



บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ในการกำหนดกรอบความคิดและแนวทางในการวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้องหลายรายการ คือ (1) แนวความคิดเกี่ยวกับมาตราประมาณค่า (2) สภาพการใช้มาตราประมาณค่าในงานวิจัย (3) แนวความคิดเกี่ยวกับมาตราประมาณค่าเชิงพฤติกรรม (Behaviorally-Anchored Rating Scales) (4) งานวิจัยที่มุ่งพัฒนาและศึกษาคุณภาพของมาตราประมาณค่า (5) แนวความคิดเกี่ยวกับการประเมินพฤติกรรมของบุคคล และ (6) พฤติกรรมการสอนและพฤติกรรมด้านการประเมินผลของครู ข้อสรุปจากการศึกษาแต่ละรายการปรากฏดังนี้

ตอนที่ 1 แนวความคิดเกี่ยวกับมาตราประมาณค่า

1.1 ความหมาย ลักษณะ และขอบเขตของการประยุกต์ใช้มาตราประมาณค่า

โดยปกติ ในการตัดสินคุณค่าของสิ่งใด ๆ ผู้ตัดสินสามารถลงข้อสรุปด้วยการบรรยายคุณค่าของสิ่งนั้น ๆ ออกมาในรูปของความเรียง ซึ่งผู้บรรยายแต่ละคนมีอิสระที่จะเขียนบรรยายออกมาในรูปแบบใดก็ได้ มีผลทำให้ข้อสรุปขาดแบบแผนที่ชัดเจนแน่นอน เพื่อแก้ไขปัญหาดังกล่าว นักวิจัยได้พยายามหาวิธีการในการตัดสินคุณค่าของสิ่งต่าง ๆ ให้เป็นไปในรูปแบบที่เป็นปรนัย สามารถตีความได้อย่างชัดเจนและเป็นที่ยอมรับกันมากยิ่งขึ้น การใช้มาตราประมาณค่า (Rating Scales) เป็นวิธีการอันหนึ่งที่ได้รับการพัฒนาขึ้นเพื่อจุดประสงค์ดังกล่าวนี้ เป็นวิธีการที่เน้นให้ผู้ประเมินตัดสินคุณค่าของสิ่งที่ต้องการลงบนสเกลหรือระดับคุณภาพที่ได้กำหนดความหมายไว้เป็นการล่วงหน้าแล้ว

มาตราประมาณค่า ถือเป็นมาตราการจัดอันดับ (Ordinal Scales) ที่ใช้ถ้อยคำหรือตัวเลขแสดงไว้อย่างเป็นลำดับขั้น (Saslow, 1982: 411) เป็นสิ่งที่จัดทำขึ้นเพื่อใช้ในการประเมินผลผลิต ทักษะ หรือบุคลิกลักษณะของบุคคล รูปแบบที่นิยมใช้กันเป็นแบบที่ได้กำหนดตำแหน่งต่าง ๆ ไว้ให้เลือก (Good, 1973: 469)

การประยุกต์ใช้มาตรฐานค่า สามารถประยุกต์ใช้ในกิจกรรมทางการศึกษาเพื่อประเมินผลการเรียนรู้และพัฒนาการต่าง ๆ ซึ่งสามารถจำแนกได้เป็น 3 กลุ่มใหญ่ ๆ ดังนี้

1. ใช้ประเมินกระบวนการ (Process Evaluation) เช่น ใช้ประเมินทักษะในการพูด ทักษะการบรรยาย การทำงานเป็นกลุ่ม ทักษะเชิงดนตรี ทักษะการเล่นกีฬา เป็นต้น
2. ใช้ในการประเมินผลผลิต (Products Evaluation) เช่น ภาพวาด โคลงกลอน การทำรายงาน สิ่งประดิษฐ์ต่าง ๆ เป็นต้น
3. ใช้ในการประเมินพัฒนาการทางบุคลิกภาพและทางสังคม (Evaluating Personal-Social Development) เช่น ความสนใจ ทักษะคิด การมีส่วนร่วมในกิจกรรม ความเป็นผู้นำ ความคิดริเริ่มสร้างสรรค์ ความรับผิดชอบ ความซื่อสัตย์ มนุษยสัมพันธ์ ความมั่นคงทางอารมณ์ การปรับตัว เป็นต้น

Tuckman (1975: 199-201) ได้เสนอตัวอย่างการใช้มาตรฐานค่าในการประเมินพฤติกรรมด้านการเรียนการสอน ดังนี้

1. ใช้ในการประเมินพฤติกรรมในชั้นเรียน ซึ่งถือเป็นพื้นฐานในการประเมินผลนักเรียน เพื่อให้ระดับผลการเรียนแก่นักเรียนในสิ่งที่เขารู้ หรือวิธีการที่เขาปฏิบัติ
2. ใช้ในการเก็บรวบรวมข้อมูลเกี่ยวกับพฤติกรรมเฉพาะอย่างเพื่อวางแผนพัฒนาหรือปรับปรุงพฤติกรรมที่ไม่ต้องการ
3. ใช้ในการวินิจฉัยเพื่อหาข้อมูลสนับสนุนปัญหาทางพฤติกรรมของเด็ก
4. ใช้ในการประเมินครูและโปรแกรมการสอน เช่น ประเมินพฤติกรรมการใช้คำถามของครู พฤติกรรมการนำเข้าสู่บทเรียน เป็นต้น

โดยสรุป กล่าวได้ว่า มาตรฐานค่าเป็นเครื่องมือที่ได้รับการพัฒนาขึ้นเพื่อประโยชน์ในการแปลความหมายของข้อมูลในเชิงคุณภาพออกมาเป็นปริมาณ มุ่งหวังให้เกิดความเป็นปรนัยในการแปลความหมายของข้อมูลด้วยวิธีการกำหนดมาตรฐานของรายการที่ต้องการประเมินให้อยู่ในรูปแบบของทางเลือกต่าง ๆ ที่เรียงเอาไว้อย่างเป็นลำดับขั้น เป็นเครื่องมือที่สามารถประยุกต์ใช้ในกิจกรรมการประเมินต่าง ๆ ได้อย่างหลากหลาย

1.2 รูปแบบของมาตรประมาณค่า

นักวัดผลหลายท่านได้พยายามวิเคราะห์และจำแนกรูปแบบของมาตรประมาณค่า โดยพิจารณาจากลักษณะเด่นชัดที่ปรากฏให้เห็น หรือพิจารณาจากลักษณะการตรวจให้คะแนน แล้วสรุปเป็นรูปแบบต่าง ๆ ดังต่อไปนี้

Guilford ได้แบ่งประเภทของมาตรประมาณค่าที่นิยมใช้ออกเป็น 5 แบบ คือ

1. มาตรเชิงตัวเลข (Numerical Scales) กำหนดตัวเลขพร้อมกับนิยามหรือคำบรรยายไว้ให้ผู้ประเมิน ดังตัวอย่างมาตรวัดที่ใช้ในการประมาณค่าของสีและกลิ่นต่อไปนี้

- 10 Most pleasant imaginable
- 9 Most pleasant
- 8 Extremely pleasant
- 7 Moderately pleasant
- 6 Mildly pleasant
- 5 Indifferent
- 4 Mildly unpleasant
- 3 Moderately unpleasant
- 2 Most unpleasant
- 1 Most unpleasant imaginable

ในบางครั้งอาจไม่มีตัวเลขให้ มีแต่เพียงข้อความ (Descriptive Cues) ดังนี้

Very Heavy

Heavy

Medium

Light

Very light

2. มาตรแบบกราฟิก (Graphic Scales) เป็นแบบที่นิยมใช้กันอย่างกว้างขวางมากที่สุด มีได้หลายรูปแบบ เน้นการจัดข้อความเข้ากับเส้นตรง ดังตัวอย่าง

ในวงสนทนาคุณเป็นอย่างไร — — — — —
 ช่างพูด พูดน้อย พูดเมื่อ ชอบฟัง ไม่เคย
 จำเป็น มากกว่า พูดเลย

เขาคิดได้ช้าหรือเร็ว _____

ช้ามาก ค่อย ๆ คิด ธรรมดา ว่องไว รวดเร็วมาก

3. มาตรแบบมาตรฐาน (Standard Scales) เป็นแบบที่ได้กำหนดค่าสเกล (Scales Value) เป็นค่าประจำของชุดมาตรฐานที่จะใช้ในการเปรียบเทียบไว้ล่วงหน้า ผู้ประเมินจะต้องทำการเปรียบเทียบรายการที่ให้ประเมินกับชุดมาตรฐาน

4. การประมาณค่าโดยใช้แต้มสะสม (Rating by Cumulative Points) แตกต่างจากแบบอื่นที่วิธีการให้คะแนน วิธีการตรวจรายการ (Check-list) เป็นตัวอย่างหนึ่งของรูปแบบนี้ได้มีผู้ใช้วิธีการตรวจรายการในการประเมินคุณลักษณะของเด็ก โดยกำหนดรายการที่มีทั้งคุณลักษณะที่ดีและไม่ดีมาให้เลือก เช่น ให้ความร่วมมือ ช่างคิด ดุร้าย เห็นแก่ตัว เป็นต้น ผู้ประเมินแต่ละคนจะเลือกทุก ๆ คุณลักษณะที่ตรงกับเด็กคนนี้ คะแนนของเด็กแต่ละคนจะเป็นผลรวมของคะแนนทั้งหมดซึ่งได้กำหนดให้ 1 สำหรับทุกลักษณะที่ดี และให้คะแนน -1 สำหรับทุกลักษณะที่ไม่ดี

5. การประมาณค่าแบบบังคับทางเลือก (Forced-Choice Rating) เป็นแบบที่พัฒนาขึ้นเพื่อประเมินบุคคล ผู้ประเมินไม่ต้องบอกว่าผู้ถูกประเมินมีลักษณะใด มากน้อยเพียงใด แต่ให้เปรียบเทียบเป็นคู่ ๆ แล้วบอกว่ามีลักษณะใดมากกว่ากัน

Turney และ Robb (1971 :136-137) มีความเห็นว่า ถึงแม้มาตรประมาณค่าที่นิยมใช้มีหลายรูปแบบ แต่แบบที่เป็นพื้นฐานนั้นมีอยู่ 3 แบบ คือ

1. มาตรเชิงตัวเลข (Numerical Scale) ใช้ตัวเลขซึ่งบ่งระดับของลักษณะหรือพฤติกรรมที่ต้องการประเมิน เช่น ให้ประเมินพฤติกรรมของนักเรียนโดยวงกลมล้อมรอบตัวเลข หมายเลข 1 เมื่ออยู่ในระดับต่ำ, 2 เมื่ออยู่ในระดับต่ำกว่าปานกลาง, 3 เมื่ออยู่ในระดับปานกลาง, 4 เมื่อสูงกว่าปานกลาง และ 5 เมื่ออยู่ในระดับสูง

ความร่วมมือกับเพื่อน	1	2	3	4	5
ความร่วมมือกับครู	1	2	3	4	5

2. มาตรประมาณค่าแบบกราฟิก (Graphic Rating Scale) ให้ทำเครื่องหมายลงบนเส้นตรงเพื่อชี้ค่าที่ประเมิน เช่น ให้ทำเครื่องหมาย / ลงบนเส้นตรงในตำแหน่งที่ตรงกับลักษณะของนักเรียนมากที่สุด

ความร่วมมือ	_____
ซัดขวาง	เต็มใจเสมอ
ความคงที่ของอารมณ์	_____
ไม่คงที่	ควบคุมได้ดี

3. มาตรเชิงบรรยาย (Descriptive Scale) มีคำบรรยายหรือข้อความแสดงระดับของลักษณะที่ต้องการประเมิน เช่น ให้ทำเครื่องหมาย / ลงในช่องที่ตรงกับพฤติกรรมของนักเรียนในรายการต่อไปนี้

	ไม่เคย	นาน ๆ ครั้ง	บางครั้ง	บ่อยครั้ง	เสมอ
(1) ฟังความคิดเห็นของคนอื่น					

Rummel (1964: 201-202) ได้กล่าวถึงมาตราประมาณค่าที่เป็นที่นิยมใช้ว่ามีหลายรูปแบบ พร้อมทั้งได้ยกตัวอย่างไว้ 3 รูปแบบ เช่นเดียวกับ Turney และ Robb ดังนี้

1. มาตราประมาณค่าแบบตัวเลข (Numerical Rating Scales) ให้ตัดสินลักษณะออกมาเป็นตัวเลข เช่น ให้ประเมินบุคคลตามข้อกระทงต่อไปนี้ โดยวงกลมล้อมรอบ 1 เมื่อไม่พอใจ, 2 เมื่ออยู่ในระดับต่ำ, 3 เมื่ออยู่ในระดับปานกลาง, 4 เมื่อดีกว่าปานกลาง และ 5 เมื่ออยู่ในระดับที่ดีกว่า

มีบุคลิกดึงดูดความสนใจ.....	1	2	3	4	5
พินิจถนกับงานที่ทำ.....	1	2	3	4	5

2. มาตราประมาณค่าแบบกราฟิก (Graphic Rating Scales) ประกอบด้วยเส้นตรงสำหรับให้ผู้ประเมินค่าทำเครื่องหมายเพื่อแสดงตำแหน่งที่เป็นจริงที่สุดเกี่ยวกับผู้ถูกประเมิน
ความกระตือรือร้น

ไม่มีเลย	ไม่ค่อยจะ	กระตือรือร้น	ตื่นตัวและ	ทำงานด้วยใจจดจ่อ
		กระตือรือร้น	เป็นบางครั้ง	กระฉับกระเฉง
				อย่างยิ่ง

ความซื่อสัตย์

ไม่ซื่อสัตย์	น่าสงสัยใน	โดยปกติ	เท่าที่ปรากฏ	ซื่อสัตย์อย่างแนบแน่น
				ซื่อสัตย์
				ไว้ใจได้
				มีความซื่อสัตย์

3. มาตราประมาณค่าแบบบรรยาย (Descriptive Rating Scales) มักเป็นคำบรรยายแสดงระดับต่าง ๆ ของบุคลิกลักษณะ ดังตัวอย่างการประเมินพฤติกรรมโดยให้เลือกคำหรือข้อความที่บรรยายถึงพฤติกรรมของนักเรียนตามที่สังเกตได้

- นักเรียนคนนี้แข็งแรงหรือไม่
-แข็งแรงและคล่องแคล่ว
-ไม่ใคร่เหนื่อย
-มีความอดทน
-เจื่องหงอย, เหนื่อยง่าย

Vockell (1983: 92) ได้ให้ตัวอย่างของมาตรฐานค่าไว้ดังนี้

แบบที่ 1 ให้เขียนตัวเลขลงในช่องว่างหน้าพฤติกรรมที่กำหนดมาให้ เช่น ให้แสดงระดับการกระทำของนักเรียน โดยใช้สเกลต่อไปนี้

- 5 หมายถึง ดีมาก
- 4 หมายถึง ดี
- 3 หมายถึง ปานกลาง
- 2 หมายถึง พอใช้
- 1 หมายถึง ยังใช้ไม่ได้

.....ฟังการบรรยายอย่างตั้งใจ

.....ยอมรับฟังความคิดเห็นของคนอื่น

แบบที่ 2 เป็นแบบธรรมดาเช่นเดียวกับแบบแรก แต่ให้ผู้ประเมินเลือกตัวเลขจากที่ให้ไว้ดังนี้

ฟังการบรรยายอย่างตั้งใจ 1 2 3 4 5

ยอมรับฟังความคิดเห็นของคนอื่น 1 2 3 4 5

แบบที่ 3 เป็นแบบกราฟิก ให้วงกลมล้อมรอบข้อความที่ต้องการ ดังตัวอย่างนักเรียนตั้งใจฟังการบรรยายเพียงใด

ดีมาก ดี ปานกลาง พอใช้ ยังใช้ไม่ได้

นักเรียนยอมรับฟังความคิดเห็นของคนอื่นเพียงใด

ดีมาก ดี ปานกลาง พอใช้ ยังใช้ไม่ได้

แบบที่ 4 เป็นแบบกราฟิกเช่นกัน แต่ให้ทำเครื่องหมายลงบนส่วนใดก็ได้บนเส้นตรงดังตัวอย่าง

นักเรียนตั้งใจฟังการบรรยายเพียงใด

สนใจเกือบ

สนใจ

เกือบไม่สนใจ

ตลอดเวลา

เลย

อุทุมพร จามรमान (2530: 105-106) ได้กล่าวถึงมาตรฐานค่าว่า เป็น เครื่องมือช่วยสังเกตพฤติกรรมที่ปรากฏ หรือไม่ปรากฏก็ได้ แต่ผู้สังเกตต้องเปรียบเทียบความ มากน้อยหรือความแตกต่างของพฤติกรรมให้ได้ และได้จำแนกแบบประเมินค่าเป็น 3 ประเภท พร้อมทั้งยกตัวอย่าง ดังนี้

1. มาตรฐานค่าเชิงตัวเลข (Numerical Rating scales) ประกอบด้วย รายการพฤติกรรมและเลขกำกับ เช่น 1,2,3,4,5, ให้ผู้ประเมินค่าทำวงกลมรอบเลขที่เลือก ตัวอย่างเช่น

บุคลิกภาพของอาจารย์ X	เลขประเมินค่า				
ก. แต่งกายเรียบร้อย	1	2	3	4	5
ข. ตรงเวลา	1	2	3	4	5

2. มาตรฐานค่าแบบกราฟิก (Graphic Rating Scales) คล้ายแบบที่ 1 แต่ใช้ภาษาแทนตัวเลขในสเกลและใช้เส้นแทนช่อง โดยมีข้อความกำกับหัวท้าย เช่น

ก. การใช้อุปกรณ์การสอนของอาจารย์ X	ไม่เหมาะสม	_____	เหมาะสม
ข. การอธิบาย	_____	_____	_____
	ไม่ชัดเจนเลย	ชัดเจนบางครั้ง	ชัดเจน

สเกลแบบนี้ให้ผู้ประเมินค่าใส่เครื่องหมายลงที่ใดที่หนึ่งบนเส้น

3. มาตรฐานค่าแบบกราฟิกเชิงบรรยาย (Descriptive Graphic Rating Scales) เป็นแบบประเมินค่าของ 1 และ 2 รวมกัน คือมีทั้งเลขและภาษากำกับหัวท้าย เช่น

น้ำเสียงของอาจารย์ X	เวลาสอน	1	2	3	4	5
		เบามาก		กำลังดี		ดังเกินไป

หรืออาจจัดสเกลเรียงลำดับจากมากไปน้อย หรือน้อยไปมาก เช่น

น้ำเสียงของอาจารย์ X	เวลาสอน	1	2	3	4	5
		ดังมาก	ดังมาก	ดัง	เบา	ไม่ได้
		เกินไป		พอสมควร		ยินเลย

ตามที่กล่าวมา จะเห็นว่า รูปแบบของมาตรประมาณค่าที่มักจะได้รับกล่าวถึงในลักษณะที่ร่วมกันว่าเป็นแบบที่นิยมใช้กันอย่างแพร่หลายทั่วไปนั้น น่าจะเป็น 3 แบบที่สำคัญคือ (1) แบบตัวเลข เป็นแบบที่กำหนดมาตร หรือสเกลเป็นตัวเลขมาให้ ผู้ประเมินสามารถตัดสินรายการที่ให้ประเมินออกมาเป็นระดับตัวเลข (2) แบบกราฟิก เป็นแบบที่กำหนดรายการที่ให้ประเมินคู่ควบกับเส้นตรงแสดงระดับความมากน้อยของคุณลักษณะ โดยในแต่ละช่วงบนมาตรวัด จะมีค่าคุณศัพท์เชิงปริมาณบ่งบอกระดับของคุณลักษณะอย่างเป็นลำดับ และ (3) แบบบรรยาย เป็นแบบที่มีคำบรรยายสั้น ๆ ในแต่ละทางเลือกเพื่อประกอบการตัดสินใจ

1.3 ปัญหา จุดอ่อน หรือ ข้อจำกัดของมาตรประมาณค่า

แม้ว่ามาตรประมาณค่าจะเป็นเครื่องมือที่พัฒนาได้ไม่ยากนัก มีความสะดวกในการใช้และสามารถประยุกต์ใช้ในสถานการณ์การประเมินต่าง ๆ อย่างหลากหลาย อย่างไรก็ตาม การที่จะสร้าง หรือพัฒนามาตรวัดรูปแบบนี้ให้สามารถวัดได้อย่างถูกต้อง เป็นปรนัย และเกิดความคลาดเคลื่อนน้อยนั้น จัดได้ว่าเป็นเรื่องที่มีความยุ่งยากมากพอสมควร ทั้งนี้เพราะว่าโดยธรรมชาติของมาตรวัดชนิดนี้แล้ว มีโอกาสที่จะเกิดความคลาดเคลื่อน หรือมีจุดอ่อนในหลายลักษณะ แหล่งของความคลาดเคลื่อนในการใช้มาตรวัดรูปแบบนี้ที่อาจเป็นไปได้ เช่น เกิดจากลักษณะของสิ่งที่มุ่งประเมิน ถ้าสิ่งที่ให้ประเมินเป็นลักษณะที่แฝงอยู่ในตัวบุคคล ไม่สามารถสังเกตได้โดยตรง เช่น ความคิดริเริ่ม ความเป็นมิตร ความเมตตา เป็นต้น โอกาสของการตีความหรือการประเมินของแต่ละคนก็ไม่เหมือนกัน Guilford (1954: 296) กล่าวว่า ลักษณะบางอย่างผู้ประเมินสามารถตัดสินได้อย่างสอดคล้อง เช่น ความสามารถ ความคิดริเริ่ม ความเป็นมิตร สติปัญญา เป็นต้น ในขณะที่บางลักษณะการตัดสินมีโอกาสขัดแย้งกันมาก เช่น ความกล้าหาญ ความร่วมมือ ความเมตตา การรู้จักกาลเทศะ เป็นต้น

