



## ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันหน่วยงานหรือองค์กรต่างๆ ได้พยายามสรรหาและคัดเลือกบุคคลที่มีคุณภาพเข้ามาทำงานในหน่วยงานหรือองค์กรของตน และหลังจากที่ได้คัดสรรและบุคลากรได้ทำงานในระยะหนึ่ง ถ้าบุคคลทำงานได้อย่างมีคุณภาพ เป็นที่น่าพึงพอใจแก่หน่วยงาน บุคลากรคนนั้นจะได้รับการพิจารณาเพื่อได้รับผลตอบแทนในรูปแบบต่างๆ ไม่ว่าจะเป็นการขึ้นเงินเดือน การเลื่อนตำแหน่ง หรือการตอบแทนในรูปแบบอื่น เพื่อเป็นกำลังใจให้กับบุคลากรในการทำงานครั้งต่อไป แต่เนื่องด้วยข้อจำกัดทางด้านงบประมาณ บุคลากรที่จะได้รับผลตอบแทนดังกล่าวจะมีจำนวนจำกัด ทั้งนี้ขึ้นอยู่กับ การตัดสินใจของกลุ่มผู้บริหาร ดังนั้นกลุ่มผู้บริหารจะต้องนำกระบวนการที่เป็นระบบและเชื่อถือได้ มาใช้ เพื่อให้ได้ข้อมูลหรือสารสนเทศประกอบการตัดสินใจ

กิจกรรมการประเมิน (evaluation activities) จึงเป็นกิจกรรมหนึ่งที่กลุ่มผู้บริหารได้นำมาใช้ประกอบการตัดสินใจในเรื่องดังกล่าว เนื่องจากกิจกรรมการประเมินเป็นกระบวนการที่เป็นระบบ ดำเนินการตามขั้นตอน โดยเฉพาะอย่างยิ่งมีการเก็บรวบรวมข้อมูล และใช้ข้อมูลที่ได้ประกอบการตัดสินใจ (Cronbach, 1983) แต่ทั้งนี้ กลุ่มผู้บริหารควรตระหนักถึงความถูกต้องและความน่าเชื่อถือของกิจกรรมการประเมินด้วย ซึ่งส่วนหนึ่งขึ้นอยู่กับความเที่ยงตรงของแหล่งสารสนเทศที่ใช้ในการประเมิน ความถูกต้องเหมาะสมของเทคนิควิธีและเครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล (สุพัตร์พิบูลย์, 2531)

ในวงการศึกษากิจกรรมการประเมิน ถือได้ว่าเป็น กิจกรรมสำคัญกิจกรรมหนึ่งที่น่ามาใช้ทั้งในด้านการบริหารและการเรียนการสอน โดยเฉพาะการประเมินการสอนของครู เป็นกิจกรรมที่กระทำอย่างต่อเนื่องและสม่ำเสมอ แหล่งสารสนเทศที่ใช้ในการประเมินการสอนของครูได้มาจากหลายแหล่ง เช่น การประเมินโดยกลุ่มผู้บริหาร (superior evaluation) การประเมินโดยกลุ่มเพื่อนร่วมงาน (colleague evaluation) การประเมินโดยกลุ่มนักเรียน (student evaluation) การประเมินตนเอง (self evaluation) เป็นต้น (Batista, 1976) และในการตัดสินใจว่าจะเลือกใช้แหล่งสารสนเทศใดนั้น จะต้องตระหนักให้มาก เพราะแหล่งสารสนเทศแต่ละแหล่งจะมีข้อดีข้อเสียต่างกันไป ควรเลือกใช้แหล่งสารสนเทศให้เหมาะกับจุดประสงค์ของการประเมิน เช่น ถ้าจะประเมินเพียงเพื่อปรับปรุงการสอนของอาจารย์ เราสามารถใช้การประเมินโดยกลุ่มนักเรียนและตนเองก็เพียงพอ แต่ถ้าจะ

ประเมินเพื่อการเลื่อนตำแหน่ง หรือขึ้นเงินเดือน ควรใช้การประเมินโดยกลุ่มผู้บริหาร และกลุ่มเพื่อนร่วมงาน เป็นต้น (Centra, 1972) ซึ่งถ้าเลือกใช้แหล่งสารสนเทศไม่ถูกต้อง สารสนเทศที่ได้จากกิจกรรมการประเมินจะขาดความน่าเชื่อถือไปอย่างมาก ดังที่สภาคณาจารย์ มหาวิทยาลัยมหิดล (2527) ได้ทำการประเมินผลการปฏิบัติงานทางวิชาการของอาจารย์ พบว่า การให้กลุ่มผู้บริหารประเมินผลการปฏิบัติงานของอาจารย์เท่านั้น อาจจะก่อให้เกิดความมีอคติหรือไม่ยุติธรรมในการพิจารณาผลงานทางวิชาการได้ จากการศึกษาของ Marsh และคณะ (1988) พบว่า การประเมินการสอนของอาจารย์โดยนักศึกษาเป็นการวัดคุณภาพการสอนได้อย่างมีความแม่นยำในระดับหนึ่ง แต่ในขณะที่เดียวกันเมื่อให้กลุ่มเพื่อนร่วมงานของอาจารย์ทำการประเมินคุณภาพการสอนด้วย จะพบว่า สารสนเทศที่ได้จากการประเมินมีความน่าเชื่อถือมากขึ้นและเป็นที่ยอมรับโดยทั่วไปทั้งนี้อาจเป็นเพราะกลุ่มเพื่อนร่วมงานเป็นผู้ที่เข้าใจงานของครูอย่างดีที่สุด และรู้จุดมุ่งหมายของการเรียนการสอน ดังที่ สัมพันธ์ พันธุ์พุกภัย (2537) พบว่า การประเมินการสอนของอาจารย์โดยเพื่อนร่วมงาน มีความเที่ยงค่อนข้างสูง โดยเฉพาะการประเมินการเรียนการสอนเพื่อนำสารสนเทศที่ได้จากการประเมินมาปรับปรุงกระบวนการเรียนการสอนนั้นควรให้ทั้งนักเรียนและกลุ่มเพื่อนร่วมงานเป็นผู้ประเมินควบคู่กันไป

