



บทที่ 4

ผลการวิเคราะห์ข้อมูล

ในบทนี้ ได้นำเสนอผลการวิเคราะห์ข้อมูลเพื่อตรวจสอบสมมติฐานการวิจัย เรียงตามลำดับดังนี้

1. ผลการวิเคราะห์ข้อมูลเพื่อตรวจสอบความตรงในการประมาณค่า
2. ผลการวิเคราะห์ข้อมูลเพื่อตรวจสอบความเที่ยงในการประมาณค่า
3. ผลการวิเคราะห์ข้อมูลเพื่อตรวจสอบลักษณะความคลาดเคลื่อนในการประมาณค่า

โดยในการนำเสนอผลการวิเคราะห์ข้อมูลแต่ละรายการ จะปรากฏสัญลักษณ์ หรือคำย่อที่ใช้ทั่วไป ดังต่อไปนี้

| | | | |
|-----------|-------|---------|---|
| \bar{X} | | หมายถึง | ค่าเฉลี่ยเลขคณิต (Arithmetic Mean) |
| S.D | | หมายถึง | ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) |
| n | | หมายถึง | จำนวนตัวอย่างผู้ให้ข้อมูล หรือผู้ประเมิน |
| BARS | | หมายถึง | มาตรประมาณค่าเชิงพฤติกรรม (Behaviorally-Anchored Rating Scales) |
| GRS | | หมายถึง | มาตรประมาณค่าแบบกราฟิก (Graphic Rating Scales) |

** $p < .01$

หรือ

* $p < .05$

..... หมายถึง ในการทดสอบความแตกต่างในประชากรในครั้งนั้น ๆ ได้พบความแตกต่าง หรือความสัมพันธ์ระหว่างตัวแปร มีนัยสำคัญทางสถิติที่ระดับ .01 หรือ .05 ตามลำดับ

ในการทดสอบความแตกต่างของผลการประเมิน ระหว่างเงื่อนไขการใช้มาตรวัดต่าง
รูปแบบกัน ด้วย The Wilcoxon Signed Ranks Test

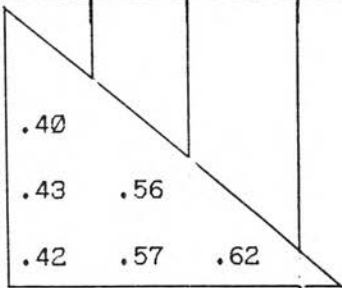
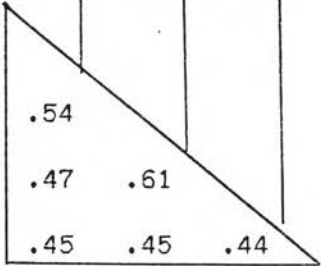
| | | | |
|----------------|-------|---------|---|
| N | | หมายถึง | จำนวนคู่ของรายการเปรียบเทียบ |
| T^+ | | หมายถึง | ผลรวมของลำดับที่ (Ranks) ที่ปรากฏค่าบวก ซึ่งเกิดขึ้นใน กรณีที่ค่าสังเกตภายใต้เงื่อนไขหลักมีค่ามากกว่าค่าสังเกตภายใต้ อีกเงื่อนไขหนึ่ง |
| μ_{T^+} | | หมายถึง | ค่าเฉลี่ยที่คาดหวังของลำดับที่มีเครื่องหมายบวก คำนวณโดยสูตร |
| | | | $\mu_{T^+} = \frac{N(N+1)}{4}$ |
| σ_{T^+} | | หมายถึง | ความคลาดเคลื่อนมาตรฐานของการเกิดลำดับที่มี เครื่องหมายบวก คำนวณโดยสูตร |
| | | | $\sigma_{T^+}^2 = \frac{N(N+1)(2N+1)}{24}$ |
| z | | หมายถึง | ค่าจำนวน Z โดยสูตร $z = \frac{T^+ - \mu_{T^+}}{\sigma_{T^+}}$ |

สัญลักษณ์อื่นใดที่นอกเหนือจากนี้ ได้อธิบายไว้แล้วในขณะนำเสนอข้อมูลแต่ละตอน
ต่อไปนี้เป็นรายละเอียดผลการวิเคราะห์ข้อมูลเพื่อตรวจสอบสมมติฐานการวิจัยแต่ละรายการ

1. ผลการวิเคราะห์ข้อมูลเพื่อตรวจสอบความตรงในการประมาณค่า

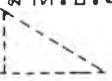
ผลการวิเคราะห์ข้อมูลในส่วนนี้ประกอบด้วย (1) ผลการวิเคราะห์เมตริกซ์สหสัมพันธ์
แบบ Multitrait - Multimethod ตามหลักการของ Campbell และ Fiske (1959)
เพื่อตรวจสอบ Convergent & Discriminant Validity ของมาตรประมาณค่าทั้งสอง
รูปแบบ ปรากฏดังตารางที่ 5-6 และ (2) ผลการเปรียบเทียบคะแนนผลการประเมินพฤติกรรม
ตนเองของครูกลุ่มที่ปฏิบัติงานเป็นระบบ กับกลุ่มที่ปฏิบัติงานยังไม่เป็นระบบ และการเปรียบเทียบ
น้ำหนักเชิงจำแนกความแตกต่างของกลุ่ม ระหว่าง กรณีประเมินโดยใช้มาตรประมาณค่าเชิงพฤติกรรม กับ
แบบกราฟิก เพื่อตรวจสอบความตรงของมาตรประมาณค่าในการจำแนกความแตกต่างระหว่างกลุ่มครู
ที่ปฏิบัติงานมีประสิทธิภาพต่างกัน ปรากฏผลดังตารางที่ 7-9

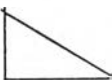
ตารางที่ 5 ค่าสหสัมพันธ์ระหว่างกันของรายการพฤติกรรม 4 ด้าน ภายใต้การใช้มาตรวัด 2 รูปแบบ เป็นเครื่องมือประเมิน (สถานการณ์การประเมินตนเองของครูผู้สอน)

| | | มาตร ประเมินค่าเชิงพฤติกรรม | | | | มาตรประเมินค่าแบบกราฟิก | | | | | |
|-------------------|---|---|-------|-------|-----|--|---|---|---|-----|-----|
| | | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | | |
| มาตร เชิงพฤติกรรม | 1 |  | | | | | | | | | |
| | 2 | | | | | | | | | .40 | |
| | 3 | | | | | | | | | .43 | .56 |
| | 4 | | | | | | | | | .42 | .57 |
| มาตรแบบกราฟิก | 1 | .29** | .16 | .11 | .07 |  | | | | | |
| | 2 | .22 | .45** | .28 | .37 | | | | | | |
| | 3 | .27 | .46 | .33** | .32 | | | | | | |
| | 4 | .21 | .23 | .30 | .37 | | | | | | |

