

## บทที่ 2

### วรรณคดีที่เกี่ยวข้อง

ในการนำเสนอวรรณคดีที่เกี่ยวข้อง ผู้วิจัยขอนำเสนอเนื้อหาออกเป็น 4 หัวข้อ คือ ความหมายและลักษณะของมาตรฐานค่า ความคลาดเคลื่อนที่เกิดจากการใช้มาตรฐานค่า สถานภาพของผู้ประเมิน และงานวิจัยที่เกี่ยวข้อง

#### 1. ความหมายและลักษณะของมาตรฐานค่า

ปัจจุบัน นักการศึกษาจำนวนมาก ได้นำมาตรฐานค่า มาใช้เป็นเครื่องมือประเมินคุณภาพการสอนและการปฏิบัติงานของครู แสดงให้เห็นว่า มาตรฐานค่า (Rating Scales) เป็นเครื่องมือประเมินคุณภาพการสอนและการปฏิบัติงานของครูที่ได้รับความนิยมอย่างกว้างขวาง ทั้งนี้เป็นเพราะมาตรฐานค่า เป็นเครื่องมือที่สร้างและใช้ง่าย ตลอดจนสามารถใช้วัดพฤติกรรมหรือคุณลักษณะได้หลายด้าน ดังนั้น ได้มีผู้ให้ความหมายของมาตรฐานค่าไว้หลายท่านดังนี้

Good (1973) กล่าวว่า มาตรฐานค่า หมายถึง เครื่องมือที่ใช้ประเมินสิ่งที่มีลักษณะเป็นนามธรรม ไม่สามารถวัดออกมาเป็นตัวเลขโดยตรงได้ เช่น เจตคติ ความเชื่อ แรงจูงใจใฝ่สัมฤทธิ์ (achievement motivation) มโนภาพเกี่ยวกับตนเอง (self - concept) ความคิดเห็นและบุคลิกลักษณะของบุคคล เป็นต้น ดังนั้นมาตรฐานค่าจึงเป็นการแปลงค่าเชิงคุณภาพให้เป็นเชิงปริมาณ ด้วยการประเมินค่านั้นในเชิงเปรียบเทียบกันเป็นกลุ่ม มีการเรียงลำดับมากน้อยด้วยตัวเลข หรือข้อความที่แสดงปริมาณ

Gronlund (1954) กล่าวว่า มาตรฐานค่า เป็นกระบวนการที่เป็นระบบ เพื่อที่จะให้ได้มาซึ่งผลการตัดสินให้คุณค่าของผู้สังเกต ซึ่งมาตรฐานค่าประกอบด้วยคุณลักษณะหรือคุณภาพของสิ่งที่จะประเมิน และระดับคุณค่าของสิ่งที่ปรากฏ

อุทุมพร จารมรمان (2538) กล่าวว่า มาตรฐานค่า หมายถึง มาตรฐานวัดทางจิตวิทยา ที่มีลักษณะเป็นเส้นตรง และมีตัวเลขหรืออักษรกำกับ อาจแบ่งเป็นจำนวนช่วงได้ตั้งแต่ 2 ช่วง อาจจะใช้จำนวนช่วงคู่หรือคี่ก็ได้ อาจมีเฉพาะบวกมากที่สุดถึงศูนย์ หรือจากศูนย์ถึงลบน้อยที่สุดก็ได้ อาจมีความหมายทั้งบวกและลบอยู่ในมาตรวัดเดียวกันก็ได้ หรือจะมีเฉพาะบวกแยก จากลบก็ได้

สรุปได้ว่า มาตรฐานค่า คือ เครื่องมือที่ใช้วัดและประเมินคุณลักษณะ หรือพฤติกรรมเชิงคุณภาพ ให้ออกมาในรูปของข้อมูลเชิงปริมาณหรือคะแนน โดยผู้ทำมาตรฐานค่า จะต้องเลือกตัวเลขหรืออักษร หรือจำนวนช่วงมาตรฐาน เพื่อแสดงถึงปริมาณมากน้อยของคุณลักษณะหรือพฤติกรรมที่ปรากฏ ซึ่งจำนวนช่วงมาตรฐานอาจมีได้ตั้งแต่ 2 ช่วงขึ้นไป

ด้วยเหตุผลที่มาตรฐานค่าเป็นเครื่องมือที่สร้างไม่ยากนัก และสามารถประยุกต์ใช้ในสถานการณ์ต่างๆ ได้อย่างกว้างขวาง จึงมีนักการศึกษาได้พยายามสร้างและพัฒนาคุณลักษณะของมาตรฐานค่าออกเป็นรูปแบบต่าง ๆ กัน ดังนี้

อุทุมพร จามรมาน (2530) ได้จำแนกมาตรฐานค่าออกเป็น 3 ประเภท พร้อมทั้งยกตัวอย่าง ดังนี้

1. มาตรฐานค่าแบบตัวเลข (numerical rating scales) เป็นมาตรฐานค่าที่กำหนดจำนวนช่วงมาตรฐานเป็นตัวเลข กำกับรายการพฤติกรรมหรือคุณลักษณะ เช่น 1 , 2 , 3 , 4 , 5 เป็นต้น โดยให้ผู้ประเมินค่าทำวงกลมล้อมรอบตัวเลขที่ระบุ ตัวอย่างเช่น

บุคลิกภาพของอาจารย์ ก	จำนวนช่วงมาตรฐาน				
1) พุดจาสุภาพ ไพเราะ	1	2	3	4	5
2) ยิ้มแย้มแจ่มใส	1	2	3	4	5

2. มาตรฐานค่าแบบกำหนดเส้น (graphic rating scales) เป็นมาตรฐานค่าที่มีลักษณะคล้ายแบบที่ 1 แต่ใช้ภาษาแทนตัวเลขในจำนวนช่วงมาตรฐาน และใช้เส้นแทนช่อง เช่น

1) ในการสนทนาทางสังคม	-----	-----	-----	-----	-----
ท่านมีพฤติกรรมอย่างไร	ช่างพูด	พูดอย่างสบายใจ	พูดขมขื่น	ชอบฟัง	เงียบ
2) ลักษณะของท่านคือ	-----				
	ชอบบริการคนอื่นมาก	ชอบบริการ	ไม่ชอบบริการ	ไม่ชอบบริการ	ไม่ชอบบริการเลย

3. มาตรฐานค่าแบบบรรยาย (Descriptive Graphic Scales) เป็นมาตรฐานค่าที่มีลักษณะแบบที่ 1 และ 2 รวมกัน คือมีทั้งตัวเลขและภาษากำกับหัวข้อ เช่น



ซึ่งการที่เราจัดลำดับที่ก่อนเช่นนี้จะช่วยให้เห็นความแตกต่างของลักษณะนิสัยระหว่างผู้ถูกประเมินได้ชัดเจนยิ่งขึ้น

5. แบบ paired comparisons ผู้ประเมินต้องเปรียบเทียบผู้ถูกประเมินแต่ละคนกับคนอื่นทุกคนทีละคู่ ถ้ามีลักษณะนิสัยดีกว่าก็ทำเครื่องหมายไว้ เมื่อครบทุกคนแล้ว ก็นับจำนวนความถี่ที่มีลักษณะนิสัยดีกว่าว่ามีจำนวนเท่าใด และด้อยกว่ามีจำนวนเท่าใด การวัดแบบนี้จะได้ค่าความเที่ยงสูงกว่าวิธีอื่น

Guilford ได้แบ่งประเภทของมาตรประมาณค่าตามที่นิยมใช้กันอยู่ออกเป็น 5 แบบ คือ

1. แบบตัวเลข (numerical scales) เป็นมาตรประมาณค่าที่กำหนดตัวเลขพร้อมกับคำนิยามหรือข้อความแสดงพฤติกรรมไว้ให้ผู้ประเมิน ดังตัวอย่างมาตรประมาณค่าที่ใช้ในการประเมินค่าของสีและกลิ่น ต่อไปนี้

10	Most pleasant imaginable
11	Most pleasant
12	Extremely pleasant
13	Moderately pleasant
14	Midly pleasant
15	Indifferent
16	Midly impleasant
17	Moderately impleasant
18	Most impleasant
19	Most impleasant imaginable

หรือบางครั้งอาจจะไม่มีตัวเลขกำกับให้ มีเพียงแต่ข้อความ (Descriptive Cues) ดังนี้

Very Heavy

Heavy

Medium

Light

Very Light

2. แบบกราฟิก (graphic scales) เป็นมาตรฐานค่าที่ใช้เส้นตรงกำกับข้อความแสดงพฤติกรรมแทนตัวเลข มีได้หลายรูปแบบ ดังตัวอย่าง

ในวงสนทนาคุณมักเป็นอย่างไ	----	----	----	----	----
	ช่างพูด	พูดบ่อย	พูดเมื่อ	ชอบฟัง	ไม่เคย
			จำเป็น	มากกว่า	พูดเลย
ลูกของคุณคิดเลขเป็นอย่างไ	----	----	----	----	----
	ช้า	ค่อยๆคิด	ธรรมดา	รวดเร็ว	

3. แบบสเกลมาตรฐาน (standard scales) มาตรฐานค่าแบบนี้ได้กำหนดค่าสเกล (Scales Value) เป็นค่าประจำของชุดมาตรฐานที่จะใช้ในการเปรียบเทียบไว้ล่วงหน้า ผู้ประเมินจะต้องทำการเปรียบเทียบรายการที่ให้ประเมินกับชุดมาตรฐาน

4. แบบแต้มสะสม (rating by cumulative points) แตกต่างจากแบบอื่นที่วิธีการให้คะแนน วิธีการตรวจรายการ (check - list) เป็นตัวอย่างหนึ่งของรูปแบบนี้ ได้มีผู้ใช้วิธีการตรวจรายการในการประเมินคุณลักษณะของเด็ก โดยกำหนดรายการที่มีทั้งคุณลักษณะที่ดีและไม่ดีมาให้เลือก เช่น ให้ความร่วมมือ ช่างคิด คุยง่าย เห็นแก่ตัว เป็นต้น ผู้ประเมินแต่ละคนจะเลือกทุก ๆ คุณลักษณะที่ตรงกับเด็กคนนี้ คะแนนของเด็กแต่ละคนจะเป็นผลรวมของคะแนนทั้งหมด ซึ่งได้กำหนดให้ 1 สำหรับทุกคุณลักษณะที่ดี และให้คะแนน -1 สำหรับทุกคุณลักษณะที่ไม่ดี

5. แบบบังคับให้เลือก (forced - choice rating) เป็นแบบที่พัฒนาขึ้นเพื่อประเมินบุคคล ผู้ประเมินไม่ต้องบอกว่าผู้ถูกประเมินมีลักษณะใด มากน้อยเพียงใด แต่ให้เปรียบเทียบเป็นคู่ ๆ แล้วบอกว่ามีลักษณะใดมากกว่ากัน