แหล่งของความคลาดเคลื่อนในการวัดโดยใช้เทคนิคมาตรประมาณค่าอีกแหล่งหนึ่งที่น่าจะเป็นได้ คือ ลักษณะของมาตรวัด การใช้คำคุณศัพท์ในเชิงปริมาณ เช่น ในกรณีของมาตรประมาณค่าแบบกราฟิก คำว่า "ดี" "ดีเยี่ยม" "น่าพอใจ" ฯลฯ เป็นคำที่ก่อให้เกิดความแตกต่างในการตีความหมายของผู้ประเมินแต่ละคน คำว่า "ดี" ของคนหนึ่งอาจจะ เป็นเพียงระดับพอใช้ของอีกคนหนึ่ง (Crane, 1979: 295) โดยนัยนี้ ในกรณีที่มาตรวัดไม่มีเกณฑ์ในการตัดสินใจที่ชัดเจน จะส่งผลให้การตีความหมายของแต่ละคนแตกต่างกัน และการที่ผู้ประเมินต่างคนกัน ตีความหรือมี

ขอบเขตในการประมาณค่าที่แตกต่างกันนี้ ได้ส่งผลทำให้ค่าความเที่ยงระหว่างผู้ประเมินลดลง (Selltiz, et al., 1959: 352)

นอกจากแหล่งความคลาดเคลื่อนอันเนื่องมาจากลักษณะของมาตร และลักษณะของสิ่งที่ประเมินดังกล่าวข้างต้นแล้ว แหล่งความคลาดเคลื่อนในการประเมินอาจเกิดจากตัวผู้ประเมิน หรือธรรมชาติของผู้ประเมิน ทั้งนี้เพราะว่าโดยทฤษฎีทางจิตวิทยาแล้วถือว่า บุคคลต่างคนกันอาจรับรู้ในสิ่งเดียวกันในลักษณะที่แตกต่างกันได้ (Mathis, 1979: 295) ซึ่งอาจเกิดจากสาเหตุหลายอย่าง อาทิ การใช้มาตรฐานในการตัดสินที่แตกต่างกัน เกิดจากการที่ผู้ประเมินยึดติดกับแบบประเมินบางลักษณะ (Rater Patterns) เป็นต้น ลักษณะความคลาดเคลื่อนอันเกิดจากตัวผู้ประเมินที่สำคัญ ๆ เช่น Halo Error Leniency Error Error of Central Tendency เป็นต้น ความคลาดเคลื่อนแต่ละลักษณะ มีรายละเอียดดังนี้

Halo Error เกิดจากการที่ผู้ประเมินมีความรู้สึกบางอย่างต่อเป้าหมายที่กำลังประเมิน หรือเกิดจากการที่เคยมีปฏิสัมพันธ์มาก่อน ทำให้ประเมินผิดไปจากความเป็นจริงอันเนื่องมาจากความรู้สึกนั้น ความคลาดเคลื่อนในลักษณะนี้อาจเกิดขึ้นได้ทั้งทางบวกและทางลบ ลักษณะผลการประเมินที่ปรากฏค่าคะแนนในระดับเดียวกันหรือมีค่าใกล้เคียงกันในทุกรายการที่ให้ประเมินเป็นหลักฐานที่สื่อถึงแนวโน้มของการเกิดความคลาดเคลื่อนในลักษณะนี้ (Mathis, 1979: 298)

Leniency Error เป็นความคลาดเคลื่อนที่ผู้ประเมินมีความเอนเอียงที่จะประเมินสูงกว่าความเป็นจริง (Positive Leniency) หรือประเมินต่ำกว่าความเป็นจริง (Negative Leniency) ทั้งนี้อาจเป็นเพราะรู้จัก หรือเกี่ยวข้องกับเป้าหมายที่กำลังประเมินในลักษณะใดลักษณะหนึ่ง เช่น นักเรียนประเมินพฤติกรรมการสอนของครู หรืออาจเกิดจากธรรมชาติของผู้ประเมินที่มักจะเป็นผู้ที่ปล่อยคะแนน หรืออดคะแนนจนเกินไป

Error of Central Tendency เกิดจากผู้ประเมินมีข้อมูลเกี่ยวกับผู้ถูกประเมินไม่เพียงพอ จึงเลือกประเมินเฉพาะในระดับกลาง ๆ เพื่อป้องกันความผิดพลาด หรือเพื่อความปลอดภัยไว้ก่อน

ความคลาดเคลื่อนต่าง ๆ ดังที่ได้กล่าวมานี้ เป็นความคลาดเคลื่อนอย่างเป็นระบบ (Systematic Error) ที่มีผลทำให้ค่าความตรงในการประเมินลดลง (Selltiz, et al., 1959: 352)

1.4 แนวทางการปรับปรุงคุณภาพของเครื่องมือวัดประเภทมาตรประมาณค่า

ในการปรับปรุงแก้ไขการประมาณค่าให้ได้ผลอย่างถูกต้อง หรือมีความแม่นยำตรงนั้น ได้มีผู้เสนอแนวคิดที่น่าสนใจหลายท่าน อาทิ

Thorndike และ Hagen (1977: 487) ได้เสนอแนะว่า

- 1) การประเมินควรทำเฉพาะลักษณะที่ปรากฏให้เห็นอย่างชัดเจน
- 2) ลักษณะที่จะประเมินต้องแสดงออกมาให้เห็นเป็นรูปพฤติกรรม และการตัดสินจะกระทำจากพฤติกรรมที่เห็นนั้น
- 3) รูปแบบของการประมาณค่า (Rating Form) จะต้องปรับปรุงเพื่อให้ผู้ประเมินบอกความแตกต่าง หรือพยายามควบคุมการที่ผู้ประเมินมีมาตรฐานในการตัดสินที่แตกต่างกัน
- 4) ผู้ประเมินจะต้องเป็นผู้ที่มีโอกาสมากที่สุดในการสังเกตพฤติกรรมในสถานการณ์ที่ต้องการ
- 5) ผู้ประเมินต้องได้รับการบอกเล่าถึงคุณค่าของการประมาณค่า และได้รับการฝึกฝนในการใช้เครื่องมือ
- 6) ถ้าสามารถมีผู้ประเมินหลายคนก็ควรรวบรวมผลการประเมินจากผู้ประเมินหลาย ๆ คน

Oppenhiem (1966: 86-87) ได้เสนอแนะว่า ในการสร้างหรือพัฒนามาตรประมาณค่า จะต้องกำหนดขอบเขต และทิศทางของลักษณะที่จะประเมินให้เป็นที่เข้าใจตรงกันเสียก่อน จากนั้นจึงค่อยกำหนดระดับในการประมาณค่า และกำหนดความหมายในแต่ละระดับ และเพื่อให้ลักษณะที่จะประเมินมีความชัดเจน ลักษณะนั้น ๆ จะต้องอธิบายไว้อย่างเฉพาะเจาะจง แต่ละลักษณะต้องหมายถึงกิจกรรมเพียงอย่างเดียว และเป็นลักษณะที่คงที่การตัดสินแต่ละลักษณะต้องมีพื้นฐานมาจากอดีต หรือ ปัจจุบันมากกว่าเป็นการมองอนาคต ถ้าเป็นการประเมินตนเอง ไม่ควรมีลักษณะที่คนส่วนมากจะตัดสินสูง หรือต่ำกว่าความเป็นจริงเสมอ

Guilford (1954: 280) กล่าวว่า การให้ความรู้และฝึกฝนผู้ประเมินอย่างดีเป็นวิธีการที่มีประสิทธิภาพมากที่สุดในการปรับปรุงการประมาณค่า โดยเฉพาะในการลดความคลาดเคลื่อนอย่างเป็นระบบ ผู้ประเมินที่ได้รับการฝึกฝนอย่างดีจะสามารถหลีกเลี่ยงไม่ให้เกิดความคลาดเคลื่อนได้

Cronbach (1960 : 510) กล่าวว่า ในกรณีที่สามารถรวบรวมความเห็นจากผู้ประเมินหลาย ๆ คน ถ้าผู้ประเมินเหล่านี้รู้จักผู้ถูกประเมินดีพอก็จะช่วยให้ค่าความเที่ยงในการประมาณค่าสูงขึ้น เนื่องจากอคติที่เกิดเป็นค่าเฉลี่ย นอกจากนั้นยังเป็นการเพิ่มเติมข้อมูลจากผู้ประเมินบางคนไม่มีโอกาสสังเกตเห็นได้อีกด้วย

Mathis (1979: 297) ได้เสนอแนะว่า ในการที่จะลดอคติของผู้ประเมิน (Rater Bias) หรือลดความคลาดเคลื่อนอย่างเป็นระบบในการประเมินพฤติกรรมการทำงานนั้น นอกเหนือจากการให้ความรู้ความเข้าใจในเรื่องความคลาดเคลื่อนในลักษณะต่าง ๆ ที่อาจเกิดขึ้นแก่ผู้ประเมินแล้ว สิ่งที่สำคัญยิ่งประการหนึ่งก็คือ การนิยาม หรือกำหนดความหมายของระดับการประเมินแต่ละระดับให้ชัดเจน อีกทั้งกำหนดขอบเขตในการพิจารณาในแต่ละรายการหรือแต่ละข้อให้ผู้ประเมินแต่ละคนสามารถเข้าใจตรงกัน ก็จะสามารถแก้ปัญหา หรือลดปัญหาของการเกิดความคลาดเคลื่อนในการประเมินได้ในการนี้ Mathis ชี้แนะว่า มาตรฐานประมาณค่าเชิงพฤติกรรม (Behaviorally-Anchored Rating Scales) เป็นรูปแบบหนึ่งของความพยายามที่จะลดความคลาดเคลื่อนอย่างเป็นระบบในการประเมินพฤติกรรมการทำงาน

โดยสรุป กล่าวได้ว่า ในการปรับปรุงการประมาณค่าให้ได้ผลการประเมินที่มีความคมชัดหรือเกิดความคลาดเคลื่อนน้อยนั้น ควรเลือกใช้การประเมินในสถานการณ์ที่สามารถสังเกตพฤติกรรม หรือคุณลักษณะ ได้อย่างชัดเจน อีกทั้งควรตระหนักในตัวแปรที่สำคัญ 2 ประการที่ส่งผลต่อคุณภาพการประเมิน คือ (1) ลักษณะของเครื่องมือประเมิน จะต้องกำหนดขอบเขตของแต่ละระดับบนสเกลให้ชัดเจน เป็นที่เข้าใจตรงกัน (2) ผู้ประเมินควรได้รับการฝึกฝนอย่างดี ได้รับการอธิบายถึงคุณค่าของการประเมิน มีโอกาสได้ทราบว่า ในการประเมินนั้น ตนเองอาจจะตกอยู่ภายใต้อิทธิพลของอคติบางประการที่ส่งผลให้เกิดความคลาดเคลื่อนในการประเมินได้ อย่างไรก็ตามในสถานการณ์การประเมินบางลักษณะ เช่น การประเมินพฤติกรรมการทำงานในลักษณะของการวิจัยเชิงสำรวจ หรือ การติดตาม-ประเมินผลผู้ผ่านการอบรมในโครงการต่าง ๆ ซึ่งเป็นสถานการณ์การประเมินที่มุ่งประเมินเป้าหมายจำนวนมาก โดยใช้ผู้ประเมินจำนวนมาก อาทิ การให้ผู้บังคับบัญชาประเมินพฤติกรรมการทำงานของมหานัณฑิต หรือให้มหานัณฑิตประเมินพฤติกรรมการทำงานของตนเอง ฯลฯ ในสถานการณ์ของการศึกษาวิจัยดังกล่าวเหล่านี้ โอกาสของการที่จะฝึกทักษะแก่ผู้ประเมินเป็นเรื่องที่ทำได้ยาก แต่ก็ก็เป็นสถานการณ์การประเมินที่มีประโยชน์อย่างยิ่งต่อการตัดสินใจ และจะยังคงเป็นสถานการณ์ของการศึกษาวิจัยที่จำเป็นต้องดำเนินการ

กันอย่างหลากหลายทั่วไป ดังนั้นในการปรับปรุงคุณภาพของการประมาณค่าในกรณีนี้คงจะต้องเน้นที่ การปรับปรุงหรือพัฒนาลักษณะของเครื่องมือประเมินเป็นสำคัญ

ตอนที่ 2 สภาพการใช้มาตรฐานประมาณค่าในงานวิจัย

2.1 สภาพความนิยมในการใช้มาตรฐานประมาณค่า

มาตราประมาณค่า (Rating Scales) เป็นเทคนิคการวัดที่ได้รับความนิยม และใช้กันอย่างแพร่หลายมากที่สุด (Guilford, 1954: 263, Turney and Robb, 1971: 36) เฉพาะในสหรัฐอเมริกา Monahan (1975: 421) กล่าวว่า "ได้มีการใช้กันอย่างแพร่หลาย เสมือนการระบาดของวิธีการวัด (Measurement Myopia)" ส่วนในประเทศไทย จะพบว่า ได้มีการนำเทคนิคการวัดรูปแบบนี้มาใช้ในการศึกษาหลายวิจัยหลายลักษณะ เช่น ในการวิจัยเชิงสำรวจ การศึกษาเชิงประเมิน หรือการประเมินผล การติดตามผลรูปแบบต่าง ๆ เป็นต้น แนวทางการประยุกต์ใช้ที่พบ จะมีทั้งใช้เพื่อประเมินกระบวนการ อาทิ ประเมินทักษะการทำงาน รูปแบบการทำงาน พฤติกรรมการปฏิบัติงาน ฯลฯ ใช้ในการประเมินผลผลิต อาทิ การประเมินผลผลิตเพื่อการแข่งขันต่าง ๆ หรือประเมินพัฒนาการทางบุคลิกภาพและทางสังคม อาทิ ความสนใจ ทัศนคติ ความเป็นผู้นำ ความคิดริเริ่มสร้างสรรค์ ความมั่นคงทางอารมณ์ มนุษยสัมพันธ์ ฯลฯ

กรณีตัวอย่างที่แสดงถึงการนิยมใช้มาตรฐานประมาณค่าในกิจกรรมการประเมินในประเทศไทยที่เห็นได้อย่างชัดเจนที่สุดกรณีหนึ่งก็คือ กรณีของงานวิจัยกลุ่มที่ประยุกต์แนวคิดเกี่ยวกับการประเมินความต้องการจำเป็น (Educational Needs Assessment) รูปแบบต่าง ๆ เช่น กลุ่มที่ศึกษาปัญหาในการปฏิบัติงาน กลุ่มที่ศึกษาความต้องการของหน่วยงานหรือของบุคลากรของหน่วยงาน หรือกลุ่มที่ศึกษาเชิงประเมินพฤติกรรมการปฏิบัติงาน ซึ่งจากการที่ผู้วิจัยได้ทำการศึกษางานวิจัยกลุ่มนี้ ประมาณ 20 เรื่อง ซึ่งดำเนินการระหว่างปี พ.ศ.2525 - พ.ศ.2531 ได้พบว่าทุกเรื่องได้ใช้เทคนิคการประมาณค่าเป็นเครื่องมือในการเก็บรวบรวมข้อมูล และรูปแบบการประมาณค่าที่นิยมใช้จะเป็นแบบตัวเลข (Numerical Scales) หรือไม้เรียวแบบกราฟิก (Graphic Scales)

2.2 ลักษณะที่น่าสังเกตบางประการจากงานวิจัยที่ใช้เทคนิคมาตรประมาณค่า

จากการที่ผู้วิจัยได้ศึกษางานวิจัยกลุ่มที่ประยุกต์แนวคิดในการประเมินความต้องการจำเป็นทางการศึกษาที่ดำเนินการในประเทศไทย ประมาณ 20 เรื่อง ซึ่งพบว่าทุกเรื่องได้ใช้มาตรประมาณค่าแบบตัวเลขหรือแบบกราฟิกเป็นเครื่องมือประเมิน ดังที่ได้กล่าวมาแล้วนั้น ผู้วิจัยจะขอยกตัวอย่างเป็นการศึกษาเพียงบางเรื่อง เพื่อให้มองเห็นประเด็นที่น่าสังเกตบางประการดังนี้

งามทรัพย์ เทศะบำรุง (2530: 30-42) ได้ศึกษาปัญหาการจัดการบริการสุขภาพในโรงเรียนประถมศึกษา สังกัดกรุงเทพมหานคร ตามการรับรู้ของครู โดยผู้วิจัยได้ใช้มาตรประมาณค่า (Rating Scales) ชนิด 5 ระดับทางเลือก เป็นเครื่องมือในการประเมินปัญหา 13 ด้าน คือ ด้านการบันทึกสุขภาพ การตรวจสุขภาพ การปฐมพยาบาลและการรักษาพยาบาล การติดตามผลการรักษา การจัดการบริการโภชนาการในโรงเรียน การป้องกันและควบคุมโรคติดต่อ การซึ่งน้ำหนักและวัดส่วนสูง การทดสอบสายตาและการได้ยิน การจัดสวัสดิการในโรงเรียน การแนะแนวสุขภาพ การส่งเสริมสุขภาพจิต การส่งเสริมสุขภาพบุคลากรในโรงเรียน และการจัดดำเนินงานบริการสุขภาพ ในแต่ละด้านจะประกอบด้วยรายการที่ให้ประเมิน (Items) ประมาณ 6-9 รายการ ผลการประเมินพบว่า ร้อยละ 80 ของรายการปัญหาที่ให้ประเมิน ได้รับการประเมินว่าเป็นปัญหาในระดับน้อย ส่วนอีกร้อยละ 20 ได้รับการประเมินว่าเป็นปัญหาในระดับปานกลาง ไม่ปรากฏรายการใดได้รับการประเมินว่าเป็นปัญหาในระดับมาก

เจิม สืบขจร (2527: 110-142) ได้ศึกษาปัญหาการบริหารงานวิชาการตามหลักสูตรมัธยมศึกษาตอนต้น พุทธศักราช 2521 ของโรงเรียนมัธยมศึกษาส่วนกลาง สังกัดกรมสามัญศึกษา กลุ่มตัวอย่างคือผู้บริหารโรงเรียน ผู้ช่วยผู้บริหารฝ่ายวิชาการ และหัวหน้าหมวดวิชา ผู้วิจัยได้ใช้มาตรประมาณค่า ชนิด 5 ระดับทางเลือก เป็นเครื่องมือในการประเมินระดับความเป็นปัญหาในด้านต่าง ๆ 10 ด้าน ในแต่ละด้านจะมีรายการคำถามประมาณ 7-9 รายการ ผลการวิจัยพบว่า ร้อยละ 95 ของรายการที่มุ่งให้ประเมิน ได้รับการประเมินว่าเป็นปัญหาในระดับน้อย หรือน้อยมาก อีกร้อยละ 1 หรือประมาณ 1-2 รายการได้รับการประเมินว่าเป็นปัญหาในระดับปานกลาง เมื่อเปรียบเทียบระหว่างกลุ่ม พบว่า กลุ่มผู้บริหาร และผู้ช่วยผู้บริหารฝ่ายวิชาการ ประเมินความเป็นปัญหาในระดับที่น้อยกว่าผลการประเมินโดยครูผู้สอนหรือหัวหน้าหมวดวิชา ในทุกรายการ ทุกด้าน

ชลอ มงคลการุญย์ (2530: 102-125) ได้ศึกษาปัญหาการบริหารงานวิชาการในโรงเรียนประถมศึกษา สังกัดสำนักงานการประถมศึกษากรุงเทพมหานคร กลุ่มตัวอย่างผู้ให้ข้อมูลคือ ผู้บริหารโรงเรียน และหัวหน้ากลุ่มประสบการณ์ รวม 252 คน ผู้วิจัยได้ใช้เทคนิคมาตรฐานประมาณค่าชนิด 5 ระดับทางเลือก เป็นเครื่องมือในการประเมินระดับความเป็นปัญหา 8 ด้าน แต่ละด้านประกอบด้วยรายการที่มุ่งให้ประเมินระหว่าง 8-14 รายการ ผลการวิจัยพบว่า ร้อยละ 100 ของรายการที่ให้ประเมินได้รับการประเมินว่าเป็นปัญหาในระดับน้อยหรือน้อยที่สุดในกรณีที่เปรียบเทียบผลการประเมินระหว่างกลุ่มผู้บริหาร กับหัวหน้ากลุ่มประสบการณ์ พบว่า ผู้บริหารประเมินระดับความเป็นปัญหาในระดับที่น้อยกว่าผลการประเมินโดยหัวหน้ากลุ่มประสบการณ์ ในทุกรายการ ทุกด้าน