หมายเหตุ 1. n = 84 คน **p<.01

2. ตัวเลขที่มีวงกลมล้อมรอบ คือ ค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินโดยมาตรประเมินค่าต่างรูปแบบที่มุ่งประเมินพฤติกรรมด้านเดียวกัน

3.  ค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินพฤติกรรมต่างด้านด้วยมาตรประเมินค่าต่างรูปแบบ

4.  ค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินพฤติกรรมต่างด้านด้วยมาตรประเมินค่ารูปแบบเดียวกัน

จากตารางที่ 5 ในสถานการณ์การประเมินพฤติกรรมตนเองของครูผู้สอนพบว่า


(1) ค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินโดยมาตรประมาณค่าต่างรูปแบบที่มุ่งประเมินพฤติกรรมด้านเดียวกัน หรือค่าสหสัมพันธ์ในแนวทแยงที่มีวงกลมล้อมรอบมีค่าสูง แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทั้ง 4 รายการพฤติกรรม ซึ่งแสดงถึงการเกิด Convergent Validity ของมาตรประมาณค่าทั้งสองรูปแบบ (2) เมื่อเปรียบเทียบค่าสหสัมพันธ์ในแนวทแยงกับค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินพฤติกรรมที่แตกต่างกันด้วยมาตรประมาณค่าต่างรูปแบบ (Heterodimensional-Heteromethod Correlations) โดยค่าสหสัมพันธ์ในแนวทแยง 1 ค่า ได้ถูกเปรียบเทียบกับค่าอื่น 6 ค่า ปรากฏว่าค่าสหสัมพันธ์ในแนวทแยงมีค่ามากกว่าจำนวน 22 ใน 24 คู่เปรียบเทียบ ซึ่งแสดงถึงการเกิด Discriminant Validity ระหว่างพฤติกรรมด้านต่าง ๆ ที่มุ่งประเมิน (3) เมื่อเปรียบเทียบค่าสหสัมพันธ์ในแนวทแยงกับสหสัมพันธ์ระหว่างคะแนนผลการประเมินพฤติกรรมต่างด้านด้วยมาตรประมาณค่ารูปแบบเดียวกัน (Heterodimensional-Monomethod Correlation) โดยค่าสหสัมพันธ์ในแนวทแยง 1 ค่า ได้ถูกเปรียบเทียบกับค่าในแนวแถวหรือสดมภ์ที่สมนัยกัน จำนวน 3 ค่า ปรากฏว่าค่าสหสัมพันธ์ในแนวทแยงมีค่ามากกว่าเพียง 1 ใน 12 คู่เปรียบเทียบ ทั้งกรณีใช้มาตรประมาณค่าเชิงพฤติกรรมเป็นเครื่องมือประเมินและกรณีใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมิน ซึ่งแสดงถึงการขาดความตรงเชิงจำแนกความแตกต่างระหว่างตัวแปรภายใต้การใช้มาตรประมาณค่าแต่ละรูปแบบ หรือผลการประเมินพฤติกรรมด้านต่าง ๆ ไม่เป็นอิสระจากกัน และ (4) เมื่อเปรียบเทียบค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินพฤติกรรมต่างด้านด้วยมาตรประมาณค่ารูปแบบเดียวกัน (Heterodimension-Monomethod Correlation) ระหว่างกรณีใช้มาตรประมาณค่าเชิงพฤติกรรมเป็นเครื่องมือประเมินกับกรณีใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมิน โดยเปรียบเทียบเป็นรายคู่พฤติกรรมจำนวน 6 คู่เปรียบเทียบ ปรากฏว่าค่าสหสัมพันธ์ภายใต้การใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมินให้ค่าสหสัมพันธ์สูงกว่าจำนวน 4 ใน 6 คู่เปรียบเทียบ ซึ่งจากการทดสอบด้วย The Wilcoxon Signed Ranks Test ไม่พบความแตกต่างมีนัยสำคัญทางสถิติ จึงสรุปได้ว่า กรณีใช้มาตรประมาณค่าเชิงพฤติกรรมกับกรณีใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมินให้ค่าสหสัมพันธ์ระหว่างตัวแปรที่มุ่งประเมินเทียบเท่ากัน หรืออีกนัยหนึ่งก็คือมาตรประมาณค่าทั้งสองรูปแบบให้ค่าน้ำหนักเชิงจำแนกความแตกต่างระหว่างตัวแปรได้เท่าเทียมกัน


ตารางที่ 6 ค่าสหสัมพันธ์ระหว่างกันของรายการพฤติกรรม 4 ด้าน ภายใต้การใช้
มาตรวัด 2 รูปแบบเป็นเครื่องมือประเมิน (สถานการณ์การประเมินพฤติกรรมของ
ครูผู้สอนโดยหัวหน้าหมวดวิชา)

| | | มาตราประมาณค่าเชิงพฤติกรรม | | | | มาตราประมาณค่าแบบกราฟิก | | | |
|-------------------|---|----------------------------|-------|------|-------|-------------------------|-----|-----|---|
| | | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| มาตราเชิงพฤติกรรม | 1 | | | | | | | | |
| | 2 | .30 | | | | | | | |
| | 3 | .41 | .34 | | | | | | |
| | 4 | .47 | .47 | .60 | | | | | |
| มาตรแบบกราฟิก | 1 | .52** | .39 | .29 | .38 | | | | |
| | 2 | .25 | .53** | .19 | .36 | .56 | | | |
| | 3 | .39 | .52 | .23* | .36 | .65 | .63 | | |
| | 4 | .36 | .38 | .44 | .60** | .49 | .60 | .54 | |

หมายเหตุ 1. n = 84 คน **p<.01, *p<.05

2. ตัวเลขที่มีวงกลมล้อมรอบ คือ ค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินโดย
มาตรประมาณค่าต่างรูปแบบที่มุ่งประเมินพฤติกรรมด้านเดียวกัน

3.  ค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินพฤติกรรมต่างด้าน
ด้วยมาตรประมาณค่าต่างรูปแบบ

4.  ค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินพฤติกรรมต่างด้านด้วยมาตร
ประมาณค่ารูปแบบเดียวกัน

