



บทที่ 5

### สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบคุณภาพของมาตรวัด 2 รูปแบบ คือ มาตรฐานค่าเชิงพฤติกรรม (Behaviorally-Anchored Rating Scales) กับมาตรฐานค่าแบบกราฟิก (Graphic Rating Scales) ในด้านความตรง ความเที่ยง และลักษณะความคลาดเคลื่อนในการประมาณค่า โดยมุ่งเน้นเปรียบเทียบคุณภาพของมาตรวัดในสถานการณ์การศึกษาเชิงสำรวจ คือ การประเมินพฤติกรรมด้านการประเมินผลของครู ด้วยการให้ครูผู้สอนประเมินพฤติกรรมของตนเอง และให้หัวหน้าหมวดวิชาประเมินพฤติกรรมของครูผู้สอนอีกชั้นหนึ่ง

ในการดำเนินการวิจัย ผู้วิจัยได้พัฒนาแบบประเมินพฤติกรรมด้านการประเมินผลของครูผู้สอนขึ้นมา 2 ชุด ชุดแรก เป็นแบบประเมินชนิดมาตรฐานค่าเชิงพฤติกรรม ชุดที่สอง เป็นแบบประเมินชนิดมาตรฐานค่าแบบกราฟิก แบบประเมินทั้งสองชุดได้ถูกบรรจุไว้ด้วยเนื้อหาสาระหรือรายการพฤติกรรมที่มุ่งประเมินที่เหมือนกันจำนวน 18 รายการพฤติกรรม แต่จะแตกต่างกันในส่วนของสเกลคำตอบที่เป็นทางเลือก 5 ระดับ (Alternatives) กล่าวคือ ชนิดมาตรฐานค่าเชิงพฤติกรรม ทางเลือกในการประมาณค่าได้ถูกจัดเรียงไว้ในลักษณะของมาตรวัดในแนวดิ่ง โดยแต่ละทางเลือกจะมีข้อความบรรยายพฤติกรรมการทำงานผูกติดเอาไว้อย่างเรียงลำดับจากมีคุณภาพมากที่สุดไปน้อยที่สุด หรือจากน้อยที่สุดไปมากที่สุด ส่วนแบบประเมินชนิดมาตรฐานค่าแบบกราฟิก ทางเลือกในการประมาณค่าได้ถูกจัดทำในลักษณะของเส้นกราฟแบ่งเป็น 5 ระดับทางเลือก มีค่าคุณศัพท์ที่แสดงถึงคุณภาพการทำงานในด้านบวก-ลบ หรือ ด้านดี-ไม่ดี ผูกติดไว้ที่หัวทั้งสองของเส้นกราฟ แบบประเมินทั้งสองชุดนี้ถือเป็นมาตรวัด 2 รูปแบบที่ใช้เป็นเครื่องมือในการเก็บรวบรวมข้อมูลเพื่อการเปรียบเทียบคุณภาพของมาตรฐานค่าเชิงพฤติกรรม กับมาตรฐานค่าแบบกราฟิก ในการวิจัยครั้งนี้ ทั้งนี้ในการประเมินพฤติกรรมใด ๆ ก็ว่าช่วงคะแนนมาตรวัดทั้งสองรูปแบบ จัดอยู่ในระบบช่วงคะแนนที่เทียบเท่ากัน

กลุ่มตัวอย่างที่เป็นเป้าหมายของการประเมินเพื่อการตรวจสอบคุณภาพของมาตรวัด (Subjects for Scale Evaluation/Target Ratees) ในการวิจัยนี้คือ ครูคณิตศาสตร์ ที่สอนระดับชั้นมัธยมศึกษาตอนต้น ในโรงเรียนมัธยมศึกษาสังกัดกรมสามัญศึกษา เขตกรุงเทพมหานคร ในภาคต้น ปีการศึกษา 2533 จำแนกเป็น (1) ครูคณิตศาสตร์ประเภทครูผู้สอนทั่วไป คัดเลือก โดยสุ่มอย่างง่ายจากครูคณิตศาสตร์ที่สอนในระดับชั้นมัธยมศึกษาปีที่ 3 โรงเรียนละ 1 คน รวม 103 คน และ (2) ครูคณิตศาสตร์ที่ปฏิบัติงานเป็นระบบ-มีประสิทธิภาพ และครูคณิตศาสตร์ที่ปฏิบัติ ปฏิบัติงานยังไม่เป็นระบบ-ขาดประสิทธิภาพ คัดเลือกโดยให้ศึกษานิเทศก์ ภาควิชาคณิตศาสตร์ ศึกษานิเทศก์กรมสามัญศึกษา กระทรวงศึกษาธิการ และหัวหน้าหมวดวิชาคณิตศาสตร์ในโรงเรียน มัธยมศึกษาเป็นผู้ระบุครูผู้สอนที่มีคุณสมบัติตามเกณฑ์ที่กำหนดขึ้น รวมกลุ่มตัวอย่างที่เป็นเป้าหมายของ การประเมินกลุ่มนี้ จำนวน 60 คน เป็นครูคณิตศาสตร์ที่ปฏิบัติงานเป็นระบบ-มีประสิทธิภาพ จำนวน 30 คน และครูคณิตศาสตร์ที่ปฏิบัติงานยังไม่เป็นระบบ-ขาดประสิทธิภาพ จำนวน 30 คน

กลุ่มผู้ประเมิน (Raters) ที่ทำหน้าที่ประเมินพฤติกรรมด้านการประเมินผลของครู กลุ่มเป้าหมายแต่ละคน ประกอบด้วย (1) ตนเอง ทำหน้าที่ประเมินพฤติกรรมของตนเอง และ (2) หัวหน้าหมวดวิชาคณิตศาสตร์ ทำหน้าที่ประเมินพฤติกรรมของครูที่อยู่ในความดูแลของตนอีก หนึ่ง

