี่ยงสิ่งหนึ่งๆ มากเพียงใด ดังนั้น แบบวัดทัศนคติด้วยวิธีของลิเชิต จึงเป็นแบบวัดชนิดมาตราประเมินรวมค่า และให้รายงานตนเองไปพร้อมกัน ซึ่งเป็นแบบวัดชนิดที่ใช้โอกาส และประโยชน์เมื่อสิ่งที่กำลังถูกวัดและวิจัยเป็นมนุษย์ เท่านั้น

การวัดแบบให้รายงานตนเอง (self report) บนมาตราประเมินเป็นที่นิยมมากตลอดมาจนถึงปัจจุบัน ทั้งนี้เพราะวิธีแบบ self rating นั้นสะดวก เป็นระบบ และผู้ตอบไม่ต้องให้ชื่อ หรือข้อมูลอื่นที่จะต้องใช้ในการหาข้อมูลในหน่วยงาน โรงเรียน หรือที่อำเภอ เพิ่มเติมในภายหลัง

ความสะดวกของการให้รายงานตนเองนั้น คือ ความเรียบง่ายของการถามให้ได้ข้อมูลกว้างขวาง ทั้งอดีต และปัจจุบันในเรื่องหนึ่งๆ ส่วนความเป็นระบบ หมายถึง การถามแบบเดียวกัน กับผู้ตอบทุกคนที่อาจมีจำนวนเป็นร้อย หรือเป็นพัน ในการวิจัยหนึ่งๆ ที่มาจากหลายที่ หลายหน่วยงาน การใช้ข้อมูลจากต่างแหล่ง ต่างหน่วยงาน แม้จะเป็นเรื่องเดียวกัน ก็อาจมีการบันทึก การประเมินที่ต่างกันไปได้มาก ดังนั้นนักวิจัยจึงต้องสร้างระบบใหม่ของตนเอง โดยใช้วิธี self report ดังกล่าวนอกจากนี้ การถามให้ผู้กระทำตอบเอง ยังเป็นการรักษาความลับ ความเป็นส่วนตัว และบางครั้งวัดด้วยการได้ข้อมูลจากแหล่งอื่นๆ จะยุ่งยากจนไม่คุ้มค่า เพราะไม่มีใครรู้เรื่องของใครได้ละเอียดครบถ้วน ทั้งในอดีต และปัจจุบัน ในหลายแง่มุม หลายสถานที่ มากเท่ากับตัวผู้กระทำ ทั้งที่เกี่ยวกับการที่ตนได้ทำอะไรแก่ใครบ้างเมื่อใด และได้รับอะไรจากใคร เมื่อใด ตัวอย่างเช่น การให้ผู้ชราอายุกว่า

65 ปี ขึ้นไป รายงานว่าตนเคยถูกทำร้ายในด้านต่างๆ มากเพียงใด เช่น การถูกปล่อยปละละเลย การถูกเอาประโยชน์ทางการเงิน การถูกทารุณทางเพศ และการถูกทารุณทางกาย เป็นต้น (Buri และคณะ, 2006) จากผู้ตอบจำนวนกว่าพันคน 21% ตอบว่าเคยถูกทำทารุณมาก่อน การใช้แบบสอบถามให้รายงานตนเองนี้ ผู้วิจัยพบว่าให้ผลเป็นที่น่าพอใจ และเสนอแนะให้มีการเปรียบเทียบกับ การวัดการถูกทารุณของผู้ตอบวัยชราด้วยวิธีอื่นๆ ที่มีใช้มาก่อน เช่น การรายงานของผู้รู้เห็น ซึ่งเป็นบุคคลที่สาม หรือจากข้อมูลด้านการไปรับการตรวจรักษา เป็นต้น