จากตารางที่ 6 ในสถานการณ์การประเมินพฤติกรรมของครูโดยหัวหน้าหมวดวิชา พบว่า

(1) ค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินโดยมาตรฐานค่าต่างรูปแบบที่มุ่งประเมินพฤติกรรมด้านเดียวกัน มีน้ำหนักสูง แตกต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับ .01 หรือ .05 ทั้ง 4 รายการพฤติกรรม ซึ่งแสดงถึงการเกิด Convergent Validity ของมาตรฐานค่าสองรูปแบบ (2) เมื่อเปรียบเทียบค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินโดยมาตรฐานค่าต่างรูปแบบที่มุ่งประเมินพฤติกรรมด้านเดียวกัน หรือค่าสหสัมพันธ์ในแนวทแยงกับค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินพฤติกรรมต่างด้านด้วยมาตรฐานค่าต่างรูปแบบ ปรากฏว่า ค่าสหสัมพันธ์ในแนวทแยงมีค่ามากกว่าจำนวน 19 ใน 24 คู่เปรียบเทียบซึ่งแสดงถึงการเกิด Discrimination Validity ระหว่างรายการพฤติกรรมที่มุ่งประเมิน (3) เมื่อเปรียบเทียบค่าสหสัมพันธ์ระหว่างคะแนนในแนวทแยง กับ ค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินพฤติกรรมต่างด้านด้วยมาตรฐานค่ารูปแบบเดียวกัน พบว่า ในกรณีของการประเมินโดยมาตรฐานค่าเชิงพฤติกรรม ปรากฏว่าค่าในแนวทแยงมีค่ามากกว่าจำนวน 9 ใน 12 คู่เปรียบเทียบ ส่วนกรณีการประเมินโดยมาตรฐานค่าแบบกราฟิก 4 ใน 12 คู่เปรียบเทียบปรากฏค่าสหสัมพันธ์ในแนวทแยงมีค่ามากกว่า ซึ่งภายใต้สัดส่วนนี้ได้แสดงถึงความแตกต่างในคุณภาพของมาตรฐานค่าสองรูปแบบ ในการจำแนกความแตกต่างระหว่างรายการพฤติกรรมที่มุ่งประเมิน และ (4) เมื่อเปรียบเทียบค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินพฤติกรรมต่างด้านด้วยมาตรฐานค่ารูปแบบเดียวกัน ระหว่างกรณีประเมินโดยใช้มาตรฐานค่าเชิงพฤติกรรมกับกรณีประเมินโดยใช้มาตรฐานค่าแบบกราฟิก ปรากฏว่า ค่าสหสัมพันธ์ภายใต้การใช้มาตรฐานค่าแบบกราฟิกเป็นเครื่องมือประเมิน 5 ใน 6 คู่เปรียบเทียบมีค่ามากกว่า และจากการทดสอบด้วย The Wilcoxon Signed Ranks Test ได้พบความแตกต่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งสรุปได้ว่า กรณีประเมินโดยมาตรฐานค่าเชิงพฤติกรรมให้ค่าสหสัมพันธ์ระหว่างด้านพฤติกรรมต่ำกว่ากรณีประเมินโดยมาตรฐานค่าแบบกราฟิกอันเป็นสภาพของการให้ผลการประเมินระหว่างด้านพฤติกรรมที่เป็นอิสระจากกันมากกว่า หรืออีกนัยหนึ่งก็คือ ให้ค่าความตรงเชิงจำแนกความแตกต่างระหว่างด้านได้อย่างคมชัดกว่า

โดยสรุป จากการวิเคราะห์ความตรงตามภาวะสัมพันธฐานด้วยวิธีการประยุกต์ แนวทาง การอธิบายเมตริกซ์สหสัมพันธ์แบบ Multitrait-Multimethod พบว่า ในสถานการณ์การประเมิน พฤติกรรมตนเองของครูผู้สอน ผลการวิเคราะห์ข้อมูลไม่เป็นไปตามภาวะสัมพันธฐาน ข้อมูลจากส่วน นี้ไม่สามารถสรุปได้ว่ามาตรฐานค่ารูปแบบใดสะท้อนถึงคุณภาพในด้านความตรงเชิงจำแนกความ แตกต่างระหว่างรายการที่มุ่งประเมินได้ดีกว่ากัน ส่วนในสถานการณ์การประเมินพฤติกรรมของครู โดยหัวหน้าหมวดวิชา ได้ปรากฏผลการวิเคราะห์ข้อมูลที่เป็นไปตามภาวะสัมพันธฐาน และปรากฏ หลักฐานยืนยันว่า กรณีใช้มาตรฐานค่าเชิงพฤติกรรมเป็นเครื่องมือประเมิน ได้ให้ผลการ ประเมินระหว่างด้านพฤติกรรมที่เป็นอิสระจากกัน หรือมีความตรงเชิงจำแนกความแตกต่างระหว่าง พฤติกรรมด้านต่าง ๆ ที่มุ่งประเมินได้อย่างคมชัดกว่ากรณีประเมินโดยใช้มาตรฐานค่าแบบกราฟิก

ตารางที่ 7 ผลการประเมินตนเองของครูผู้สอนกลุ่มปฏิบัติงานเป็นระบบ และกลุ่มปฏิบัติงาน
ยังไม่เป็นระบบ ภายใต้การใช้มาตราประมาณค่าเชิงพฤติกรรมเป็นเครื่องมือประเมิน

| รายการพฤติกรรม | | กลุ่มปฏิบัติงาน เป็นระบบ | | กลุ่มปฏิบัติงาน ไม่เป็นระบบ | | น้ำหนักเชิง จำแนกความ แตกต่าง (r_{1b}) |
|---|----|-----------------------------|------|--------------------------------|------|---|
| | | \bar{x} | S.D | \bar{x} | S.D | |
| พฤติกรรมด้าน การวางแผนการวัดผล ประเมินผล | 1 | 4.30 | 0.70 | 3.37 | 1.22 | 0.59** |
| | 2 | 4.57 | 0.73 | 3.63 | 1.37 | 0.54** |
| | 3 | 2.97 | 0.81 | 2.23 | 0.90 | 0.54** |
| | 4 | 3.97 | 1.16 | 2.83 | 1.23 | 0.60** |
| | 5 | 4.03 | 1.03 | 3.13 | 1.14 | 0.52** |
| พฤติกรรมการวัดผล ประเมินผลในชั้นสอน | 6 | 3.20 | 1.24 | 2.83 | 1.10 | 0.19 |
| | 7 | 4.37 | 0.93 | 3.50 | 1.36 | 0.47** |
| | 8 | 4.27 | 0.83 | 3.23 | 1.35 | 0.59** |
| | 9 | 3.37 | 0.99 | 2.57 | 1.13 | 0.48** |
| พฤติกรรมการวัดผล ประเมินผลขั้นสรุปผลการเรียน | 10 | 3.50 | 1.07 | 2.97 | 1.09 | 0.31 [†] |
| | 11 | 4.63 | 0.62 | 3.63 | 1.19 | 0.66** |
| | 12 | 3.50 | 1.00 | 2.40 | 0.93 | 0.71** |
| | 13 | 3.90 | 1.00 | 2.97 | 1.10 | 0.56** |
| พฤติกรรมการตรวจสอบ และนัฒนาระบบวัดผล | 14 | 3.73 | 1.17 | 2.33 | 1.32 | 0.71** |
| | 15 | 3.67 | 0.96 | 2.90 | 1.21 | 0.45** |
| | 16 | 3.26 | 1.26 | 2.03 | 0.93 | 0.70** |
| | 17 | 2.70 | 1.26 | 1.73 | 0.93 | 0.54** |
| | 18 | 3.40 | 0.97 | 2.57 | 0.73 | 0.61** |
| รวม | | 3.74 | 0.99 | 2.83 | 1.12 | |