Tumey และ Robb (1971) ได้จำแนกประเภทของมาตรฐานค่าออกเป็น 3 รูปแบบพื้นฐาน คือ

1. แบบตัวเลข (numerical scales) เป็นการให้ตัวเลขบ่งชี้ระดับของคุณลักษณะหรือพฤติกรรมที่ต้องการประเมินค่า เช่น ให้ประเมินพฤติกรรมของนักเรียน โดยวงกลมล้อมรอบหมายเลข 1 เมื่ออยู่ในระดับต่ำ, 2 เมื่ออยู่ในระดับต่ำกว่าปานกลาง, 3 เมื่ออยู่ในระดับปานกลาง, 4 เมื่ออยู่ในระดับสูงกว่าปานกลาง และ 5 เมื่ออยู่ในระดับสูง ดังตัวอย่าง

ก. ความร่วมมือกับเพื่อน ๆ	1	2	3	4	5
ข. ความร่วมมือกับครู	1	2	3	4	5



## 2. ความคลาดเคลื่อนที่เกิดจากการใช้มาตราประมาณค่า

ถึงแม้ว่า มาตราประมาณค่าเป็นเครื่องมือที่สร้างง่าย ใช้สะดวก เป็นที่นิยมกันอย่างกว้างขวาง โดยเฉพาะการนำมาตราประมาณค่ามาใช้ในการประเมินการปฏิบัติงานของบุคลากรในหน่วยงาน แต่การใช้มาตราประมาณค่ายังคงมีข้อควรระวัง เพราะสารสนเทศที่ได้จากการใช้มาตราประมาณค่า อาจเกิดความคลาดเคลื่อน หรือขาดความเที่ยงตรง ความน่าเชื่อถือได้ เนื่องจากมาตราประมาณค่า เป็นเครื่องมือวัดพฤติกรรมทางจิตวิทยา และพฤติกรรมส่วนใหญ่เป็นพฤติกรรมภายในตัวบุคคล หรือมีคุณลักษณะแฝง (latent trait) บางครั้งสังเกตได้ยาก ผู้สังเกตอาจมีความสามารถหรือประสบการณ์ในการสังเกตพฤติกรรมภายในดังกล่าวน้อย หรือแม้แต่ในส่วนของเครื่องมือมาตราประมาณค่า อาจประกอบด้วยข้อความที่ไม่ชัดเจน คลุมเครือ ก่อให้เกิดความไม่เข้าใจตรงกันระหว่างผู้สังเกต หรือผู้ประเมิน เป็นผลให้เกิดความคลาดเคลื่อนในการวัดนั่นเอง (Guilford, 1975)

Selltiz (1959) กล่าวว่า การใช้เครื่องมือมาตราประมาณค่าในการประเมินการปฏิบัติงาน หรือการเรียนการสอนของครูนั้น มีแหล่งความคลาดเคลื่อนที่อาจเกิดขึ้นได้จาก 3 แหล่งดังนี้

1. ความบกพร่องที่เกิดจากกระบวนการประเมิน เป็นความคลาดเคลื่อนที่เกิดจากกำหนดคุณลักษณะหรือสิ่งที่ต้องการให้ประเมินนั้นมีจำนวนมากเกินไป จึงส่งผลให้การประเมินเกิดความคลาดเคลื่อนได้ และข้อบกพร่องอีกประการหนึ่งคือ กรณีที่ใช้ผู้ประเมิน (rater) เพียงคนเดียว ทำการประเมิน ผู้ถูกประเมิน (ratee) หลาย ๆ คน สามารถก่อให้เกิดความคลาดเคลื่อนได้เช่นกัน ซึ่งควรแก้ไขโดยใช้ผู้ประเมินหลาย ๆ คน อันเป็นวิธีที่ช่วยลดความคลาดเคลื่อนที่เกิดจากผู้ประเมินแต่ละคนได้ ทำให้ผลการประเมินมีความแม่นยำมากขึ้น

2. ความบกพร่องของเครื่องที่ใช้ในการประเมิน เป็นความคลาดเคลื่อนที่เกิดจากการใช้เครื่องมือที่เป็นมาตราประมาณค่าที่ขาดการนิยามหรือการบรรยายคุณลักษณะ (characteristic) ของสิ่งที่จะประเมินไว้อย่างชัดเจน ทำให้เกิดความเข้าใจที่ไม่ตรงกันระหว่างผู้ประเมิน เป็นผลให้การประเมินค่านั้นขาดความเที่ยง หรืออาจเป็นเพราะภาษาที่ใช้ไม่เหมาะสม หรือไม่สามารวัดสิ่งที่ต้องการวัดได้ เป็นต้น

3. ความบกพร่องที่เกิดจากตัวผู้ประเมิน เป็นความคลาดเคลื่อนที่อาจเกิดจากความลำเอียง หรือความมีอคติของแต่ละบุคคล (personal bias) ซึ่งเป็นผลมาจากการมีปฏิสัมพันธ์มาก่อน (halo effect) และความคลาดเคลื่อนแบบใช้ตรรกผิด (logical error)

- 3.1 ความลำเอียงหรือความมีอคติส่วนบุคคลของผู้ประเมิน เป็นความคลาดเคลื่อนที่เกิดขึ้นเฉพาะลักษณะนิสัยประจำตัวของผู้ประเมิน ที่มักมีรูปแบบการประเมินตายตัว (fixed pattern) ซึ่งมี 3 ชนิด คือ

3.1.1 ความคลาดเคลื่อนที่เกิดจากการประเมินค่าสูงกว่าความเป็นจริง (generosity error) เกิดขึ้นเมื่อผู้ประเมินประเมินค่าคุณลักษณะต่าง ๆ ของผู้ถูกประเมินทุกคนในระดับที่ค่อนข้างสูงเพียงด้านเดียวอย่างสม่ำเสมอ

3.1.2 ความคลาดเคลื่อนที่เกิดจากการประเมินค่าต่ำกว่าความเป็นจริง (severity error) เกิดขึ้นเมื่อผู้ประเมิน ประเมินค่าคุณลักษณะต่าง ๆ ของผู้ถูกประเมินทุกคนในระดับที่ค่อนข้างต่ำเพียงด้านเดียวอย่างสม่ำเสมอ

3.1.3 ความคลาดเคลื่อนที่เกิดจากการประเมินค่าให้มีแนวโน้มเข้าสู่ส่วนกลาง (central tendency error) เกิดขึ้นเมื่อผู้ประเมินพยายามหลีกเลี่ยงการประเมินค่าทางระดับค่อนข้างสูงและต่ำ มาเป็นระดับปานกลางอย่างสม่ำเสมอ

3.2 ผลจากการมีปฏิสัมพันธ์มาก่อน (halo effect) เป็นความคลาดเคลื่อนที่เกิดขึ้นเมื่อความรู้สึกของผู้ประเมินที่มีต่อผู้ถูกประเมินมีอิทธิพลต่อการประเมินค่าคุณลักษณะของผู้ถูกประเมิน เช่น ถ้าผู้ประเมินมีความชอบพอต่อผู้ถูกประเมินมาก่อนแล้ว ผู้ประเมินมักจะประเมินคุณลักษณะต่าง ๆ ของผู้ถูกประเมินในทางบวก หรือในทางตรงกันข้าม ถ้าผู้ประเมินมีความไม่พึงพอใจต่อผู้ถูกประเมินมาก่อนแล้ว ผู้ประเมินมักจะประเมินคุณลักษณะต่าง ๆ ของผู้ถูกประเมินในทางลบ ความคลาดเคลื่อนแบบนี้จะเกิดขึ้นกับเฉพาะผู้ถูกประเมินบางคน

3.3 ความคลาดเคลื่อนแบบใช้ตรรกผิด (logical error) เป็นความคลาดเคลื่อนที่เกิดขึ้นเมื่อผู้ประเมินนำตัวแปรอื่นที่คิดว่ามีความสัมพันธ์กับตัวแปรที่กำลังประเมินมามีส่วนในการประเมินด้วย เช่น การประเมินความฉลาด ครูมักจะประเมินค่าความฉลาดของนักเรียนจากผลสัมฤทธิ์ทางการเรียนของนักเรียน กล่าวคือ ถ้านักเรียนมีผลสัมฤทธิ์ทางการเรียนในระดับสูง นักเรียนคนนั้นมักจะได้รับการประเมินความฉลาดในระดับสูง ทั้งนี้เพราะครูมีความเชื่อว่า คุณลักษณะ 2 คุณลักษณะ คือ ความฉลาดกับผลสัมฤทธิ์ทางการเรียนมีความสัมพันธ์กัน เป็นต้น

นอกจากนี้ Guilford (1954) ได้แบ่งประเภทของความคลาดเคลื่อนในการประเมินค่าด้วยมาตรประมาณค่าออกเป็นรูปแบบต่าง ๆ ดังนี้

1. ความคลาดเคลื่อนแบบปล่อยคะแนน (leniency error) เกิดจากความพยายามของผู้ประเมินที่จะให้คุณค่าของผู้ประเมินดีเกินความเป็นจริง หรือในทางตรงกันข้าม ผู้ถูกประเมินได้รับการประเมินคุณค่าต่ำเกินความเป็นจริงอย่างสม่ำเสมอ ไม่ว่าจะด้วยเหตุผลใดก็ตาม

2. ความคลาดเคลื่อนแบบมีแนวโน้มเข้าสู่ส่วนกลาง (central tendency error) เกิดขึ้นเมื่อผู้ประเมินพยายามหลีกเลี่ยงที่จะประเมินคุณค่าให้มากไปทางใดทางหนึ่งจนเกินไป ประกอบกับว่า ผู้ประเมินอาจจะไม่รู้จักผู้ถูกประเมินดีพอ จึงมีผลให้ผู้ประเมินพยายามประเมินให้คุณค่าในระดับปานกลางอย่างสม่ำเสมอ



3. ความคลาดเคลื่อนแบบฮาโล (halo error) เป็นความคลาดเคลื่อนที่เกิดขึ้นเมื่อผู้ถูกประเมินมีอิทธิพลต่อผู้ประเมิน คือ ผู้ประเมินจะให้คุณค่ามากหรือน้อยขึ้นอยู่กับความพอใจของผู้ประเมินต่อผู้ถูกประเมิน ซึ่งความคลาดเคลื่อนแบบนี้จะเกิดเฉพาะผู้ถูกประเมินบางคนเท่านั้น

4. ความคลาดเคลื่อนแบบใช้ตรรกผิด (logical error) เกิดขึ้นเมื่อผู้ประเมินนำตัวแปรอื่นที่ตนเองคิดว่ามีความสัมพันธ์กับตัวแปรที่ประเมินมามีส่วนในการประเมินด้วย ความคลาดเคลื่อนประเภทนี้มีได้เกิดจากความลำเอียงแต่เกิดจากการที่ผู้ประเมินได้รับรู้รับฟังเรื่องราวของผู้ถูกประเมินมาก่อน

5. ความคลาดเคลื่อนอันเนื่องมาจากความใกล้เคียง (proximity error) เกิดขึ้นเมื่อข้อความที่แสดงถึงคุณลักษณะ 2 คุณลักษณะ (trait) อยู่ใกล้กัน หรือนำมาใช้ในการประเมินในเวลาใกล้เคียงกัน ผลการประเมินจะมีค่าสูง แต่ถ้านำคุณลักษณะที่มีความคล้ายคลึงกันนี้มาวางให้ห่างกัน หรือประเมินในเวลาที่ไม่ใกล้เคียงกัน ผลการประเมินจะมีค่าต่ำลง