การเก็บรวบรวมข้อมูลในการวิจัย ได้ดำเนินการในลักษณะของการสำรวจ คือ จัดส่งแบบประเมินทางไปรษณีย์ถึงหัวหน้าหมวดวิชาคณิตศาสตร์ในโรงเรียนมัธยมศึกษา ในเขต กรุงเทพมหานคร จำนวน 103 โรงเรียน เพื่อให้หัวหน้าหมวดวิชาประเมินพฤติกรรมของครูกลุ่มเป้าหมายที่อยู่ในความดูแลของตน พร้อมทั้งขอความร่วมมือจากหัวหน้าหมวดวิชามอบแบบประเมินแก่ ครูผู้สอนกลุ่มเป้าหมายเพื่อใช้ประเมินพฤติกรรมการทำงานของตนเองด้วย โดยสรุป กลุ่ม เป้าหมาย (Subject/Target Ratee) 1 คน จะถูกประเมินโดยผู้ประเมิน (Raters) 2 คน คือ ตนเอง และหัวหน้าหมวดวิชา โดยผู้ประเมินแต่ละคนได้ทำการประเมินพฤติกรรม 2 รอบ ด้วยแบบประเมินที่แตกต่างกัน คือ แบบประเมินชนิดมาตราประมาณค่าเชิงพฤติกรรม และแบบประเมิน ชนิดมาตราประมาณค่าแบบกราฟิก โดยนัยนี้ การเก็บรวบรวมข้อมูลผลการประเมินพฤติกรรมของกลุ่ม เป้าหมายแต่ละคนจะถือว่าได้ข้อมูลครบถ้วนสมบูรณ์ก็ต่อเมื่อได้รับแบบประเมินกลับคืนครบ 4 ฉบับ ซึ่งจากการดำเนินการเก็บรวบรวมข้อมูลตามแนวทางดังกล่าวนี้ ปรากฏว่า ในกรณีของการประเมิน พฤติกรรมของกลุ่มเป้าหมายประเภทครูผู้สอนทั่วไป ได้ข้อมูลกลับคืนครบถ้วนสมบูรณ์ จำนวน 94 ราย จากจำนวนที่คาดหวัง 103 ราย ส่วนกรณีการประเมินพฤติกรรมของครูกลุ่มปฏิบัติงานเป็นระบบ กับ

กลุ่มปฏิบัติงานยังไม่เป็นระบบ ได้รับข้อมูลผลการประเมินตนเองของครูผู้สอนกลับคืนครบถ้วนจำนวน 30 และ 29 ราย ตามลำดับ

ในการวิเคราะห์ข้อมูลเพื่อตรวจสอบสมมุติฐานการวิจัย ในขั้นต้นได้ทำการวิเคราะห์ค่าสถิติพื้นฐานของผลการประเมิน คือ วิเคราะห์ค่าเฉลี่ยเลขคณิต และส่วนเบี่ยงเบนมาตรฐานของคะแนนผลการประเมินในแต่ละรายการพฤติกรรม ค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของคะแนนผลการประเมินรายบุคคล ค่าสหสัมพันธ์ของคะแนนผลการประเมินระหว่างรายการพฤติกรรมต่าง ๆ (Item Intercorrelations) และค่าสหสัมพันธ์ของคะแนนผลการประเมินระหว่างด้านพฤติกรรม (Dimension Intercorrelations) แล้วนำค่าสถิติพื้นฐานทั้งจากสถานการณ์การประเมินพฤติกรรมตนเองของครูผู้สอน และสถานการณ์การประเมินโดยหัวหน้าหมวดวิชามาทำการวิเคราะห์เปรียบเทียบในแง่มุมต่าง ๆ เพื่อตรวจสอบสมมุติฐานการวิจัยแต่ละรายการ

### สรุปผลการวิจัย

#### 1. ผลการตรวจสอบความตรงในการประมาณค่า

จากการวิเคราะห์ข้อมูลเพื่อตรวจสอบความตรงตามภาวะสัมพันธภาพ (Construct Validity) 2 ลักษณะ คือ วิเคราะห์เมตริกซ์สหสัมพันธ์แบบ Multitrait - Multimethod เพื่อตรวจสอบ Convergent & Discriminant Validity และเปรียบเทียบผลการประเมินพฤติกรรมของครูกลุ่มปฏิบัติงานเป็นระบบ กับ กลุ่มปฏิบัติงานยังไม่เป็นระบบ เพื่อตรวจสอบความตรงของมาตรฐานค่าสองรูปแบบในการจำแนกความแตกต่างระหว่างกลุ่มครูที่ทราบประสิทธิภาพการทำงานแล้ว สามารถสรุปผลการตรวจสอบได้ดังนี้

1.1 จากการวิเคราะห์เมตริกซ์สหสัมพันธ์แบบ Multitrait-Multimethod ซึ่งพบว่า ในสถานการณ์การประเมินพฤติกรรมตนเองของครูผู้สอน ผลการประเมินไม่เป็นไปตามภาวะสัมพันธภาพตามหลักการของ Campbell และ Fiske (1959) และไม่สามารถสรุปได้ว่า มาตรฐานค่ารูปแบบใดสามารถจำแนกความแตกต่างระหว่างรายการพฤติกรรมที่มุ่งประเมิน

ได้ดีกว่ากัน ส่วนในสถานการณ์การประเมินพฤติกรรมของครูผู้สอนโดยหัวหน้าหมวดวิชา ผลการประเมินสอดคล้องกับภาวะสัมพันธภาพตามหลักการของ Campbell และ Fiske และ กรณีใช้มาตรประมาณค่าเชิงพฤติกรรมเป็นเครื่องมือประเมินให้ค่าสัมพันธภาพระหว่างด้านพฤติกรรมคู่ใด ๆ ต่ำกว่ากรณีใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมินอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งแสดงถึงการให้ผลการประเมินระหว่างด้านพฤติกรรมที่เป็นอิสระจากกันมากกว่านั้น จากข้อค้นพบนี้ สรุปได้ว่า มาตรประมาณค่าเชิงพฤติกรรมและมาตรประมาณค่าแบบกราฟิกให้ผลการประเมินที่มีความตรงตามภาวะสัมพันธภาพที่เรียกว่า Convergent & Discriminant Validity โดยที่มาตรประมาณค่าเชิงพฤติกรรมสามารถจำแนกความแตกต่างระหว่างรายการพฤติกรรมที่มุ่งประเมินได้อย่างคมชัดกว่ามาตรประมาณค่าแบบกราฟิก

1.2 จากการเปรียบเทียบผลการประเมินพฤติกรรมตนเองของครูผู้สอนระหว่างกลุ่มที่ปฏิบัติงานเป็นระบบ กับ กลุ่มที่ปฏิบัติงานยังไม่เป็นระบบ ซึ่งพบว่า กรณีใช้มาตรประมาณค่าเชิงพฤติกรรมเป็นเครื่องมือประเมินให้ค่าน้ำหนักเชิงจำแนกความแตกต่างระหว่างกลุ่มสูงกว่ากรณีใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมิน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยนัยนี้สรุปได้ว่า กรณีใช้มาตรประมาณค่าเชิงพฤติกรรมเป็นเครื่องมือประเมินสามารถจำแนกความแตกต่างระหว่างกลุ่มครูที่ปฏิบัติงานมีประสิทธิภาพต่างกัน ได้อย่างคมชัดกว่ากรณีใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมิน

โดยสรุปภายใต้เงื่อนไขการตรวจสอบความตรงในการประมาณค่า 2 ลักษณะ ที่ได้ดำเนินการในการวิจัยครั้งนี้ คือ ตรวจสอบ Convergent & Discriminant Validity ด้วยวิธีการวิเคราะห์เมตริกซ์สหสัมพันธ์แบบ Multitrait - Multimethod และเปรียบเทียบผลการประเมินพฤติกรรมของครูกลุ่มที่ปฏิบัติงานมีประสิทธิภาพแตกต่างกัน ได้ปรากฏข้อมูลยืนยันว่า มาตรประมาณค่าเชิงพฤติกรรมให้ผลการประเมินที่มีความตรงตามภาวะสัมพันธภาพสูงกว่าผลการประเมินโดยมาตรประมาณค่าแบบกราฟิก เป็นไปตามสมมุติฐานการวิจัยที่ 1