การเปรียบเทียบกับวิธีวัดอื่น

การวัดแบบให้ผู้ตอบรายงานเกี่ยวกับพฤติกรรม หรือจิตลักษณะ หรือสถานการณ์ของตนนั้น เป็นทางเลือกหนึ่งในการวัดตัวแปรดังกล่าว แต่วิธีการวัดที่อาจนำมาใช้แทน หรือเปรียบเทียบกับวิธีให้รายงานตนเอง (self report) มีอีกหลายวิธี เช่น ให้คนอื่นที่เป็นผู้ใกล้ชิดและรู้เห็นรายงานแทนบุคคลๆ นั้น เช่น คนในครอบครัว หรือเพื่อน รายงานแทนผู้ชราในเรื่องการถูกทำทารุณที่กล่าวไปแล้ว ส่วนการวัดพฤติกรรมการทำงาน นอกจากผู้ปฏิบัติงานจะประเมินตนเองแล้ว วิธี 360 องศา ของข้อมูลย้อนกลับ แก่หัวหน้า หรือผู้จัดการ ในองค์การต่างๆ ยังต้องมีการประเมินหัวหน้า โดยให้ฝ่ายต่างๆ กระทำไปพร้อมกัน คือ ฝ่ายลูกน้อง ฝ่ายหัวหน้าใหญ่ ฝ่ายเพื่อนหัวหน้าด้วยกัน ฝ่ายลูกค้า และฝ่ายผู้ผลิตหรือผู้ขาย (London & Beatty, 1993) จึงจะให้ภาพการทำงานและผลงานของหัวหน้าองค์การได้ครบถ้วน และชัดเจนน่าเชื่อถือ แต่อย่างไรก็ตาม อีก 10 ปีต่อมา ก็ยังมีผู้ยืนยันว่า ผู้ปฏิบัติควรเป็นผู้ประเมินหลัก ทั้งนี้เพราะการเน้นให้ “คนอื่น” ประเมินผู้ปฏิบัติงานจะทำให้ผู้ปฏิบัติขาดความรู้สึกรับผิดชอบต่อการปฏิบัติงานของตน และจะลดคุณภาพของกระบวนการทำงานและผลผลิตของงานนั้นได้มาก (Vaughan, ,2003)

ในการวัดพฤติกรรมบางประการ การให้ผู้ปฏิบัติรายงานตนเอง กลับเป็นที่นิยมในหมู่นักวิจัยมากกว่าการตรวจจากเครื่องบันทึก เช่นที่เกี่ยวกับพฤติกรรม การขาดงาน หรือการลางาน ผู้ถูกวิจัยอาจไม่ยินยอมให้นักวิจัยใช้จดหมายรายงานของตนเป็นข้อมูล เนื่องจากอาจละเมิดสิทธิส่วนบุคคล เรื่องความลับทางด้านโรคภัยไข้เจ็บ ส่วนการใช้ข้อมูลจากจำนวนวันลา ความถี่ และระยะห่างของการลา ย่อมไม่ให้รายละเอียดเกี่ยวกับสาเหตุที่แท้จริงของการลางานได้ เช่น ในราชการไทยมีการลากิจ ลาป่วย และลาพักร้อน ถ้าผู้วิจัยต้องการจำนวนการลากิจเพื่อดูแลญาติผู้ใหญ่ที่ป่วยเรื้อรัง อาจต้องให้ผู้ตอบรายงานโดยตนเอง (self report) จึงจะได้ข้อมูลประเภทที่ต้องการ อย่างไรก็ตามนักวิชาการแนะนำว่า ถ้าใช้การรายงานโดยตนเองแล้ว ควรใช้ข้อมูลเกี่ยวกับเรื่องเดียวกันนี้จากแหล่งอื่นมาเปรียบเทียบ หรือหาความสัมพันธ์ จะทำให้เกิดความมั่นใจในข้อมูลจากการวิจัยนั้นได้มากขึ้น ด้วยวิธีวัดลักษณะหนึ่งในหลายแง่มุม โดยใช้หลายวิธีการ (Multitrait multimethod technique) (Campbell & Fiske, 1959 และดวงเดือน พันธุมนาวิน, 2519)