6. ความคลาดเคลื่อนแบบขัดแย้ง (contrast error) เกิดขึ้นเมื่อผู้ประเมินให้ระดับคุณค่าแก่ผู้ถูกประเมินตรงกันข้ามกับความรู้สึก หรือคุณลักษณะเดียวกันของตน ขณะที่ทำการประเมิน กล่าวคือ ผู้ประเมินที่คิดว่าตนเองมีคุณลักษณะ (trait) ที่ประเมินสูง มีแนวโน้มที่จะประเมินผู้ถูกประเมินให้มีลักษณะที่ต่ำกว่าที่ตนเองมีอยู่ และในทางตรงกันข้าม ผู้ประเมินที่คิดว่าตนเองมีคุณลักษณะที่ประเมินต่ำ มีแนวโน้มที่จะประเมินผู้ถูกประเมินสูงกว่าที่ตนเองมีอยู่

ในการวิเคราะห์หาค่าความคลาดเคลื่อนในการประเมินค่านี้จะเป็นการเปรียบเทียบความคลาดเคลื่อนที่เกิดจากการประเมินของผู้ประเมิน โดยใช้มาตรฐานค่าที่มีจำนวนช่วงต่างกัน (rater error on different number of intervals) ซึ่งสามารถทำการวิเคราะห์ได้ด้วยการวิเคราะห์ความแปรปรวน 2 ทาง จากข้อมูลที่ประกอบด้วยผู้ประเมิน (raters) และผู้ถูกประเมิน (ratees) หลายๆ คน ทำการประเมินข้อมูลหลาย ๆ กลุ่มคุณลักษณะ (traits) ดังนั้นประเภทของความคลาดเคลื่อนที่สามารถหาค่าได้ด้วยการวิเคราะห์ความแปรปรวนสองทางและให้ค่าคะแนนความคลาดเคลื่อนได้ทั้งค่าบวกและค่าลบ ได้แก่ ความคลาดเคลื่อนแบบปล่อยคะแนน ความคลาดเคลื่อนแบบฮาโล และความคลาดเคลื่อนแบบขัดแย้ง ซึ่งแหล่งความแปรปรวนแต่ละแหล่งมีดังนี้

ผู้ประเมิน (Raters)

ผู้ถูกประเมิน (Ratees)

ปฏิสัมพันธ์ระหว่างผู้ประเมินกับผู้ถูกประเมิน (Raters x Ratees)

คุณลักษณะ (Traits)

ปฏิสัมพันธ์ระหว่างผู้ประเมินกับคุณลักษณะ (Raters x Traits)

โดยความคลาดเคลื่อนแต่ละชนิด สามารถพิจารณาได้ดังนี้

1. ความคลาดเคลื่อนแบบปล่อยคะแนน สามารถตรวจสอบได้ด้วยการวิเคราะห์ความแปรปรวนสองทาง ซึ่งถ้าอิทธิพลหลักของผู้ประเมิน (rater main effect) มีนัยสำคัญทางสถิติ แสดงว่าคะแนนที่ได้จากการประเมินมีความคลาดเคลื่อนแบบปล่อยคะแนน
2. ความคลาดเคลื่อนแบบฮาโล สามารถตรวจสอบได้ด้วยการวิเคราะห์ความแปรปรวนสองทาง ซึ่งถ้าปฏิสัมพันธ์ระหว่างผู้ประเมินและผู้ถูกประเมิน (raters x ratees) มีนัยสำคัญทางสถิติ แสดงว่า คะแนนที่ได้จากการประเมินมีความคลาดเคลื่อนแบบฮาโล
3. ความคลาดเคลื่อนแบบขัดแย้ง สามารถตรวจสอบได้ด้วยการวิเคราะห์ความแปรปรวนสองทาง ซึ่งถ้าปฏิสัมพันธ์ระหว่างผู้ประเมินกับคุณลักษณะ (raters x traits) มีนัยสำคัญทางสถิติ แสดงว่าคะแนนที่ได้จากการประเมินมีความคลาดเคลื่อนแบบขัดแย้ง

ส่วนการประมาณค่าคะแนนความคลาดเคลื่อนแต่ละชนิด สามารถเขียนให้อยู่ในรูปสมการเชิงเส้นตรงทางคณิตศาสตร์ในรูปทั่วไปของคะแนนการประเมินได้ดังนี้

$$X_{ijk} = X_{ji} + X_{ijkc} \quad \text{----- (1)}$$

โดยที่  $X_{ijk}$  คือ คะแนนการประเมินของคนที I ในคุณลักษณะ J โดยผู้ประเมิน K  
 $X_{ji}$  คือ คะแนนการประเมินจริงของคนที I ในคุณลักษณะ J  
 $X_{ijkc}$  คือ คะแนนความคลาดเคลื่อนทั้งหมดในการประเมิน  $X_{ijk}$

สำหรับค่า  $X_{ijkc}$  สามารถแยกออกเป็นส่วนประกอบย่อยๆ ได้ดังนี้

$X_{ki}$  เป็นคะแนนความคลาดเคลื่อนแบบปล่อยคะแนน (rater K's leniency error) ซึ่งเป็นลักษณะที่ผู้ประเมินมีแนวโน้มที่จะให้คะแนนสูงกว่าหรือต่ำกว่าความเป็นจริง สามารถพิจารณาได้จากอิทธิพลหลักของผู้ประเมิน (rater main effect)

$X_{ki}$  เป็นคะแนนความคลาดเคลื่อนแบบฮาโล ระหว่างผู้ประเมินกับผู้ถูกประเมิน (rater K's halo error in connection with person I) ซึ่งผู้ประเมินมีแนวโน้มที่จะให้คะแนนสูงในผู้ถูกประเมินบางคน ในขณะที่ให้คะแนนต่ำในผู้ถูกประเมินคนอื่นในคุณลักษณะเดียวกัน สามารถพิจารณาได้จากปฏิสัมพันธ์ระหว่างผู้ประเมินกับผู้ถูกประเมินในการวิเคราะห์ความแปรปรวน (rater - ratee interaction error)

$X_{kj}$  เป็นคะแนนความคลาดเคลื่อนแบบขัดแย้ง ที่ในบางคุณลักษณะผู้ประเมินให้คะแนนสูงกว่าความเป็นจริง ในขณะที่บางคุณลักษณะให้คะแนนต่ำกว่าความเป็นจริง สามารถ

พิจารณาได้จากปฏิสัมพันธ์ระหว่างผู้ประเมินกับคุณลักษณะที่ประเมินในการวิเคราะห์ความแปรปรวน (rater - trait interaction error)

$X_{ijk}$  เป็นคะแนนความคลาดเคลื่อนที่เหลือจากการประเมินคนที่ I โดยผู้ประเมิน K (a residual error make by rater K in rating person I)

จากคะแนนความคลาดเคลื่อนทั้งหมด  $X_{ijkc}$  สามารถเขียนให้อยู่ในรูปสมการอย่างง่ายได้ดังนี้

$$X_{ijkc} = X_{ki} + X_{ki} + X_{kj} + X_{ijkr} \text{ ----- (2)}$$

และนำสมการที่ (2) แทนค่าในสมการที่ (1) ได้ดังนี้

$$X_{ijk} = X_{ij} + X_{ki} + X_{ki} + X_{kj} + X_{ijkr} \text{ ----- (3)}$$

การประมาณค่าความคลาดเคลื่อน (The Estimation of Error Contributions to Ratings) จากสมการ (3) สามารถเขียนให้อยู่ในรูปของสมการประมาณค่าได้ดังนี้

$$X'_{ijk} = X'_{ij} + X'_{ki} + X'_{ki} + X'_{kj} + X'_{ijkr} \text{ ----- (4)}$$

เพื่อที่จะกำจัดค่าความคลาดเคลื่อนออกจากคะแนนที่สังเกตได้ (observed score) ดังนั้นจึงต้องมีการประมาณค่าความคลาดเคลื่อน ในที่นี้จะใช้สัญลักษณ์แทนค่าประมาณความคลาดเคลื่อนแต่ละชนิดดังนี้

- $X'_{ki}$  แทนการประมาณค่า leniency error
- $X'_{ki}$  แทนการประมาณค่า halo error
- $X'_{kj}$  แทนการประมาณค่า contrast error
- $X'_{ijkr}$  แทนการประมาณค่า residual error

การประมาณค่าความคลาดเคลื่อนแบบปล่อยคะแนน ( $X'_{ki}$ ) มีขั้นตอนดังนี้

- 1) หาค่าเฉลี่ยของคะแนนการประเมิน ที่ผู้ประเมินแต่ละคนทำการประเมินผู้ได้รับการประเมินแต่ละคน
- 2) หาค่าเฉลี่ยของผู้ถูกประเมินที่ได้รับการประเมินจากผู้ประเมินคนเดียวกัน (mean all rates)

3) หาค่าเฉลี่ยของผู้ประเมินทั้งหมดที่ประเมินผู้ถูกประเมินแต่ละคน (mean all raters)

4) หาค่า grand mean

5) หาค่า  $X_{ki}$  จาก mean all ratees - grand mean

6) หาค่า  $d_i$  จาก mean all raters - grand mean

ข้อพิจารณา ถ้า  $X'_{ki} = 0$  แสดงว่า ผลการประเมินไม่มีความคลาดเคลื่อนแบบ  
ปล่อยคะแนน

$X'_{ki} > 0$  แสดงว่า ผลการประเมินมีความคลาดเคลื่อน  
แบบปล่อยคะแนน โดยให้คะแนนสูงกว่า  
ความเป็นจริง

$X'_{ki} < 0$  แสดงว่า ผลการประเมินมีความคลาดเคลื่อนแบบ  
ปล่อยคะแนน โดยให้คะแนนต่ำกว่าความเป็นจริง

การประมาณค่าความคลาดเคลื่อนแบบมีปฏิสัมพันธ์ ( $X'_{ki}$ ) มีขั้นตอนดังนี้

- 1) นำค่าเฉลี่ยแต่ละเซลล์ในขั้นตอนที่ 1 จากข้อ 1.1 ลบด้วยค่า  $X_{ki}$  และค่า  $d_i$
- 2) หาค่าเฉลี่ยของผลลัพธ์ที่ได้จากขั้นตอนที่ 2.1
- 3) นำผลที่ได้จากข้อ 2.2 ลบด้วยค่า Grand Mean

ข้อพิจารณา ถ้า  $X'_{ki} = 0$  แสดงว่า ผลการประเมินไม่มีความคลาดเคลื่อน  
แบบฮาโล

$X'_{ki} > 0$  แสดงว่า ผลการประเมินมีความคลาดเคลื่อนแบบ  
ฮาโล โดยให้คะแนนสูงกว่าความเป็นจริง  
(overvalue)

$X'_{ki} < 0$  แสดงว่า ผลการประเมินมีความคลาดเคลื่อนแบบ  
ฮาโล โดยให้คะแนนต่ำกว่าความเป็นจริง  
(undervalue)

การประมาณค่าความคลาดเคลื่อนแบบขัดแย้ง ( $X'_{kj}$ ) มีขั้นตอนดังนี้

- 1) หาค่าเฉลี่ยคะแนนการประเมินของผู้ประเมินแต่ละคนที่ทำการประเมินใน  
แต่ละคุณลักษณะ