## 2. ผลการตรวจสอบความเที่ยงในการประมาณค่า

จากการวิเคราะห์ค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินตนเองของครูผู้สอน กับคะแนนผลการประเมินโดยหัวหน้าหมวดวิชา ซึ่งพบว่า กรณิที่ใช้มาตรประมาณค่าเชิงพฤติกรรม เป็นเครื่องมือประเมินให้ค่าสหสัมพันธ์สูงกว่ากรณิที่ใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมิน ในเกือบทุกรายการพฤติกรรมที่มุ่งประเมิน และพบความแตกต่างมีนัยสำคัญทางสถิติที่ระดับ .05 นั้น โดยนัยนี้สรุปได้ว่า ในกรณิที่ใช้มาตรประมาณค่าเชิงพฤติกรรมเป็นเครื่องมือประเมิน คะแนนผลการประเมินตนเองของครูผู้สอน กับคะแนนผลการประเมินโดยหัวหน้าหมวดวิชาจะสอดคล้องกัน มากกว่ากรณิที่ใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมิน หรืออีกนัยหนึ่งก็คือ มาตราประมาณค่าเชิงพฤติกรรมให้ค่าความเที่ยงระหว่างผู้ประเมิน (Rater Reliability) สูงกว่ามาตรประมาณค่าแบบกราฟิก เป็นไปตามสมมุติฐานการวิจัยที่ 2

## 3. ผลการตรวจสอบลักษณะความคลาดเคลื่อนในการประมาณค่า

### 3.1 ผลการตรวจสอบ Leniency Error

จากการเปรียบเทียบค่าเฉลี่ยเลขคณิตของผลการประเมินพฤติกรรมการปฏิบัติงานของครูแต่ละรายพฤติกรรม ระหว่าง กรณิที่ใช้มาตรประมาณค่าเชิงพฤติกรรมเป็นเครื่องมือประเมินกับกรณิที่ใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมิน ซึ่งพบว่า กรณิที่ใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมินให้ค่าเฉลี่ยเลขคณิตของผลการประเมินแต่ละรายการพฤติกรรม สูงกว่ากรณิที่ใช้มาตรประมาณค่าเชิงพฤติกรรมเป็นเครื่องมือประเมินอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และการกระจายของค่าเฉลี่ยเลขคณิตของผลการประเมินระหว่างรายการพฤติกรรมมีค่าน้อยกว่า ทั้งในสถานการณ์การประเมินตนเองของครูและสถานการณ์การประเมินโดยหัวหน้าหมวดวิชานั้น โดยนัยนี้สรุปได้ว่า กรณิที่ใช้มาตรประมาณค่าเชิงพฤติกรรมเป็นเครื่องมือประเมิน ปรากฏลักษณะการปล่อยคะแนนน้อยกว่ากรณิที่ใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมิน หรืออีกนัยหนึ่งก็คือ เกิด Leniency Error น้อยกว่า

### 3.2 ผลการตรวจสอบ Halo Error

จากการวิเคราะห์คะแนนความคลาดเคลื่อนในการประมาณค่า (Halo Score) หรือส่วนเบี่ยงเบนมาตรฐานของคะแนนรายพฤติกรรมของกลุ่มเป้าหมายเป็นรายบุคคลแล้วเปรียบเทียบค่าเฉลี่ยเลขคณิตของคะแนนความคลาดเคลื่อน ระหว่าง กรณีประเมินโดยใช้มาตรฐานค่าต่างรูปแบบ ซึ่งพบว่า กรณีใช้มาตรฐานค่าเชิงพฤติกรรมเป็นเครื่องมือประเมิน ปรากฏค่าเฉลี่ยเลขคณิตของคะแนนความคลาดเคลื่อนสูงกว่ากรณีใช้มาตรฐานค่าแบบกราฟิกเป็นเครื่องมือประเมิน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทั้งในสถานการณ์การประเมินพฤติกรรมตนเองของครูผู้สอน และการประเมินพฤติกรรมของครูโดยหัวหน้าหมวดวิชา ข้อค้นพบนี้แสดงให้เห็นว่า ในการประเมินพฤติกรรมที่แตกต่างกันของกลุ่มเป้าหมายคนหนึ่ง ๆ กรณีใช้มาตรฐานค่าเชิงพฤติกรรมเป็นเครื่องมือประเมินจะปรากฏคะแนนผลการประเมินระหว่างรายการพฤติกรรมแตกต่างกันมาก ในขณะที่การใช้มาตรฐานค่าแบบกราฟิกเป็นเครื่องมือประเมิน ได้ให้คะแนนผลการประเมินระหว่างรายการพฤติกรรมแตกต่างกันน้อย หรือคะแนนผลการประเมินแต่ละรายการพฤติกรรมมีค่าใกล้เคียงกัน อย่างสม่ำเสมอ ซึ่งเป็นการสะท้อนถึงการเกิด Halo Error ในการประเมินพฤติกรรม โดยนัยนี้สรุปได้ว่า กรณีใช้มาตรฐานค่าเชิงพฤติกรรมเป็นเครื่องมือประเมินเกิด Halo Error น้อยกว่ากรณีใช้มาตรฐานค่าแบบกราฟิกเป็นเครื่องมือประเมิน

โดยสรุป จากการตรวจสอบลักษณะความคลาดเคลื่อนในการประมาณค่า 2 ลักษณะ ในการวิจัยนี้ คือ Leniency Error และ Halo Error สรุปได้ว่า กรณีใช้มาตรฐานค่าเชิงพฤติกรรมเป็นเครื่องมือประเมินปรากฏลักษณะความคลาดเคลื่อนในการวัดหรือลักษณะความคลาดเคลื่อนในการประมาณค่าน้อยกว่ากรณีใช้มาตรฐานค่าแบบกราฟิกเป็นเครื่องมือประเมิน เป็นไปตามสมมุติฐานการวิจัยที่ 3

## อภิปรายผล

การอภิปรายผลการวิจัยในที่นี้ ได้เลือกอภิปรายในส่วนของข้อค้นพบหรือประเด็นที่น่าสนใจ เกตที่สำคัญบางประการที่ปรากฏในผลการวิเคราะห์ข้อมูลเท่านั้น ดังรายละเอียดต่อไปนี้