นอกจากนี้ ในส่วนของการใช้เทคนิควิธีและเครื่องมือในการเก็บรวบรวมข้อมูล ปัจจุบันพบว่า นักการศึกษา นักวัดผล นักประเมิน หรือนักวิจัย ได้พยายามพัฒนาเทคนิควิธีในการเก็บรวบรวมข้อมูลที่ต้องการและเหมาะสมไว้หลายรูปแบบและสามารถทำให้สารสนเทศที่ได้จากกิจกรรมการประเมินมีความถูกต้องและน่าเชื่อถือมากขึ้น Mathis (1979) ได้จำแนกประเภทของเครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลเพื่อการประเมินไว้ 4 ประเภท ได้แก่

1. เครื่องมือประมาณค่าประเภทกำหนดรายการประเมินไว้เป็นรายการๆ (Category Rating Inventory) เครื่องมือในกลุ่มนี้ได้แก่ มาตรฐานค่าแบบกราฟฟิก (The Graphic Rating Scales) แบบตรวจสอบรายการ (The Checklist) และวิธีการกำหนดทางเลือกจำกัด (Forced Choice Methods)
2. วิธีการเชิงเปรียบเทียบ (Comparison Methods) เทคนิคที่ใช้ในกลุ่มนี้ได้แก่ การเรียงอันดับ (Ranking) การเปรียบเทียบ (Paired Comparison) เป็นต้น
3. วิธีการเชิงบันทึก (Written Methods) เทคนิคที่ใช้ในกลุ่มนี้ได้แก่ การบันทึกเหตุการณ์สำคัญๆ ที่เกิดขึ้น (The Critical Incident Methods) เป็นต้น
4. การใช้มาตรฐานค่าเชิงพฤติกรรม (Behaviorally-anchored Rating Scales)

แต่อย่างไรก็ตาม ในทางปฏิบัติ เนื่องจากข้อจำกัดทางด้านทรัพยากรสำหรับดำเนินกิจกรรมการประเมินประกอบกับสภาพการณ์ที่ผู้บริหารหรือผู้ทำหน้าที่ตัดสินใจต้องการข้อมูลหรือสารสนเทศอย่างรวดเร็ว ทำให้เทคนิควิธีที่ซับซ้อน ต้องใช้เวลาและการลงทุนมาก ไม่ได้รับความสนใจนัก ดังนั้นเทคนิควิธีที่ดำเนินการง่าย สร้างเครื่องมือได้ง่าย และได้ข้อมูลสารสนเทศรวดเร็ว

เช่น การใช้มาตรประมาณค่า (Rating Scales) จึงได้รับความสนใจและนำมาใช้อย่างแพร่หลายตั้งแต่อดีตจนถึงปัจจุบัน

ในวงการศึกษ ปัจจุบันพบว่า มาตรประมาณค่าส่วนใหญ่ได้ถูกนำมาใช้เป็นเครื่องมือประเมินคุณภาพการสอนและการปฏิบัติงานของครูอย่างแพร่หลาย ทั้งนี้อาจเป็นเพราะมาตรประมาณค่า เป็นเครื่องมือที่สร้างและใช้ง่าย ตลอดจนใช้วัดพฤติกรรมได้หลายลักษณะ ดังที่ มัลลิกา ยูวะนะเตมีย์ (2527) ที่ได้ทำการสำรวจเครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลในการทำวิทยานิพนธ์ของนิสิตบัณฑิตวิทยาลัย คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ระหว่างปีการศึกษา 2521-2525 โดยวิธีสุ่มอย่างง่ายจากบัตรรายการจำนวน 50 ฉบับ พบว่า มีการใช้แบบสอบถามประเภทมาตรประมาณค่าจำนวน 34 ฉบับ คิดเป็นร้อยละ 68 ซึ่งสอดคล้องกับการศึกษาของ สุปัทร์ พิบูลย์ (2534) ทำการศึกษางานวิจัยที่เน้นการศึกษาเชิงประยุกต์แนวความคิดในการประเมินความจำเป็นทางการศึกษาที่ได้ดำเนินการระหว่างปี พ.ศ. 2525-2531 ประมาณ 20 เรื่อง พบว่า ทุกเรื่องใช้แบบประเมินประเภทมาตรประมาณค่า เป็นเครื่องมือเก็บรวบรวมข้อมูล