ฤดี ภูมิภูถาวร (2526: 62-73) ได้ศึกษาความต้องการเสริมสมรรถภาพการสอนของครูสังคมศึกษา ในโรงเรียนมัธยมศึกษา เขตการศึกษา 4 กลุ่มตัวอย่างคือ ครูสังคมศึกษา จำนวน 308 คน เครื่องมือที่ใช้คือ แบบสอบถามมาตรฐานประมาณค่า ชนิด 4 ระดับทางเลือก โดยผู้วิจัยให้กลุ่มตัวอย่างประเมินความต้องการเสริมสมรรถภาพ 5 ด้าน คือ การวางแผนการสอน ทักษะการสอน การจัดกิจกรรมเสริมหลักสูตร บุคลิกภาพ มนุษยสัมพันธ์และการแนะแนวการเรียนการสอน และวิธีการเสริมสมรรถภาพการสอน ในแต่ละด้านประกอบด้วยรายการที่มุ่งให้ประเมินระหว่าง 12-19 รายการ ผลการวิจัย พบว่า ร้อยละ 100 ของรายการที่ให้ประเมินได้รับการประเมินว่าต้องการในระดับมาก

อวยชิต สือยรรยงศิริ (2528: 37-54) ได้ศึกษาความต้องการเสริมสมรรถภาพการสอนวิชาคณิตศาสตร์ ของครูคณิตศาสตร์ ในโรงเรียนมัธยมศึกษา เขตการศึกษา 10 กลุ่มตัวอย่างคือ ครูคณิตศาสตร์ จำนวน 337 คน โดยผู้วิจัยได้ใช้เทคนิคมาตรฐานประมาณค่าชนิด 5 ระดับทางเลือก เป็นเครื่องมือในการประเมินความต้องการ 6 ด้าน แต่ละด้านประกอบด้วย รายการที่มุ่งให้ประเมินระหว่าง 10-15 รายการ ผลการประเมิน พบว่า ร้อยละ 91 ของรายการที่มุ่งให้ประเมิน ได้รับการประเมินว่าต้องการในระดับมาก ส่วนที่เหลือได้รับการประเมินความต้องการในระดับปานกลาง คะแนนเฉลี่ยผลการประเมินจะอยู่ระหว่าง 3.69-3.99 คะแนน

ชาญศึก มีพร้อม (2530: 31-46) ได้ศึกษาความต้องการนิเทศโครงการสุขภาพในโรงเรียน ของผู้บริหารโรงเรียนประถมศึกษา เขตการศึกษา 11 กลุ่มตัวอย่างคือ ผู้บริหารโรงเรียน จำนวน 250 คน โดยผู้วิจัยได้ใช้แบบสอบถามมาตรฐานค่า ชนิด 5 ระดับทางเลือก เป็นเครื่องมือในการประเมินความต้องการ 6 ด้าน แต่ละด้านประกอบด้วยรายการที่ให้ประเมินระหว่าง 14-20 รายการ ผลการประเมินพบว่า ร้อยละ 100 ของรายการที่ให้ประเมิน ได้รับการประเมินว่ามีความต้องการในระดับมาก คะแนนเฉลี่ยผลการประเมินส่วนใหญ่อยู่ระหว่าง 4.05-4.91

สาหร่าย แดงนาวา (2527: 127-155) ได้ศึกษาความต้องการของครูและผู้บริหารในการพัฒนาสมรรถภาพด้านการจัดบริการและการจัดโครงการในโรงเรียนประถมศึกษา สังกัดกรุงเทพมหานคร กลุ่มตัวอย่างคือผู้บริหารและครู จำนวน 155 คน และ 272 คน ตามลำดับ ผู้วิจัยได้ใช้แบบสอบถามมาตรฐานค่า ชนิด 5 ระดับทางเลือก เป็นเครื่องมือในการประเมินความต้องการ 10 ด้าน ผลการประเมิน พบว่า ผู้บริหารและครูประเมินระดับความต้องการในการพัฒนาสมรรถภาพในระดับมากเกือบทุกรายการ ทุกด้าน เมื่อเปรียบเทียบคะแนนเฉลี่ยผลการประเมินระหว่างผู้บริหารกับครู ปรากฏว่า ผู้บริหารจะประเมินความต้องการในแต่ละรายการในระดับที่สูงกว่าครูเสมอ เช่น ในการประเมินความต้องการด้านการจัดบริการและแนวของกลุ่มผู้บริหารและครูในโรงเรียนในเขตตัวเมือง พบว่า คะแนนผลการประเมินของผู้บริหารจะอยู่ระหว่าง 4.00-4.17 ในขณะที่คะแนนผลการประเมินโดยครู ในแต่ละรายการจะอยู่ระหว่าง 3.57-3.84 หรือในกรณีของการประเมินความต้องการพัฒนาสมรรถภาพด้านการสอนซ่อมเสริม พบว่า คะแนนผลการประเมินโดยผู้บริหาร ในแต่ละรายการจะอยู่ระหว่าง 4.11-4.39 ในขณะที่คะแนนผลการประเมินความต้องการโดยครู ในแต่ละรายการจะอยู่ระหว่าง 3.65-3.95 เป็นต้น

พยุงค์ศักดิ์ สุนทร (2527: 64-81) ได้ศึกษาความคิดเห็นของครูพลศึกษาและนักเรียนเกี่ยวกับพฤติกรรมการสอนพลศึกษา ในโรงเรียนมัธยมศึกษาตอนปลาย ในเขตกรุงเทพมหานคร ผู้วิจัยได้ใช้มาตรฐานค่า ชนิด 5 ระดับทางเลือก เป็นเครื่องมือให้นักเรียนและครูพลศึกษาประเมินพฤติกรรมการสอนพลศึกษา 7 ด้าน รวมรายการที่ให้ประเมินทั้งสิ้น 62 ข้อ ผลการประเมินพบว่า 55 รายการ หรือประมาณร้อยละ 89 ของรายการที่ให้ประเมินทั้งหมดได้รับการประเมินว่าเป็นพฤติกรรมที่ได้ปฏิบัติในระดับบ่อยครั้ง หรือเป็นประจำ (ระดับ 4-5 คะแนน ส่วนที่เหลือได้รับการประเมินว่าเป็นพฤติกรรมที่ได้ปฏิบัติบางครั้ง หรือปานกลาง ผลการประเมินโดยครูมีแนวโน้ม

สูงกว่าผลการประเมินโดยนักเรียนในเกือบทุกรายการที่ให้ประเมิน

ขวัญชัย ชิงรักษา (2529: 48-65) ได้ศึกษาความคิดเห็นของผู้บริหารโรงเรียน ครู และครูพลศึกษา เกี่ยวกับสมรรถภาพของครูพลศึกษาในโรงเรียนระดับประถมศึกษา สังกัด กรุงเทพมหานคร โดยผู้วิจัยได้ใช้มาตราประมาณค่า ชนิด 4 ระดับทางเลือก 4 -มากที่สุด 3 -มาก 2-น้อย และ 1-น้อยที่สุด เป็นเครื่องมือในการเก็บรวบรวมข้อมูลด้วยการให้ผู้บริหาร ครู และครูพลศึกษาประเมินสมรรถภาพตามที่ปรากฏจริงของครูพลศึกษา 8 ด้าน คือ ด้านหลักสูตรและวิชาการ ด้านการสอน ด้านการประเมินผล ด้านความรับผิดชอบหน้าที่ ด้านมนุษยสัมพันธ์ ด้านบุคลิกภาพ ด้านสุขภาพ และ ด้านคุณลักษณะและเจตคติ รวมรายการสมรรถภาพที่ให้ประเมินทั้งสิ้น 86 รายการ ผลการศึกษา พบว่า ร้อยละ 100 ของรายการที่ให้ประเมิน ได้รับการประเมินว่ามีสมรรถภาพในระดับมาก และผลการประเมินโดยครูพลศึกษา ปรากฏแนวโน้มของค่าเฉลี่ยสูงกว่าผลการประเมินโดยกลุ่มอื่น ๆ

พวงพยอม ศรีหาบัติ (2529: 51-68) ได้ศึกษาเรื่อง "การปฏิบัติงานนิเทศทางวิชาการของหัวหน้าหมวดวิชาภาษาไทย ตามการรับรู้ของตนเองและครูผู้สอนภาษาไทยโรงเรียนมัธยมศึกษาในภาคตะวันออกเฉียงเหนือ" โดยใช้เทคนิคมาตราประมาณค่าชนิด 4 ระดับทางเลือก 4-มากที่สุด 3-มาก 2-น้อย และ 1-น้อยที่สุด ให้หัวหน้าหมวดวิชาภาษาไทยและครูภาษาไทยประเมินพฤติกรรมการปฏิบัติงานนิเทศของหัวหน้าหมวดวิชาภาษาไทย รวม 7 ด้าน ประกอบด้วยรายการที่ให้ประเมินทั้งสิ้น 71 รายการ ผลการประเมินพบว่าประมาณร้อยละ 75 ของรายการที่ให้ประเมินได้รับการประเมินว่าได้ปฏิบัติในระดับมาก และผลการประเมินโดยหัวหน้าหมวดวิชาเองมีแนวโน้มสูงกว่าผลการประเมินโดยครูภาษาไทย

จากการวิเคราะห์ทบทวนงานวิจัยที่ใช้เทคนิคการประมาณค่า แบบตัวเลขหรือแบบกราฟิก เป็นเครื่องมือในการเก็บรวบรวมข้อมูล ดังกรณีตัวอย่างที่ได้เสนอมานั้น ผู้วิจัยได้พบข้อสังเกตบางประการคือ

1) การศึกษาในกรณีของ งามทรัพย์ เทศะบำรุง (2530) เจริญ สืบขจร (2527) ชลอมงคลการุญย์ (2530) รวมทั้งกรณีวิจัยอื่น ๆ เช่น การศึกษาของฉวีวรรณ ไหวพริบ (2529) เรื่อง "ปัญหาการนิเทศวิชาสุขศึกษาในระดับมัธยมศึกษา" ศุภกนิ มหิทธิบุรินทร์ (2526) เรื่อง "ปัญหาการนิเทศและการเรียนการสอนวิชาพลศึกษา ใน 4 จังหวัดชายแดนภาคใต้"

สุชาติ จันทรทอง (2530) เรื่อง "ปัญหาการบริหารและความต้องการการนิเทศด้านสิ่งแวดล้อมทางสุขภาพ ในโรงเรียนมัธยมศึกษา เขตการศึกษา 7 " 444 ซึ่งไม่ได้แสดงรายละเอียดไว้ ในที่นี้ การศึกษาดังกล่าวเหล่านี้เป็นการศึกษาความเป็นปัญหาในเรื่องใดเรื่องหนึ่งในระบบโรงเรียน ข้อสรุปจากงานวิจัยกลุ่มนี้ พบว่า กลุ่มผู้บริหาร หรือ ครู ซึ่งเป็นประชากรกลุ่มเป้าหมายที่นักวิจัย นิยมใช้มากที่สุด มักจะรายงานสภาพความเป็นปัญหาไม่ว่าในเรื่องใด รายการใด ในระดับน้อยมาก เกือบทุกรายการ ทุกด้านที่ให้ประเมิน มีเพียงส่วนน้อยที่ปรากฏผลการประเมินว่าเป็นปัญหาในระดับ ปานกลาง หรือมาก ผลการประเมินส่วนใหญ่จะสรุปว่า ไม่มีรายการใดที่เป็นปัญหาเด่นชัด เกือบ ทุกรายการมีความเป็นปัญหาในระดับน้อยและใกล้เคียงกัน

2) การศึกษาในกรณีของ ฤดี ภูมิภูถาวร (2526) เรื่อง "ความต้องการเสริมสมรรถภาพ การสอนของครูสังคมศึกษา ในโรงเรียนมัธยมศึกษา เขตการศึกษา 4" อวยชิต สือยรยงศิริ(2528) เรื่อง "ความต้องการเสริมสมรรถภาพการสอนวิชาคณิตศาสตร์ ของครูคณิตศาสตร์ในโรงเรียน มัธยมศึกษา เขตการศึกษา 10 " ชาญุศิก มีพร้อม (2530) เรื่อง "ความต้องการการนิเทศ โครงการสุขภาพในโรงเรียน ของผู้บริหารโรงเรียนประถมศึกษา เขตการศึกษา 11 " สหราชัย แต่งนาวา (2527) เรื่อง "ความต้องการของครูและผู้บริหารในการพัฒนาสมรรถภาพ ด้านการจัดบริการและการจัดโครงการในโรงเรียนประถมศึกษา สังกัดกรุงเทพมหานคร" รวมทั้ง กรณีศึกษาในลักษณะคล้ายคลึงกันที่ไม่ได้เสนอรายละเอียดไว้ในที่นี้ เช่น การศึกษาของ สุทัศน์ เหล่าทอง (2529) เรื่อง "ความต้องการพัฒนาวิชาชีพครูของครูพลศึกษาในโรงเรียน มัธยมศึกษา เขตการศึกษา 1 " สุภา ศิลสร (2526) เรื่อง "ความต้องการพัฒนาสมรรถภาพ ในการปฏิบัติการสอนของครูประถมศึกษา ในเขตกรุงเทพมหานคร" วิภา แก้วสังข์ทอง (2530) เรื่อง "ความต้องการของผู้บริหารในการจัดและการใช้สารสนเทศทางการศึกษาของศูนย์การ ศึกษาออกโรงเรียนจังหวัด" เป็นต้น การศึกษาวิจัยในกลุ่มนี้เน้นการประเมินความต้องการตาม การรับรู้ของกลุ่มบุคคลกลุ่มใดกลุ่มหนึ่ง ข้อสรุปจากการใช้เทคนิคมาตราประมาณค่าแบบตัวเลข หรือ แบบกราฟิก ในการเก็บรวบรวมข้อมูลในลักษณะนี้ พบว่า กลุ่มผู้บริหาร หรือครู มักประเมินความ ต้องการในเรื่องนั้น ๆ ในระดับมาก เกือบทุกรายการ ทุกด้านที่ให้ประเมิน ผลการประเมินของ งานวิจัยกลุ่มนี้มักจะได้ข้อสรุปว่าประชากรกลุ่มนั้น ๆ มีความต้องการในเรื่องที่ให้ประเมินในระดับ มากเกือบทุกรายการ ทุกด้านในระดับที่ใกล้เคียงกัน และกลุ่มผู้บริหารมีแนวโน้มประเมินความต้อ การในระดับที่สูงกว่าผลการประเมินโดยกลุ่มครู

3) การศึกษาในกรณีของ พยุงศักดิ์ สมนเทศ (2527) เรื่อง "ความคิดเห็นของครูพลศึกษาและนักเรียนเกี่ยวกับพฤติกรรมการสอนพลศึกษา ในโรงเรียนมัธยมศึกษาตอนปลาย ในเขตกรุงเทพมหานคร" ขวัญชัย ชิงรักษา (2529) เรื่อง "ความคิดเห็นของผู้บริหารโรงเรียนครู และครูพลศึกษา เกี่ยวกับสมรรถภาพของครูพลศึกษาในโรงเรียนระดับประถมศึกษา สังกัดกรุงเทพมหานคร พวงพยอม ศรีหาบัตติ (2529) เรื่อง "การปฏิบัติงานนิเทศทางวิชาการของหัวหน้าหมวดวิชาภาษาไทย ตามการรับรู้ของตนเองและครูผู้สอนภาษาไทย โรงเรียนมัธยมศึกษาในภาคตะวันออกเฉียงเหนือ" งานวิจัยกลุ่มนี้ได้เน้นการประเมินพฤติกรรมการปฏิบัติงานในสภาพที่เป็นจริงของกลุ่มบุคคลกลุ่มใดกลุ่มหนึ่ง เพื่อประโยชน์ในการวางแผนพัฒนาบุคลากร หรือการตัดสินใจเกี่ยวกับการบริหารงานบุคคลในลักษณะใดลักษณะหนึ่ง งานวิจัยที่มีลักษณะใกล้เคียงกับงานวิจัยกลุ่มนี้คือ งานวิจัยที่สะท้อนแนวความคิดในการประเมินความต้องการจำเป็นด้านการพัฒนาบุคลากร (Staff Development Needs Assessment) ซึ่งทำการศึกษานวัตกรรมการปฏิบัติงานที่เป็นจริงและพฤติกรรมการปฏิบัติงานที่ควรจะเป็นเพื่อวิเคราะห์หารายการที่เกิดช่องว่าง (Gap) แล้วกำหนดเป็นความต้องการจำเป็น (Needs) ในการพัฒนาบุคลากร เช่น การศึกษาของ เรณู ศิริโรมาศกุล (2530) เรื่อง "บทบาทที่เป็นจริงและบทบาทที่คาดหวังของศึกษานิเทศก์อำเภอในการดำเนินงานการศึกษานอกระบบโรงเรียน" สุรพล ไพโรจน์ประยูร (2525) เรื่อง "ความคิดเห็นของอาจารย์และนักศึกษาเกี่ยวกับสมรรถภาพของอาจารย์วิทยาศาสตร์ในวิทยาลัยครูภาคใต้" ซึ่งทำการศึกษาทั้งสมรรถภาพที่มีอยู่จริงและสมรรถภาพที่ต้องการ ฯลฯ ข้อสรุปจากงานวิจัยกลุ่มดังกล่าวนี้ ในกรณีที่มีการศึกษาครั้งนั้น ๆ มุ่งเน้นเฉพาะการศึกษานวัตกรรมการปฏิบัติงานที่เป็นจริงหรือสมรรถภาพที่มีอยู่จริง ส่วนใหญ่จะปรากฏคะแนนผลการประเมินในระดับมาก หรือค่อนข้างมากอย่างสม่ำเสมอในเกือบทุกรายการ ทุกด้านที่ให้ประเมิน ส่วนในกรณีที่มีการศึกษาครั้งนั้น ๆ มุ่งวิเคราะห์ทั้งสภาพที่เป็นจริงและสภาพที่ควรจะเป็นควบคู่กันไป ข้อสรุปจากงานวิจัย มักจะพบว่า ในส่วนของสภาพการปฏิบัติงานที่เป็นจริงหรือสภาพที่เป็นจริงจะได้รับการประเมินในระดับปานกลางหรือค่อนข้างมากอย่างสม่ำเสมอ ในขณะที่ ผลการประเมินในส่วนของสภาพที่ควรจะเป็นจะได้รับการประเมินในระดับมาก หรือ มากที่สุดเกือบทุกรายการ ทุกด้านที่ให้ประเมิน

จากข้อสังเกตเกี่ยวกับผลการวิจัยที่นิยมใช้มาตรประมาณค่าแบบตัวเลขหรือแบบกราฟิก เป็นเครื่องมือในการเก็บรวบรวมข้อมูล ดังกล่าวข้างต้น ซึ่งพบว่าในการพิจารณาประเมินสภาพ ความเป็นปัญหา มักจะได้รับการรายงานว่าเป็นปัญหาในระดับน้อยอย่างสม่ำเสมอเกือบทุกรายการ ที่ให้ประเมิน ส่วนกรณีการศึกษาความต้องการ มักจะได้รับการรายงานว่ามีความต้องการในระดับ มากหรือมากที่สุด อย่างสม่ำเสมอเกือบทุกรายการที่ให้ประเมิน หรือในกรณีของการประเมิน พฤติกรรมการปฏิบัติงานที่เป็นจริง มักจะได้รับการรายงานว่าได้ปฏิบัติในระดับค่อนข้างมากอย่าง สม่ำเสมอเกือบทุกรายการที่ให้ประเมิน นั้น การที่ผลการวิจัยปรากฏข้อสรุปในลักษณะนี้ อาจเป็น เพราะเครื่องมือที่ใช้ในการประเมินไม่สามารถจำแนกความแตกต่างในสภาพที่เป็นจริงได้ ทั้งนี้ เพราะโดยลักษณะของความน่าจะเป็นแล้ว ในการประเมินเป้าหมายใดเป้าหมายหนึ่งในแง่มุม ที่แตกต่างกัน หรือประเมินเป้าหมายนั้น ๆ ในหลาย ๆ ด้าน หลาย ๆ มิติ ไม่น่าจะเป็นไปได้ที่ผล การประเมินจะออกมาในลักษณะที่ใกล้เคียงกันอย่างสม่ำเสมอทุกครั้ง และการที่ผลการประเมินไม่ สามารถจำแนกความแตกต่างของรายการทั้งหลายที่ให้ประเมิน เช่นนี้ได้ก่อให้เกิดปัญหาในการตัดสินใจ หรือตัดสินใจไม่ได้อย่างสมเหตุสมผล ในบางครั้ง ความแตกต่างในค่าเฉลี่ยเพียงเล็กน้อย ได้ถูกนำ มาเสนอในรูปของการเรียงอันดับสิ่งที่ถูกประเมิน ซึ่งก่อให้เกิดผลได้ผลเสียแก่สิ่งที่ถูกประเมิน (Showers, 1973: 5731-A) โดยเฉพาะอย่างยิ่งในกรณีที่มีการประเมินครั้งนั้น ๆ เป็นการประเมิน ผลการปฏิบัติงานของบุคคล