### 1. ความเป็นไปได้ หรือความสมเหตุสมผลของข้อค้นพบ

หลักการที่สำคัญประการหนึ่งของการวิจัยครั้งนี้คือ มุ่งเน้นการพัฒนาเทคนิควิธีและตรวจสอบคุณภาพของวิธีการที่พัฒนาขึ้น ซึ่งจากหลักการดังกล่าวนี้ ในการดำเนินการวิจัย ผู้วิจัยได้พัฒนามาตราประมาณค่าเชิงพฤติกรรมขึ้นมาชุดหนึ่งตามหลักการหรือแนวทางที่เสนอโดย Landy และ Schwab (1974) และ Heneman และ DeCotter (1975) แล้วจัดทำเป็นแบบประเมินพฤติกรรมด้านการประเมินผลของครูผู้สอน ใช้เป็นเครื่องมือในการเก็บรวบรวมข้อมูลเพื่อหาข้อสรุปเกี่ยวกับคุณภาพของมาตราประมาณค่ารูปแบบนี้ด้วยการเปรียบเทียบกับมาตราประมาณค่าแบบกราฟิกอันเป็นรูปแบบหนึ่งของมาตราประมาณค่าที่ได้รับความนิยมใช้กันอย่างแพร่หลายทั่วไปและเป็นรูปแบบที่ได้รับการตรวจสอบยืนยันคุณภาพในด้านความเที่ยงในงานวิจัยในประเทศไทยว่ามีความเที่ยงสูงกว่ารูปแบบอื่น ๆ (มัลลิกา ยวนะเดมิย์, 2527)

จากการดำเนินการพัฒนามาตราประมาณค่าเชิงพฤติกรรมอย่างมีลำดับขั้นตอน ทำให้ได้มาตราประมาณค่าสำหรับใช้ประเมินพฤติกรรมของครูผู้สอนชุดหนึ่งที่มีลักษณะสำคัญบางประการสอดคล้องกับข้อเสนอแนะ เพื่อการปรับปรุงคุณภาพของการประมาณค่ารองนักวัดผลหลายท่าน อาทิ Thorndike และ Hagen (1977: 487) เสนอแนะว่า ในการประเมินนั้นควรทำการประเมินเฉพาะลักษณะที่ปรากฏให้เห็นชัดเจน ให้ผู้ประเมินสามารถตัดสินได้โดยพิจารณาจากรายการพฤติกรรมที่กำหนดให้เท่านั้น" หรือ Oppenheim (1966) ได้เสนอแนะว่า ในการสร้างมาตราประมาณค่าจะต้องกำหนดขอบเขตและทิศทางของลักษณะที่จะประเมินให้เป็นที่เข้าใจตรงกันเสียก่อน เป็นต้นตามข้อเสนอแนะดังกล่าวเหล่านี้ จะพบว่า ในการพัฒนามาตราประมาณค่าเชิงพฤติกรรมครั้งนี้ การใช้ตัวอย่างพฤติกรรมหรือลักษณะการปฏิบัติงานผูกติดไว้ที่ระดับทางเลือกในการประมาณค่า (Alternatives) ทั้ง 5 ระดับ ทำให้ผู้ประเมินสามารถตัดสินได้โดยพิจารณาจากรายการพฤติกรรมที่ผูกติดไว้กับที่ระดับทางเลือกนั้น ๆ เป็นไปตามข้อเสนอของ Thorndike และ Hagen หรือการที่ผู้ประเมินได้สารสนเทศ (Information) ที่เพียงพอจากการอ่านพิจารณารายการพฤติกรรมหลักที่มุ่งประเมิน พร้อมทั้งพิจารณารายละเอียดของพฤติกรรมในแต่ละทางเลือกทั้ง 5 ระดับ จะ

ทำให้ผู้ประเมินเข้าใจขอบเขตและทิศทางของลักษณะที่กำลังประเมินได้อย่างสอดคล้องตรงกัน แม้จะเป็นผู้ประเมินต่างบุคคลกัน ในประเด็นนี้ก็ถือได้ว่าเป็นไปตามเงื่อนไขข้อเสนอแนะของ Oppenheim ดังนั้น ภายใต้การพิจารณาด้วยวิธีการเชิงเหตุผลดังกล่าวแล้วนี้ มาตรฐานค่าเชิงพฤติกรรมที่ได้รับการพัฒนาขึ้นในการวิจัยครั้งนี้ น่าจะมีคุณภาพดีกว่ามาตรฐานค่าแบบกราฟิกที่ใช้เป็นมาตรวัดคู่เปรียบเทียบ

ในการตรวจสอบคุณภาพของมาตรฐานค่าเชิงพฤติกรรม เปรียบเทียบกับมาตรฐานค่าแบบกราฟิกด้วยอาศัยข้อมูลเชิงประจักษ์ในการวิจัยนี้ ดัชนีบ่งชี้คุณภาพ รายการแรก ที่มุ่งตรวจสอบก็คือ ความตรงในการประมาณค่า ซึ่งพบว่า มาตรฐานค่าเชิงพฤติกรรมมีความตรงในการจำแนกความแตกต่างระหว่างรายการพฤติกรรมที่ให้ประเมิน (Discriminant Validity) สูงกว่ามาตรฐานค่าแบบกราฟิก ข้อค้นพบนี้สอดคล้องกับข้อค้นพบในงานวิจัยในต่างประเทศที่ได้ทำการศึกษาในลักษณะที่ใกล้เคียงกับการศึกษาครั้งนี้ อาทิ Keaveny และ McGann (1975) ได้ทำการเปรียบเทียบคุณภาพของมาตรคาดหมายเชิงพฤติกรรม (Behavioral Expectation Scales) กับมาตรฐานค่าแบบกราฟิก (Graphic Rating Scales) ในสถานการณ์การประเมินพฤติกรรมการสอนของศาสตราจารย์ 4 คน โดยให้นักศึกษาเป็นผู้ประเมิน ที่พบว่า มาตรคาดหมายเชิงพฤติกรรมปรากฏ Discriminant Validity สูงกว่ามาตรฐานค่าแบบกราฟิก Campbell และคณะ (1973) ได้เปรียบเทียบคุณภาพของมาตรฐานค่าเชิงพฤติกรรม (Behaviorally Based Rating Scales) กับมาตรฐานค่าแบบรวมสรุป (Summated Rating Scales) อันเป็นรูปแบบหนึ่งของมาตรฐานค่าที่ไม่มีตัวอย่างพฤติกรรมผูกติดบนมาตรวัด โดยใช้ประเมินพฤติกรรมของผู้จัดการฝ่ายในห้างขายส่งสินค้า ด้วยการให้ผู้จัดการร้านค้าและผู้ช่วยผู้จัดการเป็นผู้ประเมิน ซึ่งพบว่า มาตรฐานค่าเชิงพฤติกรรม ปรากฏ Discriminant Validity สูงกว่า เช่นเดียวกัน ดังนั้นภายใต้ความสอดคล้องของข้อค้นพบดังกล่าวเหล่านี้ น่าจะเป็นข้อมูลยืนยันในความเป็นไปได้ของข้อค้นพบในงานวิจัยครั้งนี้ได้เป็นอย่างดี