- 2) หาค่าเฉลี่ยคะแนนการประเมินคุณลักษณะของผู้ประเมินแต่ละคน (mean all traits)
- 3) หาค่าเฉลี่ยคะแนนการประเมินคุณลักษณะแต่ละคุณลักษณะของผู้ประเมินทุกคน (mean all raters)
- 4) หาค่า grand mean
- 5) หาค่า  $X_{ki}$  จาก mean all traits - grand mean
- 6) หาค่า  $d_i$  จาก mean all raters - grand mean
- 7) พิจารณาค่า  $X_{ki}$  เช่นเดียวกับข้อ 1
- 8) นำค่าเฉลี่ยแต่ละเซลล์จากข้อ 3.1 ลบด้วยค่า  $X_{ki}$  และ  $d_i$  ที่คำนวณได้ในข้อ 3.5 และ 3.6
- 9) นำผลลัพธ์ที่ได้จากขั้นตอนที่ 8 ลบด้วยค่า grand mean แต่ละเซลล์

ข้อพิจารณา ถ้า  $X'_{kj} = 0$  แสดงว่า ผลการประเมินไม่มีความคลาดเคลื่อนแบบจัดแย้ง

$X'_{kj} > 0$  แสดงว่า ผลการประเมินมีความคลาดเคลื่อนแบบจัดแย้งในลักษณะที่ผู้ประเมินให้คะแนนต่ำกว่าความเป็นจริง

$X'_{kj} < 0$  แสดงว่า ผลการประเมินมีความคลาดเคลื่อนแบบจัดแย้งในลักษณะที่ผู้ประเมินให้คะแนนสูงกว่าความเป็นจริง

ความคลาดเคลื่อนดังที่ได้กล่าวมานี้เป็นความคลาดเคลื่อนที่เกิดขึ้นอย่างเป็นระบบ (systematic error) ที่ทำให้ความตรงลดลง ดังนั้นการปรับปรุงแก้ไขให้การใช้มาตรฐานค่าเป็นเครื่องมือประเมินที่ให้ผลถูกต้องและน่าเชื่อถือได้นั้น Thomdike และ Hagen (1977) ได้สรุปไว้ว่า ควรปฏิบัติตามเกณฑ์ต่อไปนี้

1. การประเมินควรจะทำเฉพาะลักษณะที่ปรากฏให้เห็นชัดเจน
2. ลักษณะที่จะประเมินต้องแสดงออกมาให้เห็นในรูปของพฤติกรรม และการตัดสินใจจะกระทำจากพฤติกรรมที่เห็นนั้น
3. รูปแบบของมาตรฐานค่า (rating form) ต้องปรับปรุงเพื่อบังคับให้ผู้ประเมินบอกความแตกต่าง หรือเพื่อควบคุมความแตกต่างของมาตรฐานในการตัดสินใจระหว่างผู้ประเมินแต่ละคน
4. ผู้ประเมินต้องเป็นผู้ที่มีโอกาสมากที่สุดในการสังเกตพฤติกรรมในสถานการณ์ตามที่ต้องการ

5. ผู้ประเมินต้องได้รับการบอกเล่าถึงคุณค่าของการประเมินค่าและได้รับการฝึกฝนการใช้เครื่องมือ

6. ถ้าสามารถมีผู้ประเมินหลายคนก็ควรรวบรวมผลการประเมินจากผู้ประเมินหลายคน

ในการประเมินค่านั้น แม้จะมีความคลาดเคลื่อนเกิดขึ้นได้ แต่ก็มีวิธีที่จะช่วยลดความคลาดเคลื่อนได้โดยออกแบบมาตรฐานค่าให้ถูกต้องและเหมาะสม Guilford (1954) ได้เสนอแนะวิธีการออกแบบและสร้างมาตรฐานค่าให้มีคุณภาพ ซึ่งแบ่งเป็น 3 ขั้นตอนดังนี้

1. เลือกคุณลักษณะหรือพฤติกรรมที่จะประเมินค่า ซึ่งควรมีลักษณะดังต่อไปนี้
  - 1.1 ต้องมีความสำคัญและตรงกับสิ่งที่ต้องการประเมินมากที่สุด
  - 1.2 ต้องเป็นอิสระต่อกัน ไม่กำกวมกับคุณลักษณะอื่น ๆ
  - 1.3 ต้องเป็นสิ่งที่สามารถสังเกตได้โดยตรง
  - 1.4 แต่ละองค์ประกอบของพฤติกรรมมีความเป็นเอกพันธ์ (Homogeneity)
  - 1.5 คุณลักษณะหรือพฤติกรรมต่าง ๆ ไม่ควรมีมากเกินไป ปกติควรมี 5 - 6

รายการ

2. นิยามหรือบรรยายคุณลักษณะหรือพฤติกรรมที่จะประเมินค่าให้ชัดเจนเพื่อให้เกิดความเข้าใจที่ตรงกัน และสามารถตัดสินได้อย่างเป็นปรนัย (objectivity) และเที่ยงตรง (reliability)

3. ออกแบบจำนวนช่วงมาตรฐาน (number of intervals) ที่จะใช้ประเมิน ดังนี้

3.1 ปกติมักจะแบ่งออกเป็น 3 - 7 ช่วง การแบ่งจำนวนช่วงยิ่งมากจะช่วงให้ผู้ประเมินแยกแยะความแตกต่างได้มาก การแบ่งจำนวนช่วงเป็นจำนวนคี่จะมีข้อดีที่ทำให้มีค่ากลางและในบางครั้งอาจใช้ในกรณีที่การประเมินค่านั้นตัดสินใจไม่ได้ ซึ่งโดยทั่วไปแล้วมักจะนิยมใช้มาตรฐานค่าที่มีจำนวนช่วงมาตรฐาน 5 ช่วง

3.2 มีการนิยามหรืออธิบายความต่อเนื่องในแต่ละจำนวนช่วงมาตรฐานให้ชัดเจน

3.3 ควรมีการสลับตัวเลขกำกับช่วงมาตรฐานจากน้อยไปมาก และจากมากไปน้อยบ้างในบางครั้ง ทั้งนี้เพื่อป้องกันมิให้ผู้ประเมินประเมินค่าโดยอัตโนมัติ

Anderson (1988) ได้ให้ข้อเสนอแนะในการพัฒนาจำนวนช่วงมาตรฐานในการใช้มาตรฐานค่าไว้ว่า การใช้จำนวนช่วงมาตรฐานเป็นจำนวนคู่โดยตัดค่าตรงกลางออกไปนั้นมีจุดมุ่งหมายเพื่อบังคับให้ผู้ประเมินต้องเลือกคำตอบไปในทิศทางใดทิศทางหนึ่งเท่านั้น เพราะการกำหนดให้มีค่าตรงกลางเพื่อแสดงความไม่แน่ใจไว้ในมาตรฐานค่านั้น จะทำให้ผู้ประเมินหลีกเลี่ยงการเลือกคำตอบที่แท้จริงโดยหันมาเลือกคำตอบไม่แน่ใจมากขึ้น ในด้านจำนวนช่วงมาตรฐานนั้น การใช้จำนวนช่วงมาตรฐานน้อยก็เพื่อให้เหมาะกับผู้ประเมินที่มีอายุน้อย หรือมีระดับการศึกษาต่ำ ในขณะที่เดียวกันก็เป็นการลดการกระจายของคะแนนซึ่งมีแนวโน้มที่จะทำให้ความเที่ยงลดลงด้วย ส่วนการใช้จำนวน

ช่วงมาตรฐานนั้นก็เพื่อเพิ่มความแปรปรวนของคะแนนและทำให้ความคงที่ภายในเพิ่มขึ้นด้วย แต่มีข้อเสียคือ ผู้ประเมินอาจไม่สามารถแยกแยะความแตกต่างเชิงจิตวิทยาของคำตอบที่อยู่ติดกันได้

### 3. สถานภาพของผู้ประเมิน

กระบวนการประเมินการเรียนการสอน ถือได้ว่าเป็นกระบวนการเก็บรวบรวมข้อมูลจากแหล่งต่าง ๆ มาสรุปรวมเข้าด้วยกัน และข้อมูลที่ได้จะมีความน่าเชื่อถือมากน้อยเพียงใด ย่อมขึ้นอยู่กับความเหมาะสมของเครื่องมือและแหล่งสารสนเทศที่ใช้ในการประเมินการเรียนการสอน (วิจิตรศรีสอาน, 2525)

แหล่งสารสนเทศที่ใช้ในการประเมิน หรือเรียกอีกอย่างหนึ่งว่า สถานภาพของผู้ประเมิน (rater status) มีหลายแหล่ง แต่ละแหล่งสามารถนำไปใช้ด้วยจุดประสงค์ที่แตกต่างกัน มีผู้ได้แบ่งสถานภาพของผู้ประเมิน (rater status) ไว้หลายประเภทดังนี้

Batista (1976) ได้แบ่งสถานภาพของผู้ประเมินไว้ 6 ประเภท ดังนี้

1. self evaluation หรือ การประเมินตนเอง เป็นการตรวจสอบ วิเคราะห์การสอนของครูผู้สอนด้วยตนเองอย่างสม่ำเสมอโดยปราศจากอคติ ซึ่งย่อมได้ข้อมูลที่สามารถนำมาปรับปรุงแก้ไขข้อบกพร่องในการเรียนการสอนเป็นอย่างดี เนื่องจากข้อมูลบางส่วนไม่มีใครทราบได้นอกจากตัวผู้สอน สำหรับการนำข้อมูลการประเมินตนเองไปใช้ในการบริหารนั้น อาจก่อให้เกิดผลดีโดยที่ครูผู้สอนรู้สึกว่าได้มีส่วนร่วมในการประเมินด้วย การประเมินตนเองนี้สามารถนำเทคโนโลยีที่ทันสมัยมาใช้ได้ด้วย เช่น เทปบันทึกเสียง วิดีโอเทป เป็นต้น

2. student evaluation หรือ การประเมินโดยนักเรียน ถือเป็นแหล่งข้อมูลหลักในการให้ข้อมูลที่เกี่ยวข้องกับการสอนในรายวิชานั้น ๆ ได้เป็นอย่างดี เช่น บุคลิกลักษณะของครูผู้สอน เอกสารตำราที่ใช้ การบ้านที่มอบให้ ความสมบูรณ์ของเนื้อหาสาระ วิธีการสอน ความเอาใจใส่ต่อนักเรียน ความตั้งใจของนักเรียน เจตคติของนักเรียนต่อรายวิชานั้น เป็นต้น เนื่องจากเป็นผู้ได้สัมผัสกับตัวแปรนั้นโดยตรง ครูผู้สอนจึงควรเปิดโอกาสให้นักเรียนได้แสดงความคิดเห็นถึงการสอนของครูสำหรับเป็นข้อมูลในการปรับปรุงการสอนให้มีคุณภาพและประสิทธิภาพมากขึ้น