ตอนที่ 3 แนวความคิดเกี่ยวกับมาตรประมาณค่าเชิงพฤติกรรม

3.1 จุดมุ่งหมายและลักษณะของมาตรประมาณค่าเชิงพฤติกรรม

มาตรประมาณค่าเชิงพฤติกรรมได้รับการพัฒนาขึ้นเป็นครั้งแรกโดย Smith และ Kendall (1963; อ้างถึงใน Witkin, 1984: 15) ต่อมาได้รับความสนใจและพัฒนาอย่างต่อเนื่องภายใต้ชื่อเรียกเป็นภาษาอังกฤษที่แตกต่างกันหลายชื่อ เช่น Behaviorally-Anchored Rating Scales- BARS (Witkin, 1984: 79) Behavioral Scales หรือ Behavioral Based Rating Scales (Campbell, 1973) Behavioral Expectation Scales (Keaveny, 1975: 695) เป็นต้น การพัฒนามาตรวัดตามแนวความคิดนี้ถือว่าเป็นวิธีการใหม่ที่ออกแบบขึ้นเพื่อแก้ปัญหาในเรื่องของการจัดระดับชั้นคุณภาพของรายการที่ให้ประเมินทั้งนี้ด้วยวิธีการกำหนด หรือนิยามพฤติกรรมการปฏิบัติงานให้เห็นเป็นรูปธรรมอย่างเรียงลำดับชัดเจน (Mathis, 1979: 305) โดยมุ่งหวังว่าจะสามารถลดโอกาสการเกิดความคลาดเคลื่อนอันเนื่องมาจากการตีความหมายข้อความหรือการรับรู้ข้อความที่แตกต่างกันระหว่างผู้ประเมินต่างกลุ่ม-ต่างคนกัน

มาตรประมาณค่าเชิงพฤติกรรม ถือเป็นรูปแบบหรือลักษณะหนึ่งของมาตรประมาณค่า (Rating Scales) ลักษณะของมาตรวัดในแต่ละมิติที่ให้ประเมินจะประกอบด้วยข้อความที่กระชับ ชัดเจน อธิบายถึงพฤติกรรมการปฏิบัติงานต่าง ๆ ที่เกี่ยวข้องกับมิตินั้น ๆ ข้อความจะถูกวางในตำแหน่งอย่างเป็นลำดับและมีความต่อเนื่องบนมาตรวัด ดังตัวอย่างมาตรวัดประสิทธิภาพในการให้คำปรึกษาแนะนำแก่นักงานของผู้จัดการฝ่าย ต่อไปนี้ (Witkin, 1984: 80-81)

- 9-สามารถให้คำปรึกษาแนะนำพร้อมทั้งฝึกปฏิบัติแก่พนักงานชายใหม่ 2 คน ภายในเวลา 1 วัน จนมีผลทำให้กลายเป็นพนักงานชายระดับแนวหน้าของฝ่าย
- 8-
- 7-ไม่เคยขาดการร่วมประชุม และประสบความสำเร็จในการจัดประชุมเพื่อพัฒนาบุคลากรของฝ่ายเสมอ สามารถชักนำให้บุคลากรมองเห็นและคล้อยตามสิ่งที่เขาต้องการ
- 6-
- 5-สามารถเตือนพนักงานชายให้คอยเอาใจใส่ลูกค้าแทนการนั่งคุยกันเอง
- 4-
- 3-บอก/แนะนำให้พนักงานคนหนึ่งมาปฏิบัติหน้าที่ให้จงได้ ทั้ง ๆ ที่ทราบว่าเขาป่วย
- 2-
- 1-ให้คำสัญญาต่อบุคคลใดบุคคลหนึ่งเกี่ยวกับอัตราเงินเดือนในตำแหน่งพนักงานชาย ทั้ง ๆ ที่รู้ว่าการกระทำนั้นขัดต่อนโยบายของบริษัท

แนวความคิดในการพัฒนามาตรฐานประมาธค่าเชิงพฤติกรรม นั้น ถือว่าวิธีการนี้เป็นวิธีการส่งเสริมหรือเพิ่มประสิทธิภาพของการบริหารตามแนวทางของการบริหารโดยยึดวัตถุประสงค์ (The MBO Approach) ทั้งนี้เพราะข้อความในมาตรฐานประมาธค่าจะแสดงให้เห็นว่า "อย่างไรจึงเรียกว่าเป็นพฤติกรรมที่มีประสิทธิภาพในการปฏิบัติงาน" ลักษณะสำคัญที่ทำให้มาตรฐานประมาธค่าเชิงพฤติกรรมเป็นวิธีการที่แตกต่างไปจากวิธีการเดิม ๆ โดยทั่วไปคือ (Crane, 1985: 413)

- 1) เป็นวิธีการที่เน้นเป้าหมายเชิงพัฒนา หรือชี้แนะแนวทางการพัฒนา
- 2) เน้นในเรื่องพฤติกรรมปฏิบัติงานของบุคคล

- 3) เป็นวิธีการที่เน้นการนิยามพฤติกรรมอย่างชัดเจน สังเกตได้ และวัดได้
- 4) เป็นวิธีการที่แสดงออกซึ่งความแตกต่างระหว่างคำว่า "พฤติกรรมโดยทั่วไป (Behaviors) พฤติกรรมการปฏิบัติงาน (Performance) และประสิทธิภาพของการปฏิบัติงาน (Effectiveness)

3.2 การประยุกต์ใช้มาตรฐานค่าเชิงพฤติกรรม

นักวิจัย และนักประเมินได้พยายามประยุกต์ใช้มาตรฐานค่าเชิงพฤติกรรมในกิจกรรมการประเมินในต่างประเทศ หลายลักษณะ โดยเฉพาะในการประเมินพฤติกรรมการปฏิบัติงานของบุคคล อาทิ ใช้ประเมินศักยภาพทางการบริหารธุรกิจ (Campbell and Others, 1973) ประเมินพฤติกรรมการปฏิบัติงานของลูกจ้าง (Kearney, 1979) ประเมินพฤติกรรมการปฏิบัติงานของเภสัชกร แรงจูงใจในการทำงานของวิศวกร ประเมินการสอนในระดับวิทยาลัย หรือ ประเมินผลการฝึกอบรม เป็นต้น (Witkin, 1994: 79) รวมทั้งใช้ในกิจกรรมการประเมินเพื่อกำหนดความต้องการจำเป็นในเรื่องต่าง ๆ (Witkin: 83-85)

3.3 ขั้นตอนการพัฒนามาตรประมาณค่าเชิงพฤติกรรม

การพัฒนามาตรประมาณค่าเชิงพฤติกรรม ต้องอาศัยกระบวนการในการพิจารณาตัดสินหลายขั้นตอน Landy และ Schwab (1974), Heneman และ DeCotiis (1975) ได้สรุปขั้นตอนการพัฒนามาตรประมาณค่าเชิงพฤติกรรม โดยได้เสนอแนะให้ดำเนินการตามลำดับขั้นดังนี้

(1) รวบรวมรายการเหตุการณ์หรือรายการพฤติกรรมที่แสดงออกเกี่ยวกับการปฏิบัติงานในหน้าที่อย่างเฉพาะเจาะจง ทั้งพฤติกรรมที่แสดงถึงความมีประสิทธิภาพ และพฤติกรรมด้อยประสิทธิภาพ จากกลุ่มบุคคลที่มีความรู้ ความเข้าใจเกี่ยวกับงานที่ประเมินเป็นอย่างดี

(2) จัดกลุ่มเหตุการณ์หรือพฤติกรรมการแสดงออกต่าง ๆ ออกเป็นกลุ่ม ๆ กำหนดชื่อ หรือนิยามความหมายของแต่ละกลุ่ม ถือเป็นมิติหรือด้าน (Dimensions) ที่ต้องการประเมิน ขณะเดียวกันก็กำหนดระดับของคุณลักษณะนั้น ๆ อย่างคร่าว ๆ ว่า มาก ปานกลาง หรือน้อยคืออย่างไร

(3) ตรวจสอบยืนยันการจัดกลุ่ม การกำหนดนิยาม หรือการให้ชื่อกลุ่ม ในข้อ (2) โดยใช้กลุ่มผู้ตัดสินอีกกลุ่มหนึ่ง ทั้งนี้เพื่อขจัดความกำกวม-ความสับสนของข้อความ และจัดรายการข้อความที่เห็นว่าไม่มีสาระออกไป พร้อมทั้งให้กลุ่มตัวอย่างกลุ่มนี้จัดกลุ่มรายการข้อความใหม่อีกครั้งหนึ่ง โดยตัดสินรายการข้อความแต่ละรายการว่าควรอยู่ในกลุ่มใด ในขั้นนี้อาจใช้เกณฑ์ความสอดคล้องของความคิดเห็น ระหว่างร้อยละ 50-80 ของจำนวนผู้ตัดสินเป็นเกณฑ์ในการตัดสินว่ารายการข้อความใดควรจะอยู่ในกลุ่มใด

(4) แปลงรายการพฤติกรรมให้อยู่ในรูปแบบของมาตรวัด โดยให้กลุ่มผู้ตัดสินกลุ่มที่ 3 ประเมินค่ารายการข้อความแต่ละรายการว่า ได้แสดงออกซึ่งคุณลักษณะ (Trait) ที่มุ่งประเมินในระดับมากน้อยเพียงใด แล้วทำการวิเคราะห์ค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของคะแนนผลการประเมิน ทำการเปรียบเทียบค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานระหว่างรายการข้อความต่าง ๆ แล้วเรียงลำดับรายการข้อความอย่างต่อเนื่องตามลำดับจากมากไปน้อย

(5) สร้างแบบประเมิน ซึ่งจะประกอบด้วยชุดของมาตรวัดตามแนวตั้งจำนวนหนึ่งตามจำนวนมิติที่สรุปได้ไปข้อ (3) ในแต่ละมาตรวัดไม่จำเป็นจะต้องเรียงลำดับรายการข้อความในลักษณะที่เป็นช่วงเท่ากัน (Equal Intervals) การทดสอบประสิทธิภาพของมาตรวัดอาจทำได้โดยใช้ผู้ประเมิน 3 คน ทำการประเมินพฤติกรรมของบุคคลใดบุคคลหนึ่ง แล้วนำผลการประเมินมาวิเคราะห์ค่าสหสัมพันธ์ และถือเป็นค่าความเที่ยงระหว่างผู้ประเมิน (Interrater Reliability)

ตอนที่ 4 งานวิจัยที่มุ่งพัฒนาและศึกษาคุณภาพของมาตรประมาณค่า

นักวิจัยหลายท่านได้พยายามพัฒนาและศึกษาคุณภาพของมาตรประมาณค่ารูปแบบต่าง ๆ ภายใต้สถานการณ์การศึกษาที่แตกต่างกัน เช่น ศึกษาในแง่ของจำนวนรายการคำตอบที่เหมาะสม ผลของการใช้คำคุณศัพท์ที่แตกต่างกัน การใช้แบบฟอร์มการตอบที่แตกต่างกัน เป็นต้น แนวทางและผลการศึกษาบางรายการ อาทิ

Campbell และคณะ (1973: 15-22) ได้พัฒนามาตรประมาณค่าเชิงพฤติกรรม สำหรับประเมินการปฏิบัติงานของผู้จัดการฝ่ายในห้างขายส่งสินค้า โดยมีมุ่งหวังผลในหลายลักษณะ คือ ต้องการพิสูจน์คุณค่าของมาตรวัดชนิดนี้ในแง่ของการใช้เพื่อการวิจัย ใช้ในการประเมินองค์ประกอบในการปฏิบัติงาน และการวิเคราะห์เพื่อกำหนดความต้องการจำเป็นในการพัฒนาบุคลากร จุดมุ่งหมายเฉพาะของการศึกษาครั้งนี้คือ ต้องการตรวจสอบว่าการใช้มาตรประมาณค่าเชิงพฤติกรรมประเมินการปฏิบัติงานของบุคคลนั้น จะเกิดความคลาดเคลื่อนเชิงปล่อยคะแนน (Leniency Error) และความคลาดเคลื่อนอันเนื่องมาจากการมีปฏิสัมพันธ์มาก่อน (Halo Error) น้อยกว่ามาตรประมาณค่ารูปแบบที่ไม่มีตัวอย่างพฤติกรรมผูกติดหรือไม่ ขณะเดียวกันก็ศึกษาความตรงตามภาวะสันนิษฐานที่เรียกว่า Convergent and Discriminant Validity ของแบบประเมินทั้งสองด้วย ในขั้นของการดำเนินงานวิจัย คณะผู้วิจัยได้พัฒนามาตรประมาณค่าเชิงพฤติกรรมขึ้นมาชุดหนึ่ง ประกอบด้วยมาตรวัด 9 มิติ (Dimensions) ตัวอย่างเช่น ด้านการให้คำปรึกษาแก่พนักงานขาย การจัดการแก้ไขข้อข้องใจของลูกค้า การดำเนินงานตามข้อตกลงที่วางไว้ เป็นต้น ในขณะเดียวกันก็พัฒนามาตรประมาณค่ารวมสรุปรูปแบบ (Summate Rating Scales) ตามแนวคิดของ Likert ชนิด 4 ทางเลือก ทั้งนี้ด้วยวิธีการกระจายรายการพฤติกรรมในแต่ละด้านของมาตรประมาณค่าเชิงพฤติกรรมและใช้เป็นรายการข้อความ (Items) ให้ผู้ตอบประเมินความถี่ของการเกิดพฤติกรรมนั้น ๆ โดยกำหนดระดับการประเมินตั้งแต่ 1-แทบไม่ปรากฏเลย ถึง 4-เกิดขึ้นบ่อยที่สุด กลุ่มเป้าหมายของการประเมิน (Subject for Scale Evaluation) คือ ผู้จัดการฝ่ายของร้านขายส่งสินค้าทั่วประเทศสหรัฐอเมริกา จำนวน 537 คน ผู้ประเมิน (Raters) คือ ผู้จัดการร้านค้าและผู้ช่วยผู้จัดการ กลุ่มเป้าหมายแต่ละคนจะถูกประเมินโดยผู้ประเมิน 2 คน ด้วยแบบประเมินทั้ง 2 ชุด แล้วนำคะแนนผลการประเมินมาวิเคราะห์หาค่าเฉลี่ยเลขคณิต (Mean) ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) แยกตามกลุ่มผู้ประเมิน และตามชนิดของมาตรประมาณค่า

วิเคราะห์ค่าสหสัมพันธ์ระหว่างคะแนนของผู้ประเมิน 2 กลุ่ม วิเคราะห์องค์ประกอบของคะแนน ผลการประเมิน และวิเคราะห์เมตริกซ์สหสัมพันธ์แบบ Multitrait-Multimethod ตามเกณฑ์ของ Campbell และ Fiske (1959) โดยถือว่ารายการพฤติกรรมทั้ง 9 ด้านที่ให้ประเมิน เป็นรายการพฤติกรรมที่แตกต่างกัน (Heterodimension) เสมือนเป็นคุณลักษณะที่แตกต่างกัน (Heterotrait) ผลการวิจัยพบว่า มาตรฐานค่าเชิงพฤติกรรมเกิดความคลาดเคลื่อนเชิงปล่อยคะแนน (Leniency Error) น้อยกว่ามาตรฐานค่ารวมสรุป กล่าวคือ ในการประเมินด้วยมาตรฐานค่ารวมสรุป ซึ่งในแต่ละด้านได้กำหนดคะแนนเต็ม 4.00 คะแนน ปรากฏว่า 6 ใน 9 ด้าน ให้ค่าเฉลี่ยผลการประเมินระหว่าง 3.00-4.00 คะแนน ในขณะที่คะแนนจากมาตรฐานค่าเชิงพฤติกรรม ในแต่ละด้านจะมีค่าเฉลี่ยประมาณ 6.00 คะแนน ซึ่งใกล้เคียงกับค่าเฉลี่ยปานกลางคือ 5.00 คะแนน และจากการวิเคราะห์องค์ประกอบ พบว่าคะแนนจากมาตรฐานค่าเชิงพฤติกรรม ให้ตัวประกอบที่แตกต่างกันถึง 9 ตัวประกอบ แต่ละตัวประกอบจะปรากฏด้านที่มีน้ำหนักบนตัวประกอบนั้นสูงเพียง 1 ด้านเท่านั้น ในขณะที่คะแนนจากมาตรฐานค่ารวมสรุป ไล่ตัวประกอบที่ต่างกัน 5 ตัวประกอบ และปรากฏลักษณะของตัวประกอบทั่วไป (General Factors) รวมอยู่ด้วย ซึ่งโดยนัยนี้ แสดงว่ามาตรฐานค่าเชิงพฤติกรรม ให้ผลการประเมินระหว่างด้านเป็นอิสระจากกัน มากกว่ามาตรฐานค่ารวมสรุป แสดงว่าเกิดความคลาดเคลื่อนอันเนื่องมาจากการมีปฏิสัมพันธ์มาก่อน (Halo Error) ในระดับที่น้อยกว่า ส่วนในด้านความตรงตามภาวะสันนิษฐานซึ่งดำเนินการโดยวิเคราะห์เมตริกซ์สหสัมพันธ์แบบ Multitrait - Multimethod เพื่อตรวจสอบ Convergent and Discriminant Validity นั้น พบว่ามาตรฐานค่าทั้งสองรูปแบบให้ผลการประเมินที่คมชัดเป็นไปตามเกณฑ์ของ Campbell และ Fiske โดยที่มาตรฐานค่าเชิงพฤติกรรมปรากฏ Discriminant Validity สูงกว่ามาตรฐานค่ารวมสรุป

Showers (1973: 5731-A) พบว่าในการประเมินพฤติกรรมการสอนของครู นักศึกษามักจะเลือกประเมินในระดับสเกลสูง ๆ เพียง 2-3 สเกลเท่านั้น ซึ่งเป็นสถานการณ์ที่เรียกว่าได้เกิด Leniency Bias ขึ้นในการประเมินครั้งนั้น ๆ ส่งผลให้การประเมินแต่ละครั้งไม่สามารถจำแนกความแตกต่างระหว่างครูแต่ละคนได้ และจะเกิดผลได้ผลเสียแก่ผู้ถูกประเมิน ถ้าผลการประเมิน ซึ่งปรากฏความแตกต่างเพียงเล็กน้อย ได้ถูกนำมารายงานในรูปของลำดับที่ (Ranking)

ที่แสดงถึงความแตกต่างอย่างชัดเจน เพื่อต้องการแก้ไขปัญหาดังกล่าวนี Showers จึงได้วางแผนศึกษาเปรียบเทียบเพื่อตรวจสอบข้อแอบอ้างที่สำคัญ 2 ประการคือ (1) มาตรฐานค่าประเภทที่ใช้คำคุณศัพท์เชิงประเมิน (Evaluative Cues) มีแนวโน้มที่จะเกิดอคติในการประเมินมากกว่ารูปแบบอื่น ๆ และ (2) แบบประเมินประเภทที่ใช้ตัวเลือกที่มีโอกาสเกิดความคลาดเคลื่อนมากกว่าการใช้ตัวเลือกแบบบรรยาย โดยในการศึกษาครั้งนี้ได้ทดลองใช้แบบประเมินชนิดมาตรฐานค่าที่แตกต่างกัน 3 รูปแบบ คือ (1) แบบตัวเลือกของลิเคอร์ต (Fixed Alternative Likert Cues) (2) แบบตัวเลือกที่ใช้คำคุณศัพท์เชิงประเมิน (Fixed Alternative Evaluation Cues - Superior - Inferior) และ (3) แบบทางเลือกเชิงบรรยาย (Descriptive Cues) ผลการศึกษาพบว่า แบบที่ 2 คือแบบตัวเลือกที่ใช้คำคุณศัพท์เชิงประเมินเกิดอคติน้อยที่สุด ส่วนแบบลิเคอร์ตซึ่งเป็นแบบที่นิยมใช้กันมากมีแนวโน้มที่จะเกิดอคติในการประเมินสูงสุด อย่างไรก็ตาม เมื่อเปรียบเทียบค่าความเที่ยงของผู้ประเมิน (Rater Reliability) พบว่าไม่แตกต่างกัน

Pierson (1975: 2718-A) ได้ทำการศึกษาเปรียบเทียบเพื่อตรวจสอบว่าในการวัดทัศนคติโดยใช้เทคนิคนี้จำแนก (The Semantic Differential Technique) กับ มาตรฐานจำแนกแบบกราฟิก (Graphic Differential) รูปแบบใดให้ค่าความเที่ยงสูงกว่า โดยได้ทดลองใช้ประเมินทัศนคติของนักเรียนเกรดหก จำนวน 585 คน ที่ได้จำแนกระดับความสามารถในการอ่านเป็น 3 กลุ่ม คือ สูง กลาง และ ต่ำ ผลการศึกษาพบว่าค่าความเที่ยงของแบบวัดทั้งสองรูปแบบนี้ขึ้นอยู่กับทักษะในการอ่านของเด็ก และแบบวัดที่ใช้เทคนิคนี้จำแนกให้ค่าความเที่ยงสูงกว่า ในการนี้ Pierson ได้อภิปรายว่า น่าจะเป็นเพราะว่าในการประเมินโดยใช้มาตรฐานจำแนกแบบกราฟิก นั้น ผู้ประเมินอาจไม่เข้าใจอย่างถ่องแท้ในความเป็นนามธรรมของระดับประเมิน ส่งผลให้ไม่สามารถประเมินอย่างคงเส้นคงวาได้ และได้เสนอแนะว่า น่าจะทำการศึกษาวิจัยเพื่อพัฒนามาตรวัดแบบกราฟิกในลักษณะที่เป็นรูปธรรม (Concrete Graphic) ใช้ในกิจกรรมการประเมิน และในการพัฒนามาตรวัดนั้น น่าจะเน้นการใช้คำศัพท์ที่อยู่ในโดเมนความรู้ของกลุ่มผู้ที่จะประเมินด้วย