ผลวิจัยเกี่ยวกับการอบรมเลี้ยงดูเด็ก

นอกจากการให้ผู้ตอบรายงานเกี่ยวกับจิตและพฤติกรรมของตนเองแล้ว ในการวิจัยไทยทางจิตพฤติกรรมศาสตร์ (ดวงเดือน พันธุมนาวิน และเพ็ญแข ประจักษ์จณี, 2520) ยังให้เด็กอายุตั้งแต่ 12 ปี (ชั้นประถมศึกษาปีที่ 4) เป็นต้นมา รายงานเกี่ยวกับการถูกอบรมเลี้ยงดูในครอบครัว การให้บุคคลรายงานตนเองนี้ ในต่างประเทศมีหลักฐานว่าใช้ได้กับเด็กอายุตั้งแต่ 8 ปี ขึ้นไป (Walden และคณะ, 2003)

การวิจัยที่ศึกษาการอบรมเลี้ยงดูของครอบครัวที่เกี่ยวข้องกับลักษณะทางจิต และพฤติกรรมของเด็ก ที่มักพบผลสำคัญคือ การวัดจากรายงานของเด็กเพียง

ฝ่ายเดียว โดยให้เด็กรายงานว่าได้รับการอบรมเลี้ยงดูแบบรักสนับสนุนมากเพียงใด และแบบใช้เหตุผลมากเพียงใด จากบิดามารดาของตน เด็กที่รายงานว่าได้รับการอบรมเลี้ยงดูสองแบบนี้มาก มักเป็นเด็กที่มีจิต และพฤติกรรมที่น่าปรารถนามากตามไปด้วย ผลนี้พบมากในเด็กเพศชาย และเด็กจากครอบครัวฐานะค่อนข้างต่ำในประเทศไทย ตลอดช่วงเวลา 25 ปีมานี้ (ดวงเดือน พันธุมนาวิน และเพ็ญแข ประจวบจันติก 2520, 2524 ดวงเดือน พันธุมนาวิน งามตา วนินทานนท์ และคณะ, 2536 และงามตา วนินทานนท์, 2548) ซึ่งแสดงถึงความคงที่ และความน่าเชื่อถือ ของผลวิจัยเหล่านี้ในประเทศไทย

ผลวิจัยที่เก็บข้อมูลจากฝ่ายบุตรเพียงแหล่งเดียว และพบผลที่ชัดเจนดังกล่าว ยังปรากฏมาก่อนในการวิจัยการถูกอบรมเลี้ยงดูกับการปรับตัวของเด็กอเมริกัน (Serot & Teevan, 1961) และต่อมาพบในการวิจัยครอบครัวกับความรู้สึกสุขใจในเด็กชาวจีนจากฮ่องกงด้วย (Shek, 1997)

ส่วนการที่พบผลวิจัยว่า การรายงานตนเองของบิดามารดา เกี่ยวกับการอบรมเลี้ยงดูบุตร มีความสอดคล้องน้อยกับการรับรู้การถูกอบรมเลี้ยงดูของบุตรเอง และเกี่ยวข้องไม่มากนักกับลักษณะทางจิตและพฤติกรรมของบุตร ก็ปรากฏมากในการวิจัยต่างๆ ในต่างประเทศ ซึ่งก็อาจจะอธิบายได้จากการที่พบข้อมูลจากการรายงานตนเองว่าอบรมเลี้ยงดูบุตรแบบรักใกล้ชิดกับการปฏิบัติจริงต่อบุตรในการเล่นเกมนร่วมกันในห้องปฏิบัติการ สังเกตและประเมินโดยนักวิจัย ว่ามีความสอดคล้องกันน้อย และข้อมูลสองประเภทนี้ก็ยังสอดคล้องกับรายงานของบุตรเกี่ยวกับการปฏิบัติทางบวกของบิดามารดาต่อบุตรไม่มากนัก ความสอดคล้องระหว่างรายงานตนเองของบิดามารดา กับรายงานของบุตรในการรับรู้การปฏิบัติของบิดามารดาจะสอดคล้องกันมาก ในกรณีที่บิดามารดาที่รายงานว่าตนรักใกล้ชิดบุตรมาก ในขณะที่เดียวกันก็เป็นผู้มีบุคลิกภาพแบบชอบเข้าสังคมต่ำ (less