หมายเหตุ n = 84 คน

** $p < .01$

* $p < .05$

r_{1b} คือ สัมประสิทธิ์สหสัมพันธ์แบบไบซีเรียล (Biserial Coefficient of Correlation) กรณีประเมินโดยใช้มาตราประมาณค่าเชิงพฤติกรรม

จากตารางที่ 7 พบว่า ค่าเฉลี่ยเลขคณิตของผลการประเมินพฤติกรรมตนเองของครูผู้สอนกลุ่มที่ปฏิบัติงานเป็นระบบ สูงกว่ากลุ่มที่ปฏิบัติงานยังไม่เป็นระบบ โดยมีน้ำหนักเชิงจำแนกสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 16 ใน 18 รายการพฤติกรรม ยกเว้นรายการพฤติกรรมที่ 6 ซึ่งเป็นเรื่องที่เกี่ยวข้องกับการปรับแก้ทักษะพื้นฐานของนักเรียนกลุ่มอ่อน ก่อนที่จะสอนเนื้อหาใหม่ และพฤติกรรมที่ 10 เกี่ยวกับการดำเนินการพัฒนาเครื่องมือวัดผลปลายภาคเรียน ที่ปรากฏค่าเฉลี่ยเลขคณิตของผลการประเมินในระดับปานกลางทั้งสองกลุ่มทำให้ความแตกต่างระหว่างกลุ่มไม่ปรากฏชัดเจน และเมื่อคำนวณน้ำหนักเชิงจำแนกด้วยสัมประสิทธิ์สหสัมพันธ์แบบไบซีเรียลพบว่า ได้ค่าน้ำหนักเชิงจำแนกต่ำ

ตารางที่ 8 ผลการประเมินตนเองของครูผู้สอนกลุ่มปฏิบัติงานเป็นระบบ และกลุ่มปฏิบัติงานยังไม่เป็นระบบ ภายใต้การใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมิน

| รายการพฤติกรรม | | กลุ่มปฏิบัติงานเป็นระบบ | | กลุ่มปฏิบัติงานไม่เป็นระบบ | | น้ำหนักเชิงจำแนกความแตกต่าง (r_{rs}) |
|--|----|-------------------------|------|----------------------------|------|--|
| | | \bar{x} | S.D | \bar{x} | S.D | |
| พฤติกรรมด้านการวางแผนการวัดผล-ประเมินผล | 1 | 3.20 | 0.89 | 3.14 | 0.74 | 0.05 |
| | 2 | 3.90 | 0.71 | 3.79 | 0.86 | 0.09 |
| | 3 | 3.06 | 0.98 | 2.79 | 0.73 | 0.20 |
| | 4 | 3.70 | 0.99 | 3.52 | 0.99 | 0.11 |
| | 5 | 4.40 | 0.72 | 4.38 | 0.77 | 0.02 |
| พฤติกรรมการวัดผล-ประเมินผลในชั้นสอนหรือระหว่างดำเนินการสอน | 6 | 3.20 | 0.76 | 2.89 | 0.90 | 0.23 |
| | 7 | 3.93 | 0.64 | 3.79 | 0.94 | 0.11 |
| | 8 | 4.23 | 0.73 | 3.69 | 0.81 | 0.44** |
| | 9 | 3.20 | 0.71 | 2.97 | 0.91 | 0.18 |
| พฤติกรรมการวัดผลประเมินผลขั้นสรุปผลการเรียน | 10 | 3.87 | 0.73 | 3.59 | 0.95 | 0.21 |
| | 11 | 3.90 | 0.92 | 3.69 | 1.04 | 0.13 |
| | 12 | 3.57 | 0.86 | 2.97 | 0.94 | 0.42** |
| | 13 | 4.10 | 0.85 | 3.62 | 0.82 | 0.36** |
| พฤติกรรมการตรวจสอบและพัฒนาระบบวัดผล | 14 | 3.47 | 0.86 | 2.55 | 1.02 | 0.61** |
| | 15 | 3.77 | 0.68 | 3.21 | 0.99 | 0.41** |
| | 16 | 2.93 | 0.87 | 2.52 | 1.15 | 0.26* |
| | 17 | 3.17 | 0.87 | 2.79 | 1.18 | 0.23 |
| | 18 | 3.80 | 0.81 | 3.65 | 0.90 | 0.10 |
| รวม | | 3.63 | 0.81 | 3.31 | 0.92 | |

หมายเหตุ ** $p < .01$

* $p < .05$

r_{rs} คือ สัมประสิทธิ์สหสัมพันธ์แบบไบซีเรียล (Biserial Coefficient of Correlation) กรณีประเมินโดยใช้มาตรประมาณค่าแบบกราฟิก

จากตารางที่ 8 พบว่า คะแนนเฉลี่ยของผลการประเมินตนเองของครูผู้สอนภายใต้การใช้มาตราประมาค่าแบบกราฟิกเป็นเครื่องมือประเมินทั้งกลุ่มครูที่ปฏิบัติงานเป็นระบบและกลุ่มที่ปฏิบัติงานยังไม่เป็นระบบ ปรากฏคะแนนในระดับปานกลางหรือมากเกือบทุกรายการ การกระจายของคะแนนเฉลี่ยกลุ่มปฏิบัติงานเป็นระบบมีค่าระหว่าง 2.93 ถึง 4.40 คะแนน โดยมีค่าเฉลี่ยรวมเท่ากับ 3.63 คะแนน ส่วนการกระจายของค่าเฉลี่ยของกลุ่มปฏิบัติงานยังไม่เป็นระบบ มีค่าเฉลี่ยรวมเท่ากับ 3.31 คะแนน เมื่อวิเคราะห์น้ำหนักเชิงจำแนกความแตกต่างระหว่างกลุ่มด้วยสัมประสิทธิ์สหสัมพันธ์แบบไบซีเรียล พบว่า มีรายการพฤติกรรมที่ปรากฏน้ำหนักเชิงจำแนกระดับสูงหรือมีค่าสหสัมพันธ์ระดับสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 หรือ .05 จำนวน 6 ใน 18 รายการพฤติกรรม