3. peer evaluation หรือการประเมินโดยเพื่อนร่วมงาน เพื่อนร่วมงาน เป็นผู้มีความรู้ความเข้าใจในหลักวิธีการสอน รู้จุดมุ่งหมายของการสอนเป็นอย่างดีเช่นเดียวกับตัวผู้สอน การยอมรับความคิดเห็นของเพื่อนร่วมงานที่วิพากษ์วิจารณ์การสอนอย่างยุติธรรมย่อมเป็นประโยชน์ในการแลกเปลี่ยนประสบการณ์สำหรับการปรับปรุงการสอนให้ดีขึ้น การประเมินโดยเพื่อนร่วมงานสามารถกระทำได้โดยการให้เข้าไปสังเกตการสอนในชั้นเรียน หรืออาศัยจากประสบการณ์ ความสัมพันธ์ การสังเกตเป็นเวลานาน Centra (1975) ได้เสนอแนะว่า การประเมินโดยเพื่อนร่วมงานที่เข้าไปสังเกตการสอน ต้องได้รับการฝึกและปฏิบัติอย่างถูกต้อง ผู้สังเกตควรสังเกตมากกว่า 1 ครั้ง

และควรใช้ผู้สังเกตมากกว่า 1 คน และรายการที่สังเกตพฤติกรรมต้องมีจำนวนมากพอสมควร จึงทำให้ผลการประเมินมีความน่าเชื่อถือได้

4. superior evaluation หรือการประเมินโดยผู้บริหาร ผู้บริหาร เป็นผู้ดูแลแก้ไข ปรับปรุง และส่งเสริมให้การปฏิบัติงานของครูเป็นไปตามนโยบายหรือแผนงานที่โรงเรียนกำหนดไว้ ผู้บริหารควรเอาใจใส่ในการติดตามการสอนของครูโดยสม่ำเสมอ มีทักษะในการสังเกตหรือการสัมภาษณ์เป็นอย่างดี การประเมินโดยผู้บริหารเพื่อการพิจารณาความคิดความชอบของครู ผู้บริหารต้องปฏิบัติไปโดยไม่ลำเอียงหรืออาศัยความชอบส่วนตัว การสังเกตการสอนโดยตรงของผู้บริหาร และสามารถนำไปพิจารณาร่วมกับแหล่งข้อมูลอื่น ๆ การประเมินคุณภาพการสอนของครูเป็นส่วนหนึ่ง que ผู้บริหารสามารถใช้ในการตรวจสอบคุณภาพของโรงเรียน

5. ผลสัมฤทธิ์ทางการเรียนของนักเรียน (student achievement) เป็นการวัดความรู้ในสิ่งที่นักเรียนได้รับจากการสอนของครูซึ่งสะท้อนให้เห็นถึงคุณภาพการสอน ปัญหาที่พบคือ แบบทดสอบวัดผลสัมฤทธิ์ของนักเรียนไม่สามารถวัดได้ครอบคลุมจุดประสงค์ทั้งหมด การสร้างแบบทดสอบวัดผลสัมฤทธิ์ของครูยังมีข้อบกพร่องอยู่ แบบทดสอบที่เป็นมาตรฐานยังมีน้อยและมีปัญหาในการขอใช้ นอกจากนี้จุดประสงค์ของการเรียนการสอนไม่ต้องการเพียงความรู้ในเนื้อหาวิชานั้นอย่างเดียว นักเรียนต้องมีคุณธรรมและจริยธรรมด้วย การที่ครูมุ่งสอนให้นักเรียนสอบได้โดยตัวข้อสอบจึงไม่เหมาะสม อย่างไรก็ตามผลสัมฤทธิ์ทางการเรียนของนักเรียนสามารถใช้เป็นข้อมูลที่ี้ดีได้ในระดับหนึ่ง เพื่อให้ครูได้ทราบว่า นักเรียนเรียนรู้อะไรบ้าง มีการเปลี่ยนแปลงพฤติกรรมอย่างไร ซึ่งอาจใช้ปรับปรุงการสอนให้ดีขึ้นต่อไป

6. ศิษย์เก่า ( alumni evaluation ) นักเรียนที่จบการศึกษาแล้ว และออกไปศึกษาต่อหรือประกอบอาชีพทำให้เขามีประสบการณ์มากขึ้น สามารถให้ข้อมูลเพื่อนำมาเปรียบเทียบบวิพากษ์วิจารณ์ในสิ่งที่ผ่านมา ข้อมูลจากศิษย์เก่าเป็นวิธีการหนึ่งที่ให้ภาพการสอนของครูได้เป็นอย่างดี ดังที่ Howard (1985) พบว่า การประเมินคุณภาพการสอนโดยศิษย์เก่า เป็นวิธีการที่มีประสิทธิภาพและมีความตรง แต่ปัญหาคือ การติดตามเก็บข้อมูลค่อนข้างทำได้ลำบาก

Mathis (1979) ได้เสนอสถานภาพของผู้ประเมินที่ใช้ในการประเมินพฤติกรรมกาปฏิบัติงานที่นิยมใช้มี 4 รูปแบบ คือ

1. การประเมินโดยกลุ่มผู้ใต้บังคับบัญชา (subordinate rating approach) เป็นการให้ผู้ใต้บังคับบัญชาหรือกลุ่มสมาชิกเป็นผู้ประเมินผู้บังคับบัญชา

2. การประเมินโดยกลุ่ม (group rating approach) เป็นการประเมินบุคคลใดบุคคลหนึ่งโดยให้สมาชิกที่ประชุม หรือสมาชิกแต่ละคนประเมินอย่างเป็นอิสระแยกจากกัน แล้วจึงสรุปผลการประเมิน

3. การให้ผู้บังคับบัญชาประเมินผู้ใต้บังคับบัญชา (superior rating of subordinates) เป็นการให้ผู้บังคับบัญชาหรือกลุ่มผู้บังคับบัญชาเป็นผู้ประเมินผู้ใต้บังคับบัญชา ซึ่งโดยทั่วไปมักจะใช้ประเมินการปฏิบัติงานของผู้ใต้บังคับบัญชา เพื่อพิจารณาความคิดความชอบ



4. การประเมินตนเองโดยยึดเป้าหมายงานเป็นเกณฑ์ (self - appraisal through the use of management by objectives) เป็นวิธีการที่เปิดโอกาสให้บุคลากรสามารถประเมินผลงานของตนเองโดยยึดผลงานเป็นเกณฑ์

Crane (1985) กล่าวว่า ในทางธุรกิจ หรือในระบบอุตสาหกรรมนั้น โดยปกติ หัวหน้างาน หรือผู้นิเทศงานระดับต้น (immediate supervisor) เป็นผู้ทำหน้าที่ในการประเมินพฤติกรรม การปฏิบัติงานของเจ้าหน้าที่ที่อยู่ในความควบคุมดูแลของตน ส่วนการประเมินพฤติกรรม การปฏิบัติงานของบุคคลที่อยู่ในตำแหน่งสูงกว่า ก็อาจให้ผู้ที่บังคับบัญชา ลูกจ้าง หรือลูกค้า เป็นผู้ประเมินก็ได้ ซึ่งเรียกการประเมินในลักษณะนี้ว่า Inverted Appraisal อันเป็นแนวความคิดที่ปรากฏให้เห็นในการดำเนินงานของประเทศในยุโรปหลายประเทศ เช่น การประเมินคุณภาพของผู้สอน โดยให้นักศึกษาเป็นผู้ประเมินเป็นต้น อีกทั้งในการประเมินพฤติกรรมในการปฏิบัติงานนั้น อาจดำเนินการในลักษณะอื่น เช่น การใช้กลุ่มเพื่อนเป็นผู้ประเมิน (buddy rating or peer rating) เป็นการให้สมาชิกในกลุ่มประเมินซึ่งกันและกัน การใช้กลุ่มผู้บริหารหรือหัวหน้าเป็นผู้สรุปเชิงประเมินสภาพการปฏิบัติงานของบุคลากรที่อยู่ในความรับผิดชอบ (The Group Summary Appraisal) ซึ่งเป็นวิธีการที่เป็นประโยชน์อย่างมากต่อการวางแผนพัฒนาบุคลากรขององค์กร หรืออาจใช้วิธีการให้บุคลากรประเมินตนเอง (The Self - Appraisal Process) ซึ่งจะเป็นประโยชน์ในแง่ของการเกิดความตระหนักในผลการปฏิบัติงานของตนเองและจะนำไปสู่การปรับปรุงตนเองให้มีประสิทธิภาพมากยิ่งขึ้นในโอกาสต่อไป

ในด้านการประเมินพฤติกรรมการสอนหรือคุณภาพของครูนั้น ได้มีผู้แสดงทัศนะไว้ดังนี้

Dressel (1977) , Harris (1986) และ Lemlech (1976) มีความเห็นสอดคล้องกันว่า ผู้ที่ทำการประเมินการสอนของครูควรประกอบด้วย (1) นักเรียน (2) เพื่อนครู หรือเพื่อนร่วมงาน (3) ครูผู้สอนประเมินตนเอง และ (4) ผู้บริหารหรือผู้นิเทศ ซึ่งแนวคิดนี้สอดคล้องกับข้อเสนอของนักการศึกษาของไทยหลายท่าน อาทิ อุทุมพร จามรมาน (2530) ทองจันทร์ หงส์ลดาธรมภ์ (2524) บุญธรรม กิจปริดาภิสุทธิ (2526) ซึ่งเห็นว่า บุคคลที่ควรทำหน้าที่ประเมินการสอน หรือประเมินอาจารย์ ได้แก่ (1) ผู้เรียน / นักศึกษา (2) เพื่อนร่วมงาน / เพื่อนอาจารย์ (3) ตัวผู้สอน / อาจารย์เอง (4) ผู้เชี่ยวชาญทางด้านการสอน หรือ คณะกรรมการเฉพาะกิจ หรือ ผู้บริหาร ผู้ที่เกี่ยวข้อง รวมทั้ง ไพฑูริย์ สีนลารัตน์ (2524) ได้เสนอแหล่งที่มาของข้อมูลการประเมินที่เหมาะสมสำหรับการประเมินการสอนที่สอดคล้องกับนักการศึกษาอื่น ๆ ที่กล่าวมาแล้ว โดยให้เหตุผลประกอบสำหรับการใช้ข้อมูลจากแต่ละแหล่งดังนี้

1. อาจารย์ผู้สอนเอง เนื่องจากผู้สอนเป็นบุคคลที่สำคัญที่สุดในการสอน และการปรับปรุงการสอน ผู้สอนสามารถมองเห็น เข้าใจ และรู้จักตนเองดีพอในการประเมินตนเอง แต่ก็อาจมีบางครั้งที่ทำให้เกิดความลำเอียงได้

2. เพื่อนร่วมงาน โดยหลักการแล้ว การได้ข้อมูลจากเพื่อนร่วมงานเป็น สิ่งที่ดีและควรกระทำ เพราะเพื่อนร่วมงานเข้าใจงานสอนของเราได้ดี

3. ผู้บังคับบัญชา ข้อมูลที่ได้จากผู้บังคับบัญชามักจะเป็นการประเมินเพื่อพิจารณาความดีความชอบ ถ้าผู้บริหารสนใจ ติดตามการสอน โดยตลอด ก็จะเป็นแหล่งข้อมูลที่ดี

4. ผลจากการสอบ เมื่อการสอนดำเนินไปอย่างมีประสิทธิภาพ มีจุดมุ่งหมายชัดเจน มีระบบการประเมินที่ดีแล้ว ผลการสอบหรือสารสนเทศที่ปรากฏก็สามารถใช้อ้างอิงได้

5. ตัวผู้เรียน ซึ่งอาจให้ข้อมูลที่น่าสนใจ และเป็นประโยชน์มาก เมื่อผู้เรียนเป็นผู้รู้ และรับผลจากผู้สอน โดยตรงตลอดเวลา