extravert) ที่บุตรจะรายงานว่ามีบิดามารดาของตนมีพฤติกรรมทางบวกมาก และสร้างบรรยากาศที่ดีในครอบครัวมากด้วย (Metsapelto & Fulkkinen, 2005)

สรุปได้ว่า การรายงานตนเองทางด้านการอบรมเลี้ยงดูบุตรของบิดามารดา กับพฤติกรรมจริง และตามการรับรู้ของบุตร มีความสอดคล้องกันไม่มากนัก เช่นเดียวกับการประเมินการทำงานโดยตนเอง กับให้หลายฝ่ายประเมินแบบ 360 องศา ก็ไม่ประสบความสำเร็จ ส่วนหัวหน้าก็มักประเมินลูกน้องได้ยาก ไม่ว่าจะเป็นงานในสถานีดำรวจภูธร งานนักเรียนนายร้อยตำรวจไปฝึกปฏิบัติงานกับครอบครัวชนบท หรืองานสอนนักเรียนมัธยมของครูคณิตศาสตร์ จนกระทั่งต้องกลับมาฟังการประเมินตนเอง (self rating) ของผู้กระทำเป็นสำคัญ (Vanghan, 2003)

การควบคุมโดยใช้ Social Desirability Scale

การใช้วิธีการให้รายงานตนเองในบางเรื่อง บางพฤติกรรม อาจมีข้อผิดพลาดได้ ซึ่งอาจเกิดจากสาเหตุหลายประการ เช่น ด้านเนื้อหาของตัวแปรที่ให้รายงานตนเอง เช่น ลักษณะและปริมาณการลาหยุดงานที่กล่าวมานี้ หรือเกิดจากการวัดตัวแปรเชิงเหตุ หรือตัวแปรอิสระ ด้วยวิธีให้รายงานโดยตนเองอีกเช่นเดียวกันจากการประมวลและวิเคราะห์วิธีวิจัย และผลวิจัย มีผู้พบว่า (Johns, 1994) การรายงานตนเองอาจผิดพลาด บิดเบือน จากความจำและความเข้าใจความหมายของสิ่งที่ถาม ของผู้ตอบ นอกจากนั้นอาจพบความสัมพันธ์ของสิ่งที่เป็นสาเหตุ กับสิ่งที่เป็นผลในเรื่องนั้นได้มากเกินไปจนจริงได้ถ้าให้ตอบในเวลาใกล้เคียงกัน และต่อเนื่องกัน อาจทำให้ผู้ตอบ ตอบในลักษณะยอมรับ หรือปฏิเสธเหมือนกัน หรือตอบสิ่งที่ตนเห็นว่าเป็นคำถามเกี่ยวกับสาเหตุ กับพฤติกรรมเพื่อให้เกิดความสอดคล้องกันมากกว่าที่เป็นจริง หรืออาจเกิดจากการตอบเพื่อให้เป็นที่น่านิยมในสังคม (Social desirability) ทั้งที่เกี่ยวกับการวัดด้านสาเหตุ และด้านผล จึงมีการใช้แบบ

วัดการตอบตามความนิยมของสังคม (Social Desirability Scale) มาใช้เพื่อการควบคุมข้อมูลจากการวัดแบบ self rating อีกทีหนึ่ง