Slater (1976: 6203-A) ได้พยายามศึกษาเพื่อตรวจสอบว่า Leniency Error เป็นผลเกี่ยวเนื่องกับรูปแบบในการประเมินหรือไม่ โดยได้พัฒนามาตรประมาณค่า 2 รูปแบบ คือ แบบแรกเป็นแบบที่ใช้คำคุณศัพท์ผูกติดกับระดับการประเมิน (Adjective-Anchored Scales) และแบบที่สอง เป็นแบบที่ใช้ตัวอย่างพฤติกรรมผูกติดกับระดับการประเมินในแต่ละสเกล (Behaviorally-Anchored Scales) ผู้วิจัยได้นำแบบประเมินทั้งสองรูปแบบไปใช้ในการประเมินพฤติกรรมการสอนของศาสตราจารย์จำนวน 15 คน ผู้ประเมินคือ นักศึกษาในความรับผิดชอบของศาสตราจารย์เหล่านั้น โดยสุ่มครึ่งหนึ่งให้ประเมินโดยใช้เครื่องมือรูปแบบแรก และอีกครึ่งหนึ่งใช้รูปแบบหลัง นำข้อมูลมาวิเคราะห์ความแปรปรวนแบบ 3 ทาง ผลการวิจัย พบว่า ไม่มีรูปแบบใดที่สามารถลดอคติได้สำเร็จ ส่วนความเที่ยงของคะแนนจากรูปแบบทั้งสอง พบว่า แบบที่มีคำคุณศัพท์บนสเกลให้ค่าความเที่ยงสูงกว่าเล็กน้อย 3 ใน 5 ของจำนวนมิติพฤติกรรมที่ให้ประเมิน

Keaveny และ McGann (1975: 695-703) ได้พัฒนาและศึกษาเปรียบเทียบประสิทธิภาพของมาตรคาดหวังเชิงพฤติกรรม (Behavioral Expectation Scales) กับมาตรประมาณค่าแบบกราฟิกในการประเมินพฤติกรรมการสอนของศาสตราจารย์ 4 คน โดยนักศึกษาที่ลงทะเบียนเรียนกับศาสตราจารย์เหล่านั้นเป็นผู้ประเมิน ในขั้นตอนการดำเนินการวิจัย ผู้วิจัยได้ พัฒนามาตรวัดขึ้นมา 2 รูปแบบ คือ มาตรคาดหวังเชิงพฤติกรรม กับ มาตรประมาณค่าแบบกราฟิก โดยมาตรวัดทั้งสองรูปแบบ ประกอบด้วยรายการที่ให้ประเมินที่เหมือนกัน จำนวน 13 รายการพฤติกรรม จะแตกต่างกันเฉพาะในส่วนของสเกลคำตอบเท่านั้น แล้วนำมาตรวัดทั้งสองรูปแบบไปให้นักศึกษา 4 กลุ่ม ทำการประเมินพฤติกรรมของผู้สอน แล้วทำการวิเคราะห์ค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของคะแนนผลการประเมิน วิเคราะห์เมตริกซ์สหสัมพันธ์แบบ Multitrait - Multimethod เพื่อตรวจสอบ Convergent & Discriminant Validity ของมาตรวัดทั้ง 2 รูปแบบ โดยถือว่า รายการพฤติกรรมที่ต่างกัน 13 รายการพฤติกรรมเป็นเสมือนคุณลักษณะที่ต่างกัน (Heterotrait) ผลการวิจัยพบว่า มาตรคาดหวังเชิงพฤติกรรมเกิด Halo Error น้อยกว่ามาตรประมาณค่าแบบกราฟิก ทั้งนี้โดยพิจารณาจากค่าสหสัมพันธ์ระหว่างคู่รายการพฤติกรรมใด ๆ เปรียบเทียบระหว่างกรณีใช้มาตรคาดหวังเชิงพฤติกรรมเป็นเครื่องมือประเมินกับกรณีใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมิน ซึ่ง

ปรากฏว่าค่าสหสัมพันธ์ภายใต้การใช้มาตรคาคหมายเชิงพฤติกรรมมีค่าน้อยกว่า อันแสดงถึงความ เป็นอิสระจากกันของผลการประเมินแต่ละรายการพฤติกรรม ในส่วนการตรวจสอบ Leniency Error จากการเปรียบเทียบค่าเฉลี่ยเลขคณิตของผลการประเมินรายการพฤติกรรมต่าง ๆ ระหว่าง กรณีใช้มาตรวัดต่างรูปแบบกัน ปรากฏข้อสรุปว่ามาตรคาคหมายเชิงพฤติกรรมไม่สามารถลด Leniency Error ได้ และจากการวิเคราะห์เมตริกซ์สหสัมพันธ์แบบ Multitrait- Multimethod พบว่า มาตรคาคหมายเชิงพฤติกรรมให้ความตรงเชิงจำแนก (Discriminant Validity) สูง กว่ามาตรประมาณค่าแบบกราฟิก

Cook (1988: 31-45) ได้ศึกษาผลของการใช้มาตรประมาณค่าที่มีรูปแบบแตกต่างกัน 2 รูปแบบ คือ มาตรประมาณค่าเชิงพฤติกรรม (Behaviorally-Anchored Rating Scales) กับ มาตรประมาณค่าแบบลิเคอร์ต (Likert Scales) และผลของการฝึกทักษะแก่ผู้ประเมินต่อค่า ความคลาดเคลื่อนเชิงปล่อยคะแนน (Leniency Error) และความคลาดเคลื่อนอันเนื่องมาจากการเคยมีปฏิสัมพันธ์มาก่อน (Halo Error) โดยผู้วิจัยได้ใช้มาตรวัด 2 รูปแบบดังกล่าว ให้ นักศึกษา 2 กลุ่ม คือ กลุ่มที่ได้รับการฝึกทักษะกับกลุ่มที่ไม่ได้รับการฝึกทักษะ ทำการประเมิน พฤติกรรมการสอนของศาสตราจารย์ 1 ท่าน แล้วทำการวิเคราะห์หาค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของคะแนนผลการประเมินแต่ละมิติ เปรียบเทียบระหว่างกลุ่มที่ได้รับการฝึกทักษะ กับกลุ่มที่ไม่ได้รับการฝึกทักษะ พร้อมทั้งเปรียบเทียบคะแนนผลการประเมินระหว่างมาตรวัดต่าง รูปแบบกันเฉพาะในกลุ่มที่ได้รับการฝึกทักษะ โดยในการตรวจสอบ Halo Error ผู้วิจัยได้ใช้ ส่วนเบี่ยงเบนมาตรฐานของคะแนนผลการประเมิน (Standard Deviation) เป็นคะแนนบ่งชี้ ความคลาดเคลื่อน (Halo Score) ภายใต้เงื่อนไขว่า ถ้าคะแนนความคลาดเคลื่อน (Halo Score) มีค่ามาก แสดงถึงการเกิด Halo Error น้อยกว่ากรณีที่คะแนนความคลาดเคลื่อนมีค่าน้อย ผลการวิจัยพบว่า การฝึกทักษะมีผลต่อการลด Halo Error การประเมิน โดยมาตรประมาณค่าเชิงพฤติกรรมเกิด Leniency Error น้อยกว่า มาตรประมาณค่าแบบ ลิเคอร์ต ในขณะที่มาตรประมาณค่าแบบลิเคอร์ตเกิด Halo Error น้อยกว่า

สำหรับในประเทศไทย ได้มีผู้ศึกษาเพื่อตรวจสอบคุณภาพของมาตรฐานค่าในสถาน การณ์การประเมินบางลักษณะ อาทิ

ธารทิพย์ ประเสริฐสม (2522: 38-42) ได้ศึกษาความเที่ยงของมาตรวัดทัศนคติ แบบลิเคอร์ท์ที่มีจำนวนรายการคำตอบแตกต่างกัน คือ 2 3 4 5 6 และ 7 รายการ ในสถานการณ์การประเมินทัศนคติต่อวิชาชีพครู และทัศนคติต่อการสอบสัมภาษณ์ของนักศึกษาระดับ ประกาศนียบัตรวิชาชีพการศึกษาระดับสูง ปีที่ 1 ผลการวิจัยสรุปว่า จำนวนรายการคำตอบ 3 ถึง 7 รายการให้ค่าความเที่ยงไม่ต่างกัน รายการคำตอบ 6 หรือ 7 รายการมีแนวโน้มให้ค่าความ เที่ยงสูงกว่าแบบประเมินที่มีรายการคำตอบ 2 ถึง 5

มัลลิกา ยูวะนะเตมีย์ (2527: 40-41) ได้ศึกษาค่าความเที่ยงของแบบสอบถาม มาตรฐานค่าที่มีรูปแบบคำตอบแตกต่างกัน 6 รูปแบบ คือ (1) แบบทางเลือกเป็นตัวอักษร ก. ถึง จ. (2) ทางเลือกเป็นรูปสี่เหลี่ยมมีขนาดต่างกันจากใหญ่ไป เล็ก (3) แบบบรรยายด้วย ข้อความเชิงปริมาณแบบคงที่ จาก "มากที่สุด" ถึง "น้อยที่สุด" (4) แบบทางเลือกเป็นตัวเลข 5 ระดับ จาก 5-มากที่สุด/น้อยที่สุด ถึง 1-น้อยที่สุด/ไม่เคยเลย (5) แบบทางเลือกที่เป็นตัวเลข 5 ระดับ แต่กลับทิศทางของความหมายในทางตรงข้ามกับแบบที่ 4 และ (6) ทางเลือกแบบกราฟิก สถานการณ์ในการประเมินเป็นการประเมินประสิทธิภาพการสอนของครู และการสำรวจนิสัยและ ทัศนคติในการเรียนระดับมัธยมศึกษาตอนปลาย ผลการวิจัยพบว่า แบบกราฟิก เป็นแบบที่ให้ค่า ความเที่ยงสูงที่สุด

นินา พานสมบัติ (2527: 50-57) ได้ศึกษาความแม่นยำตรงในการประเมินค่าพฤติกรรม ใฝ่สัมฤทธิ์ของครูในโรงเรียนประถมศึกษาที่มีทัศนคติต่อการวัดผลและประเมินผลการเรียนแตกต่างกัน สถานการณ์ในการประเมินคือให้ครู 15 คน ประเมินพฤติกรรมใฝ่สัมฤทธิ์ของนักเรียนจำนวน 360 คน แล้วนำผลการประเมินมาวิเคราะห์ความสัมพันธ์กับคะแนนผลสัมฤทธิ์ทางการเรียนของ นักเรียน ถือเอาค่าสหสัมพันธ์เป็นดัชนีบ่งชี้ความแม่นยำ แล้วเปรียบเทียบความแม่นยำระหว่าง กลุ่มครูที่มีทัศนคติต่อการวัดผลและประเมินผลแตกต่างกัน 3 ระดับ คือ สูง ปานกลาง และต่ำ

ผลการวิจัยพบว่า การประเมินค่าพฤติกรรมไม่สัมพันธ์โดยครูที่มีทัศนคติต่อการวัดผลและประเมินผลในระดับสูง มีความแม่นยำสูงกว่าการประเมินโดยครูที่มีทัศนคติต่อการวัดผลและประเมินผลการเรียนในระดับปานกลาง และต่ำ

อารยา ตั้งคินิวาส (2529: 30-39) ได้ศึกษาเปรียบเทียบค่าความเที่ยงและความตรงร่วมสมัยของมาตรวัดทัศนคติแบบลิเคอร์ที่มีสัดส่วนจำนวนข้อกระทงที่เป็นนิเสธแตกต่างกัน 3 ระดับ คือ ประกอบด้วยข้อกระทงที่เป็นนิเสธ 75 % 50% และ 25 % สถานการณ์การประเมินคือ ประเมินทัศนคติต่อวิชาชีพครูของนิสิตชั้นปีที่ 1 ของคณะครุศาสตร์ หรือ ศึกษาศาสตร์ ของมหาวิทยาลัย ส่วนกลาง 4 แห่ง ผลการวิจัย พบว่า มาตรวัดทัศนคติต่อวิชาชีพครูที่มีจำนวนรายการข้อกระทงที่เป็นนิเสธแตกต่างกัน ให้ค่าความเที่ยงและความตรงในการประเมินไม่แตกต่างกัน

จากการศึกษางานวิจัยที่มุ่งพัฒนาหรือตรวจสอบคุณภาพของมาตรประมาณค่าภายใต้เงื่อนไขการประเมินต่าง ๆ ทั้งในต่างประเทศและในประเทศไทย ผู้วิจัยเห็นว่ามิประเด็นที่น่าสังเกตดังนี้

1) งานวิจัยในต่างประเทศจำนวนมาก มุ่งเน้นการพัฒนาความคมชัดของรูปแบบสเกล (Scales Format) และตรวจสอบคุณภาพในเชิงเปรียบเทียบกับมาตรประมาณค่าโดยทั่วไปที่เป็นที่นิยมใช้ เช่น Campbell (1973) มุ่งพัฒนาและตรวจสอบคุณภาพของมาตรประมาณค่าเชิงพฤติกรรมเปรียบเทียบกับแบบลิเคอร์ท Showers (1973) ได้พยายามศึกษาคุณภาพมาตรประมาณค่าแบบทางเลือกเชิงบรรยาย Slater (1976) ได้พยายามตรวจสอบคุณภาพของมาตรประมาณค่าแบบทางเลือกที่ใช้คำคุณศัพท์ผูกติดไว้บนสเกลกับแบบที่ใช้ตัวอย่างพฤติกรรมผูกติดไว้บนสเกล Keaveny และ McGann (1975) ได้พัฒนาและศึกษาคุณภาพของมาตรประมาณค่าเชิงพฤติกรรมเปรียบเทียบกับมาตรประมาณค่าแบบกราฟิก เป็นต้น ซึ่งภายใต้ความพยายามที่จะพัฒนาหรือตรวจสอบคุณภาพของสเกลแบบต่าง ๆ นี้ พบว่า มาตรประมาณค่าเชิงพฤติกรรมเป็นรูปแบบที่ได้รับความสนใจจากนักวิจัยจำนวนมาก

2) งานวิจัยในต่างประเทศส่วนใหญ่ได้เน้นการศึกษาวิจัยหรือตรวจสอบคุณภาพของมาตรประมาณค่าในสถานการณ์ของการให้บุคคลหนึ่งประเมินพฤติกรรมของอีกบุคคลหนึ่ง โดยสถานการณ์ที่ได้รับความสนใจศึกษากันอย่างแพร่หลายที่สุด คือ การให้นักศึกษาประเมินพฤติกรรมของอาจารย์

3) ตัวแปรตาม (Dependent Variable) ที่กล่าวได้ว่าได้รับความสนใจค่อนข้างสูงในการเปรียบเทียบคุณภาพของมาตรประมาณค่า ในงานวิจัยในต่างประเทศ คือ ความคลาดเคลื่อนในการประมาณค่าที่เรียกว่า Leniency Error และ Halo Error ซึ่งจากผลของการทบทวนงานวิจัยในต่างประเทศของ Cook (1988: 32) ก็ได้ให้ข้อสังเกตในลักษณะนี้เช่นกัน

4) งานวิจัยในต่างประเทศจำนวนมากที่มุ่งพัฒนาและศึกษาคุณภาพของมาตรประมาณค่าเชิงพฤติกรรมในสถานการณ์ของการประเมินที่แตกต่างกัน ยังปรากฏข้อสรุปที่ไม่สอดคล้องกัน ซึ่งในประเด็นนี้ จากการศึกษาของ Kingstrom และ Bass (1981, quoted in Cook, 1989: 32) ที่ได้ทบทวนงานวิจัยในต่างประเทศจำนวน 21 เรื่อง ที่ดำเนินการระหว่างปี 1965 ถึง 1980 ก็ได้ข้อสรุปว่า "ยังไม่ปรากฏหลักฐานที่ชัดเจนว่ามาตรประมาณค่าเชิงพฤติกรรมจะมีคุณภาพเหนือกว่ามาตรประมาณค่ารูปแบบอื่น ๆ "

5) ในส่วนของงานวิจัยในประเทศไทย พบว่า งานวิจัยที่ได้ทำการศึกษาเกี่ยวกับคุณภาพของมาตรประมาณค่า ยังมีไม่มากนัก เป้าหมายของการศึกษายังกระจัดกระจาย เช่น ศึกษาจำนวนรายการคำถามที่เหมาะสม (ธารทิพย์ ประเสริฐสม, 2522) ศึกษารูปแบบของคำตอบที่เน้นหนักไปในเชิงของการให้สัญลักษณ์ (มัลลิกา ยวนะเทมีย์, 2527) ศึกษาผลการประเมินที่เกิดจากผู้ประเมิน ที่มีภูมิหลังต่างกัน (นินา พานสมบัติ, 2527) เป็นต้น งานวิจัยในเชิงพัฒนาความคมชัดของสเกลยังไม่ปรากฏให้เห็นอย่างชัดเจน

6) การศึกษาคุณภาพของมาตรประมาณค่า ในงานวิจัยในประเทศไทย ส่วนใหญ่จะเน้นการศึกษาความเที่ยง (Reliability) เป็นสำคัญ การศึกษาคุณภาพในด้านความตรง (Validity) มีปรากฏบ้าง โดยการใช้ดัชนีค่าสหสัมพันธ์ระหว่างคะแนนจากเครื่องมือที่พัฒนาขึ้นกับคะแนนจากเครื่องมืออื่นที่มีผู้พัฒนาไว้แล้ว เป็นดัชนีบ่งชี้ค่าความตรง

ตอนที่ 5 แนวความคิดเกี่ยวกับการประเมินผลกิจกรรมการปฏิบัติงานของบุคคล

5.1 บทบาทของการประเมินผลกิจกรรมการปฏิบัติงานในการบริหารบุคคลโดยทั่วไป
การบริหารบุคคลเป็นส่วนหนึ่งของศาสตร์ที่ว่าด้วยการบริหารและการจัดการ ทั้งนี้ด้วยมีผู้ให้นิยามการบริหารหรือการจัดการ (Management) ว่า "เป็นการดำเนินงานเพื่อให้ได้มาซึ่งจุดประสงค์ที่ต้องการด้วยการใช้ความสามารถของบุคคลในองค์การ" (Crane, 1985: 19) ซึ่งโดยนัยของนิยามนี้ได้ชี้ให้เห็นว่า "บุคคล" ถือเป็นส่วนสำคัญยิ่งของกระบวนการบริหาร ยิ่งในภาวะปัจจุบันที่การบริหารกิจการใด ๆ ต้องดำเนินการภายใต้ความจำกัดในด้านทรัพยากร รวมทั้งภาวะการแข่งขันที่ค่อนข้างเข้มงวดด้วยแล้ว การบริหารงานบุคคลที่มีประสิทธิภาพก็เป็นที่ต้องการขององค์การมากยิ่งขึ้น จนมีการกล่าวว่า "การบริหารบุคคลเป็นสิ่งจำเป็นอย่างยิ่งสำหรับการอยู่รอดขององค์การในปัจจุบัน" (Crane, 1985: 4)

โดยทั่วไปแล้ว วงจรการบริหารงานบุคคลประกอบด้วย (1) การสรรหาบุคคลมาปฏิบัติงาน (2) การบรรจุและแต่งตั้ง (3) การประเมินผลการปฏิบัติงาน และ (4) การให้บุคคลพ้นจากตำแหน่ง (สมาน รั้งสิโยกฤษฏ์, 2530: 1) ซึ่งตามวงจรดังกล่าวนี้ กิจกรรมการประเมินได้เข้าไปมีบทบาทต่อกระบวนการบริหารงานบุคคลในเกือบทุกขั้นตอน นับตั้งแต่การคัดเลือกบุคคลที่จะต้องมีการประเมินความสามารถหรือทักษะต่าง ๆ หรือในการที่จะมอบหมายงาน เลื่อนตำแหน่ง ให้รางวัลตอบแทน ตลอดจนการพัฒนาความสามารถของบุคคล ก็ล้วนต้องอาศัยกิจกรรมการประเมินเข้าไปมีส่วนช่วยทั้งสิ้น อย่างไรก็ตามเมื่อกล่าวถึงการประเมินบุคคลในแง่ของการประเมินผลกิจกรรมการปฏิบัติงานแล้ว ส่วนใหญ่เราจะนึกถึงการประเมินเพื่อบ่งชี้ว่าบุคลากรปฏิบัติงานได้ผลดีหรือไม่อย่างไร ซึ่งเป็นการเน้นเฉพาะผลกิจกรรมการปฏิบัติงานเป็นสำคัญ โดยมีจุดมุ่งหมายที่จะนำผลการประเมินไปใช้ในการตัดสินใจเชิงบริหารในลักษณะใดลักษณะหนึ่งต่อไปนี้ (Mathis, 1979: 290-291)