ดัชนีบ่งชี้คุณภาพรายการที่สองที่ได้ทำการตรวจสอบในการศึกษาครั้งนี้คือ ความเที่ยงในการประมาณค่า ซึ่งพบว่า กรณีใช้มาตรฐานค่าเชิงพฤติกรรมเป็นเครื่องมือประเมินให้ค่าความเที่ยงระหว่างผู้ประเมิน (Rater Reliability) สูงกว่ากรณีใช้มาตรฐานค่าแบบกราฟิกเป็นเครื่องมือประเมิน ข้อค้นพบนี้ในเชิงเหตุผลน่าจะเป็นไปได้สูง ทั้งนี้เพราะว่าในการใช้มาตรฐานค่าเชิงพฤติกรรมประมาณค่าพฤติกรรมใด ๆ ของกลุ่มเป้าหมายแต่ละคนนั้น



ผู้ประเมิน 2 คน ซึ่งในที่นี้คือ ครูผู้สอนเอง และ หัวหน้าหมวดวิชา สามารถตัดสินใจได้จากการเปรียบเทียบพฤติกรรมการทำงานของกลุ่มเป้าหมายกับข้อความเชิงบรรยายพฤติกรรมการทำงานที่ผูกติดไว้ในแต่ละระดับทางเลือกบนมาตรวัด ซึ่งการตัดสินใจภายใต้เกณฑ์เปรียบเทียบที่ชัดเจน เช่นนี้ ผลการประมาณค่าของผู้ประเมิน 2 คน จะมีโอกาสสอดคล้องกันมากกว่ากรณีที่ขาดเกณฑ์ในการตัดสินใจซึ่งจะเกิดขึ้นในกรณีของการใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมิน ยิ่งถ้ารายการพฤติกรรมหลักที่ให้ประเมินเป็นพฤติกรรมที่ชัดเจนและสามารถสังเกตได้โดยง่ายด้วยแล้ว คะแนนผลการประเมินของผู้ประเมิน 2 คน จะมีโอกาสสอดคล้องกันมากยิ่งขึ้น ดังปรากฏให้เห็นในกรณีของการประเมินพฤติกรรมหลักรายการที่ 14 เกี่ยวกับการวิเคราะห์คุณภาพของข้อสอบและแบบสอบทั้งฉบับ ซึ่งพบว่าค่าสหสัมพันธ์ระหว่างคะแนนของผู้ประเมิน 2 กลุ่มมีค่าเท่ากับ 0.71 ในกรณีที่ใช้มาตรประมาณค่าเชิงพฤติกรรมเป็นเครื่องมือประเมิน ในขณะที่ปรากฏค่าสหสัมพันธ์เท่ากับ 0.32 ในกรณีที่ใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมิน

ดัชนีบ่งชี้คุณภาพรายการที่สามที่ได้รับการตรวจสอบในการวิจัยครั้งนี้คือ ลักษณะความคลาดเคลื่อนในการประมาณค่า ซึ่งพบว่า มาตรประมาณค่าเชิงพฤติกรรมให้ผลการประเมินที่สะท้อนถึงการเกิดลักษณะความคลาดเคลื่อนในการประมาณค่าในระดับที่น้อยกว่ามาตรประมาณค่าแบบกราฟิก ข้อค้นพบนี้เป็นไปได้ว่า ในกรณีที่ขาดเกณฑ์ในการตัดสินใจประมาณค่าพฤติกรรม ดังที่ปรากฏในกรณีที่ใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมินนั้น โดยลักษณะนิสัยของคนไทยที่มีความเมตตา กรุณา เอื้อเฟื้อเผื่อแผ่ ชอบช่วยเหลือซึ่งกันและกัน และไม่เข้มงวดในระเบียบวินัยจนเกินไป (จารุวรรณ ไวยเจตน์, 2523 : 164) จึงมีโอกาที่จะเกิดการประเมินในลักษณะของการปล่อยคะแนนหรือเลือกประเมินในระดับสูง ๆ ก่อให้เกิดความคลาดเคลื่อนที่เรียกว่า Leniency Error หรือ เปิดโอกาสให้ปฏิสัมพันธ์ส่วนบุคคล เช่น การรู้จักกันมาก่อน การเคยเห็นผลงานของกลุ่มเป้าหมายทั้งในทางบวก-ดี หรือในทางลบ-ไม่ดี มีอิทธิพลต่อการประเมินพฤติกรรมในครั้งนั้น ๆ อันเป็นสภาพของการเกิด Halo Error ในการประเมินพฤติกรรม ซึ่งนอกเหนือจากความเป็นไปได้ของข้อค้นพบในเชิงเหตุเชิงผลดังกล่าวนี้แล้ว ผลการวิจัยในครั้งนี้ก็สอดคล้องกับผลการศึกษาในต่างประเทศหลายกรณี อาทิ Campbell และคณะ (1973) พบว่า มาตรประมาณค่าเชิงพฤติกรรมเกิด Halo Error น้อยกว่ามาตรประมาณค่าชนิดที่ไม่มีตัวอย่างพฤติกรรมผูกติดบนมาตรวัด Keaveny และ McGinn (1975) พบว่า ในการประเมินพฤติกรรมของผู้สอนโดยนักศึกษา มาตรประมาณค่าเชิงพฤติกรรมเกิดความคลาดเคลื่อนชนิด Halo Error น้อยกว่ามาตรประมาณค่าแบบกราฟิก

Cook (1988) พบว่า การใช้ประเมินโดยใช้มาตรฐานค่าเชิงพฤติกรรม เกิด Leniency Error น้อยกว่าการใช้มาตรฐานค่าแบบลิเคอร์ท ซึ่งถือเป็นรูปแบบหนึ่งของมาตรฐานค่าที่ไม่มีตัวอย่างพฤติกรรมผูกติดบนมาตรวัดเป็นเครื่องมือประเมิน เป็นต้น

โดยสรุป จากการวิเคราะห์ความเป็นไปได้ในเชิงเหตุผลประกอบกับข้อค้นพบในงานวิจัยในต่างประเทศหลายเรื่อง หลายลักษณะที่ปรากฏข้อสรุปอย่างสอดคล้องกับผลการวิจัยครั้งนี้ น่าจะเป็นหลักฐานยืนยันได้ว่า มาตรฐานค่าเชิงพฤติกรรมมีคุณภาพมากกว่ามาตรฐานค่าแบบกราฟิก ทั้งในด้านความตรง ความเที่ยง และลักษณะความคลาดเคลื่อนในการประมาณค่า