รูปแบบของมาตรประมาณค่ามีหลายลักษณะ เช่น มาตรประมาณค่าแบบตัวเลข (Numerical Rating Scales) มาตรประมาณค่าแบบกราฟิก (Graphic Rating Scales) มาตรประมาณค่าแบบบรรยาย (Descriptive Rating Scales) เป็นต้น ซึ่งโดยทั่วไปจะมีส่วนประกอบที่สำคัญ 2 ส่วน คือ ส่วนที่เป็นชุดข้อความแสดงพฤติกรรมหรือคุณลักษณะ (attribute) หรือเรื่องที่จะให้ผู้ประเมินพิจารณาตามความคิดเห็นหรือความรู้สึก และส่วนที่เป็นตัวเลขแสดงระดับความเข้มของพฤติกรรมหรือคุณลักษณะที่ปรากฏ หรือ จำนวนช่วงมาตรในมาตรประมาณค่า (the number of rating intervals) (Gronlund, 1981) เพื่อให้ผู้ประเมินเลือกระดับความเข้มพฤติกรรม ซึ่งมีทั้งด้านบวกและด้านลบในข้อเดียวกันก็ได้ หรือมีเฉพาะด้านบวกหรือด้านลบ โดยอีกด้านหนึ่งจะเป็นศูนย์หรือระดับน้อยมาก และข้อความที่สร้างในรูปมาตรประมาณคานี้จะอิงข้อตกลงเบื้องต้นว่า ลักษณะที่ต้องการจะวัดหรือประเมินนั้น ต้องมีความต่อเนื่องกัน จึงจะสามารถทำเป็นมาตรประมาณค่าได้ อุทุมพร จามรมาน (2530) กล่าวว่า มาตรประมาณค่าสามารถนำมาใช้เป็นเครื่องมือประเมินการปฏิบัติงานของครูได้ง่าย และสะดวก เพียงแต่ผู้ประเมินหรือผู้สังเกตพฤติกรรมให้ระดับคุณค่าเป็นตัวเลขนบนสเกลที่กำหนดให้เป็นลำดับขั้น แต่ทั้งนี้ผู้ใช้ต้องสามารถเปรียบเทียบความมากน้อยหรือความแตกต่างของพฤติกรรมได้

จำนวนช่วงมาตรในมาตรประมาณค่ามีใช้กันตั้งแต่ 2 ช่วงขึ้นไป ดังเช่นใน California Psychological Inventory ใช้จำนวน 2 ช่วงคือ ถูกและผิด ส่วนในมาตรประมาณค่าวัดทัศนคติของ Likert (Likert, 1932) ใช้จำนวน 5 ช่วง ในขณะที่มาตรวัดของ California F-scale หรือ

Semantic Differential ใช้จำนวน 7 ช่วง เป็นต้น ในการจะเลือกใช้จำนวนช่วงมาตรฐานในมาตรฐานค่าเท่าใดนั้น Komorita และ Graham (1965) แนะนำว่า ผู้สร้างมาตรฐานค่าต้องการให้ผู้ประเมินใช้ความสามารถมากน้อยเพียงใด ถ้าต้องการให้ผู้ประเมินใช้ความสามารถมาก ผู้สร้างควรใช้มาตรฐานค่าหลายช่วง แต่ถ้าต้องการให้ผู้ประเมินใช้ความสามารถน้อย ผู้สร้างควรใช้มาตรฐานค่า 2 - 3 ช่วง มีงานวิจัยของ Symond (1924) พบว่า จำนวนช่วงที่เหมาะสมที่ทำให้ความเที่ยงของมาตรฐานค่าสูงสุด คือ 7 ช่วง แต่ในขณะที่การศึกษาของ Bending (1953) และ Guilford (1954) พบว่า จำนวนช่วง 7 ช่วง ไม่ได้เหมาะสมเสมอไป ขึ้นอยู่กับว่าผู้ประเมินมีความรู้ในเรื่องที่จะประเมินมากน้อยเพียงใด และยังพบว่า จำนวนช่วง 3, 5, 7 และ 9 ช่วง จะมีความเที่ยงไม่แตกต่างกัน และความเที่ยงจะลดลงเมื่อใช้มาตรฐานค่า 11 ช่วง มาตรฐานค่าโดยมากกำหนดจำนวนช่วงคำตอบไว้ไม่ต่ำกว่า 3 ช่วง และไม่เกิน 11 ช่วง ส่วนใหญ่จะใช้กันเพียง 5 ช่วง (สุภาพ วาดเขียน, 2524) การแบ่งช่วงยิ่งละเอียดมาก ยิ่งแสดงความแตกต่างไปมาก แต่ไม่ควรมากเกินไปจนแยกความแตกต่างไม่ได้ นอกจากนี้ได้มีการศึกษาที่ให้ผลตรงกันว่า จำนวนช่วงคำตอบที่เหมาะสมและไม่ทำให้ความเที่ยงเปลี่ยนแปลงไปมากนัก คือ จำนวน 3 - 7 ช่วง (Gronlund, 1981) และจากการศึกษาผลของจำนวนช่วงมาตรฐานค่าที่มีต่อค่าความเที่ยงระหว่างผู้ประเมิน (interrater reliability) ของ Cicchetti (1985) พบว่า ความเที่ยงระหว่างผู้ประเมินจะเพิ่มขึ้นอย่างต่อเนื่องจนกระทั่งจำนวนช่วงมากกว่า 7 ช่วง ค่าความเที่ยงระหว่างผู้ประเมินจะเพิ่มขึ้นไม่มากนักและมีแนวโน้มจะลดลงเรื่อยๆ ในขณะที่งานวิจัยของ Cronbach (1946) ได้พบข้อสังเกตว่า ถ้าเพิ่มจำนวนช่วงมากขึ้น อาจทำให้ผู้ประเมินตอบมาตรฐานค่าด้วยความเอนเอียงไปทางใดทางหนึ่งอย่างมาก นั่นคือ ก่อให้เกิดความคลาดเคลื่อนในการประเมินค่าได้ และหนทางที่อาจจะลดความคลาดเคลื่อนได้ ต้องใช้จำนวนช่วงมาตรฐาน 2 ช่วง