1. เพื่อการตอบแทนผลการดำเนินงาน (Compensation) ซึ่งถือว่าผู้ที่สามารถปฏิบัติงานได้ตามเป้าหมายหรือเกินกว่าเป้าหมายควรจะได้รับผลตอบแทนในผลการดำเนินงาน
2. เพื่อจัดวางตัวบุคคล (Placement) เป็นการนำข้อมูลผลการประเมินเพื่อการจัดบุคคล หรือมอบหมายงานให้ตรงกับความสามารถ หรือเพื่อการเลื่อนตำแหน่ง ปรับเปลี่ยนหน้าที่การปฏิบัติงาน

3. เพื่อการฝึกอบรม หรือพัฒนาบุคลากร (Training and Development) ซึ่งถือเป็นแนวทางที่สำคัญยิ่งประการหนึ่งในการปรับปรุงประสิทธิภาพการทำงาน หรือเพิ่มคุณภาพของผลผลิต การประเมินพฤติกรรมการทำงานจะช่วยให้ฝ่ายส่งเสริมหรือพัฒนาบุคลากรทราบลักษณะเด่นในการปฏิบัติงานของบุคลากร ทราบจุดอ่อนในการปฏิบัติงาน ทำให้สามารถวิเคราะห์ความต้องการจำเป็นในการพัฒนาบุคลากร (Training Needs) ได้

ในปัจจุบัน พบว่า ได้มีความพยายามที่จะใช้ผลการประเมินเพื่อการตัดสินใจในเชิงบริหาร เกี่ยวกับงานบุคคลในหลายรูปแบบ โดยเฉพาะในระบบราชการของไทย ได้ปรากฏลักษณะการประเมินเพื่อตอบแทนผลการดำเนินงาน และการประเมินเพื่อวิเคราะห์ความต้องการจำเป็นในการพัฒนาบุคลากรกันอย่างแพร่หลายมากขึ้น ในส่วนของการประเมินเพื่อตอบแทนผลการดำเนินงาน ซึ่งถือเป็นลักษณะหนึ่งของการประเมินในระบบคุณธรรม ที่ยึดถือความสามารถและผลงานเป็นสำคัญนั้น ได้ปรากฏให้เห็นอย่างชัดเจนในการประเมินผลการดำเนินงานเพื่อเลื่อนระดับทั้งหลาย อาทิ การประเมินเพื่อเลื่อนระดับอาจารย์ 2 เป็นอาจารย์ 3 การประเมินเพื่อเลื่อนระดับเป็นผู้ช่วยศาสตราจารย์ รองศาสตราจารย์ หรือ ศาสตราจารย์ เป็นต้น หรือในส่วนของ การประเมินสภาพการปฏิบัติงาน ปัญหาการปฏิบัติงานของบุคคล เพื่อการวิเคราะห์ความต้องการจำเป็นในการพัฒนา และจัดทำเป็นโครงการพัฒนาบุคลากรของหน่วยฝึกอบรม กองการเจ้าหน้าที่ทั้งหลาย ก็มีปรากฏให้เห็นกันอย่างแพร่หลายทั่วไป

5.2 รูปแบบการประเมินพฤติกรรมการทำงาน/ผลการปฏิบัติงาน

(1) ผู้ประเมิน

ในการประเมินพฤติกรรมการทำงานของบุคลากรนั้น คำถามสำคัญประการหนึ่งที่เป็นที่สนใจของนักประเมิน นักการศึกษาโดยทั่วไปคือ ใครควรเป็นผู้ประเมิน ซึ่งในประเด็นนี้ได้มีผู้เสนอแนวคิดไว้หลายท่าน อาทิ

Crane (1985: 401-402, 407-408) กล่าวว่า ในทางธุรกิจ หรือในระบบอุตสาหกรรมนั้น โดยปกติ หัวหน้างานหรือผู้บังคับงานระดับต้น (Immediate Supervisor) เป็นผู้ทำหน้าที่ในการประเมินพฤติกรรมการทำงานของเจ้าหน้าที่ที่อยู่ในความควบคุมดูแลของตน ส่วนการประเมินพฤติกรรมการทำงานของบุคคลที่อยู่ในตำแหน่งสูงกว่า ก็อาจให้ผู้ใต้บังคับบัญชา

ลูกจ้าง หรือลูกค้า เป็นผู้ประเมินก็ได้ ซึ่งเรียกการประเมินในลักษณะนี้ว่า Inverted Appraisal อันเป็นแนวความคิดที่ปรากฏให้เห็นในการดำเนินงานของประเทศในยุโรปหลายประเทศ เช่น การประเมินประสิทธิภาพของผู้สอนโดยให้นักศึกษาเป็นผู้ประเมิน เป็นต้น อีกทั้งในการประเมินพฤติกรรมการทำงานนั้นอาจดำเนินการลักษณะอื่น เช่น การใช้กลุ่มเพื่อนเป็นผู้ประเมิน (Buddy Rating or Peer Rating) เป็นการให้สมาชิกในกลุ่มประเมินซึ่งกันและกัน การใช้กลุ่มผู้บริหารหรือหัวหน้างานเป็นผู้สรุปเชิงประเมินสภาพการทำงานของบุคลากรที่อยู่ในความรับผิดชอบ (The Group Summary Appraisal) ซึ่งเป็นวิธีการที่เป็นประโยชน์อย่างมากต่อการวางแผนพัฒนาบุคลากรขององค์การ หรืออาจใช้วิธีการให้บุคลากรประเมินตนเอง (The Self-appraisal Process) ซึ่งจะเป็นประโยชน์ในแง่ของการเกิดความตระหนักในผลการปฏิบัติงานของตนเอง และจะนำไปสู่การปรับปรุงตนเองให้มีประสิทธิภาพมากยิ่งขึ้นในโอกาสต่อไป

Mathis (1979: 298-301) ได้สรุปวิธีการในการประเมินพฤติกรรมการทำงานที่นิยมใช้มี 4 รูปแบบคือ

1. วิธีการประเมินโดยกลุ่มผู้ใต้บังคับบัญชา (Subordinate Rating Approach) เป็นการประเมินผู้บังคับบัญชา หรือผู้ที่อยู่เหนือกว่าโดยกลุ่มผู้ใต้บังคับบัญชา หรือกลุ่มสมาชิกในความดูแล เช่น การประเมินพฤติกรรมของอาจารย์ผู้สอนโดยกลุ่มนักศึกษามีประโยชน์ในการเปิดโอกาสให้ผู้บังคับบัญชาได้พัฒนาตนเอง โดยเฉพาะอย่างยิ่งในแง่ของศักยภาพความเป็นผู้นำ
2. วิธีการประเมินโดยกลุ่ม (Group Rating Approach) ซึ่งได้จำแนกรูปแบบการประเมินโดยยึดเกณฑ์การตัดสินใจเป็นกลุ่มเป็น 2 ลักษณะที่สำคัญคือ (1) สมาชิกแต่ละคนประเมินอย่างเป็นอิสระแยกจากกัน หรือต่างคนต่างประเมิน แล้วค่อยรวมสรุปผลการประเมิน เรียกวิธีการนี้ว่า Multiple Rating หรือ (2) การประเมินโดยคณะกรรมการ (Committee Appraisal) เป็นการประเมินบุคคลใดบุคคลหนึ่งโดยใช้มติที่ประชุมอย่างเป็นทางการของคณะกรรมการ การประเมินโดยกลุ่มบุคคลนี้อาจจำแนกเป็นรูปแบบย่อยได้อีกโดยยึดกลุ่มผู้ประเมิน เช่น ใช้กลุ่มผู้บังคับบัญชา หรือกลุ่มเพื่อน เป็นต้น

3. วิธีการให้ผู้บังคับบัญชาประเมินผู้ใต้บังคับบัญชา (Superior Rating of Subordinates) โดยถือว่าผู้บังคับบัญชาคือผู้ที่เหมาะสมที่สุดที่จะทำหน้าที่ประเมินได้อย่างถูกต้อง เป็นจริง มีความเป็นปรนัย และยุติธรรม เป็นวิธีการที่นิยมใช้กันอย่างแพร่หลายมากที่สุดในการธุรกิจ

4. วิธีการประเมินตนเองโดยยึดเป้าหมายงานเป็นเกณฑ์ (Self-Appraisal through the use of Management by Objectives) เป็นการเปิดโอกาสให้บุคลากรประเมินผลงานของตนเองโดยพิจารณาเป้าหมายผลงานที่ตั้งไว้เป็นเกณฑ์

ในด้านการประเมินพฤติกรรมการสอน หรือประสิทธิภาพการสอนของครูนั้น ได้มีผู้แสดงทัศนะไว้ดังนี้

Dressel (1977: 1486-1487), Harris (1986: 102) และ Lemlech (1976: 262) มีความเห็นสอดคล้องกันว่า ผู้ทำการประเมินการสอนของครูควรประกอบด้วย (1) นักเรียน (2) เพื่อนครู หรือเพื่อนร่วมงาน (3) ครูผู้สอนประเมินตนเอง และ (4) ผู้บริหาร หรือ ผู้มีเทศก์ ซึ่งแนวคิดนี้สอดคล้องกับข้อเสนอของนักการศึกษาของไทยหลายท่าน อาทิ อุทุมพร จามรมาน (2530: 101) ทองจันทร์ หงส์ลดาธรมภ์ (2524: 2) บุญธรรม กิจปรีดาบริสุทธิ์ (2526: 12) ซึ่งเห็นว่าบุคคลที่ควรทำหน้าที่ประเมินการสอนหรือประเมินอาจารย์ ได้แก่ (1) ผู้เรียน/นักศึกษา (2) เพื่อนร่วมงาน/เพื่อนอาจารย์ (3) ตัวผู้สอน/ตัวอาจารย์เอง และ (4) ผู้เชี่ยวชาญทางด้าน การสอน หรือคณะกรรมการเฉพาะกิจ หรือผู้บริหาร ผู้ที่เกี่ยวข้อง รวมทั้ง ไนสุรีย์ สีนลารัตน์ (2524: 96-99) ได้เสนอแหล่งที่มาของข้อมูลการประเมินที่เหมาะสมสำหรับการประเมินการสอนในลักษณะที่สอดคล้องกับนักการศึกษาอื่น ๆ ที่กล่าวมาแล้ว โดยให้เหตุผลประกอบสำหรับการใช้ข้อมูลจากแต่ละแหล่ง ดังนี้

(1) อาจารย์ผู้สอนเอง เนื่องจากผู้สอนเป็นบุคคลที่สำคัญที่สุดในการสอน และการปรับปรุงการสอน ผู้สอนสามารถมองเห็น เข้าใจ และรู้จักตนเองดีพอในการประเมินตนเอง แต่ที่อาจมีบางครั้งที่ทำให้เกิดความลำเอียงได้

(2) เพื่อนร่วมงาน โดยหลักการแล้ว การได้ข้อมูลจากเพื่อนร่วมงานเป็นสิ่งที่ดีและควรกระทำ เพราะเพื่อนร่วมงานเข้าใจงานสอนของเราได้ดี

(3) ผู้บังคับบัญชา ข้อมูลที่ได้จากผู้บังคับบัญชา มักจะเป็นการประเมินเพื่อพิจารณาความดีความชอบ ถ้าผู้บริหารสนใจติดตามการสอนโดยตลอดก็จะเป็นแหล่งข้อมูลที่ดี

(4) ผลจากการสอบ เมื่อการสอนดำเนินไปอย่างมีประสิทธิภาพ มีจุดมุ่งหมายชัดเจน มีระบบการประเมินที่ดีแล้ว ผลการสอบ หรือสารสนเทศที่ปรากฏก็สามารถใช้อ้างอิงได้

(5) ตัวผู้เรียน ซึ่งอาจให้ข้อมูลที่น่าสนใจและเป็นประโยชน์มาก เมื่อผู้เรียนเป็นผู้รู้ และรับผลจากผู้สอนโดยตรงตลอดเวลา

ในกรณีของการใช้แหล่งข้อมูลการประเมินจากนักเรียนนั้น ไนทอร์ย ลินลาร์ตัน (2524: 99) เห็นว่า ถึงแม้ว่าการประเมินโดยผู้เรียนจะให้ข้อมูลที่ดีและน่าสนใจ แต่ก็มีคำวิพากษ์วิจารณ์อยู่หลายประการ เป็นต้นว่า ผู้เรียนยังไม่มีวุฒิภาวะเพียงพอ ผู้เรียนมักจะประเมินผู้สอนที่มีลักษณะใจดีออกมาในทางบวก ทำให้เกิดการปกป้องตนเองของผู้สอนเมื่อผลการประเมินออกมาต่ำ หรือทำให้ผู้เรียนไม่มีสัมมาคารวะต่อผู้สอน เป็นต้น อีกทั้ง Aleamoni (1983: 111-115) ก็ได้เสนอแนะว่า วิธีการประเมินโดยผู้เรียนแม้จะเป็นแหล่งข้อมูลที่สำคัญของวิทยาลัย หรือมหาวิทยาลัย แต่ในการใช้ผลการประเมินก็ต้องระมัดระวังในการแปลผล ทั้งนี้เนื่องจากปรากฏผลการวิจัยเกี่ยวกับการใช้การประเมินด้วยนักเรียนได้เสนอแนะว่า นักเรียนไม่สามารถตัดสินได้อย่างคงเส้นคงวา เนื่องจากยังไม่บรรลุวุฒิภาวะ แบบแผนการประเมินของนักเรียนไม่ค่อยได้พิจารณาสาระที่เพียงพอ ครูที่แสดงออกในลักษณะที่ให้ความอบอุ่น มีความเป็นเพื่อน หรือสนุกสนานร่าเริง มักจะเป็นผู้ชนะตลอดเวลา เจื่อนไซและตัวแปรแทรกซ้อนจากภายนอกอาจส่งผลต่อการประเมินค่าของนักเรียน เช่น เศษของนักเรียนและผู้สอน ตำแหน่งของผู้สอน เป็นต้น อีกทั้งเกรดหรือคะแนนที่นักเรียนได้รับในรายวิชาจะสัมพันธ์กับการประเมินค่าผู้สอน

ในกรณีของการใช้กลุ่มเพื่อนร่วมงานเป็นผู้ประเมินนั้น Harris (1986: 130) เสนอแนะว่า โดยทั่วไป จำนวนของผู้ประเมินตัดสินไม่ควรจะน้อยกว่า 3 คน

จากที่กล่าวมา จะเห็นว่า ไม่ว่าจะเป็นวงการธุรกิจ หรือทางการศึกษา ในการประเมินผลการปฏิบัติงานของบุคลากรในระดับปฏิบัติการนั้น กลุ่มบุคคลที่ได้รับการยอมรับว่าน่าจะเป็นผู้ประเมินที่เหมาะสมคือ ผู้บังคับบัญชา กลุ่มเพื่อนร่วมงาน การประเมินตนเอง และการใช้กลุ่มผู้ใต้บังคับบัญชา หรือกลุ่มบุคคลในความดูแลเป็นผู้ประเมิน สำหรับการประเมินพฤติกรรมการทำงานครั้งนี้ เป็นการประเมินเพื่อตรวจสอบยืนยันคุณภาพของเครื่องมือประเมิน ไม่ได้มุ่งมันที่การทราบผลการดำเนินงานของครูเป็นสำคัญ อย่างไรก็ตามในการตรวจสอบยืนยันคุณภาพของเครื่องมือ

วิจัยนั้น ก็ควรได้รับการตรวจสอบยืนยันภายใต้สถานการณ์การใช้เครื่องมือที่เป็นไปได้ เหมาะสม และเกิดประโยชน์สูงสุด ดังนั้นในการศึกษาค้างนี้ ผู้วิจัยจึงเลือกกลุ่มผู้ที่ทำหน้าที่ประเมิน (Raters) 2 กลุ่ม คือ (1) กลุ่มครู ผู้สอนเป็นผู้ประเมินตนเอง ซึ่งเป็นเงื่อนไขการประเมินที่นิยมใช้มาก ในการศึกษาเชิงสำรวจและเป็นสถานการณ์ที่เกิดประโยชน์มากที่สุดสถานการณ์หนึ่งในการ บริหารบุคคล เพราะจะนำไปสู่การปรับปรุงเปลี่ยนแปลงพฤติกรรมการทำงานได้เป็นอย่างดี และ (2) กลุ่มผู้บังคับบัญชาชั้นต้น คือหัวหน้าหมวดวิชาเป็นผู้ประเมินพฤติกรรมการสอนของครู เป็นสถานการณ์ของการประเมินผู้อื่น และเป็นภาระงานที่อยู่ภายใต้กรอบบทบาทหน้าที่อยู่แล้วที่จะต้องทำหน้าที่ ควบคุม ดูแล นิเทศแนะนำ และประเมินผลงานของผู้ใต้บังคับบัญชา โดยถือว่าเป็นกลุ่มบุคคลที่ ใกล้ชิดและรู้เห็นพฤติกรรมการทำงานของครูได้ดีที่สุดกลุ่มหนึ่ง

(2) เครื่องมือที่ใช้ในการประเมินการปฏิบัติงานของบุคคล

ความน่าเชื่อถือของผลการประเมินพฤติกรรมการทำงานนั้น ส่วนหนึ่งจะขึ้นอยู่กับชนิดของเครื่องมือและแนวทางในการเก็บรวบรวมข้อมูล Mathis (1979: 301-307) ได้จำแนกประเภทของเครื่องมือที่นิยมใช้ในการประมาณค่าพฤติกรรมการทำงานหรือประเมินผลงานของบุคคล ออกเป็น 4 ประเภท คือ (1) เครื่องมือประมาณค่าประเภทกำหนดรายการประเมินไว้เป็นรายการ ๆ (Category Rating Inventory) เครื่องมือในกลุ่มนี้ อาทิ มาตรฐานค่าแบบกราฟิก (The Graphic Rating Scales) ซึ่งเป็นแบบที่นิยมใช้กันอย่างแพร่หลายมากที่สุด แบบตรวจสอบรายการ (The Checklist) และวิธีการกำหนดทางเลือกจำกัด (Forced Choice Methods) (2) วิธีการเชิงเปรียบเทียบ (Comparative Methods) เทคนิคที่ใช้ในกลุ่มนี้ อาทิ การจัดอันดับ (Ranking) การเปรียบเทียบ (Paired Comparison) เป็นต้น (3) วิธีการเชิงบันทึก (Written Methods) เทคนิคที่ใช้ในกลุ่มนี้ เช่น การบันทึกเหตุการณ์สำคัญ ๆ ที่เกิดขึ้น (The Critical Incident Method) ซึ่งจะบันทึกทั้งพฤติกรรมที่ดี-ทำงานได้ผลดี และเหตุการณ์ที่ไม่ดี-ปฏิบัติงานบกพร่อง แล้วค่อยสรุปผลเชิงประเมิน (4) การใช้มาตรฐานค่าเชิงพฤติกรรม (Behaviorally-anchored Rating Scales) ซึ่งถือเป็นวิธีการใหม่ในการ ประเมินพฤติกรรมการทำงาน เป็นวิธีการที่พัฒนาขึ้นเพื่อแก้จุดอ่อนของวิธีการในกลุ่มที่ (1)

Crane (1985: 404-409) ได้อ้างถึงการสำรวจสภาพการดำเนินการด้านการประเมินผลการปฏิบัติงานของ Locher และ Teel ซึ่งพบว่าเทคนิคที่นิยมใช้มากที่สุดในการประเมินผลการปฏิบัติงานของบุคลากรก็คือ เทคนิคมาตราประมาณค่า (Rating Scales) อีกทั้ง Crane ได้แบ่งกลุ่มเทคนิควิธีที่ใช้ในการประเมินการปฏิบัติงานออกเป็น 2 กลุ่มหลัก ๆ คือ (1) วิธีการตามประเพณีนิยม (Traditional Methods) เทคนิควิธีที่ใช้ในกลุ่มนี้ อาทิ มาตราประมาณค่าแบบกราฟิก (Graphic Scales) การเรียงอันดับ (Ranking) การบันทึกเหตุการณ์สำคัญ (The Critical Method) เป็นต้น และ (2) วิธีการเชิงพัฒนา (Developmental Approaches to Appraisal) ตามแนวคิดนี้ เชื่อว่าการประเมินจะเกิดประโยชน์ต่อการพัฒนาความสามารถของบุคลากร ดังนั้น ในกระบวนการประเมินควรจะแฝงไว้ด้วยแนวทางที่เป็นช่องทางสำหรับการพัฒนาตนเองของกลุ่มบุคลากรไปในตัวด้วย กล่าวคือ ในแบบประเมินควรจะแสดงออกซึ่งแนวทางการปฏิบัติงานที่มีประสิทธิภาพ ซึ่งจะเป็นการกระตุ้นให้ผู้ถูกประเมินมองเห็นแนวทางในการปฏิบัติงานในโอกาสต่อไป ดังนั้นจึงถือว่าวิธีการในกลุ่มนี้เป็นการมุ่งเน้นการพัฒนาพฤติกรรมควบคู่ไปด้วย ไม่ใช่เพียงแต่ประเมินผลการปฏิบัติงานเพียงอย่างเดียว เทคนิควิธีในกลุ่มนี้ เช่น การประเมินโดยอิงการบริหารด้วยจุดประสงค์ (Appraisal through Management by Objectives) และ การใช้เทคนิคมาตราประมาณค่าเชิงพฤติกรรม (The Behaviorally-Anchored Rating Scales) เป็นต้น