## 2. เหตุผลที่เป็นไปได้ที่ทำให้ผลการวิเคราะห์ข้อมูลบางรายการไม่เป็นไปตามสมมติฐานการวิจัย

จากการวิเคราะห์ข้อมูลเพื่อตรวจสอบสมมติฐานการวิจัยในการวิจัยครั้งนี้ ได้พบว่าผลการวิเคราะห์ข้อมูลรายการหนึ่ง คือ การวิเคราะห์ผลการประเมินตนเองของครูผู้สอน ด้วยการบรรยายเมตริกซ์สหสัมพันธ์แบบ Multitrait-Multimethod เพื่อตรวจสอบ Convergent & Discriminant Validity ของมาตรฐานค่าทั้งสองรูปแบบ ปรากฏว่าผลการวิเคราะห์ข้อมูลไม่เป็นไปตามภาวะสันนิษฐานตามหลักการของ Campbell และ Fiske กล่าวคือ ค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินพฤติกรรมต่างด้านภายใต้การใช้มาตรวัดรูปแบบเดียวกัน (Heterodimensional-Monomethod Correlations) มีค่าสูงกว่าค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินรายการพฤติกรรมด้านเดียวกันด้วยมาตรฐานค่าต่างรูปแบบ ซึ่งแสดงถึงการขาดคุณสมบัติในด้านความตรงที่เรียกว่า Discriminant Validity ของผลการประเมินโดยมาตรฐานค่าทั้ง 2 รูปแบบ และไม่สามารถสรุปได้ว่า มาตรฐานค่ารูปแบบใดจำแนกความแตกต่างระหว่างด้านพฤติกรรมที่มุ่งประเมินได้ดีกว่ากัน

การที่ผลการวิเคราะห์ข้อมูล ในสถานการณ์การประเมินตนเองของครูผู้สอน ปรากฏข้อสรุปดังกล่าวข้างต้น ในขณะที่ผลการวิเคราะห์ข้อมูลในสถานการณ์การประเมินโดยหัวหน้าหมวดวิชาเป็นไปตามภาวะสันนิษฐานตามหลักการของ Campbell และ Fiske (1959) อาจเป็นเพราะความตั้งใจในการตอบแบบประเมินหรือความตั้งใจในการประมาณค่าของผู้ประเมินทั้ง 2 กลุ่มนี้แตกต่างกัน ทั้งนี้ด้วยสาเหตุที่เป็นไปได้คือ (1) ในการเก็บรวบรวมข้อมูลครั้งนี้ ผู้วิจัยได้ใช้วิธีส่งแบบประเมินถึงหัวหน้าหมวดวิชา 2 ชุด ขอความร่วมมือให้มอบแบบประเมินให้ครูผู้สอน 1 ชุด

สำหรับใช้ประเมินพฤติกรรมของตนเอง และหัวหน้าหมวดวิชาเก็บไว้ 1 ชุด เพื่อใช้ประเมินพฤติกรรมของครูอีกชั้นหนึ่ง เป็นสถานะเสมือนการใช้คู่ประเมินเปรียบเทียบเพื่อตรวจสอบ ผลการประเมินของกันและกัน ซึ่งโดยนัยนี้เป็นไปได้ว่า ครูผู้สอนซึ่งทำหน้าที่ประเมินพฤติกรรมของตนเอง และส่งแบบประเมินกลับคืนทางไปรษณีย์อย่างอิสระนั้น ไม่มีโอกาสได้ทราบว่าในการประเมินครั้งนี้ มีคู่ประเมินเปรียบเทียบด้วย จึงทำการประเมินอย่างตามสบาย ไม่ได้พิถีพิถันหรืออย่างจริงจัง ในขณะที่หัวหน้าหมวดวิชาทราบว่าการประเมินครั้งนี้มีคู่ประเมินเปรียบเทียบคือครูผู้สอน มีผลทำให้ความตั้งใจในการตอบมีมากขึ้น และ (2) ด้วยความสัมพันธ์ส่วนตัวระหว่างผู้วิจัยกับกลุ่มหัวหน้าหมวดวิชา เนื่องจากเคยพบและรู้จักในที่ประชุมสัมมนาต่าง ๆ ที่กรมสามัญศึกษาจัดขึ้น อาจมีผลทำให้กลุ่มหัวหน้าหมวดวิชาให้ความร่วมมือในการตอบแบบประเมินอย่างจริงจังยิ่งขึ้น และมีความตั้งใจมากกว่าเมื่อเทียบกับกลุ่มครูผู้สอน การที่กลุ่มครูผู้สอนมีความตั้งใจในการตอบน้อยกว่า อาจเป็นสาเหตุสำคัญทำให้เกิดความคลาดเคลื่อนในการประมาณค่ามากกว่า ส่งผลให้ผลการตรวจสอบความตรงในการประมาณค่าไม่เป็นไปตามสมมุติฐานการวิจัย

### 3. วิเคราะห์เชิงเปรียบเทียบประสิทธิภาพของเครื่องมือวิจัยและประเด็นที่น่าสังเกตเกี่ยวกับอัตราการตอบแบบประเมินกลับคืน

ในการวิจัยนี้ ได้ใช้แบบประเมินพฤติกรรมการปฏิบัติงานของครูผู้สอน 2 ชุดคือแบบประเมินชนิดมาตราประมาณค่าเชิงพฤติกรรมกับแบบประเมินชนิดมาตราประมาณค่าแบบกราฟิก ทำการประเมินพฤติกรรมด้านการประเมินผลของครู แล้วนำผลการประเมินมาวิเคราะห์เปรียบเทียบคุณภาพของมาตราประมาณค่า 2 รูปแบบ ซึ่งพบว่า ภายใต้การใช้ดัชนีบังคับชี้คุณภาพหลายรายการ ทั้งด้านความตรง ความเที่ยง และลักษณะความคลาดเคลื่อนในการประมาณค่า ได้ปรากฏผลการวิเคราะห์ข้อมูลสนับสนุนว่า มาตราประมาณค่าเชิงพฤติกรรมมีคุณภาพมากกว่ามาตราประมาณค่าแบบกราฟิก อย่างไรก็ตามภายใต้ความมีคุณภาพก็มีข้อจำกัดในบางลักษณะ อาทิ ใช้เวลาในการพัฒนา ผู้ประเมินต้องใช้สมาธิและใช้เวลาในการอ่านพิจารณา มาก จึงจะสามารถประเมินได้แล้วเสร็จหรือไม่สะดวกในการตอบ เป็นต้น ในขณะที่มาตราประมาณค่าแบบกราฟิกพัฒนาได้ง่ายกว่า และเป็นเครื่องมือที่สะดวกสำหรับผู้ตอบสามารถตอบแบบประเมินได้ในเวลาอันสั้น ซึ่งในประเด็นหลังนี้ ในกรณีที่ใช้วิธีส่งแบบประเมินถึงผู้ตอบทางไปรษณีย์อาจมีผลต่อการได้รับแบบประเมินกลับคืนในอัตราส่วนที่สูงกว่า ดังเช่นที่ปรากฏในการวิจัยครั้งนี้ (ตารางที่ 3)