แบบวัดที่เรียกว่า Social Desirability Scale (SDS) ใช้วัดปริมาณการมีความโน้มเอียงของบุคคลที่จะตอบแบบวัดต่างๆ ให้สอดคล้องกับปทัสฐานของสังคมของตนในปริมาณต่างๆ จึงทำให้คะแนนการตอบแบบวัดต่างๆ มีความบิดเบือนไปได้ แบบวัด SDS ดังกล่าวนี้นี้ มีลักษณะเป็นการวัดบุคลิกภาพแบบหนึ่งซึ่งเสนอโดย (Crowne และ Marlowe, 1960 อ้างจาก Gravdal & Sandal, 2006) มี 33 ข้อ แต่อาจใช้แบบ 10 ข้อ มีคำตอบให้เลือกว่า “ใช่” และ “ไม่ใช่” ในแต่ละข้อ ใช้คะแนนรวม 10 ข้อ พิสัยของคะแนนระหว่าง 0 ถึง 10 แบบวัด SDS นี้ผู้สร้างคาดว่า มี 2 องค์ประกอบ คือ การจัดการกับความรับรู้ลึกของผู้อื่น (impression management) และการหลอกตนเอง (self – deception) ทั้งที่อยู่ ในสามัญสำนึก และที่อยู่ใต้สามัญสำนึกของบุคคลด้วย แต่จากการทำการวิเคราะห์ องค์ประกอบใหม่ กลับพบว่าข้อต่างๆ ในแบบวัด SDS ตกอยู่ในองค์ประกอบที่ เรียกว่า การหลอกคนอื่น (Other deception)

แบบวัด SDS ได้มีการนำมาใช้เพื่อตรวจสอบและควบคุมผลการตอบแบบวัดอื่นๆ ประเภทที่ให้ผู้ตอบรายงานตนเองอยู่ตลอดมา เช่น ในการวิจัยที่พัฒนา เครื่องมือวัด ความกลัวความสนิทใกล้ชิด (Fear – of – intimacy Scale, Descutner & Thelen, 1991) และอื่นๆ อีกมาก

นอกจากนั้น ยังมีการใช้ SDS เพื่อจำแนกคนประเภทที่ตอบเอาใจสังคม ออก หรือนำคะแนนจากแบบวัดนี้มาหาความสัมพันธ์กับคะแนนจากแบบวัดตัวแปรหลักในการวิจัยนั้นๆ เพื่อศึกษาว่าปัญหาการตอบตามความน่านิยมในสังคมมีปรากฏในการวัดในการวิจัยนั้นๆ หรือไม่ (Furham และคณะ, 2002) หรือใช้คะแนนจากแบบวัดการตอบตามความน่านิยมในสังคมมาควบคุมอิทธิพลของลักษณะนี้ ต่อค่าของตัวแปรสำคัญในการวิจัย (Descutner & Thelen, 1991) ส่วนเมื่อไม่

นานมานี้มีผู้ทำการวิเคราะห์เพื่อหาองค์ประกอบของแบบวัด Social Desirability Scale (Gravdal & Sandal, 2006) ซึ่งจะช่วยให้ นักวิจัยตัดสินใจว่าควรใช้แบบวัดนี้ เพื่อควบคุมการตอบอย่างบิดเบือนของผู้ตอบหรือไม่

จึงอาจกล่าวได้ว่า การวัดโดยวิธีใช้มาตรประเมินรวมค่า และการให้รายงานตนเอง ซึ่งเมื่อใช้ร่วมกัน เรียกว่าวิธีการประเมินตนเอง (self ratings) นี้ ยังเป็นที่นิยมมากในปัจจุบัน วิธีการนี้เปิดโอกาสให้นักวิจัยที่มีความพยายามมาก สามารถสร้างและใช้วิธีวัดนี้ได้เหมาะสมตามไปด้วย จึงเป็นเครื่องมือวัดที่ท้าทายความพยายาม และความสามารถของนักวิจัยรุ่นใหม่ๆ ที่จะช่วยค้นหาแนวทางที่จะทำให้ข้อมูลจากการวัดและการวิจัยทางจิตพฤติกรรมศาสตร์ มีความน่าเชื่อถือให้มากที่สุด เพื่อความมั่นใจในการที่จะทำผลจากการวัดและการวิจัยไปใช้ในการพัฒนาบุคคลและสังคม ให้ได้ผลดีอย่างจริงจังและยั่งยืนต่อไป