โดยสรุป จะเห็นว่า ในด้านเครื่องมือ หรือเทคนิคที่นิยมใช้ในการประเมินพฤติกรรม การปฏิบัติงาน หรือประเมินผลการปฏิบัติงานของบุคคลนั้น เทคนิคมาตราประมาณค่าเป็นเทคนิคที่ได้รับการยอมรับและนิยมใช้กันอย่างแพร่หลายมากที่สุด โดยเฉพาะมาตราประมาณค่าแบบกราฟิก ขณะเดียวกัน มาตราประมาณค่าเชิงพฤติกรรมก็ถือได้ว่าเป็นวิธีการทางเลือกใหม่ที่ได้รับพิจารณาว่าน่าจะแก้ปัญหาที่เป็นจุดอ่อนบางประการของมาตราประมาณค่าแบบตัวเลขหรือแบบกราฟิก อันเป็นแบบที่เป็นที่นิยมได้

ตอนที่ 6 พฤติกรรมการสอนของครู

เนื่องจากการศึกษาครั้งนี้ มุ่งศึกษาคุณภาพของเทคนิคมาตรฐาน ประมวลค่าที่มีลักษณะการจัดสเกลคำตอบแตกต่างกัน ภายใต้การประเมินพฤติกรรมการปฏิบัติงานบุคคลและเห็นว่าพฤติกรรมการปฏิบัติงานของครูผู้สอนเป็นพฤติกรรมที่มีรูปแบบชัดเจน เป็นพฤติกรรมที่ได้รับความสนใจจาก นักการศึกษา และบุคคลอื่นอีกเป็นจำนวนมาก อีกทั้งถือว่าเป็นตัวแปรที่สำคัญยิ่งตัวแปรหนึ่งในการที่จะจัดการเรียนการสอนให้บรรลุตามเจตนารมณ์ของหลักสูตร ดังนั้น ผู้วิจัยจึงได้กำหนดให้พฤติกรรมการสอนของครู เป็นเนื้อหาเป้าหมายสำหรับการตรวจสอบคุณภาพของเทคนิควิธีการประมวลค่าครั้งนี้ ในการนี้ เพื่อกำหนดกรอบเนื้อหาในการประเมิน ผู้วิจัยจึงได้ศึกษาแนวคิดเกี่ยวกับพฤติกรรมการสอนของครูที่มีประสิทธิภาพ วิธีการจัดการเรียนการสอนอย่างเป็นระบบ และพฤติกรรมด้านการวัดผล-ประเมินผลของครู ซึ่งปรากฏข้อสรุปดังนี้

6.1 รูปแบบพฤติกรรมการสอนของครูโดยทั่วไป

นักการศึกษาหลายท่านได้พยายามศึกษาและรวบรวมรายการพฤติกรรมที่เป็นแนวทางสำหรับการจัดการเรียนการสอนของครู อาทิ

Mathis (1979: 8-25) ได้เสนอแบบของพฤติกรรมการสอนของครูไว้ 18 ลักษณะ คือ (1) การยอมรับความรู้สึกและทัศนคติของนักเรียน (2) การเอาใจใส่ในความต้องการ ความสนใจและปัญหาของนักเรียน (3) การคาดการณ์ในผลการเรียนที่อาจเกิดขึ้น (4) การรู้ระดับพัฒนาการและความสนใจของเด็ก (5) การให้ความสม่ำเสมอและความเท่าเทียมกันแก่นักเรียน (6) การให้ความกระจ่างชัดในการสอน (7) การสนับสนุนให้เกิดพฤติกรรมความร่วมมือ (8) การมีพฤติกรรมที่เป็นประชาธิปไตย (9) การกระตุ้นและยกย่องชมเชยเพื่อจูงใจให้เกิดความเจริญงอกงาม (10) การเป็นบ่อเกิดของพฤติกรรมทางบวก (11) การมีความมั่นคงในงานด้านการสอน และมีความยืดหยุ่นและการปรับเปลี่ยนให้เหมาะสมกับความต้องการของเด็ก (12) การมีลักษณะการสอนเป็นรายบุคคลและเป็นการส่วนตัว ตามความต้องการของนักเรียน (13) มีการเคลื่อนไหวของครูและนักเรียน (14) มีการกำกับ ติดตามความก้าวหน้าของนักเรียน (15) การเปิดโอกาสให้นักเรียนได้มีส่วนร่วมและส่วนเกี่ยวข้อง (16) มีการกะระยะเวลาของกิจกรรมได้เหมาะสม (17) มีระเบียบวินัยในชั้นเรียน และ (18) การสร้างความอบอุ่นใจและความกระตือรือร้น

Gerhard (1971: 239-242) ได้เสนอแนะว่า กิจกรรมที่ควรจะเป็นภายใต้กระบวนการเรียนรู้ การคิด หรือการสอน (Teaching-thinking-learning Process) ว่า ควรประกอบด้วย (1) เลือกขอบข่ายของเนื้อหาวิชา (2) พัฒนาทักษะที่จะเอื้อประโยชน์ต่อกระบวนการทางด้านความรู้-ความคิด (Cognitive Process) (3) สร้างสรรค์สิ่งแวดล้อมเพื่อการสนองตอบ (4) สร้างหน่วยการเรียนรู้ที่ย่อ (5) เตรียมนักเรียนและทำการประเมินผลก่อนเรียน (6) นำเสนอหน่วยการเรียนรู้ (7) ทดสอบหลังเรียน (8) จัดวาง (Provide) นักเรียนตามความสามารถ (9) ดำเนินการในกระบวนการของโครงสร้างหน่วยการเรียนรู้ต่อไป

Erickson (1980: 59) ได้กล่าวถึงทักษะที่จำเป็นสำหรับการสอนที่มีประสิทธิภาพว่า ควรประกอบด้วย (1) การให้คำชี้แนะเพื่อสร้างความสนใจผู้เรียนและแนะนำกรอบงานด้านการเรียนการสอน (2) การใช้ตัวอย่าง หรือการให้รายละเอียดเพื่ออธิบายและทำความเข้าใจในเนื้อหาวิชา (3) การถามคำถามเพื่อกระตุ้นและชี้นำความคิด (4) การใช้วัสดุสื่อการสอนและวิธีการสอนที่หลากหลาย (5) การติดต่อสื่อสารเพื่อรับฟังและเอาใจใส่ในส่วนที่เกี่ยวข้องกับผู้เรียน (6) การดำเนินการสอบที่มีความตรงและเชื่อถือได้ (7) การให้ข้อมูลย้อนกลับเพื่อให้นักเรียนจะสามารถติดตามความก้าวหน้าของตนเองได้

Peter (quoted in Nelli, 1981: 39) ได้กล่าวถึงลักษณะของครูที่มีประสิทธิภาพว่า ควรมีลักษณะดังนี้ คือ (1) เป็นผู้ที่มีความยืดหยุ่น (Flexible) (2) สามารถที่จะรับรู้ความคิด-ความรู้สึของผู้เรียน (3) ปรับการสอนให้เหมาะสมกับตนเองได้ (Personalize Teaching) (4) มีความตั้งใจ-มีความคิดของการเป็นนักทดลอง (5) มีทักษะในการใช้คำถาม (6) มีความรู้ในเนื้อหาวิชา (7) มีการวางแผน-จัดทำกระบวนการตรวจสอบที่ดี (Well-Established Examination Procedures) (8) มีการจัดหาเครื่องมือช่วยในการเรียน (9) มีทัศนคติในเชิงสนับสนุนให้การช่วยเหลือแก่ผู้เรียน และ (10) ใช้ทำทางการสนทนาคู่ในกิจกรรมการสอน

Verner (1973: 1-13) ได้ศึกษาเชิงสำรวจสมรรถภาพของการเป็นครูในโรงเรียนมัธยมศึกษา โดยสรุปว่า ครูมัธยมศึกษาควรมีสมรรถภาพ 4 ด้าน ในแต่ละด้านประกอบด้วยพฤติกรรมการทำงานต่าง ๆ ดังนี้

- (1) ด้านการปฏิสัมพันธ์ (Interaction) ประกอบด้วย
 - ให้คำแนะนำที่ชัดเจนแก่นักเรียนได้
 - ตอบสนองต่อพฤติกรรมของนักเรียนได้
 - แก้ไขพฤติกรรมของนักเรียนได้
 - สนองตอบตามความต้องการและความรู้สึกของผู้อื่นอย่างมีเหตุผล
 - มีความสามารถในการสัมภาษณ์สอบถามนักเรียนได้หลาย ๆ วิธี
 - ทำงานเป็นคณะ ได้อย่างมีประสิทธิภาพ
- (2) ด้านการจัดการ และดำเนินการสอน (Management and Instruction)
 - มีระบบการจัดการเรียนการสอนหลายแบบ
 - นำอุปกรณ์การสอนและทรัพยากรจากแหล่งต่าง ๆ มาใช้ให้เกิดประโยชน์
 - สอนโดยคำนึงถึงความแตกต่างระหว่างบุคคล
 - วางแผนการจัดกิจกรรมร่วมกับนักเรียนได้
 - มีเทคนิคในการหาข้อมูลจากนักเรียน
 - ใช้คำถามได้อย่างเหมาะสมกับความสามารถของนักเรียน
 - ใช้วิธีการเสริมแรงได้อย่างเหมาะสม
 - กำหนดแผนการสอนได้
 - เรียงลำดับกิจกรรมการเรียนการสอนและจัดประสบการณ์ได้เหมาะสมตามหลักจิตวิทยา
- (3) ด้านการประเมินผล (Evaluation) - ประกอบด้วย
 - วินิจฉัยเพื่อการแนะนำนักเรียน
 - สร้างเครื่องมือวัดผลเพื่อตรวจสอบและประเมินผลได้
 - ทำการประเมินและแปลผลได้

- (4) ด้านความก้าวหน้าในวิชาชีพ (Professional Growth) ประกอบด้วย
- สามารถปรับปรุงพฤติกรรมการสอนโดยอาศัยข้อมูลจากการประเมินได้
 - เข้าร่วมในโปรแกรมการพัฒนาทางด้านวิชาชีพได้
 - มีส่วนร่วมในการพัฒนาหลักสูตร

Squires และ คณะ (quoted in Abruscato, 1985: 8-9) ได้อ้างถึงการศึกษางานวิจัยจำนวนมากที่ทำการศึกษาเกี่ยวกับประสิทธิภาพการสอนของครู หรือศึกษาเกี่ยวกับลักษณะของครูที่มีประสิทธิภาพ (Effective Teachers) ซึ่งปรากฏข้อสรุปว่า พฤติกรรมการสอนของครูที่มีประสิทธิภาพนั้น สามารถจัดได้เป็น 3 กลุ่ม คือ

- (1) ด้านการวางแผน (Planning) อาทิ
- การวิเคราะห์และกำหนดจุดประสงค์การเรียนรู้
 - การวิเคราะห์ความสามารถและลักษณะของผู้เรียน
 - การกำหนดทางเลือกอย่างเหมาะสมว่า จะสอนอะไร-สอนอย่างไร
- (2) ด้านการจัดการชั้นเรียน (Classroom Management) อาทิ
- การวิเคราะห์ห้องและกำหนดแนวทางการสอนอย่างละเอียดล่วงหน้า อย่างน้อย 1-2 สัปดาห์ พร้อมทั้งคาดคะเน ผลการสอน หรือประเด็นที่อาจเป็นปัญหาในการเรียนรู้
 - อธิบายกฎเกณฑ์ แนวปฏิบัติ สิ่งที่คาดหวัง และมอบหมายงานแก่นักเรียนอย่างชัดเจน
 - กำหนดระบบที่แสดงออกซึ่งความรับผิดชอบของนักเรียนเกี่ยวกับงาน หรือ พฤติกรรมทางวิชาการ
 - กำกับ ติดตามการปฏิบัติงานของนักเรียน พร้อมทั้งให้ข้อมูลป้อนกลับอย่างสม่ำเสมอ
- (3) ด้านการสอน (Instruction) พฤติกรรมด้านนี้ อาทิ
- การกล่าวนำทั่วไป (Overview) เช่น ทบทวนความรู้เดิมที่เรียนไปแล้ว การอธิบายถึงความสำคัญของเนื้อหาใหม่ที่กำลังจะสอน

- การอธิบาย (Explanation) เช่น อธิบายความคิดรวบยอดเกี่ยวกับเรื่องที่สอนให้นักเรียนสะท้อนความเข้าใจของตนเอง ให้ข้อมูลป้อนกลับเพื่อให้นักเรียนเกิดการปรับปรุง-พัฒนาความคิด พร้อมทั้งให้นักเรียนมีโอกาสฝึกทักษะความชำนาญในความคิดใหม่ที่ได้รับ
- การกำกับพฤติกรรม (Monitoring Performance) โดยครูจะต้องทำการประเมินพฤติกรรมของนักเรียนอย่างสม่ำเสมอทั้งในระหว่างสอน และเมื่อสอนจบบทเรียน

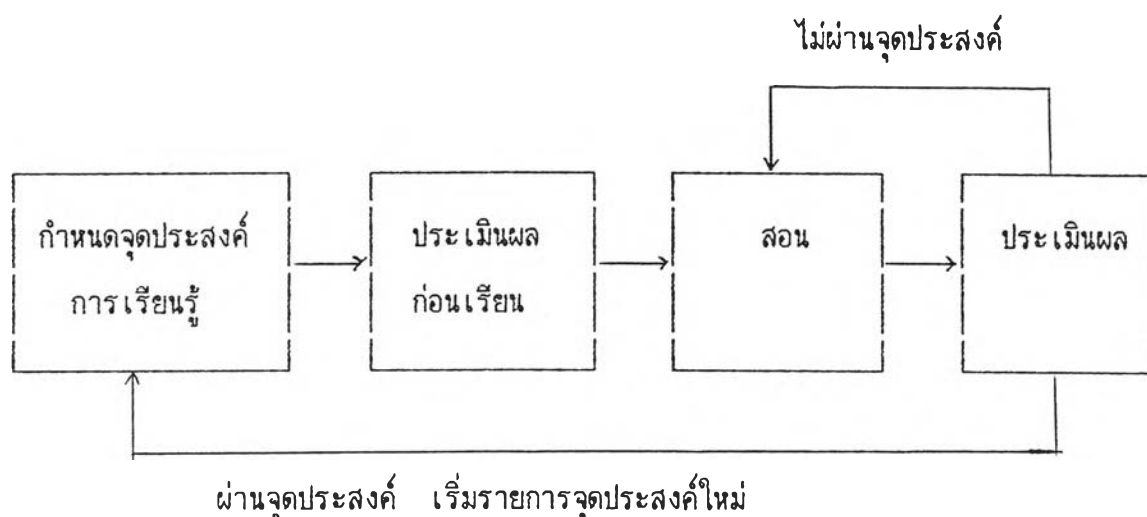
สุรัสวดี ศิลปอนันต์ (2532: 10) ให้ทัศนะว่า คุณลักษณะและสมรรถภาพของครูที่จะก้าวไปสู่วิชาชีพนั้น ควรประกอบด้วย (1) มีความรู้ดี ทั้งความรู้ที่จะสอนและในวิชาที่จะถ่ายทอด (2) มีทักษะในการสอนและการปฏิบัติงาน (3) มีคุณธรรมนิยมของความเป็นครูในด้านจริยธรรม คุณธรรม และคตินิยมเกี่ยวกับครู

สุมน อมรวิวัฒน์ และสมพงษ์ จิตระดับ (2531: 71-72) ได้สรุปพฤติกรรมการสอนของครูที่หลักสูตรต้องการว่า ประกอบด้วย (1) ศึกษาหลักสูตรและเอกสารหลักสูตรให้เข้าใจถ่องแท้ (2) เตรียมการสอนอย่างดีทุกครั้งก่อนทำการสอน (3) จัดกิจกรรมการเรียนการสอนที่ยึดนักเรียนเป็นศูนย์กลางในการเรียนรู้ (4) ใช้สื่อการสอนอย่างเหมาะสมและประเมินผลการสอนทุกขั้นตอน (5) ครูและนักเรียนมีส่วนร่วมและมีปฏิสัมพันธ์ในการจัดกิจกรรมโดยตลอด และ (6) ครูจัดสถานการณ์การเรียนการสอนให้นักเรียนเกิดการเรียนรู้ตามจุดประสงค์ของหลักสูตร และสามารถนำไปใช้ได้จริง

6.2 พฤติกรรมการสอนของครูภายใต้การจัดกิจกรรมการเรียนการสอนอย่างเป็นระบบ

นักการศึกษาหลายท่านได้พยายามประยุกต์แนวความคิดเกี่ยวกับวิธีการเชิงระบบ (System Approach) มาใช้กับกิจกรรมการเรียนการสอน โดยได้เสนอแนะแนวทางการจัดกิจกรรมในแต่ละขั้นตอนของระบบการสอนไว้เพื่อครูผู้สอนสามารถยึดถือ เป็นแนวปฏิบัติ อาทิ

Popham และ Baker (1970: 11-17) ได้เสนอแนวคิดเกี่ยวกับการจัดการเรียนการสอนโดยยึดจุดประสงค์เป็นสำคัญ (Goal-Reference Instructional Model) โดยเสนอแนะว่า กิจกรรมแรกที่ผู้สอนควรจะทำคือการคิด จะต้องวิเคราะห์เหตุการณ์การเรียนรู้ที่ต้องการให้เกิดกับผู้เรียน หรือผู้เรียนจำเป็นต้องได้รับการตอบสนอง ถือเป็นเป้าหมายหรือจุดประสงค์ที่ต้องการ และหลังจากดำเนินการสอนแล้วก็ต้องมีการประเมินผลอีกขั้นหนึ่ง รูปแบบการสอนเป็นดังนี้

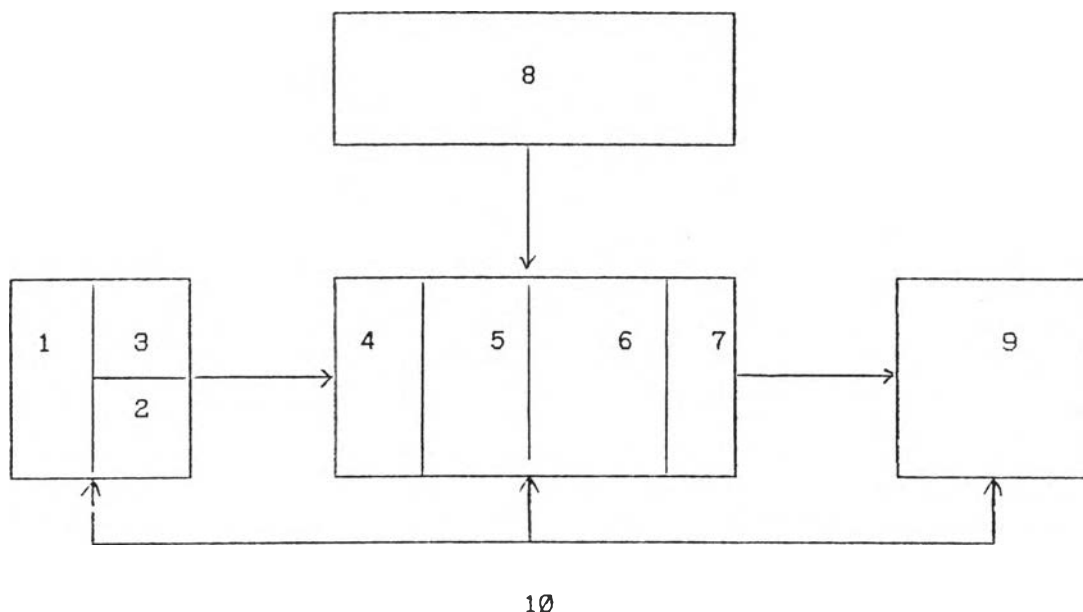


ตามแผนภาพดังกล่าวข้างต้น Popham เน้นว่า ตามรูปแบบนี้ การประเมินได้เข้ามา มีบทบาทอย่างมากในการสอน ทั้งนี้เพราะเน้นการตรวจสอบความรู้ ทั้งก่อนสอนและหลังจากที่ได้ดำเนินการสอนแล้ว ทำให้ดูเหมือนว่ารูปแบบนี้เป็นรูปแบบการประเมิน หรือ Assessment Model มากกว่าเป็นกระบวนการสอน (Teaching Procedures) อีกทั้งในการที่ผู้สอนจะสามารถดำเนินการตามแนวคิดนี้ได้ต้องมีประสิทธิภาพนั้น ครูผู้สอนควรจะเป็นนักบริโภคผลการวิจัย (The Research Consumer) ซึ่งจะก่อให้เกิดประโยชน์ในด้านการศึกษาผลการวิจัย แล้วนำเทคนิควิธีการใหม่ ๆ มาใช้ในกิจกรรมการเรียนการสอน ขณะเดียวกัน ครูจะต้องเป็นนักทดสอบสมมุติฐาน (The Hypothesis Tester) ทำหน้าที่ตรวจสอบแนวคิด วิธีการ หรือเทคนิคการสอนอยู่ตลอดเวลาว่าเกิดผลดีหรือไม่ เพื่อการใช้ข้อมูลสำหรับการปรับปรุง หรือเพิ่มประสิทธิภาพการสอนในโอกาสต่อไป