ในประเด็นของความสะดวกในการตอบสำหรับกรณีที่ใช้มาตรฐานค่าแบบกราฟิก เป็นเครื่องมือประเมิน โดยพบว่าสัดส่วนของแบบประเมินที่ได้รับกลับคืน สูงกว่าแบบประเมินชนิดมาตรฐานค่าเชิงพฤติกรรม ดังที่ปรากฏในการวิจัยครั้งนี้ ทั้ง ๆ ที่เมื่อพิจารณาในเชิงคุณภาพแล้ว ผลการประเมินโดยมาตรฐานค่าเชิงพฤติกรรมให้ข้อสรุปที่น่าเชื่อถือกว่า นั้น ข้อสังเกตในประเด็นนี้ควรจะเป็นข้อที่ควรระวังสำหรับนักวิจัยโดยทั่วไปว่า ในการเก็บรวบรวมข้อมูลเพื่อการศึกษาวิจัยในเรื่องใดก็ตาม ควรจะเน้นในเรื่องของคุณภาพของข้อมูลที่ได้รับควบคุมไปกับการเน้นในเรื่องสัดส่วนของผู้ตอบแบบประเมิน ไม่ควรเน้นเฉพาะในด้านสัดส่วนผู้ตอบแบบประเมินเพียงอย่างเดียว

#### 4. ความแตกต่างของผลการประเมินระหว่างกรณีใช้ผู้ประเมินต่างกลุ่มกัน

ในการวิจัยนี้ ได้ใช้กลุ่มผู้ประเมิน 2 กลุ่ม คือ ครูผู้สอนทำหน้าที่ประเมินพฤติกรรมของตนเอง และหัวหน้าหมวดวิชาทำหน้าที่ประเมินพฤติกรรมของครูอีกชั้นหนึ่ง หากพิจารณาในเชิงเปรียบเทียบผลการประเมินพฤติกรรมโดยบุคคลทั้ง 2 กลุ่มนี้ พบว่า กรณีใช้มาตรฐานค่าเชิงพฤติกรรมเป็นเครื่องมือประเมิน คะแนนผลการประเมินรายการพฤติกรรมต่าง ๆ ทั้ง 2 กลุ่มแตกต่างกันไม่มากนัก ค่าสหสัมพันธ์ระหว่างคะแนนของทั้ง 2 กลุ่ม มีค่าสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 จำนวน 17 ใน 18 รายการพฤติกรรม และความสอดคล้องของคะแนนจะมากขึ้นถ้ารายการพฤติกรรมที่มุ่งประเมินเป็นรายการพฤติกรรมที่กลุ่มบุคคลทั้ง 2 กลุ่ม สามารถสังเกตได้อย่างชัดเจน เช่น รายการพฤติกรรมที่ 14 เกี่ยวกับการวิเคราะห์คุณภาพของข้อสอบ ในขณะที่รายการพฤติกรรมที่ 8 เกี่ยวกับการให้ข้อมูลป้อนกลับในการตรวจผลงานหรือการบ้าน และรายการพฤติกรรมที่ 17 การเปิดโอกาสให้ผู้เรียนแสดงความคิดเห็นเกี่ยวกับกระบวนการวัดผล-ประเมินผลที่ครูใช้ ผลการประเมินพฤติกรรมทั้ง 2 รายการนี้ ปรากฏค่าสหสัมพันธ์ระหว่างคะแนนของผู้ประเมินทั้ง 2 กลุ่ม อยู่ในระดับที่ต่ำเมื่อเทียบกับค่าสหสัมพันธ์ในรายการอื่น ๆ ที่เป็นเช่นนี้อาจเป็นเพราะว่ารายการพฤติกรรมทั้งสองนี้ เป็นรายการพฤติกรรมที่หัวหน้าหมวดวิชาสังเกตได้ยาก หรือไม่มีโอกาสได้พบเห็นอย่างชัดเจน การประมาณค่าในกรณีเช่นนี้ จึงต้องใช้วิธีการคาดคะเน มีผลทำให้คะแนนผลการประเมินไม่สอดคล้องกับผลการประเมินตนเองของครูผู้สอน โดยนัยนี้เป็นการแสดงให้เห็นว่า ภายใต้อาการใช้มาตรฐานค่าเชิงพฤติกรรมเป็นเครื่องมือประเมิน ซึ่งถือว่ามีคุณภาพมากกว่ากรณีใช้มาตรฐานค่าแบบกราฟิกนั้น ก็ยังมีจุดอ่อน หรือข้อจำกัดในเรื่องของความชัดเจนของรายการพฤติกรรมที่มุ่งประเมิน ถ้าพฤติกรรมที่มุ่งประเมินเป็นพฤติกรรมที่ไม่

ชัดเจนหรือผู้ประเมิน 2 กลุ่ม ไม่มีโอกาสมองเห็นพฤติกรรมอย่างเท่าเทียมกัน ความน่าเชื่อถือของผลการประเมินก็จะลดลง

เมื่อเปรียบเทียบคะแนนผลการประเมินตนเองของครูผู้สอนกับคะแนนผลการประเมินโดยหัวหน้าหมวดวิชา ภายใต้การใช้มาตราประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมิน พบว่าคะแนนผลการประเมินไม่ค่อยสอดคล้องกัน ทั้งนี้โดยยพิจารณาจากค่าสหสัมพันธ์ระหว่างคะแนนซึ่งปรากฏว่ามีค่าต่ำ โดยเป็นที่น่าสังเกตว่า ค่าเฉลี่ยเลขคณิตของผลการประเมินตนเองของครูผู้สอนในรายการพฤติกรรมต่าง ๆ 13 ใน 13 รายการพฤติกรรม มีค่าสูงกว่าผลการประเมินโดยหัวหน้าหมวดวิชา ความแตกต่างนี้จะเห็นได้ชัดเจนยิ่งขึ้น ในกรณีของการประเมินพฤติกรรมด้านการวางแผนการวัดผล-ประเมินผลและด้านการตรวจสอบและพัฒนาระบบการวัดผลกล่าวคือ 9 ใน 10 รายการพฤติกรรมของทั้ง 2 ด้านนี้ ผลการประเมินตนเองของครูสูงกว่าผลการประเมินโดยหัวหน้าหมวดวิชา แสดงถึงการปล่อยคะแนนในระดับที่สูงกว่า เหตุผลที่เป็นเช่นนี้อาจเป็นเพราะหัวหน้าหมวดวิชาซึ่งเป็นผู้รับผิดชอบโดยตรงในด้านการวางแผนและพัฒนาระบบการเรียนการสอนได้ตั้งความคาดหวังค่อนข้างสูงที่จะให้ครูผู้สอนมีการวางแผนการทำงานและพัฒนางานอย่างจริงจัง ความคาดหวังเช่นนี้ส่งผลให้ประเมินพฤติกรรมของครูในปัจจุบันในระดับไม่สูงนัก ในขณะที่ครูผู้สอนอาจคิดว่าตนเองได้ทำหน้าที่ดีที่สุดแล้ว จึงได้เลือกประเมินในระดับที่สูงกว่า