ตารางที่ 9 เปรียบเทียบน้ำหนักเชิงจำแนกความแตกต่างระหว่างกลุ่มครูที่ปฏิบัติงานมี
ประสิทธิภาพต่างกัน ระหว่าง กรณีประเมินใช้มาตรฐานค่าเชิงพฤติกรรม
กับ กรณีประเมินโดยใช้มาตรฐานค่าแบบกราฟิก ในสถานการณ์การ
ประเมินตนเองของครูผู้สอน

| รายการพฤติกรรม ลำดับที่ | น้ำหนักเชิง จำแนกโดย BARS (r_{1b}) | น้ำหนักเชิง จำแนกโดย GRS (r_{1g}) | ความแตกต่าง ($r_{1b} - r_{1g}$) | ลำดับที่ของ ความ แตกต่าง | ค่าคำนวณในการ ทดสอบความแตกต่าง ระหว่างมาตรวัด |
|----------------------------|--|---|--------------------------------------|--------------------------------|---|
| 1 | 0.59 | 0.05 | 0.54 | 18 | $N = 18$ $\mu_T = 85.5$ $T^+ = 169.5$ $\sigma_{T^+} = 22.96$ $Z = 3.65^*$ |
| 2 | 0.54 | 0.09 | 0.45 | 13 | |
| 3 | 0.54 | 0.20 | 0.34 | 10 | |
| 4 | 0.60 | 0.11 | 0.49 | 14 | |
| 5 | 0.52 | 0.02 | 0.50 | 15 | |
| 6 | 0.19 | 0.23 | -0.04 | -1.5 | |
| 7 | 0.47 | 0.11 | 0.36 | 11 | |
| 8 | 0.59 | 0.44 | 0.15 | 5 | |
| 9 | 0.48 | 0.18 | 0.30 | 8 | |
| 10 | 0.31 | 0.21 | 0.10 | 3.5 | |
| 11 | 0.66 | 0.13 | 0.53 | 17 | |
| 12 | 0.71 | 0.42 | 0.29 | 7 | |
| 13 | 0.56 | 0.36 | 0.20 | 6 | |
| 14 | 0.71 | 0.61 | 0.10 | 3.5 | |
| 15 | 0.45 | 0.41 | 0.04 | 1.5 | |
| 16 | 0.70 | 0.26 | 0.44 | 12 | |
| 17 | 0.54 | 0.23 | 0.31 | 9 | |
| 18 | 0.61 | 0.10 | 0.51 | 16 | |

* $p < .05$

จากตารางที่ 9 เมื่อเปรียบเทียบน้ำหนักเชิงจำแนกหรือดัชนีบ่งชี้ความแตกต่างของกลุ่ม
ระหว่างกรณีใช้มาตรฐานค่าเชิงพฤติกรรมเป็นเครื่องมือประเมิน กับกรณีใช้มาตรฐานค่า
แบบกราฟิก ทีละรายการพฤติกรรม พบว่า ผลการประเมินโดยใช้มาตรฐานค่าเชิงพฤติกรรม
ปรากฏน้ำหนักเชิงจำแนกสูงกว่า จำนวน 17 ใน 18 รายการพฤติกรรม และเมื่อทำการทดสอบ
ความแตกต่างระหว่างกรณีใช้มาตรวัดต่างรูปแบบกัน ด้วย The Wilcoxon Signed Ranks
Test ปรากฏว่าได้ค่าคำนวณ $Z = 3.65$ ซึ่งสรุปได้ว่า กรณีใช้มาตรฐานค่าเชิงพฤติกรรม

เป็นเครื่องมือประเมินให้ค่าน้ำหนักเชิงจำแนกความแตกต่างระหว่างกลุ่ม สูงกว่ากรณีใช้มาตรฐาน
 ประมวลค่าแบบกราฟิกเป็นเครื่องมือประเมินอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($Z_{.05}$
 กรณีทดสอบหางเดียวมีค่าเท่ากับ 1.64)

โดยสรุป จากการตรวจสอบความตรงของเครื่องมือในการจำแนกความแตกต่างระหว่าง
 กลุ่มครู ที่ปฏิบัติงานมีประสิทธิภาพต่างกัน พบว่า ในสถานการณ์ของการประเมินตนเองของครูผู้สอน
 การประเมินโดยมาตรฐานประมวลค่าเชิงพฤติกรรมสามารถจำแนกความแตกต่างระหว่างกลุ่มครูที่ปฏิบัติ
 งานเป็นระบบกับกลุ่มครูที่ปฏิบัติงานยังไม่เป็นระบบได้ชัดเจนกว่ากรณีใช้มาตรฐานประมวลค่ากราฟิก
 เป็นเครื่องมือประเมิน

2. ผลการวิเคราะห์ข้อมูลเพื่อตรวจสอบความเที่ยงในการประมวลค่า

ผลการวิเคราะห์ข้อมูลส่วนนี้ เป็นผลการเปรียบเทียบค่าความเที่ยงของผู้ประเมิน
 (Rater Reliability) ระหว่างกรณีใช้มาตรฐานประมวลค่าเชิงพฤติกรรม กับ กรณีใช้มาตรฐาน
 ประมวลค่าแบบกราฟิกเป็นเครื่องมือประเมิน โดยที่ดัชนีบ่งชี้ความเที่ยงในการประเมินพฤติกรรม
 แต่ละรายการในที่นี้คือ ค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินของผู้ประเมิน 2 กลุ่ม
 (Intraclass Correlation Coefficient) ผลการเปรียบเทียบ ปรากฏดังตารางที่ 10