นอกจากนี้ มีนักวิจัยหลายท่านได้พยายามศึกษาถึงแหล่งที่จะก่อให้เกิดความคลาดเคลื่อน เมื่อใช้มาตรฐานค่าเพื่อเป็นเครื่องมือในการประเมินการปฏิบัติงานหรือประเมินการเรียนการสอน ซึ่ง Selltiz (1959) ได้แบ่งแหล่งที่ก่อให้เกิดความคลาดเคลื่อนไว้ 3 แหล่ง คือ ความบกพร่องที่เกิดจากกระบวนการประเมิน ความบกพร่องของเครื่องมือที่ใช้ในการประเมิน และความบกพร่องที่เกิดจากตัวผู้ประเมิน โดยเฉพาะความบกพร่องที่เกิดจากตัวผู้ประเมินนั้น เป็นแหล่งที่ก่อให้เกิดความคลาดเคลื่อนหลายประเภท ดังที่ Guilford (1954) ได้แบ่งประเภทของความคลาดเคลื่อนที่เกิดจากการใช้มาตรฐานค่าไว้ดังนี้

1. ความคลาดเคลื่อนแบบปล่อยคะแนน (leniency error) เกิดจากความพยายามของผู้ประเมินที่จะให้คุณค่าของผู้ประเมินเกินเกินความเป็นจริง หรือในทางตรงกันข้าม ผู้ถูกประเมินได้รับการประเมินคุณค่าต่ำจนเกินความเป็นจริงอย่างสม่ำเสมอ ไม่ว่าจะด้วยเหตุผลใดก็ตาม

2. ความคลาดเคลื่อนแบบมีแนวโน้มเข้าสู่ส่วนกลาง (central tendency error) เกิดขึ้นเมื่อผู้ประเมินพยายามหลีกเลี่ยงที่จะประเมินคุณค่าให้มากไปทางใดทางหนึ่งจนเกินไป ประกอบกับว่าผู้ประเมินอาจจะไม่รู้จักรู้จักผู้ถูกประเมินดีพอ จึงมีผลให้ผู้ประเมินพยายามประเมินให้คุณค่าในระดับปานกลางอย่างสม่ำเสมอ

3. ความคลาดเคลื่อนแบบฮาโล (halo error) เป็นความคลาดเคลื่อนที่เกิดขึ้นเมื่อผู้ถูกประเมินมีอิทธิพลต่อผู้ประเมิน คือ ผู้ประเมินจะให้คุณค่ามากหรือน้อยขึ้นอยู่กับความพอใจของผู้ประเมินต่อผู้ถูกประเมิน ซึ่งความคลาดเคลื่อนแบบนี้จะเกิดเฉพาะผู้ถูกประเมินบางคนเท่านั้น

4. ความคลาดเคลื่อนแบบใช้ตรรกะผิด (logical error) เกิดขึ้นเมื่อผู้ประเมินนำตัวแปรอื่นที่ตนเองคิดว่ามีความสัมพันธ์กับตัวแปรที่ประเมินมามีส่วนในการประเมินด้วย ความคลาดเคลื่อนประเภทนี้มีได้เกิดจากความลำเอียงแต่เกิดจากการที่ผู้ประเมินได้รับรู้บริบทเรื่องราวของผู้ถูกประเมินมาก่อน

5. ความคลาดเคลื่อนอันเนื่องมาจากความใกล้เคียงกัน (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

จะเห็นได้ว่า ความคลาดเคลื่อนที่ Guilford (1954) ได้แบ่งไว้ข้างต้นนั้น มีความคลาดเคลื่อนอยู่ 3 ประเภทที่สามารถบอกทิศทางได้ว่า ความคลาดเคลื่อนที่เกิดขึ้นเป็นไปในลักษณะให้คะแนนสูงหรือต่ำกว่าความเป็นจริง ได้แก่ ความคลาดเคลื่อนแบบปล่อยคะแนน ความคลาดเคลื่อนแบบฮาโล และความคลาดเคลื่อนแบบขัดแย้ง ซึ่งความคลาดเคลื่อนทั้ง 3 ประเภทนี้ สามารถนำมาวิเคราะห์และประมาณค่าได้ตามโมเดลของ Guilford (1954) ด้วยการวิเคราะห์ความแปรปรวนสองทาง โดยมีข้อตกลงเบื้องต้นว่า คะแนนที่ได้จากการประเมินต้องประกอบด้วยผู้ประเมิน (rater) และผู้ถูกประเมิน (ratee) หลาย ๆ คน ทำการประเมินข้อมูลหลาย ๆ กลุ่มคุณลักษณะ (trait) ซึ่งมีแหล่งความแปรปรวนแต่ละแหล่ง คือ ผู้ประเมิน (raters) ผู้ถูกประเมิน (ratees) ปฏิสัมพันธ์ระหว่างผู้ประเมินและผู้ถูกประเมิน (raters x ratees) คุณลักษณะ (traits) และปฏิสัมพันธ์ระหว่างผู้ประเมินกับคุณลักษณะ (raters x traits) โดยความคลาดเคลื่อนทั้ง 3 ประเภทสามารถพิจารณาได้ดังนี้