จากการวิเคราะห์ผลการประเมินพฤติกรรมของครูโดยกลุ่มบุคคลที่แตกต่างกันดังที่ได้อภิปรายข้างต้นนี้เป็นการชี้ให้เห็นว่า ความแตกต่างของผลการประเมินดังที่ได้พบในการวิจัยครั้งนี้ นั้น นอกจากจะเกิดจากลักษณะของเครื่องมือวัดแล้ว ลักษณะของเนื้อหา หรือรายการพฤติกรรมที่มุ่งประเมินอาจเป็นตัวแปรที่สำคัญที่ส่งผลให้เกิดความแตกต่างได้อีกด้วย

#### ข้อเสนอแนะ

##### 1. การพัฒนาและใช้มาตราประมาณค่าในโอกาสต่อไป

จากข้อค้นพบและประเด็นอภิปรายที่ได้เสนอไปแล้วนั้น ผู้วิจัยเห็นว่าในการพัฒนาหรือใช้มาตราประมาณค่าในโอกาสต่อไป ควรพิจารณาหรือคำนึงถึงประเด็นต่าง ๆ ต่อไปนี้

1.1 ภายใต้ข้อค้นพบเกี่ยวกับคุณภาพของมาตราประมาณค่าเชิงพฤติกรรมดังที่ปรากฏในการวิจัยนี้ จึงเห็นว่ามาตราประมาณค่ารูปแบบนี้ควรได้รับการพัฒนาและใช้ในสถานการณ์การศึกษาเชิงประเมินต่าง ๆ มากยิ่งขึ้น

1.2 ในกรณีที่พบว่ามาตรฐานค่าเชิงพฤติกรรมสามารถจำแนกความแตกต่างระหว่างกลุ่มที่ปฏิบัติงานมีประสิทธิภาพต่างกันได้อย่างคมชัดภายใต้การประเมินพฤติกรรมตนเองของครุ นั้น เป็นการบ่งชี้ถึงประโยชน์ของมาตรฐานค่ารูปแบบนี้ได้เป็นอย่างดี หากนำมาตรฐานค่ารูปแบบนี้ไปใช้ในลักษณะของการให้ครูประเมินพฤติกรรมตนเอง จะทำให้ครูทราบและยอมรับว่าขณะนี้ตนเองมีพฤติกรรมการปฏิบัติงานอยู่ในระดับใด พร้อมทั้งมีโอกาสได้ทราบด้วยว่าควรพัฒนางานหรือปฏิบัติตนอย่างไร จึงจะถือว่ามีความก้าวหน้ายิ่งขึ้น ทั้งนี้ด้วยการพิจารณาจากแนวทางการทำงานบนมาตรฐานวัด โดยนัยนี้ หน่วยงานทางการศึกษาต่าง ๆ น่าจะพัฒนามาตรฐานค่าเชิงพฤติกรรมสำหรับบุคลากรสามารถใช้ประเมินพฤติกรรมการทำงานของตนเองได้อย่างต่อเนื่องเป็นระยะ ๆ

1.3 ในการพัฒนาเครื่องมือวัดประเภทมาตรฐานค่า เพื่อให้เป็นเครื่องมือเก็บรวบรวมข้อมูลในการศึกษาวิจัยต่าง ๆ ในโอกาสต่อไป ควรจะเน้นการตรวจสอบลักษณะความคลาดเคลื่อนในการประเมินค่าคู่ควบไปกับการตรวจสอบความเที่ยงของเครื่องมือ ไม่ควรเน้นเฉพาะการตรวจสอบความเที่ยง (Reliability) เพียงอย่างเดียว

## 2. ประเด็นที่ควรทำการศึกษาวิจัยต่อเนื่อง

2.1 เนื่องจากการพัฒนามาตรฐานค่าเชิงพฤติกรรมอย่างเต็มรูปแบบ ต้องใช้ความพยายามและใช้เวลาในการพัฒนามากโดยเฉพาะในขั้นของการจัดเรียงลำดับ ตัวอย่างพฤติกรรมบนมาตรฐานวัด ดังนั้นน่าจะพัฒนามาตรฐานค่าเชิงพฤติกรรมในลักษณะของการประยุกต์รูปแบบ เช่น ใช้ตัวอย่างพฤติกรรมผูกติดไว้บนทางเลือก (Alternatives) เฉพาะในระดับสูงสุดและต่ำสุดเท่านั้น หรือระบุตัวอย่างพฤติกรรมระดับปานกลางเท่านั้น เป็นต้น แล้วเปรียบเทียบคุณภาพกับมาตรฐานค่าเชิงพฤติกรรมอย่างเต็มรูปแบบเพื่อหาข้อสรุปเกี่ยวกับคุณภาพหรือประสิทธิภาพต่อไป

2.2 ควรทำการศึกษาวิจัยเพื่อพัฒนาเกณฑ์หรือดัชนีบ่งชี้การเกิดลักษณะความคลาดเคลื่อนของผลการประเมิน ซึ่งจะ เป็นประโยชน์อย่างยิ่งสำหรับการคัดเลือกหรือกลั่นกรองแบบประเมินว่าแบบประเมินชุดใดปรากฏข้อมูลที่น่าเชื่อถือหรือไม่น่าเชื่อถือ อีกทั้งจะเป็นประโยชน์ในการแปลผลการวิจัย จะทำให้นักวิจัยสามารถแปลผลได้อย่างแม่นยำยิ่งขึ้น

2.3 ควรทำการศึกษาวิจัยเพื่ออธิบายลักษณะความคลาดเคลื่อนในการประมาณค่า ในสถานการณ์การประเมินต่าง ๆ เช่น ในกรณีที่ผู้ประเมินมีโอกาสสังเกตหรือมองเห็นพฤติกรรม ได้ไม่เท่าเทียมกัน กรณีการประเมินพฤติกรรมโดยกลุ่มเพื่อน กรณีการประเมินพฤติกรรมของครู โดยผู้เรียน เป็นต้น

2.4 ควรศึกษาผลของการใช้คู่ประเมินเปรียบเทียบกับคุณภาพของการประมาณค่า กล่าวคือ ตรวจสอบว่า ในกรณีที่กลุ่มผู้ประเมินได้รับทราบว่ามีการใช้ข้อมูลจากแหล่งอื่น หรือมี ผู้ประเมินคนอื่นกำลังทำการประเมินเป้าหมายเดียวกับตน คู่ควบไปด้วย จะมีผลต่อความตั้งใจใน การตอบหรือส่งผลต่อความน่าเชื่อถือของผลการประเมินหรือไม่