ในกรณีการใช้แหล่งข้อมูลการประเมินจากนักเรียนนั้น ไพฑูรย์ ลินลารัตน์ (2524) เห็นว่า ถึงแม้การประเมินโดยผู้เรียนจะให้ข้อมูลที่ดี และน่าสนใจ แต่ก็มีคำวิพากษ์วิจารณ์อยู่หลายประการ เป็นต้นว่า ผู้เรียนยังไม่มีวุฒิภาวะเพียงพอ ผู้เรียนมักจะประเมินผู้สอนที่มีลักษณะใจดีออกมาในทางบวก ทำให้เกิดการปกป้องตนเองของผู้สอน เมื่อผลการประเมินออกมาต่ำ หรือทำให้ผู้เรียนไม่มีสัมมาคารวะต่อผู้สอน เป็นต้น อีกทั้ง Aleamoni (1983) ก็ได้เสนอแนะว่า วิธีการประเมินโดยผู้เรียน แม้จะเป็นแหล่งข้อมูลที่สำคัญของวิทยาลัย หรือมหาวิทยาลัย แต่ในการใช้ผลการประเมินก็ต้องระมัดระวังในการแปลผล ทั้งนี้เนื่องจาก ปรากฏผลการวิจัยเกี่ยวกับ การใช้การประเมินด้วยนักเรียน ได้เสนอแนะว่า นักเรียนไม่สามารถตัดสินใจได้อย่างคงเส้นคงวา เนื่องจากยังไม่บรรลุวุฒิภาวะ แบบแผนการประเมินของนักเรียนไม่ค่อยได้พิจารณาสาระที่เพียงพอ ครูที่แสดงออกในลักษณะที่ทำให้ความอบอุ่น มีความเป็นเพื่อน หรือสนุกสนานร่าเริง มักจะเป็นผู้ชนะตลอดเวลา เงื่อนไขและตัวแปรแทรกซ้อนจากภายนอกอาจส่งผลต่อการประเมินค่าของนักเรียน เช่น เพศของนักเรียนและผู้สอน ตำแหน่งของผู้สอน เป็นต้น อีกทั้งเกรดหรือคะแนนที่นักเรียนได้รับในรายวิชา ก็จะสัมพันธ์กับการประเมินค่าผู้สอน ในกรณีของการใช้กลุ่มเพื่อนร่วมงานเป็นผู้ประเมินนั้น Harris (1986) เสนอแนะว่า โดยทั่วไป จำนวนของผู้ประเมินตัดสินไม่ควรจะน้อยกว่า 3 คน

จากที่กล่าวมาจะเห็นว่า ไม่ว่าจะป็นวงการธุรกิจหรือทางการศึกษา ในการประเมินผลการปฏิบัติงานของบุคลากรในระดับปฏิบัติการนั้น กลุ่มบุคคลที่ได้รับการยอมรับว่า น่าจะเป็นผู้ประเมินที่เหมาะสมคือ ผู้บังคับบัญชา กลุ่มเพื่อนร่วมงาน การประเมินตนเอง และการใช้กลุ่มผู้ใต้บังคับบัญชา หรือกลุ่มบุคคลในความดูแลเป็นผู้ประเมิน

#### 4. งานวิจัยที่เกี่ยวข้อง

เท่าที่ผ่านมามีนักวิจัยหลายท่าน ได้พยายามศึกษาและพัฒนามาตรฐานค่าในรูปแบบต่างๆ ภายใต้สถานการณ์การศึกษาที่แตกต่างกัน เช่น การศึกษาในแง่ของความเที่ยงของมาตรฐานค่าเมื่อใช้จำนวนช่วงมาตรฐานต่างกัน การศึกษาในแง่ความคลาดเคลื่อนในการประเมินค่า

เมื่อใช้ผู้ประเมินต่างกัน การใช้แบบฟอร์มการตอบที่แตกต่างกัน เป็นต้น การศึกษาดังกล่าวแสดงให้เห็นว่า นักวิจัยได้ให้ความสำคัญกับการปรับปรุงรูปแบบของมาตรฐานค่าให้มีคุณภาพมากขึ้น จึงขอเสนอผลการศึกษานักวิจัยได้ดังต่อไปนี้

#### 4.1 งานวิจัยที่ศึกษาเกี่ยวกับจำนวนช่วงมาตรในมาตรฐานค่า

Aitken (1982) ได้ทำการศึกษาความสัมพันธ์ระหว่างจำนวนช่วงมาตรกับความแปรปรวนของมาตรฐานค่าที่ใช้ประกอบการสอนของครู 10 ข้อ ซึ่งมีลักษณะช่วงมาตรเป็นแบบกราฟฟิค 5 ช่วง ที่เป็นมาตรฐานโดยที่ 1 หมายถึง น้อยที่สุด และ 5 หมายถึง มากที่สุด ผู้วิจัยได้เปลี่ยนรูปแบบช่วงมาตรเป็น 2 ช่วง, 3 ช่วง, 4 ช่วง, 5 ช่วง, 6 ช่วง และ 7 ช่วง โดยให้นักเรียนจำนวน 624 คน ประเมินการสอนของครู ผลการวิจัยพบว่า ค่าเฉลี่ยและความแปรปรวนจะเพิ่มขึ้นเมื่อจำนวนช่วงมาตรของมาตรฐานค่ามีมากขึ้น และยังพบว่า มาตรฐานค่า 4 ช่วงมาตร และ 6 ช่วงมาตร มีค่าความแปรปรวนต่ำกว่ามาตรฐานค่า 5 ช่วงมาตร และ 7 ช่วงมาตร

Masters (1974) ได้ทำการศึกษาความสัมพันธ์ระหว่างจำนวนช่วงมาตรคู่และคี่ กับค่าความเที่ยงของแบบวัดทัศนคติแบบลิเคอร์ท พบว่าถ้าใช้จำนวนช่วงมาตรคู่ ค่าความแปรปรวนของคะแนนทั้งหมดจะมีค่าต่ำ ในขณะที่ใช้จำนวนช่วงมาตรคี่ ค่าความแปรปรวนของคะแนนทั้งหมดจะมีค่าสูงกว่า

ธารทิพย์ ประเสริฐสม (2522) ได้เปรียบเทียบค่าความเที่ยงของแบบวัดทัศนคติแบบลิเคอร์ท ที่มีจำนวนช่วงมาตรเป็น 2, 3, 4, 5, 6 และ 7 ช่วง โดยเครื่องมือที่ใช้คือ แบบวัดทัศนคติต่อวิชาชีพครู และแบบวัดทัศนคติต่อการสัมภาษณ์เพื่อคัดเลือกนักศึกษาครูคือ นักศึกษาระดับประกาศนียบัตรการศึกษาระดับชั้นปีที่ 1 วิทยาลัยครูหมู่บ้านจอมบึงจำนวน 312 คน ผลการวิจัยพบว่า ความเที่ยงของแบบวัดทัศนคติต่อวิชาชีพครู ที่มีจำนวนช่วงมาตรเป็น 5 และ 7 ช่วง จะมีค่าสูงสุดและต่ำสุดตามลำดับ โดยค่าความเที่ยงของแบบวัดที่มีจำนวนช่วงมาตร 2 ช่วง แตกต่างกับค่าความเที่ยงของแบบวัดที่มีจำนวนช่วงมาตรเป็น 3-7 ช่วงอย่างมีนัยสำคัญ และค่าความเที่ยงของแบบวัดที่มีจำนวนช่วงมาตร 3-7 ช่วง ไม่มีความแตกต่างกัน ส่วนค่าความเที่ยงของแบบวัดทัศนคติต่อการสัมภาษณ์เพื่อคัดเลือกนักศึกษาครู ที่มีจำนวนช่วงมาตร 5 ช่วงมีค่าสูงสุด การเพิ่มค่าความเที่ยงไม่ขึ้นอยู่กับจำนวนช่วงมาตร และค่าความเที่ยงของแบบวัดที่มีจำนวนช่วงมาตรเป็น 2-7 ช่วง ไม่มีความแตกต่างกันเช่นเดียวกับการศึกษาของ วิยะดา หอธรรมอนันต์ (2526) ที่ได้เปรียบเทียบความเที่ยงแบบสอบซ้ำของแบบวัดทัศนคติแบบลิเคอร์ท ที่มีจำนวนช่วงมาตรเป็น 3, 4, 5, 6 และ 7 ช่วง โดยเว้นช่วงเวลาระหว่างการสอบครั้งที่ 1 และครั้งที่ 2 เป็น 1, 2, 3 และ 4 สัปดาห์ เครื่องมือที่ใช้เป็นแบบวัดทัศนคติต่อวิชาชีพครู กลุ่มตัวอย่างเป็นนักศึกษาระดับประกาศนียบัตรการศึกษาระดับชั้นปีที่ 1 ของวิทยาลัย

ครูบ้านสมเด็จเจ้าพระยา วิทยาลัยครูธนบุรี วิทยาลัยครูสวนสุนันทา และวิทยาลัยครูจันทระเกษม จำนวน 620 คน ผลการวิจัยพบว่า ค่าความเที่ยงแบบความคงที่ภายในของแบบวัดทัศนคติแบบลิเคอร์ท ในการทดสอบครั้งที่ 1 ค่าความเที่ยงที่ได้จากแบบวัดที่มีจำนวนช่วงมาตร 3-7 ช่วง ไม่มีความแตกต่างกัน แต่ค่าความเที่ยงแบบความคงที่ภายในที่ได้จากแบบวัดที่มีจำนวนช่วงมาตร 5 และ 7 ช่วง ให้ค่าสูงกว่าความเที่ยงของแบบวัดที่มีจำนวนช่วงมาตร 3 และ 4 ช่วง ในการทดสอบครั้งที่ 2 ค่าความเที่ยงแบบความคงที่ภายในของแบบวัดที่มีจำนวนช่วงมาตร 3-7 ช่วง ไม่มีความแตกต่างกัน และความเที่ยงแบบความคงที่ภายในที่ได้จากแบบวัดที่มีจำนวนช่วงมาตร 3-7 ช่วง เป็นอิสระจากจำนวนช่วงมาตร ค่าความเที่ยงแบบสอบซ้ำของแบบวัดที่มีจำนวนช่วงมาตร 3-7 ช่วง ไม่มีความแตกต่างและค่าที่ได้จะเป็นอิสระจากจำนวนช่วง แต่ค่าความเที่ยงแบบสอบซ้ำของแบบวัดที่ใช้ช่วงเวลาในการสอบซ้ำ 2 และ 3 สัปดาห์ ให้ค่าสูงกว่าความเที่ยงแบบสอบซ้ำของแบบวัดที่ใช้ช่วงเวลาสอบซ้ำ 1 และ 4 สัปดาห์