บรรณานุกรม

- งามตา วรินทร์านนท์. 2548. การวิจัยเพื่อสร้างระบบดัชนีคุณภาพชีวิตสมรสในครอบครัวไทย. วารสารจิตพฤติกรรมศาสตร์ : ระบบพฤติกรรมไทย, (1), 38-42.
- ดวงเดือน พันธุมนาวิน. 2519. **หลักและวิธีวิจัยทางจิตวิทยาสังคม**. โรงพิมพ์องค์การค้ำค้ำสุสภา กรุงเทพมหานคร.
- ดวงเดือน พันธุมนาวิน และงามตา วรินทร์านนท์. 2536. **ลักษณะทางจิตและพฤติกรรมของนักเรียนวัยรุ่นที่อยู่ในสภาวะเสี่ยงในครอบครัวและทางป้องกัน**. รายงานการวิจัย สนับสนุนโดย สำนักงานส่งเสริมและประสานงานเยาวชนแห่งชาติ สำนักนายกรัฐมนตรี.

- ดวงเดือน พันธุมนาวิน และ บุญยิ่ง เจริญยิ่ง. 2517. **อิทธิพลของสังคมต่อทัศนคติต่อวัยรุ่น**. รายงานการวิจัย ฉบับที่ 18 สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.
- ดวงเดือน พันธุมนาวิน และ เพ็ญแข ประจันปัจฉนีก. 2520. **จริยธรรมของเยาวชนไทย**. รายงานการวิจัยฉบับที่ 21. สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.
- ดวงเดือน พันธุมนาวิน และ เพ็ญแขประจันปัจฉนีก. 2524. **ความสัมพันธ์ภายในครอบครัว สุขภาพจิต และจริยธรรมของนักเรียนวัยรุ่นไทย**. รายงานการวิจัยฉบับที่ 26. สถาบันวิจัยพฤติกรรมศาสตร์. มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.
- ดุจเดือน พันธุมนาวิน. 2550. **การสังเคราะห์งานวิจัยเกี่ยวกับคุณธรรมจริยธรรมในประเทศไทยและต่างประเทศ**. รายงานการวิจัยทุนอุดหนุนการวิจัย โดย ศูนย์ส่งเสริมและพัฒนาพลังแผ่นดินเชิงคุณธรรม สำนักงานบริหารและพัฒนาองค์ความรู้ (องค์การมหาชน).
- ภมรินทร์ เขาวนจินดา. 2542. **ลักษณะทางพุทธศาสนาที่เกี่ยวข้องกับคุณภาพชีวิตของผู้สูงอายุ**. ภาคนิพนธ์ ศิลปศาสตรมหาบัณฑิต คณะพัฒนาลัทธิ ศาสนา สถาบันบัณฑิตพัฒนบริหารศาสตร์.

- Buri, H., Daly, J.M., Hartz, A., & Jogerst, G. 2006. Factors associated with self-reported elder mistreatment in Iowa's frailest elders. *Research on Aging*, 25(5), 562-581.
- Campbell, D.T. & Fiske, D.W. 1959. Convergent and discriminant validation by multitrait - multimethod matrix. *Psychological Bulletin*, 56, 81-105.