ตารางที่ 10 ค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินตนเองของครูผู้สอน กับคะแนนผลการประเมินโดยหัวหน้าหมวดวิชา เปรียบเทียบระหว่างกรณีประเมินโดยใช้มาตรฐานค่าเชิงพฤติกรรม กับ ประเมินโดยใช้มาตรฐานค่าแบบกราฟิก

| รายการพฤติกรรมลำดับที่ | ค่าสหสัมพันธ์ของคะแนนผลการประเมิน | | ความแตกต่างของค่าสหสัมพันธ์ $ICC_{\text{B}} - ICC_{\text{G}}$ | ลำดับที่ของความแตกต่าง | ค่าคำนวณในการเปรียบเทียบความแตกต่างระหว่างมาตรฐานวัด |
|------------------------|--------------------------------------|-------------------------------------|--|------------------------|---|
| | กรณีใช้ BARS (ICC_{B}) | กรณีใช้ GRS (ICC_{G}) | | | |
| 1 | 0.45** | 0.20* | 0.25 | 9 | $T^* = 170$ $\mu_{T^*} = 85.5$ $\sigma_{T^*} = 22.96$ $z = \frac{T^* - \mu_{T^*}}{\sigma_{T^*}}$ $= 3.68^*$ |
| 2 | 0.44** | 0.36** | 0.08 | 3 | |
| 3 | 0.46** | 0.20** | 0.26 | 10 | |
| 4 | 0.39** | 0.20** | 0.19 | 7 | |
| 5 | 0.46** | 0.08 | 0.38 | 15 | |
| 6 | 0.48** | 0.24** | 0.24 | 8 | |
| 7 | 0.49** | 0.12 | 0.37 | 14 | |
| 8 | 0.30** | 0.23* | 0.07 | 2 | |
| 9 | 0.22** | 0.26* | -0.04 | -1 | |
| 10 | 0.49** | 0.20* | 0.29 | 11.5 | |
| 11 | 0.49** | 0.20* | 0.29 | 11.5 | |
| 12 | 0.31** | 0.22* | 0.09 | 4 | |
| 13 | 0.51** | 0.06 | 0.45 | 18 | |
| 14 | 0.71** | 0.32** | 0.39 | 16 | |
| 15 | 0.42** | 0.24** | 0.18 | 6 | |
| 16 | 0.46** | 0.14 | 0.32 | 13 | |
| 17 | 0.16 | 0.05 | 0.11 | 5 | |
| 18 | 0.53** | 0.12 | 0.41 | 17 | |

หมายเหตุ

1. n = 84 คน

2. ** p < .01

* p < .05

3. ICC_{B} , ICC_{G} คือ Intraclass Correlation Coefficient ในกรณีประเมินโดยใช้มาตรฐานค่าเชิงพฤติกรรม และมาตรฐานค่าแบบกราฟิก ตามลำดับ

จากตารางที่ 10 พบว่า ค่าสหสัมพันธ์ระหว่างคะแนนผลการประเมินตนเองของครูผู้สอนกับคะแนนผลการประเมินโดยหัวหน้าหมวดวิชา ภายใต้การใช้มาตรฐานค่าเชิงพฤติกรรมเป็นเครื่องมือประเมิน มีค่าสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 หรือ .05 เกือบทุกรายการพฤติกรรม ยกเว้นรายการพฤติกรรมที่ 17 เกี่ยวกับการเปิดโอกาสให้ผู้เรียนแสดงความคิดเห็นเกี่ยวกับระบบการวัดผล-ประเมินผลที่ครูใช้ ในขณะที่ค่าสหสัมพันธ์ภายใต้การใช้มาตรฐานค่าแบบกราฟิกเป็นเครื่องมือประเมิน มีเพียง 12 รายการพฤติกรรมเท่านั้นที่ปรากฏว่าสหสัมพันธ์สูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 หรือ .05 เมื่อเปรียบเทียบค่าสหสัมพันธ์หรือค่าความเที่ยงระหว่างผู้ประเมินเป็นรายการพฤติกรรม พบว่า ผลการประเมินโดยมาตรฐานค่าเชิงพฤติกรรมให้ค่าความเที่ยงสูงกว่าในทุกรายการพฤติกรรมที่มุ่งประเมิน ยกเว้นรายการพฤติกรรมที่ 9 ปรากฏว่า ผลการประเมินโดยมาตรฐานค่าแบบกราฟิกให้ค่าความเที่ยงสูงกว่า เมื่อทำการทดสอบความแตกต่างในภาพรวมระหว่างกรณีใช้มาตรวัดต่างรูปแบบกัน ด้วย The Wilcoxon Signed Ranks Test ปรากฏว่า ได้ค่าคำนวณ $Z = 3.68$ ซึ่งสรุปได้ว่า เจื่อนไขการใช้มาตรฐานค่าเชิงพฤติกรรมเป็นเครื่องมือประเมินให้ค่าความเที่ยงสูงกว่ากรณีใช้มาตรฐานค่าแบบกราฟิก อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($Z_{.05}$ กรณีทดสอบหางเดียวมีค่าเท่ากับ 1.64)

3. ผลการวิเคราะห์ข้อมูลเพื่อตรวจสอบลักษณะความคลาดเคลื่อนในการประมาณค่า

ผลการวิเคราะห์ข้อมูลในส่วนนี้ ประกอบด้วย (1) ผลการเปรียบเทียบผลการประเมินพฤติกรรมของครู ระหว่างกรณีประเมินได้โดยใช้มาตรฐานค่าต่างรูปแบบกัน เพื่อตรวจสอบ Leniency Error หรือความคลาดเคลื่อนในลักษณะของการปล่อยคะแนนหรือประเมินในระดับคะแนนสูง ๆ อย่างสม่ำเสมอ ผลการเปรียบเทียบ ปรากฏดังตารางที่ 11-12 และ (2) ผลการวิเคราะห์ข้อมูลเพื่อตรวจสอบ Halo Error ซึ่งดำเนินการโดย เปรียบเทียบคะแนนความคลาดเคลื่อน (Halo Score) ระหว่างกรณีประเมินโดยใช้มาตรฐานค่าเชิงพฤติกรรมกับกรณีประเมินโดยใช้มาตรฐานค่าแบบกราฟิก ผลปรากฏดังตารางที่ 13

ตารางที่ 11 เปรียบเทียบผลการประเมินตนเองของครูผู้สอน ระหว่างกรณีใช้มาตรฐานค่าเชิงพฤติกรรม กับ กรณีใช้มาตรฐานค่าแบบกราฟิกเป็นเครื่องมือประเมิน