ศิรินารด สิงหาแก้ว (2535) ได้เปรียบเทียบค่าสถิติของข้อมูลที่ได้จากมาตรประมาณค่าที่มีรูปแบบคำตอบแบบกราฟฟิคที่มีจำนวนช่วงมาตรแตกต่างกัน 5 แบบ คือ แบบ 3, 4, 5, 6 และ 7 ช่วง ในการประเมินการสำรวจนิสัยในการเรียนของกลุ่มนักเรียนชั้น ม.3 ที่สังกัดกรมสามัญศึกษา จังหวัดสระบุรี จำนวน 695 คน ผลการวิจัยพบว่า มาตรประมาณค่าทั้ง 5 แบบ มีค่าเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญ โดยค่าเฉลี่ยของแบบ 7 ช่วงมีค่าสูงสุด และแบบ 3 ช่วงมีค่าต่ำสุด ความแปรปรวนแตกต่างกันอย่างมีนัยสำคัญ โดยความแปรปรวนของแบบ 7 ช่วงมีค่าสูงสุด และแบบ 4 ช่วงมีค่าต่ำสุด และค่าความเที่ยงแบบวัดซ้ำแตกต่างกันอย่างมีนัยสำคัญ โดยแบบ 3 ช่วงมีค่าความเที่ยงสูงสุด และแบบ 6 ช่วงมีค่าความเที่ยงต่ำสุด

#### 4.2 งานวิจัยที่ศึกษาเกี่ยวกับความคลาดเคลื่อนที่เกิดจากการใช้มาตรประมาณค่า

Campbell และคณะ (1973) ได้ศึกษาเปรียบเทียบมาตรประมาณค่า 2 แบบ คือ มาตรประมาณค่าแบบรวม (summated rating technique) ของ Likert และมาตรประมาณค่าเชิงพฤติกรรม (Behavioral Anchore Scale) ของ Smith และ Kendall (1963) ซึ่งมีจำนวนช่วงมาตร 7 ช่วง โดยให้เพื่อนร่วมงานเป็นผู้ประเมินผลการปฏิบัติงานของครู ผลการวิจัยพบว่า มาตรประมาณค่าเชิงพฤติกรรมเกิดความคลาดเคลื่อนแบบปล่อยหรืออดคะแนน (leniency error) และความคลาดเคลื่อนอันเนื่องมาจากการมีปฏิสัมพันธ์มาก่อน (halo error) น้อยกว่ามาตรประมาณค่าแบบรวม สำหรับด้านความตรงพบว่า มาตรประมาณค่าเชิงพฤติกรรม มีความคมชัดในแง่ของ convergent และ discriminant validity สูงกว่ามาตรประมาณค่าแบบรวม

Showers (1974) พบว่า ในการประเมินการสอนของครูนั้น นักเรียนมักจะเกิด leniency bias โดยจะเลือกเฉพาะระดับสเกลที่มีค่าสูงเพียง 2-3 ค่าเท่านั้น ทำให้การประมาณค่าไม่ได้ผล เนื่องจากไม่สามารถจะบอกความแตกต่างของครูแต่ละคนได้ ดังนั้นจึงได้ทำการวิจัยเพื่อปรับปรุงแบบประเมินการสอนของครูโดยทดลองใช้รูปแบบคำตอบเป็น 3 แบบ คือ แบบลิเคอร์ท (fix alternative Likert cues) แบบตัวเลือกคงที่ที่ใช้คำคุณศัพท์เชิงประเมิน (fix alternative evaluation cues – superior - inferior) และแบบทางเลือกเชิงบรรยาย (descriptive cues) ซึ่งทั้ง 3 แบบมีจำนวนช่วงมาตรฐาน 4 ช่วง โดยให้นักเรียนเป็นผู้ประเมิน ผลการวิจัยพบว่าแบบที่ 2 คือ แบบตัวเลือกคงที่ที่ใช้คำคุณศัพท์เชิงประเมินเกิด leniency bias น้อยที่สุด ส่วนแบบลิเคอร์ทซึ่งเป็นแบบที่นิยมใช้กันมากมีแนวโน้มที่จะเกิดอคติในการประเมินสูงสุด อย่างไรก็ตามเมื่อเปรียบเทียบค่าความเที่ยงของผู้ประเมิน (rater reliability) ทั้ง 3 รูปแบบไม่แตกต่างกัน

Keaveny และ Megann (1975) ได้ศึกษาและเปรียบเทียบมาตรฐานประเมินค่าแบบพฤติกรรมที่คาดหวัง (behavioral expectation scales) ตามแบบของ Smith และ Kendall กับมาตรฐานประมาณค่าแบบกราฟฟิกในการประเมินพฤติกรรมการสอนของศาสตราจารย์ 4 คน โดยนักศึกษาเป็นผู้ประเมินและใช้เครื่องมือประเมินทั้ง 2 รูปแบบ ซึ่งมีจำนวนช่วงมาตรฐาน 7 ช่วง ผลการวิจัยพบว่ามาตรฐานประมาณค่าแบบพฤติกรรมที่คาดหวังเกิด halo error น้อยกว่ามาตรฐานประมาณค่าแบบกราฟฟิก และมาตรฐานประมาณค่าเชิงพฤติกรรมที่คาดหวังไม่สามารถลด leniency error ได้ ส่วนในแง่ของความตรงเชิงจำแนกพบว่า มาตรฐานประมาณค่าเชิงพฤติกรรมที่คาดหวังให้ค่าความตรงเชิงจำแนกสูงกว่ามาตรฐานประมาณค่าแบบกราฟฟิก

Slater (1977) ได้พัฒนามาตรฐานประมาณค่า 2 รูปแบบ คือ แบบที่ใช้คำคุณศัพท์ผูกติดกับระดับการประเมิน (Adjective - Anchored Scales) และแบบที่ 2 เป็นแบบที่ใช้ตัวอย่างพฤติกรรมผูกติดกับระดับการประเมินในแต่ละสเกล (Behaviorally - Anchored Scales) ซึ่งมีจำนวนช่วงมาตรฐาน 6 ช่วง โดยผู้วิจัยได้นำแบบประเมินทั้งสองรูปแบบไปใช้ในการประเมินพฤติกรรมการสอนของศาสตราจารย์จำนวน 15 คน โดยนักศึกษาเป็นผู้ประเมิน ผลการวิจัยพบว่า ไม่มีรูปแบบใดที่สามารถลดอคติได้สำเร็จ สำหรับค่าความเที่ยงของคะแนนเฉลี่ยในการประมาณค่าของแต่ละรูปแบบพบว่า แบบที่มีสเกลเป็นคำคุณศัพท์ให้ค่าความเที่ยงสูงกว่าเล็กน้อย

Cook (1988) ได้ศึกษาผลของการใช้มาตรฐานประมาณค่าที่มีรูปแบบแตกต่างกัน 2 รูปแบบคือ มาตรฐานประมาณค่าเชิงพฤติกรรม (Behaviorally-Anchored Rating Scales) กับมาตรฐานประมาณค่าแบบลิเคอร์ท (Likert Scales) ซึ่งมีจำนวนช่วงมาตรฐาน 6 ช่วง และผลของการฝึกทักษะแก่ผู้ประเมินต่อค่าความคลาดเคลื่อนแบบปล่อยคะแนน (leniency error) และความคลาดเคลื่อนอันเนื่องมาจากการเคยมีปฏิสัมพันธ์มาก่อน (halo error) โดยผู้วิจัยได้ใช้มาตรวัด 2 รูปแบบดังกล่าวให้นักศึกษา 2 กลุ่ม

คือ กลุ่มที่ได้รับการฝึกทักษะ กับกลุ่มที่ไม่ได้รับการฝึกทักษะ ทำการประเมินพฤติกรรมการสอนของศาสตราจารย์ 1 ท่าน แล้วทำการวิเคราะห์หาค่าเฉลี่ยเลขคณิต และส่วนเบี่ยงเบนมาตรฐานของคะแนนผลการประเมินแต่ละมิติ เปรียบเทียบระหว่างกลุ่มที่ได้รับการฝึกทักษะกับกลุ่มที่ไม่ได้รับการฝึกทักษะ พร้อมทั้งเปรียบเทียบคะแนนผลการประเมิน ระหว่างมาตรวัดต่างรูปแบบกัน เฉพาะในกลุ่มที่ได้รับการฝึกทักษะ โดยในการตรวจสอบ halo error ผู้วิจัยได้ใช้ส่วนเบี่ยงเบนมาตรฐานของคะแนนผลการประเมิน (standard deviation) เป็นคะแนนบ่งชี้ความคลาดเคลื่อน (halo score) ภายใต้งื่อนไข่ว่า ถ้าคะแนนความคลาดเคลื่อน (halo score) มีค่ามาก แสดงถึงการเกิด halo error น้อยกว่ากรณีที่คะแนนความคลาดเคลื่อนมีค่าน้อย ผลการวิจัยพบว่า การฝึกทักษะมีผลต่อการลด halo error การประเมินโดยมาตรฐานค่าเชิงพฤติกรรมเกิด leniency error น้อยกว่า มาตรฐานค่าแบบลิเคอร์ท ในขณะที่มาตรฐานค่าแบบลิเคอร์ทเกิด halo error น้อยกว่า

สุพักตร์ พิบูลย์ (2534) ได้เปรียบเทียบคุณภาพของมาตรฐานค่าเชิงพฤติกรรมกับมาตรฐานค่าแบบกราฟฟิก ซึ่งมีจำนวนช่วงมาตร 7 ช่วง ในการประเมินพฤติกรรมด้านการประเมินผลของครู โดยให้นักเรียนเป็นผู้ประเมิน ผลการวิจัยพบว่า มาตรฐานค่าเชิงพฤติกรรมปราศจากลักษณะความคลาดเคลื่อนแบบปล่อยหรือกดคะแนน และความคลาดเคลื่อนอันเนื่องมาจากปฏิสัมพันธ์น้อยกว่า การใช้มาตรฐานค่าแบบกราฟฟิกเป็นเครื่องมือประเมิน

สุภาวดี ตั้งบุปผา (2535) ได้ศึกษาเปรียบเทียบความคลาดเคลื่อนของการประเมินค่าที่ใช้รูปแบบมาตรฐานค่าซึ่งมีจำนวนช่วงมาตร 5 ช่วง และผู้ประเมินต่างกัน ผลการวิจัยพบว่ากรณีที่ทำให้นิสิตเป็นผู้ประเมินคุณลักษณะอาจารย์ที่ปรึกษาพบว่ามาตรฐานค่าแบบตัวเลขและแบบบรรยาย มีความคลาดเคลื่อนแบบปล่อยหรือกดคะแนนและความคลาดเคลื่อนแบบขัดแย้ง ในขณะที่มาตรฐานค่าแบบกราฟฟิกมีความคลาดเคลื่อนเกิดขึ้นทั้ง 3 ชนิด คือ ความคลาดเคลื่อนแบบปล่อยหรือกดคะแนน ความคลาดเคลื่อนจากปฏิสัมพันธ์ และความคลาดเคลื่อนแบบขัดแย้ง ส่วนกรณีที่ผู้บริหารเป็นผู้ประเมินพบว่า มาตรฐานค่าทั้ง 3 รูปแบบมีความคลาดเคลื่อนแบบปล่อยหรือกดคะแนนเท่านั้น นอกจากนี้กรณีที่นิสิตเป็นผู้ประเมินจะพบความคลาดเคลื่อนแบบปล่อยหรือกดคะแนนสูงกว่ากรณีที่ผู้บริหารเป็นผู้ประเมิน