- Davison, M.L., & Sharma, A.R. 1990. Parametric statistics and level of measurement factorial designs and multiple regression. **Psychological Bulletin**, 107(3), 394-400.
- Descutner, C.J., & Thelen, M.H. 1991. Development and validation of a fear-of-intimacy scale. **Psychological Assessment**, 3(2), 218-225.
- Furnham, A., Petrides, K.V., & Spence-Bowdage, S. 2002. The effects of different types of social desirability on the identification of repressors. **Personality and Individual Differences**, 33, 119-130.
- Gravdal, L., & Sandal, G.M. 2006. The two-factor model of social desirability: Relation to coping and defense, and implications for health. **Personality and Individual Differences**, 40, 1051-1061.
- Guilford, J.P. 1954. **Psychometric methods**. New York: McGraw-Hill.
- Johns, G. 1994. How often were you absent? A review of the use of self-reported absence data. **Journal of Applied Psychology**, 79(4), 574-591.
- Kerlinger, F.N., & Lee, H.B. 2000. **Foundations of behavioral research** (4th Ed.). Albert Complex, Singapore: Wadsworth.
- Kohlberg, L. 1964. Development of moral character and moral ideology. In M.L. Hoffman and I.W. Hoffman (Eds.), **Review of Child Development Research** (Vol. 1, pp. 83-432). Hartford Connecticut Printer.

- Kohlberg, L. 1976. Moral stages and moralization : The cognitive development approach. In T. Lickona (ed.) **Moral development and behavior : Theory, research and social issues**. New York : Holt, Rinehart and Wintson. pp. 31-53.
- Leelakulthanit, O., & Day, R. 1993. Cross-cultural comparisons of quality of life of Thais and Americans. **Social Indicator Research**, 30, 49-70.
- Likert, R. 1932. A technique for the measurement of attitudes. **Archives of Psychology**, 22 : No. 140.
- London, M., & Beatty, R.W. 1993. 360-degree feedback as a competitive advantage. **Human Resource Management**, 32(2-3), 353-372.
- Meekun, K.1999. **Summated ratings : Method for measuring moral reasoning ability**. Poster presentation at the 3rd Conference of the Asian Association of Social Psychology, 4-7 August, 1999, Academia Sinica, Taipei, Taiwan.
- Metsäpelto, R.L., & Pulkkinen, L. 2005. The moderating effect of extraversion on the relation between self-reported and observed parenting. **Applied Developmental Psychology**, 26, 371-384.
- Michell, J. 1986. Measurement scales and statistics: A clash of paradigms. **Psychological Bulletin**, 100(3), 398-407.
- Rest, J.R., Thoma, S.T., Narvaez, D., & Edwards, M.J. 1997. Alchemy and beyond: Indexing the defining issues test. **Journal of Educational Psychology**, 89(3), 498- 507.

- Saal, F.E., Downey, R.G., & Lahey, M.A. 1980. Rating the ratings: Assessing the psychometric quality of rating data. **Psychological Bulletin**, 88(2), 413-428.
- Schmidt, F.L., & Hunter, J.E. 1996. Measurement error in psychology research: Lessons from 26 research scenarios. **Psychological Methods**, 1(2), 199-223.
- Serot, N., & Teevan, R. 1961. Perception of the parent child relationship and its relation to child adjustment. **Child Development**, 32, 373-378.
- Shek, D.T.L. 1997. The relation of family functioning to adolescent psychological well being, school adjustment and problem behavior. **Journal of Genetic Psychology**, 158(4), 467-479.
- Spector, P.E. 1992. **Summated rating scale construction: An introduction**. London: Sage Publications.
- Spilker, B. 1990. "Introduction" in B. Spilker (Ed.). **Quality of life assessment in clinical trials** (pp. 3-9). New York: Raven Press
- Vaughan, S. 2003. Performance self as the principal evaluator. **Human Resource Development International**, 6(3), 371-385.
- Walden, T.A., Harris, V.S., & Cartron, T.F. 2003. How I feel: A self-report measure of emotional arousal and regulation for children. **Psychological Assessment**, 15(3), 399-412