| รายการพฤติกรรม ลำดับที่ | ผลการประเมิน โดยใช้ BARS | | ผลการประเมิน โดยใช้ GRS | | ความ แตกต่าง ของค่าเฉลี่ย ($\bar{x}_2 - \bar{x}_1$) | ลำดับที่ ของความ แตกต่าง | ค่าคำนวณในการ เปรียบเทียบความ แตกต่างระหว่าง มาตรฐานวัด |
|----------------------------|-----------------------------|------|----------------------------|------|--|--------------------------------|---|
| | \bar{x}_1 | S.D | \bar{x}_2 | S.D | | | |
| 1 | 3.65 | 1.14 | 3.27 | 0.93 | -0.38 | -12 | $T^+ = 137$ $\mu_{T^+} = 85.5$ $\sigma_{T^+} = 22.96$ $z = \frac{T^+ - \mu_{T^+}}{\sigma_{T^+}}$ $= 2.24^*$ |
| 2 | 4.06 | 1.09 | 3.90 | 0.79 | -0.16 | - 6 | |
| 3 | 2.43 | 0.97 | 2.95 | 0.94 | +0.52 | +14 | |
| 4 | 3.37 | 1.31 | 3.67 | 0.98 | +0.30 | + 8 | |
| 5 | 3.52 | 1.21 | 4.36 | 0.71 | +0.84 | +18 | |
| 6 | 3.08 | 1.28 | 3.05 | 0.81 | -0.03 | - 2 | |
| 7 | 3.85 | 1.30 | 3.79 | 0.80 | -0.06 | - 4 | |
| 8 | 3.74 | 1.25 | 4.00 | 0.80 | +0.26 | + 7 | |
| 9 | 3.02 | 1.12 | 3.05 | 0.88 | +0.03 | + 2 | |
| 10 | 3.08 | 1.19 | 3.66 | 0.85 | +0.58 | +16 | |
| 11 | 4.11 | 1.07 | 3.77 | 0.94 | -0.34 | -10 | |
| 12 | 2.97 | 1.14 | 3.34 | 1.03 | +0.37 | +11 | |
| 13 | 3.47 | 1.14 | 3.88 | 0.87 | +0.41 | +13 | |
| 14 | 3.00 | 1.44 | 3.03 | 1.04 | +0.03 | + 2 | |
| 15 | 3.20 | 1.14 | 3.52 | 0.89 | +0.32 | + 9 | |
| 16 | 2.67 | 1.36 | 2.74 | 1.09 | +0.07 | + 5 | |
| 17 | 2.36 | 1.30 | 2.92 | 1.16 | +0.56 | +15 | |
| 18 | 2.93 | 0.96 | 3.75 | 0.85 | +0.82 | +17 | |
| รวม | 3.27 | 1.18 | 3.48 | 0.91 | | | |

* $p < .05$

จากตารางที่ 11 พบว่า ผลการประเมินพฤติกรรมตนเองของครูผู้สอนภายใต้การใช้
 มาตรฐานค่าแบบกราฟิกเป็นเครื่องมือประเมิน จำนวน 11 ใน 18 รายการพฤติกรรม
 ปรากฏค่าเฉลี่ยเลขคณิตสูงกว่ากรณีใช้มาตรฐานค่าเชิงพฤติกรรมเป็นเครื่องมือประเมิน การ
 กระจายของคะแนนเฉลี่ยรายพฤติกรรม ปรากฏค่าระหว่าง 2.74 ถึง 4.36 และมีค่าเฉลี่ย
 เลขคณิตรวมเท่ากับ 3.48 โดยมีส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ยเท่ากับ 0.45 ส่วนการ
 กระจายของค่าเฉลี่ยเลขคณิตของผลการประเมินโดยใช้มาตรฐานค่าเชิงพฤติกรรม ปรากฏ
 ค่าเฉลี่ยระหว่าง 2.36 ถึง 4.11 มีค่าเฉลี่ยเลขคณิตรวม และส่วนเบี่ยงเบนมาตรฐานของ
 ค่าเฉลี่ยเท่ากับ 3.27 และ 0.53 ตามลำดับ เป็นที่น่าสังเกตว่าผลการประเมินรายการ
 พฤติกรรมที่ 14-18 ซึ่งเป็นพฤติกรรมด้านการพัฒนาและปรับปรุงระบบการวัดผล-ประเมินผล
 ปรากฏค่าเฉลี่ยเลขคณิตของผลการประเมินทั้งโดยมาตรฐานค่าแบบกราฟิกและมาตรฐานค่า
 เชิงพฤติกรรม ในระดับปานกลางหรือค่อนข้างน้อยเกือบทุกรายการ และภายใต้สภาพการณ์ที่ผล
 การประเมินมีค่าเฉลี่ยในระดับปานกลางหรือค่อนข้างน้อยนี้ ผลการประเมินโดยมาตรฐานค่า
 แบบกราฟิกได้ปรากฏคะแนนเฉลี่ยในระดับที่สูงกว่าอย่างสม่ำเสมอทุกรายการ เมื่อทำการทดสอบ
 ความแตกต่างในภาพรวมระหว่างกรณีใช้มาตรวัดหรือเครื่องมือประเมินต่างรูปแบบกัน ด้วย The
 Wilcoxon Signed Ranks Test ปรากฏว่าได้ค่าคำนวณ $Z = 2.24$ ซึ่งสรุปได้ว่า กรณีใช้
 มาตรฐานค่าแบบกราฟิกเป็นเครื่องมือในการประเมิน ให้ค่าคะแนนเฉลี่ยสูงกว่ากรณีใช้มาตร
 ฐานค่าเชิงพฤติกรรมเป็นเครื่องมือประเมินอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($Z_{.05}$
 กรณีทดสอบแบบหางเดียว มีค่าเท่ากับ 1.64)

ตารางที่ 12 เปรียบเทียบผลการประเมินพฤติกรรมครูโดยหัวหน้าหมวดวิชา ระหว่างกรณีใช้
 มาตรฐานค่าเชิงพฤติกรรม กับ กรณีใช้มาตรฐานค่าแบบกราฟิกเป็น
 เครื่องมือประเมิน