#### 4.3 งานวิจัยที่ศึกษาเกี่ยวกับการประเมินการสอน

สัมพันธ์ พันธุ์พุกภัย (2537) ได้ศึกษาความตรงและความเที่ยงของการประเมินอาจารย์มหาวิทยาลัยโดยเพื่อนร่วมงาน โดยใช้มาตรฐานค่าที่มีจำนวนช่วงมาตร 5 ช่วง พบว่า โดยส่วนรวมการประเมินอาจารย์โดยเพื่อนร่วมงานมีความตรงเชิงคู่เข้า ความตรงเชิงจำแนก และความตรงร่วมสมัยต่ำ ในขณะที่การประเมินอาจารย์โดยเพื่อนร่วมงานมีความเที่ยงค่อนข้างสูง และเมื่อ

พิจารณาโดยจำแนกตามภารกิจ พบว่าในการประเมินภารกิจด้านการวิจัย ด้านการสอน และด้านการบริการทางวิชาการแก่สังคมมีความเที่ยงค่อนข้างสูง ปานกลาง และปานกลางตามลำดับ

Marsh (1984) ได้ศึกษาความตรงตามทฤษฎีของการประเมินคุณภาพการสอนของอาจารย์มหาวิทยาลัย 20 คนในวิชาต่าง ๆ จำนวน 329 วิชา โดยให้กลุ่มเพื่อนอาจารย์มหาวิทยาลัย และนักศึกษาเป็นผู้ประเมิน โดยใช้เมตริกซ์ลักษณะหลากหลาย-วิธีหลาย และใช้มาตรประมาณค่าที่มีจำนวนช่วงมาตร 7 ช่วง ทำการวัดลักษณะ 9 ลักษณะที่ ประกอบด้วยข้อคำถาม 35 ข้อ ด้วยวิธีการวัด 2 วิธีคือการให้กลุ่มเพื่อนอาจารย์เป็นผู้ประเมิน และการให้นักศึกษาเป็นผู้ประเมิน ในการวิเคราะห์ความสัมพันธ์ตามเมตริกซ์ลักษณะหลากหลาย-วิธีหลายได้เมตริกซ์แบบ  $18 \times 18$  พบว่า ค่าสัมประสิทธิ์ความตรงเชิงคู่เข้า (convergent validity coefficient) แต่ละตัว มีลักษณะดังนี้

1. แตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติ
2. เกือบทั้งหมด (143 จาก 144 ค่า) มีค่ามากกว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างลักษณะที่วัดต่างกัน และวิธีการวัดต่างกันในแต่ละลักษณะใด ๆ เป็นอิสระกับลักษณะอื่น ๆ
3. ส่วนมาก (68 จาก 72 ค่า) มีค่ามากกว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการวัดลักษณะต่างกันด้วย โดยวิธีให้กลุ่มเพื่อนอาจารย์เป็นผู้ประเมิน แสดงว่าไม่มีความลำเอียงหรือมีน้อยมากในการประเมินโดยกลุ่มเพื่อนอาจารย์
4. มีหลายตัว (30 จาก 72 ค่า) ที่มีค่าไม่มากกว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการวัดลักษณะต่างกัน โดยให้นักศึกษาเป็นผู้ประเมิน แสดงว่ามีความเป็นไปได้ในการเกิดความลำเอียงบางอย่าง (method/halo effect) ในการประเมินโดยนักศึกษา
5. แบบแผนของสหสัมพันธ์ ระหว่างลักษณะที่วัดต่างกันเหมือนกันทั้งในการประเมินโดยกลุ่มเพื่อนอาจารย์ และการประเมินโดยนักศึกษา แสดงว่ามีความแปรปรวนร่วมในลักษณะจริง (true trait covariation) ที่เป็นอิสระกับวิธีการวัด

จากการพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ดังกล่าวสรุปได้ว่า การประเมินโดยนักศึกษามีความลำเอียง มากกว่าการประเมินโดยกลุ่มเพื่อนอาจารย์

Donald (1990) ศึกษากระบวนการตรวจสอบความถูกต้อง และเกณฑ์ที่ใช้ในการประเมินอาจารย์สาขาวิชาต่าง ๆ โดยการสัมภาษณ์แบบมีโครงสร้างกับศาสตราจารย์มหาวิทยาลัย 36 คน ที่มีความเชี่ยวชาญทั้งด้านการสอนและการวิจัยมากพอที่จะเป็นโมเดลของสาขาวิชาที่ปฏิบัติงาน และใช้มาตรประมาณค่าที่มีจำนวนช่วงมาตร 7 ช่วงเป็นเครื่องมือในการวิจัย ได้ข้อสรุปว่า การตรวจสอบความถูกต้อง ต้องยึดกรอบแนวคิดสาขาวิชาธรรมชาติและสังคมศาสตร์และการประเมิน โดยเพื่อนร่วมงานมีความสำคัญต่อการประเมินในสาขามนุษยศาสตร์ ทั้งนี้ศาสตราจารย์ในสาขาวิชาบริสุทธิ์ใช้หลักฐานที่ขัดแย้งกันมาทำการตรวจสอบความถูกต้องในงานที่ปฏิบัติมากกว่า ในสาขาวิชาประยุกต์ (applied fields)

พยุงศักดิ์ สนั่นเทศ (2527) ได้ศึกษาความคิดเห็นของครูพลศึกษา และนักเรียนเกี่ยวกับพฤติกรรมการสอนพลศึกษาในโรงเรียนมัธยมศึกษาตอนปลายในเขตกรุงเทพมหานคร ผู้วิจัยได้ใช้มาตรประมาณค่าซึ่งมีจำนวนช่วงมาตรฐาน 5 ช่วง เป็นเครื่องมือให้นักเรียนและครูพลศึกษาประเมินพฤติกรรมการสอนพลศึกษา 7 ด้าน รวบรวมรายการที่ให้ประเมินทั้งสิ้น 62 ข้อ ผลการประเมินพบว่า 55 รายการหรือ ประมาณร้อยละ 89 ของรายการที่ให้ประเมินทั้งหมด ได้รับการประเมินว่าเป็นพฤติกรรมที่ได้ปฏิบัติในระดับบ่อยครั้งหรือเป็นประจำ (ระดับ 4-5คะแนน) ส่วนที่เหลือได้รับการประเมินว่าเป็นพฤติกรรมที่ได้ปฏิบัติบางครั้งหรือปานกลาง ผลการประเมินโดยครูมีแนวโน้มสูงกว่าผลการประเมินโดยนักเรียน ในเกือบทุกรายการที่ประเมิน

ขวัญชัย อึ้งรักษา (2529) ได้ศึกษาความคิดเห็นของผู้บริหารโรงเรียน ครู และครูพลศึกษาเกี่ยวกับสมรรถภาพของครูพลศึกษา ในโรงเรียนระดับประถมศึกษาสังกัดกรุงเทพมหานคร โดยผู้วิจัยได้ใช้มาตรประมาณค่าจำนวน 4 ช่วงมาตรฐาน คือ 4 - มากที่สุด 3- มาก 2- น้อย และ 1- น้อยที่สุด เป็นเครื่องมือในการเก็บรวบรวมข้อมูล ด้วยการให้ผู้บริหาร ครู และครูพลศึกษา ประเมินสมรรถภาพตามที่ปรากฏจริงของครูพลศึกษา 8 ด้าน คือ ด้านหลักสูตรและวิชาการ ด้านการสอน ด้านการประเมินผล ด้านความรับผิดชอบต่อหน้าที่ ด้านมนุษยสัมพันธ์ ด้านบุคลิกภาพ ด้านสุขภาพ และด้านคุณลักษณะและเจตคติ รวบรวมรายการสมรรถภาพที่ให้ประเมินทั้งสิ้น 86 รายการ ผลการศึกษาพบว่า ร้อยละ 100 ของรายการที่ให้ประเมินได้รับการประเมินว่ามีสมรรถภาพในระดับมาก และผลการประเมินโดยครูพลศึกษา ปรากฏแนวโน้มของค่าเฉลี่ยสูงกว่าผลการประเมินโดยกลุ่มอื่น ๆ

พวงพยอม ศรีหาบัติ (2529) ได้ศึกษาเรื่อง “การปฏิบัติงานนิเทศทางวิชาการของหัวหน้าหมวดวิชาภาษาไทย ตามการรับรู้ของตนเองและครูผู้สอนภาษาไทยโรงเรียนมัธยมศึกษาในภาคตะวันออกเฉียงเหนือ” โดยใช้เทคนิคมาตรประมาณค่าจำนวน 4 ช่วงมาตรฐาน คือ 4 - มากที่สุด 3 - มาก 2 - น้อย และ 1 - น้อยที่สุด ให้หัวหน้าหมวดวิชาภาษาไทยและครูภาษาไทยประเมินพฤติกรรมการปฏิบัติงานนิเทศของหัวหน้าหมวดวิชาภาษาไทย รวม 7 ด้าน ประกอบด้วยรายการที่ให้ประเมินทั้งสิ้น 71 รายการ ผลการประเมินพบว่า ประมาณร้อยละ 75 ของรายการที่ให้ประเมิน ได้รับการประเมินว่าได้ปฏิบัติในระดับมาก และผลการประเมินของหัวหน้าหมวดวิชาเอง มีแนวโน้มสูงกว่าผลการประเมินโดยครูภาษาไทย

จากการศึกษางานวิจัยที่กล่าวข้างต้น จะเห็นได้ว่า การจะใช้มาตรประมาณค่าเพื่อการประเมินคุณภาพการสอน หรือการปฏิบัติงานได้อย่างมีคุณภาพนั้น ต้องคำนึงถึงองค์ประกอบหลัก 2 ประการคือ รูปแบบของมาตรประมาณค่า และ สถานภาพของผู้ประเมิน ในส่วนของรูปแบบมาตรประมาณค่านั้น พบว่า ถ้าใช้จำนวนช่วงมาตรฐานในมาตรประมาณค่า ต่างกัน ผลของการ



ประเมินจะเกิดความคลาดเคลื่อนต่างกัน เช่นเดียวกับการใช้สถานภาพของผู้ประเมินต่างกัน ความเที่ยงตรงของการประเมินจะเกิดขึ้นแตกต่างกันด้วย องค์ประกอบหลัก 2 องค์ประกอบที่กล่าวข้างต้น ส่วนใหญ่นักวิจัยจะศึกษาแยกกัน ไม่ได้ศึกษาทั้ง 2 องค์ประกอบพร้อมกัน ดังนั้นผู้วิจัยจึงสนใจที่จะศึกษาว่า ถ้าใช้จำนวนช่วงมาตรในมาตรประมาณค่า และสถานภาพของผู้ประเมินต่างกัน จะเกิดความคลาดเคลื่อนในการประเมินค่า ได้แก่ ความคลาดเคลื่อนแบบปล่อยคะแนน ความคลาดเคลื่อนแบบฮาโล และความคลาดเคลื่อนแบบขัดแย้ง แตกต่างกันหรือไม่ ซึ่งข้อค้นพบที่ได้จะเป็นประโยชน์อย่างมากต่อนักวิจัยในการเลือกจำนวนช่วงมาตรในมาตรประมาณค่า และสถานภาพของผู้ประเมิน ที่เหมาะสมในการประเมินคุณภาพการสอน หรือการปฏิบัติงาน อันจะส่งผลให้การประเมินนั้นมีคุณภาพเชื่อถือได้