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

หลังจากที่ Guilford (1954) ได้แบ่งประเภทของความคลาดเคลื่อนไว้ดังที่กล่าวข้างต้น ทำให้มีนักวิจัยหลายท่านได้พยายามศึกษาลักษณะการเกิดความคลาดเคลื่อนประเภทต่าง ๆ ในการใช้มาตรประมาณค่า โดยเฉพาะความคลาดเคลื่อนแบบปล่อยคะแนน ความคลาดเคลื่อนแบบฮาโล และความคลาดเคลื่อนแบบขัดแย้ง ดังเช่น Shower (1974) ได้ใช้มาตราประมาณค่าแบบลิเคอร์ท (Fix Alternative Likert Cues) แบบตัวเลือกคงที่ที่ใช้คำคุณศัพท์เชิงประเมิน (fix alternative evaluation cues) และแบบทางเลือกเชิงบรรยาย (descriptive cues) ซึ่งทั้ง 3 แบบมีจำนวนช่วงมาตรฐาน 4 ช่วง ทำการประเมินการสอนของครู พบว่า แบบลิเคอร์ท ซึ่งเป็นที่นิยมใช้กันมากมีแนวโน้มเกิดอคติในการประเมินสูงสุด ในขณะที่แบบตัวเลือกคงที่ที่ใช้คำคุณศัพท์เชิงประเมินเกิดความคลาดเคลื่อนแบบปล่อยคะแนน (leniency error) น้อยที่สุด นอกจากนี้ สุภาวดี ตั้งบุปผา (2535) ได้ใช้มาตรประมาณค่าแบบตัวเลข แบบบรรยาย และแบบกราฟฟิคซึ่งมีจำนวน 5 ช่วง ทำการประเมินคุณลักษณะอาจารย์ที่ปรึกษากิจการนิสิต โดยให้นิสิตและผู้บริหารเป็นผู้ประเมิน ผลของการประเมินพบว่า มาตรประมาณค่าแบบตัวเลขและแบบบรรยาย มีความคลาดเคลื่อนแบบปล่อยคะแนน และความคลาดเคลื่อนแบบขัดแย้ง ในขณะที่มาตรประมาณค่าแบบกราฟฟิคมีความคลาดเคลื่อนแบบปล่อยคะแนน ความคลาดเคลื่อนแบบขัดแย้ง และความคลาดเคลื่อนแบบฮาโล นอกจากนี้ยังพบว่า ในกรณีที่นิสิตเป็นผู้ประเมิน ปรากฏความคลาดเคลื่อนแบบปล่อยคะแนนสูงกว่าในกรณีที่ผู้บริหารเป็นผู้ประเมิน ส่วนการศึกษาของ Slater (1977) ได้ใช้มาตรประมาณค่าแบบเชิงพฤติกรรม (Behaviorally Anchored Scales) ที่มีจำนวนช่วงมาตรฐาน 6 ช่วง ทำการประเมินพฤติกรรมการสอนของศาสตราจารย์จำนวน 15 คน พบว่า การประเมินเกิดความคลาดเคลื่อนแบบฮาโล (halo error) ซึ่งสอดคล้องกับการศึกษาของ Cook (1988) ซึ่งใช้มาตรประมาณค่าเชิง พฤติกรรมที่มีจำนวนช่วงมาตรฐาน 6 ช่วง พบว่า การประเมินการปฏิบัติงานของครู เกิดความคลาดเคลื่อนแบบฮาโล (halo error) เช่นเดียวกัน ในขณะที่ Campbell (1973) Keaveny (1975) และสุพักตร์ พิบูลย์ (2534) ทำการประเมินคุณภาพการสอนของอาจารย์ โดยใช้มาตรประมาณค่าที่มีจำนวนช่วงมาตรฐาน 7 ช่วง ได้ผลการประเมิน

ที่สอดคล้องกันว่า การประเมินคุณภาพการสอนของครู โดยนักเรียนเป็นผู้ประเมิน มีแนวโน้มที่จะเกิดความคลาดเคลื่อนแบบปล่อยคะแนน และความคลาดเคลื่อนแบบฮาโล

จากที่กล่าวมา จะพบว่า มีดัชนีที่จะบ่งชี้ว่า สารสนเทศที่ได้จากการใช้มาตรฐานค่าในการประเมินการเรียนการสอน และการปฏิบัติงานมีความน่าเชื่อถือมากน้อยเพียงใดนั้น นักวิจัยกลุ่มหนึ่งพิจารณาจากความเที่ยง (reliability) ของมาตรฐานค่า ในขณะที่อีกกลุ่มหนึ่งพิจารณาจากความคลาดเคลื่อนในการประเมินค่า (rating error) ได้แก่ ความคลาดเคลื่อนแบบปล่อยคะแนน (leniency error) ความคลาดเคลื่อนแบบฮาโล (halo error) ความคลาดเคลื่อนแบบขัดแย้ง (contrast error) เป็นต้น (Guilford, 1954) ดังนั้นนักวิจัยในกลุ่มแรกจึงพยายามศึกษาว่า จำนวนช่วงมาตรฐาน (number of interval) ในมาตรฐานค่าเท่าใดที่จะทำให้ค่าความเที่ยง (reliability) มีค่าสูงสุด แต่ยังไม่พบนักวิจัยที่พยายามศึกษาว่า จำนวนช่วงมาตรฐานในมาตรฐานค่าเท่าใดที่จะทำให้เกิดความคลาดเคลื่อนในการประเมินค่าแต่ละชนิดน้อยที่สุด Bardo (1982) กล่าวว่า สาเหตุที่ทำให้การประเมินการเรียนการสอนและการปฏิบัติงานของครู เกิดความคลาดเคลื่อนในการประเมินค่า ได้แก่ ความบกพร่องของตัวผู้ประเมิน (subject - related error) และความบกพร่องของเครื่องมือ (form - related error) ในส่วนของความบกพร่องของตัวผู้ประเมิน มีนักวิจัยได้ทำการศึกษาความคลาดเคลื่อนในการประเมินค่า โดยใช้ผู้ประเมินที่มีสถานภาพแตกต่างกัน และพบว่า ความคลาดเคลื่อนในการประเมินค่าจะเกิดขึ้นแตกต่างกัน ในส่วนของความบกพร่องของเครื่องมือ มีนักวิจัยได้ทำการศึกษาความคลาดเคลื่อนในการประเมินค่า โดยใช้รูปแบบมาตรฐานค่าแตกต่างกัน และพบว่า ความคลาดเคลื่อนในการประเมินค่าเกิดขึ้นแตกต่างกัน ในประเด็นหลังนี้ ผู้วิจัยพบข้อสังเกตว่า ในแต่ละงานวิจัยจะใช้จำนวนช่วงมาตรฐานในมาตรฐานค่าแตกต่างกัน และความคลาดเคลื่อนในการประเมินค่าที่วิเคราะห์ได้จะเกิดขึ้นแตกต่างกันด้วยตามที่เสนอข้างต้น ซึ่งในประเด็นนี้ยังไม่มีการวิจัยที่ศึกษาอย่างจริงจัง ดังนั้นผู้วิจัยจึงมีความสนใจที่จะศึกษาว่า เมื่อใช้จำนวนช่วงมาตรฐานในมาตรฐานค่า และผู้ประเมินที่มีสถานภาพแตกต่างกัน จะมีผลทำให้เกิดความคลาดเคลื่อนในการประเมินค่าแตกต่างกันหรือไม่ ซึ่งข้อค้นพบที่ได้ ผู้วิจัยคิดว่าจะเป็นประโยชน์ต่อนักวิจัยและนักการศึกษาอย่างมากในการเลือกใช้จำนวนช่วงมาตรฐานในมาตรฐานค่าและผู้ประเมินที่เหมาะสม และก่อให้เกิดความคลาดเคลื่อนในการประเมินค่าน้อยที่สุด เป็นผลให้การประเมินการสอนและการปฏิบัติงานของครู ตลอดจนบุคลากรในหน่วยงาน มีความน่าเชื่อถือมากที่สุด

### วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบความคลาดเคลื่อนแบบปล่อยคะแนน ความคลาดเคลื่อนแบบฮาโล และความคลาดเคลื่อนแบบขัดแย้ง ระหว่างมาตรฐานค่าที่มีจำนวนช่วงมาตรฐาน 4 ช่วง 5 ช่วง 6 ช่วง และ 7 ช่วง เมื่อสถานภาพของผู้ประเมินเป็นนักเรียน

2. เพื่อศึกษาความคลาดเคลื่อนแบบปล่อยคะแนน ความคลาดเคลื่อนแบบฮาโล และความคลาดเคลื่อนแบบขัดแย้ง ระหว่างมาตรฐานค่าที่มีจำนวนช่วงมาตรฐาน 4 ช่วง 5 ช่วง 6 ช่วง และ 7 ช่วง เมื่อสถานภาพของผู้ประเมินเป็นเพื่อนร่วมงาน

3. เพื่อเปรียบเทียบความคลาดเคลื่อนแบบปล่อยคะแนน ความคลาดเคลื่อนแบบฮาโล และความคลาดเคลื่อนแบบขัดแย้ง เมื่อใช้จำนวนช่วงมาตรฐานในมาตรฐานค่า และสถานภาพของผู้ประเมินต่างกัน

### สมมติฐานการวิจัย

จากงานวิจัยที่กล่าวมาในความเป็นมาและความสำคัญของปัญหานั้น จะเห็นว่า เมื่อใช้จำนวนช่วงมาตรฐานในมาตรฐานค่าต่างกัน ผลการประเมินที่วิเคราะห์ได้ จะเกิดความคลาดเคลื่อนในการประเมินค่าแตกต่างกันไป ผู้วิจัยจึงขอสรุปเป็นข้อสังเกต ดังนี้

ผู้วิจัย	จำนวนช่วงมาตรฐานในมาตรฐานค่าที่ใช้	สถานภาพของผู้ประเมิน	ความคลาดเคลื่อนในการประเมินค่าที่พบ
Shower (1974)	4 ช่วง	นักเรียนระดับมัธยมปีที่ 6	แบบปล่อยคะแนน
สุภาวดี (2535)	5 ช่วง	นิสิต	แบบปล่อยคะแนน แบบขัดแย้ง แบบฮาโล
Slater (1977) และ Cook (1988)	6 ช่วง	นิสิต	แบบฮาโล
Keaveny (1975) และ สุปัทร์ (2534)	7 ช่วง	นิสิต	แบบปล่อยคะแนน แบบฮาโล
Campbell และคณะ (1973)	7 ช่วง	เพื่อนร่วมงาน	แบบปล่อยคะแนน



นอกจากนี้ ผลการศึกษาของ สุภาวดี ตั้งบุปผา (2535) ยังพบว่า กรณีที่นักเรียนเป็นผู้ประเมิน เกิดความคลาดเคลื่อนแบบปล่อยคะแนน สูงกว่า กรณีที่ผู้บริหารเป็นผู้ประเมิน นอกจากนี้ จากการศึกษาของ Marsh (1979) ได้ศึกษาความตรงตามทฤษฎีของการประเมินคุณภาพการสอนของอาจารย์ โดยใช้มาตรประมาณค่าที่มีจำนวนช่วงมาตรฐาน 7 ช่วง พบว่า การประเมินโดยนักเรียนมีความลำเอียงมากกว่าการประเมินโดยกลุ่มเพื่อนอาจารย์ เช่นเดียวกับที่ สัมพันธ์ พันธุ์พุกภัย (2537) ที่ได้ศึกษาเปรียบเทียบความเที่ยงของการประเมินการสอนของอาจารย์โดยเพื่อนร่วมงานและนักเรียน โดยใช้มาตรประมาณค่าที่มีจำนวนช่วงมาตรฐาน 5 ช่วง พบว่า การประเมินการสอนของอาจารย์โดยเพื่อนร่วมงาน มีความเที่ยงค่อนข้างสูง ซึ่งสอดคล้องกับการศึกษาของ Donald (1990) ที่พบว่า การประเมินอาจารย์ในสาขามนุษยศาสตร์ โดยใช้มาตรประมาณค่าที่มีจำนวนช่วงมาตรฐาน 7 ช่วง ควรใช้ กลุ่มเพื่อนร่วมงานเป็นผู้ประเมินเป็นหลัก ดังนั้น ในการศึกษาครั้งนี้ ผู้วิจัย จึงกำหนดสมมุติฐานของการวิจัย ดังนี้