สงัด อุทรานันท์ (2529: 30-34) ได้เสนอแนวคิดเกี่ยวกับการจัดการเรียนการสอน
 อย่างเป็นระบบในลักษณะที่ขยายแนวคิดในระดับรายละเอียดมากกว่า Popham และ Baker
 โดยจำแนกองค์ประกอบในการเรียนการสอนดังนี้

1. ลักษณะของผู้เรียน ประกอบด้วยความสามารถทางสติปัญญา ลักษณะการเรียนรู้
 และประสบการณ์เดิม
2. จุดประสงค์ของการสอน ประกอบด้วยด้านความรู้ความเข้าใจ ด้านเจตคติและ
 ค่านิยม และด้านทักษะทางกาย
3. เนื้อหาสาระที่จะนำมาสอน
4. การเตรียมความพร้อม ประกอบด้วย การเตรียมสถานที่และสื่อ และ การเร้า
 ความสนใจ
5. การดำเนินการสอน ประกอบด้วย การเลือกกลวิธีการสอน การเลือกกิจกรรม
 การสอน และการเลือกสื่อประกอบการสอน
6. การสร้างเสริมทักษะ ประกอบด้วย การช่วยเหลือเด็กเรียนรู้ และการส่งเสริม
 เด็กเรียนเก่ง
7. การจัดกิจกรรมสนับสนุน ประกอบด้วย การทบทวน และการสั่งงาน
8. การควบคุมและการตรวจสอบ ประกอบด้วย การวัดความรู้ก่อนสอน การใช้คำถาม
 การสร้างเสริมกำลังใจ และการวัดผลหลังสอน
9. สัมฤทธิผลของการสอน
10. การปรับปรุงแก้ไข

ความสัมพันธ์ระหว่างองค์ประกอบต่าง ๆ ทั้ง 10 รายการ แสดงได้ดังแผนภาพข้างล่าง



ตามแผนภาพ จะเห็นว่า องค์ประกอบที่ 8 คือการควบคุมและการตรวจสอบนั้นเป็นองค์ประกอบที่สำคัญประการหนึ่งในบทบาทหน้าที่ของครูผู้สอน เป็นหน้าที่ในลักษณะของนักทดสอบสมมุติฐาน โดยทำการวัดผลทั้งก่อนสอน และหลังสอน เป็นองค์ประกอบที่มีประโยชน์ในแง่ของการหาข้อมูลเพื่อประกอบการตัดสินใจในกระบวนการเรียนการสอนลักษณะต่าง ๆ ส่วนองค์ประกอบที่ 10 หรือการปรับปรุงแก้ไขนั้น ครูจะต้องทำการวิเคราะห์และปรับปรุงระบบการเรียนการสอนอย่างสม่ำเสมอในทุก ๆ ส่วน เป็นองค์ประกอบที่จะส่งผลต่อการพัฒนาวิชาชีพครู (Professional Growth)

จากข้อเสนอแนะ หรือแนวคิดเกี่ยวกับพฤติกรรมการสอนของครูที่มีผู้เสนอแนะไว้อย่างหลากหลายในหัวข้อ 6.1 และ แนวความคิดเกี่ยวกับการจัดการเรียนการสอนอย่างเป็นระบบในหัวข้อ 6.2 ผู้วิจัยเห็นว่า หากจะยึดถือกรอบแนวคิดเชิงระบบซึ่งประกอบด้วยกิจกรรมหลักที่สำคัญ คือ การวางแผน ดำเนินการตามแผน และประเมินผล ก็น่าจะทำให้สามารถมองเห็นแนวทางการจัดกิจกรรมของครู หรือพฤติกรรมการทำงานของครูที่เป็นระบบและชัดเจนยิ่งขึ้น ดังนั้น พฤติกรรมการสอนของครูเมื่อนิยามอย่าง เป็นระบบแล้วควรจะเป็นดังนี้

(1) พฤติกรรมด้านการวางแผนการสอน เป็นพฤติกรรมในขั้นตอนการวางแผน อาทิ การวิเคราะห์หลักสูตรเพื่อกำหนดจุดประสงค์การสอน/จุดประสงค์การเรียนรู้ การจัดทำแผนการสอนหรือแนวทางการสอน-แนวทางในการวัดผล การประเมินความสามารถของนักเรียนก่อนเริ่มสอน การชี้แจงแนวปฏิบัติหรือสิ่งที่คาดหวังแก่นักเรียนเป็นการล่วงหน้า เป็นต้น

(2) พฤติกรรมการจัดการชั้นเรียน เป็นพฤติกรรมการดำเนินการในขณะที่สอน เช่น การกำกับ ติดตามผลการปฏิบัติงานของนักเรียนและการให้ข้อมูลป้อนกลับ ใช้วิธีสอนอย่างหลากหลายเหมาะสมกับเนื้อหา สอนโดยคำนึงถึงความแตกต่างระหว่างบุคคล มีเทคนิคในการใช้คำถาม มีวิธีการในการตรวจสอบความเข้าใจของนักเรียน ใช้ข้อมูลผลการประเมินเพื่อแก้ปัญหาการเรียนการสอน ใช้วิธีการเสริมแรงได้อย่างเหมาะสม ประเมินผลการเรียนอย่างสม่ำเสมอทั้งในระหว่างสอนและเมื่อจบเนื้อหาแต่ละเรื่อง-แต่ละบท การให้นักเรียนฝึกทักษะเพิ่มเติม เป็นต้น

(3) พฤติกรรมด้านการตรวจสอบระบบการเรียนการสอน เป็นพฤติกรรมที่แสดงออกซึ่งความเป็นนักทดสอบสมมุติฐานเพื่อการพัฒนาวิชาชีพการสอน อาทิ การทดลองวิธีสอนใหม่ ๆ ศึกษาค้นคว้าวิธีการเรียนการสอน วิเคราะห์ข้อสอบ พัฒนาแบบสอบ สังเกตหรือเก็บรวบรวมข้อมูลความคิดเห็นของผู้เรียนเกี่ยวกับประสิทธิภาพของวิธีสอน เทคนิคการสอน และระบบการวัดผล-ประเมินผลที่ครูใช้ เป็นต้น

6.3 พฤติกรรมด้านการประเมินผลของครูผู้สอน

จากรายการพฤติกรรมที่ควรจะเป็นของครูผู้สอน จะพบว่ารายการพฤติกรรมของครูส่วนหนึ่งเป็นรายการที่แสดงออกซึ่งความสามารถหรือทักษะด้านการวัดและประเมินผลของครูผู้สอน ซึ่งบางรายการ เป็นพฤติกรรมด้านการวัด-ประเมินผลที่ได้กำหนดไว้อย่างชัดเจนในระเบียบว่าด้วยการประเมินผลตามหลักสูตร อาทิ ตามระเบียบกระทรวงศึกษาธิการว่าด้วยการประเมินผลการเรียนตามหลักสูตรมัธยมศึกษาตอนปลายพุทธศักราช 2524 (ฉบับปรับปรุง พ.ศ.2533) (กรมวิชาการ 2532: 1-10) สามารถจำแนกรายการพฤติกรรมด้านการวัดผล-ประเมินผลที่ครูจะต้องดำเนินการดังนี้ (1) แจ้งให้ผู้เรียนทราบจุดประสงค์การเรียนรู้ วิธีการประเมินผลการเรียนเกณฑ์การผ่านจุดประสงค์การเรียนรู้ และเกณฑ์ขั้นต่ำของการผ่านรายวิชา (2) ประเมินผลก่อนเรียนเพื่อศึกษาความรู้พื้นฐานของผู้เรียน (3) จัดดำเนินการสอนซ่อมเสริมถ้าพบว่าผู้เรียนมีพื้นฐานความรู้ไม่เพียงพอในการที่จะเรียนเนื้อหาใหม่ (4) วัดผลและประเมินผลระหว่างเรียนเป็นระยะ ๆ

- (5) ประเมินคุณลักษณะด้านคุณธรรม จริยธรรม และค่านิยม ตลอดภาคเรียน โดยเน้นความรู้สึก และคุณภาพของการแสดงออก (6) ใช้ข้อมูลผลการประเมินในการจัดสอนซ่อมเสริม หรือปรับปรุง กระบวนการสอนของตนเอง การสอนซ่อมเสริมอาจดำเนินการเป็นรายบุคคล หรือเป็นกลุ่ม
- (7) วัดผลปลายภาคเรียนเพื่อตรวจสอบผลการเรียน โดยวัดให้ครอบคลุมจุดประสงค์ที่สำคัญตามที่ กลุ่มโรงเรียนกำหนด (8) ตัดสินผลการเรียน และให้ระดับผลการเรียน (9) จัดสอนซ่อมเสริม ก่อนที่จะมีการสอบแก้ตัวสำหรับผู้ที่ได้คะแนน "๐"

เนื่องจากพฤติกรรมทางด้านการวัดผล-ประเมินผลของครูผู้สอน เป็นพฤติกรรมที่มีรูปแบบและแนวทางการปฏิบัติที่ชัดเจน และถือเป็นรายการหรือแนวทางการปฏิบัติที่สำคัญที่ก่อให้เกิดผลได้ผลเสียแก่เด็ก อีกทั้งเป็นเครื่องมือที่สำคัญยิ่งสำหรับการตรวจสอบความสำเร็จของการจัดการเรียนการสอน ดังนั้น ในการศึกษาครั้งนี้จึงใช้กรอบพฤติกรรมด้านการประเมินผลของครูเป็นกรอบเนื้อหาในการตรวจสอบคุณภาพของเครื่องมือวัดประเภทมาตรประมาณค่า ซึ่งกรอบพฤติกรรมด้านการประเมินผลของครูในที่นี่จะรวมถึงพฤติกรรมที่แสดงออกซึ่งความสามารถในการวางแผนประเมินผลของครู และพฤติกรรมที่แสดงออกซึ่งแนวความคิดในการพัฒนาระบบการวัดและประเมินผลด้วย กรอบพฤติกรรมโดยสรุปในการศึกษาครั้งนี้ คือ

1. พฤติกรรมด้านการวางแผนการวัดผล-ประเมินผลการเรียน เช่น การวิเคราะห์จุดประสงค์การเรียนรู้ในรายวิชาที่สอน การกำหนดแนวทางการวัดผล-ประเมินผลการเรียน เป็นต้น
2. พฤติกรรมการวัดผล-ประเมินผลในชั้นสอน หรือระหว่างดำเนินการสอน เช่น การปรับแก้ทักษะพื้นฐานของผู้เรียนก่อนสอนเนื้อหาใหม่ การตรวจสอบความสามารถของผู้เรียนเมื่อจบการสอนแต่ละหน่วย/แต่ละเรื่อง เป็นต้น
3. พฤติกรรมการวัดผล-ประเมินผล ในขั้นสรุปผลการเรียน เช่น การพัฒนาเครื่องมือวัดผลปลายภาคเรียน การดำเนินการสอนซ่อมเสริมเมื่อนักเรียนสอบไม่ผ่านในรายวิชา เป็นต้น
4. พฤติกรรมในเชิงตรวจสอบและพัฒนาระบบการวัดผล-ประเมินผลของตนเอง เช่น การวิเคราะห์คุณภาพของข้อสอบ ความสนใจที่จะศึกษาค้นคว้าเพื่อพัฒนาแนวทางการวัดผล เป็นต้น

กรอบความคิดโดยสรุปที่ได้จากการศึกษาแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

จากการศึกษาแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้องทั้ง 6 รายการดังกล่าวข้างต้น สามารถสรุปสาระสำคัญและกรอบความคิดที่ได้จากการศึกษา ดังนี้

การศึกษาแนวความคิดเกี่ยวกับมาตรฐานค่าในตอนที่ 1 ทำให้ได้ข้อสรุปว่ามาตรฐานค่าเป็นเครื่องมือหรือเทคนิคที่สามารถประยุกต์ใช้ในกิจกรรมการประเมินต่าง ๆ หลายลักษณะ มีรูปแบบให้เลือกใช้หลายรูปแบบ รูปแบบที่เป็นที่นิยมใช้กันอย่างกว้างขวางคือ แบบตัวเลข และแบบกราฟิก มาตรฐานค่าเป็นเครื่องมือวัดที่ยังมีจุดอ่อนหรือข้อจำกัดอยู่หลายประการ ซึ่งนักวิจัยหลายท่านได้พยายามศึกษาและชี้แนะช่องทางเพื่อการปรับปรุงและขจัดข้อจำกัดหรือจุดอ่อนเหล่านั้น ช่องทางในการปรับปรุงการประเมินค่าให้มีความคมชัดนั้น โดยสรุปน่าจะจำแนกเป็น 3 ประเด็น คือ (1) ควรเลือกใช้เครื่องมือวัดรูปแบบนี้ในสถานการณ์การประเมินที่สามารถสังเกตพฤติกรรม หรือคุณลักษณะที่มุ่งประเมินได้อย่างชัดเจน (2) ปรับปรุงลักษณะของตัวมาตรฐานค่าให้มีความชัดเจนมากยิ่งขึ้น และ (3) ปรับปรุงในส่วนของผู้ประเมิน เช่น ใช้วิธีการฝึกทักษะแก่ผู้ประเมิน ชี้แนะแนวทางในการประเมิน เป็นต้น

ในสภาพเชิงประจักษ์เกี่ยวกับการประยุกต์ใช้มาตรฐานค่าในงานวิจัย ซึ่งได้นำเสนอไว้ในตอนที่ 2 นั้น สรุปได้ว่ามาตรฐานค่าเป็นเครื่องมือที่ได้รับความนิยมใช้ในงานวิจัยมากที่สุดประเภทหนึ่ง แต่ในขณะเดียวกันภายใต้การได้รับความนิยมซึ่งส่วนใหญ่เป็นการใช้มาตรฐานค่าแบบตัวเลขหรือแบบกราฟิกนั้น ข้อสรุปจากงานวิจัยจำนวนหนึ่ง ซึ่งได้ยกตัวอย่างไว้ในตอนที่ 2 ได้ชี้ให้เห็นถึงปัญหาเกี่ยวกับผลการวิจัยหลายประการ สภาพปัญหาหรือจุดอ่อนที่พบนี้เป็นการยืนยันแนวคิดที่ว่าด้วยจุดอ่อนหรือข้อจำกัดของมาตรวัดรูปแบบนี้ที่กล่าวแล้วในตอน 1 ได้เป็นอย่างดี

ภายใต้ข้อ เสนอแนะ เพื่อการปรับปรุงคุณภาพของการประเมินค่าให้มีความคมชัด ดังปรากฏข้อสรุปในตอนต้นแล้วนั้น การปรับปรุงลักษณะของมาตรวัดให้มีความชัดเจนมากยิ่งขึ้น เป็นประเด็นที่ผู้วิจัยเห็นว่าน่าจะเป็นแนวทางเลือกที่สำคัญที่มีคุณค่ายิ่ง ทั้งนี้ เพราะว่าเป็นเรื่องของการปรับปรุงในส่วนของผู้ประเมิน เช่น การฝึกทักษะแก่ผู้ประเมินหรือชี้แนะแนวทางในการประเมินแก่ผู้ประเมิน เป็นเรื่องที่ทำได้ยาก การปรับปรุงในส่วนของมาตรวัดเป็นทางเลือกที่สะดวกกว่า จากประเด็นความคิดนี้ ผู้วิจัยจึงได้ศึกษาแนวคิดว่าด้วยมาตรฐานค่าเชิงพฤติกรรม ซึ่งทำให้มอง

เห็นว่า มาตรฐานค่าเชิงพฤติกรรมน่าจะเป็นทางเลือกที่สำคัญ เป็นทางเลือกที่น่าสนใจทางเลือกหนึ่งในการที่จะนำมาใช้ในกิจกรรมการประเมิน (ดังรายละเอียดในตอนี่ 3)

ในส่วนของงานวิจัยที่มุ่งพัฒนา หรือตรวจสอบคุณภาพของมาตรฐานค่า ซึ่งได้นำเสนอไว้ในตอนที่ 4 นั้น ข้อสรุปจากส่วนนี้ทำให้มองเห็นว่างานวิจัยในต่างประเทศได้มุ่งเน้นในเรื่องของการพัฒนาลักษณะของมาตรวัด โดยเฉพาะอย่างยิ่งมาตรฐานค่าเชิงพฤติกรรมได้กลายเป็นรูปแบบของการประเมินค่าที่ได้รับความสนใจจากนักวิจัยจำนวนมาก มีการพัฒนาและตรวจสอบคุณภาพของมาตรวัดรูปแบบนี้ในงานวิจัยต่างประเทศหลายเรื่อง หลายลักษณะ ในขณะที่งานวิจัยในประเทศไทยยังไม่ปรากฏการพัฒนาหรือตรวจสอบคุณภาพของมาตรวัดรูปแบบนี้ อย่างไรก็ตาม ในกรณีของงานวิจัยในต่างประเทศนั้น แม้ว่าจะมีผู้สนใจศึกษาคุณภาพของมาตรวัดรูปแบบนี้กันอย่างหลากหลาย แต่สถานการณ์ของการศึกษาส่วนใหญ่จะเป็นสถานการณ์ของการให้นักเรียนหรือนักศึกษาประเมินพฤติกรรมของครู เป็นสำคัญ ประเด็นความคิดของผู้วิจัยที่ได้จากการศึกษาในตอนนั้นก็คือ เห็นว่าควรมีการพัฒนาและตรวจสอบคุณภาพของมาตรฐานค่าเชิงพฤติกรรมในภาวะแวดล้อมการประเมินของคนไทย โดยเน้นการใช้มาตรวัดรูปแบบนี้ในสถานการณ์ของการประเมินพฤติกรรม การปฏิบัติงานและเป็นสถานการณ์ของการศึกษาเชิงสำรวจ ทั้งนี้ด้วยเห็นว่าสถานการณ์การประเมินในลักษณะนี้เป็นสถานการณ์ของการศึกษาวิจัยที่มีคุณค่าในหลายแง่มุม อาทิ เป็นประโยชน์ในการบริหารงานบุคคล เป็นประโยชน์ต่อกิจกรรมการประเมินในลักษณะของการติดตาม-ประเมินผล ผู้ผ่านการอบรมหลักสูตรต่าง ๆ เป็นต้น

จากประเด็นความคิดที่เห็นว่าควรจะให้พัฒนาและตรวจสอบคุณภาพของมาตรฐานค่าเชิงพฤติกรรมในสถานการณ์ของการประเมินพฤติกรรม การปฏิบัติงานของบุคคลในเชิงสำรวจ ผู้วิจัยจึงได้ศึกษาแนวคิดเกี่ยวกับการประเมินพฤติกรรมของบุคคล (ตอนที่ 5) ซึ่งในที่สุดได้วางแผนใช้มาตรวัดรูปแบบนี้ประเมินพฤติกรรม การปฏิบัติงานของครูผู้สอน โดยเน้นการให้ครูผู้สอนประเมินพฤติกรรมของตนเอง และให้ผู้บังคับบัญชาชั้นต้น คือหัวหน้าหมวดวิชาเป็นผู้ประเมินพฤติกรรมของครูผู้สอนอีกชั้นหนึ่ง เพื่อใช้ข้อมูลส่วนนี้เป็นการอธิบายคุณภาพของมาตรวัดในสถานการณ์ของการประเมินผู้อื่น ในส่วนของพฤติกรรมเป้าหมายที่มุ่งประเมินนั้น ได้เลือกพฤติกรรม การปฏิบัติงานด้านการประเมินผลของครู เป็นพฤติกรรมเป้าหมายทั้งนี้เพราะ เห็นว่าเป็นพฤติกรรมที่สำคัญและเป็นพฤติกรรมการทำงานที่มีรูปแบบที่ชัดเจน (จากการศึกษาแนวคิดว่าด้วยพฤติกรรม การสอนของครู ในตอนที่ 6)

อนึ่ง เพื่อความคมชัดในการหาข้อสรุปเกี่ยวกับคุณภาพของมาตรฐานค่าเชิงพฤติกรรมในการศึกษาครั้งนี้จึงได้เลือกมาตรฐานค่าแบบกราฟิกซึ่งเป็นรูปแบบที่นิยมใช้กันโดยทั่วไป และเป็นรูปแบบที่เคยได้รับการยืนยันคุณภาพในด้านความเที่ยงจากงานวิจัยในประเทศไทยมาแล้ว (มัลลิกา ยูวะนะ เตมีย์, 2527) เป็นมาตรฐานค่าคู่เปรียบเทียบ