1. กรณีที่สถานภาพของผู้ประเมินเป็นนักเรียน เมื่อใช้จำนวนช่วงมาตรฐานในมาตรประมาณค่าต่างกัน จะเกิดความคลาดเคลื่อนแบบปล่อยคะแนน ความคลาดเคลื่อนแบบฮาโล และความคลาดเคลื่อนแบบขัดแย้ง แตกต่างกันไป ซึ่งสังเกตได้จากการศึกษาของ Shower (1974), Slater (1977), Cook (1988), Keaveny (1975), สุภาวดี ตั้งบุปผา (2535) และสุพัตร์ พิบูลย์ (2534) ที่ใช้นักเรียนเป็นผู้ประเมินคุณภาพการสอนของครูเช่นกัน แต่ใช้จำนวนช่วงมาตรฐานต่างกัน ผลการศึกษาพบว่า ความคลาดเคลื่อนแต่ละชนิดเกิดขึ้นแตกต่างกัน

2. กรณีที่สถานภาพของผู้ประเมินเป็นเพื่อนร่วมงาน เมื่อใช้จำนวนช่วงมาตรฐานในมาตรประมาณค่าต่างกัน จะเกิดความคลาดเคลื่อนแบบปล่อยคะแนน ความคลาดเคลื่อนแบบฮาโล และความคลาดเคลื่อนแบบขัดแย้งแตกต่างกัน ซึ่งสังเกตได้จากการศึกษาของ Campbell (1973), Marsh (1979), Donald (1990) และสัมพันธ์ พันธุ์พุกภัย (2537) ที่ใช้เพื่อนร่วมงานเป็นผู้ประเมินคุณภาพการสอนของครูเช่นกัน แต่ใช้จำนวนช่วงมาตรฐานต่างกัน ผลการศึกษาพบว่า คลาดเคลื่อนแต่ละชนิดเกิดขึ้นแตกต่างกัน

3. เมื่อใช้จำนวนช่วงมาตรฐานและสถานภาพของผู้ประเมินต่างกัน จะเกิดความคลาดเคลื่อนแบบปล่อยคะแนน ความคลาดเคลื่อนแบบฮาโล และความคลาดเคลื่อนแบบขัดแย้ง แตกต่างกันไป ซึ่งสังเกตได้จากการศึกษาของนักวิจัยทุกท่านที่กล่าวมาในสมมุติฐานการวิจัยข้อที่ 1 และ 2 ที่ใช้จำนวนช่วงมาตรฐานและสถานภาพของผู้ประเมินแตกต่างกัน ผลการศึกษาพบว่า ความคลาดเคลื่อนแต่ละชนิดเกิดขึ้นแตกต่างกัน

## ขอบเขตของการวิจัย

### ตัวแปรที่ศึกษา

1. ตัวแปรต้น คือ จำนวนช่วงมาตรในมาตรฐานค่า ได้แก่ จำนวน 4 ช่วง 5 ช่วง 6 ช่วง และ 7 ช่วง และสถานภาพของผู้ประเมิน ได้แก่ นักเรียน และเพื่อนร่วมงาน
2. ตัวแปรตาม คือ ประเภทของความคลาดเคลื่อนในการประเมินค่า ได้แก่ ความคลาดเคลื่อนแบบปล่อยคะแนน ความคลาดเคลื่อนแบบฮาโล และความคลาดเคลื่อนแบบขัดแย้ง

### ข้อตกลงเบื้องต้น

การศึกษาครั้งนี้ นักเรียนชั้นมัธยมศึกษาปีที่ 5 เป็นผู้ประเมินคุณภาพการสอนของครู เนื่องจาก นักเรียนชั้นมัธยมศึกษาปีที่ 5 มีความสามารถเพียงพอในการประเมินคุณภาพการสอนของครู โดยใช้มาตรฐานค่าที่มีจำนวนช่วงมาตร 4 ช่วง 5 ช่วง 6 ช่วง และ 7 ช่วง

### นิยามเชิงปฏิบัติการ

มาตรฐานค่า ประเมินคุณภาพ การสอนของครู	หมายถึง	แบบประเมินคุณภาพการสอนของครูในคุณลักษณะ ทั้งหมด 4 ด้าน ได้แก่ ด้านวิธีสอน ด้านความสัมพันธ์ ระหว่างครูกับนักเรียน ด้านบุคลิกภาพ และด้านการ ประเมินผล ซึ่งมีจำนวนช่วงมาตรของมาตรฐาน ค่าแตกต่างกัน 4 รูปแบบ คือ 4 ช่วง 5 ช่วง 6 ช่วง และ 7 ช่วง
จำนวนช่วงมาตรใน มาตรฐานค่า	หมายถึง	มาตรฐานค่าที่มีจำนวนช่วงมาตร 4 ช่วง 5 ช่วง 6 ช่วง และ 7 ช่วง

ความคลาดเคลื่อนแบบ ปล่อยคะแนน	หมายถึง	คะแนนของการประเมินที่ผู้ประเมินมีความพยายามที่จะ ประเมินสูงหรือต่ำกว่าความเป็นจริง ซึ่งพิจารณาได้จาก การมีนัยสำคัญทางสถิติระหว่างผู้ประเมิน (rater main effect) ด้วยวิธีวิเคราะห์ความแปรปรวน (analysis of variance)
ความคลาดเคลื่อน แบบฮาโล	หมายถึง	คะแนนของการประเมินที่ผู้ประเมินมีความรู้สึกบางอย่าง ต่อผู้ถูกประเมินทำให้ผลการประเมินไม่ตรงตามความ เป็นจริงอันเนื่องมาจากความรู้สึกนั้น อาจเป็นได้ทั้งทาง บวก และทางลบ ซึ่งพิจารณาได้จากการมีนัยสำคัญทาง สถิติของปฏิสัมพันธ์ระหว่างผู้ประเมินกับ ผู้ถูกประเมิน (rater x ratee interaction) ด้วยวิธีวิเคราะห์ความ แปรปรวน (analysis of variance)
ความคลาดเคลื่อน แบบขัดแย้ง	หมายถึง	คะแนนของการประเมินที่เกิดจากผู้ประเมินพยายาม ประเมินคุณลักษณะบางคุณลักษณะในทิศทางที่สูงหรือ ต่ำกว่าที่เป็นจริง ซึ่งพิจารณาได้จากการมีนัยสำคัญทาง สถิติของปฏิสัมพันธ์ระหว่างผู้ประเมินกับคุณลักษณะ (rater x trait interaction) ด้วยวิธีวิเคราะห์ความ แปรปรวน (analysis of variance)
ค่าประมาณของความ คลาดเคลื่อนแบบปล่อย คะแนน	หมายถึง	ผลต่างของคะแนนเฉลี่ยที่ผู้ประเมินแต่ละคนทำการ ประเมินผู้ได้รับการประเมินทุกคนที่แตกต่างไปจาก คะแนนเฉลี่ยทั้งหมด (grand mean)
ค่าประมาณของความ คลาดเคลื่อนแบบฮาโล	หมายถึง	คะแนนของผู้ที่ได้รับการประเมินแต่ละคนลบด้วยผลต่าง ของค่าเฉลี่ยของผู้ที่ได้รับการประเมินทุกคนที่ประเมิน จากผู้ประเมินคนเดียวกับค่าเฉลี่ยทั้งหมด (mean all ratees – grand mean) และลบด้วยผลต่างของค่าเฉลี่ยของ ผู้ประเมินทั้งหมดที่ประเมินผู้ได้รับการประเมินแต่ละคน กับค่าเฉลี่ยทั้งหมด (mean all raters – grand mean)

ค่าประมาณของความคลาดเคลื่อนแบบขจัดแย้ง	หมายถึง คะแนนของคุณลักษณะแต่ละคุณลักษณะลบด้วยผลต่างของค่าเฉลี่ยของคะแนนคุณลักษณะที่ได้รับการประเมินจากผู้ประเมินคนเดียวกับค่าเฉลี่ยทั้งหมด (mean all traits) และลบด้วยผลต่างของค่าเฉลี่ยของผู้ประเมินทั้งหมดที่ประเมินคุณลักษณะแต่ละคุณลักษณะ กับค่าเฉลี่ยทั้งหมด (mean all raters – grand mean)
สถานภาพของผู้ประเมิน	หมายถึง ความสัมพันธ์ระหว่างผู้ประเมินกับผู้ถูกประเมิน ซึ่งได้แก่เพื่อนร่วมงานซึ่งเป็นครูผู้สอนวิชาวิชาชีววิทยา และเคมี ระดับชั้นมัธยมศึกษาปีที่ 5 และ นักเรียนที่เรียนอยู่ในสายวิทย์ – คณิต ระดับชั้นมัธยมศึกษาปีที่ 5
เพื่อนร่วมงาน	หมายถึง ครูที่ทำหน้าที่เป็นผู้ประเมินการสอนของครูที่อยู่ในหมวดวิชาเดียวกัน ซึ่งได้ปฏิบัติงานสอน และงานบริการวิชาการร่วมกันอย่างสม่ำเสมอ
นักเรียน	หมายถึง นักเรียนที่ได้รับกระบวนการเรียนการสอนจากครูที่จะได้รับการประเมิน

### ประโยชน์ที่คาดว่าจะได้รับ

1. ได้แนวทางให้นักวิจัยได้เลือกใช้จำนวนช่วงมาตรในมาตรประมาณค่าที่เหมาะสมและมีความคลาดเคลื่อนน้อยที่สุด
2. ได้แนวทางให้นักวิจัยได้เลือกสถานภาพของผู้ประเมินได้อย่างเหมาะสม และมีความคลาดเคลื่อนน้อยที่สุด
3. ได้แนวทางให้นักวิจัยได้เลือกใช้จำนวนช่วงมาตรในมาตรประมาณค่า และสถานภาพของผู้ประเมินได้อย่างเหมาะสม เพื่อลดความคลาดเคลื่อนจากแหล่งต่างๆ ให้มากที่สุด
4. เป็นแนวทางให้แก่ักวิจัยในการพัฒนามาตรประมาณค่าให้มีคุณภาพดียิ่งขึ้นต่อไป