| รายการพฤติกรรม ลำดับที่ | ผลการประเมิน โดยใช้ BARS | | ผลการประเมิน โดยใช้ GRS | | ความ แตกต่าง ของค่าเฉลี่ย ($\bar{X}_2 - \bar{X}_1$) | ลำดับที่ ของความ แตกต่าง | ค่าคำนวณในการ เปรียบเทียบความ แตกต่างระหว่าง มาตรวัด |
|----------------------------|-----------------------------|------|----------------------------|------|--|--------------------------------|---|
| | \bar{X}_1 | S.D | \bar{X}_2 | S.D | | | |
| 1 | 3.67 | 0.80 | 3.14 | 0.74 | -0.50 | -13.5 | $T^+ = 138.5$ $\mu_{T^+} = 85.5$ $\sigma_{T^+} = 22.96$ $z = \frac{T^+ - \mu_{T^+}}{\sigma_{T^+}}$ $= 2.30^*$ |
| 2 | 4.06 | 1.01 | 3.79 | 0.74 | -0.27 | - 7 | |
| 3 | 2.46 | 1.23 | 2.74 | 0.76 | 0.28 | + 8 | |
| 4 | 3.20 | 1.26 | 3.63 | 1.02 | 0.43 | +10 | |
| 5 | 3.44 | 1.25 | 4.25 | 0.84 | 0.81 | +18 | |
| 6 | 2.83 | 1.10 | 2.81 | 0.86 | -0.02 | - 1 | |
| 7 | 3.79 | 1.33 | 3.59 | 0.82 | -0.20 | - 5 | |
| 8 | 3.78 | 1.02 | 3.85 | 0.97 | 0.07 | + 3 | |
| 9 | 2.81 | 1.12 | 3.29 | 0.97 | 0.48 | +12 | |
| 10 | 3.17 | 1.20 | 3.56 | 0.82 | 0.39 | + 9 | |
| 11 | 4.08 | 1.11 | 3.84 | 0.75 | -0.24 | - 6 | |
| 12 | 2.89 | 1.06 | 3.42 | 0.89 | 0.53 | +15 | |
| 13 | 3.47 | 1.13 | 4.03 | 0.76 | 0.56 | +16 | |
| 14 | 2.90 | 1.38 | 2.95 | 1.02 | 0.05 | + 2 | |
| 15 | 2.92 | 1.14 | 3.38 | 0.71 | 0.46 | +11 | |
| 16 | 2.79 | 1.32 | 2.92 | 0.85 | 0.13 | + 4 | |
| 17 | 2.32 | 1.21 | 2.82 | 0.99 | 0.50 | +13.5 | |
| 18 | 3.08 | 1.01 | 3.65 | 0.86 | 0.57 | +17 | |
| รวม | 3.20 | 1.15 | 3.42 | 0.85 | | | |

* $p < .05$

จากตารางที่ 12 พบว่า ผลการประเมินพฤติกรรมการสอนของครูโดยหัวหน้าหมวดวิชา ภายใต้การใช้มาตรฐานค่าแบบกราฟิกเป็นเครื่องมือประเมิน จำนวน 13 ใน 18 รายการ ปรากฏค่าเฉลี่ยเลขคณิตสูงกว่ากรณีใช้มาตรฐานค่าเชิงพฤติกรรมเป็นเครื่องมือประเมิน การกระจายของค่าเฉลี่ยเลขคณิตกรณีประเมินโดยมาตรฐานค่าแบบกราฟิกมีค่าระหว่าง 2.74 ถึง 4.25 มีค่าเฉลี่ยเลขคณิตรวม และส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ยเท่ากับ 3.42 และ 0.45 ตามลำดับ ส่วนการกระจายของคะแนนเฉลี่ยผลการประเมินกรณีใช้มาตรฐานค่าเชิงพฤติกรรมมีค่าระหว่าง 2.33 ถึง 4.06 มีค่าเฉลี่ยเลขคณิตรวม และ ส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ยเท่ากับ 3.20 และ 0.52 ตามลำดับ เมื่อทำการทดสอบความแตกต่างระหว่างผลการประเมินกรณีใช้มาตรฐานวัดหรือเครื่องมือประเมินต่างรูปแบบกัน ปรากฏว่าได้ค่าคำนวณ $Z = 2.30$ ซึ่งสรุปได้ว่า กรณีการประเมินโดยใช้มาตรฐานค่าแบบกราฟิกให้ค่าเฉลี่ยของผลการประเมินในระดับที่สูงกว่าอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($Z_{.05}$ กรณีทดสอบแบบหางเดียว มีค่าเท่ากับ 1.64)

โดยสรุป จากการเปรียบเทียบผลการประเมินพฤติกรรมการสอนของครู ทั้งในสถานการณ์การประเมินตนเอง และการประเมินโดยหัวหน้าหมวดวิชา ซึ่งพบว่า กรณีใช้มาตรฐานค่าแบบกราฟิกเป็นเครื่องมือประเมินปรากฏคะแนนเฉลี่ยของผลการประเมินโดยภาพรวมในระดับที่สูงกว่ากรณีใช้มาตรฐานค่าเชิงพฤติกรรมอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และการกระจายของค่าเฉลี่ยระหว่างรายการพฤติกรรมมีค่าน้อยกว่านั้น โดยนัยนี้ สรุปได้ว่า ผลการประเมินโดยมาตรฐานค่าแบบกราฟิกปรากฏลักษณะของการปล่อยคะแนนในระดับที่สูงกว่า หรืออีกนัยหนึ่งก็คือ มาตรฐานค่าเชิงพฤติกรรมให้ผลการประเมินที่สะท้อนถึงการเกิดความคลาดเคลื่อนอันเนื่องมาจากการปล่อยคะแนนน้อยกว่าผลการประเมินโดยมาตรฐานค่าแบบกราฟิก

ตารางที่ 13 ค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของคะแนนความคลาดเคลื่อน^{*} (Halo Score) เปรียบเทียบระหว่าง กรณีใช้มาตรประมาณค่าเชิงพฤติกรรม กับ กรณีใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมิน

| สถานการณ์ประเมิน | กรณีใช้ BARS | | กรณีใช้ GRS | | ค่าจำนวน t ในการทดสอบความแตกต่าง |
|------------------------------------|--------------|------|-------------|------|----------------------------------|
| | \bar{X} | S.D | \bar{X} | S.D | |
| 1. ครูผู้สอนประเมินตนเอง | 1.14 | 0.23 | 0.87 | 0.21 | 6.51 [*] |
| 2. หัวหน้าหมวดวิชาประเมินครูผู้สอน | 1.12 | 0.22 | 0.83 | 0.21 | 6.47 [*] |

หมายเหตุ N = 84 คน

* $p < .05$

จากตาราง พบว่า ค่าเฉลี่ยเลขคณิตของคะแนนความคลาดเคลื่อน (Halo Score) ภายใต้การใช้มาตรประมาณค่าเชิงพฤติกรรมเป็นเครื่องมือประเมิน มีค่าสูงกว่า กรณีใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมินอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทั้งในสถานการณ์การประเมินพฤติกรรมตนเองของครูผู้สอน และการประเมินพฤติกรรมของครูโดยหัวหน้าหมวดวิชา ซึ่งสรุปได้ว่า กรณีใช้มาตรประมาณค่าเชิงพฤติกรรมเป็นเครื่องมือประเมินให้ข้อมูลผลการประเมินที่มีการกระจายของคะแนนรายบุคคลสูงกว่ากรณีใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือประเมิน หรืออีกนัยหนึ่งก็คือ ผลการประเมินโดยใช้มาตรประมาณค่าแบบกราฟิกเป็นเครื่องมือ มีการกระจายของคะแนนรายบุคคลต่ำกว่าคะแนนระหว่างรายการพฤติกรรมต่าง ๆ แตกต่างกันอย่างน้อยมาก หรือมีค่าใกล้เคียงกันอย่างสม่ำเสมอ ซึ่งเป็นการสะท้อนถึงการเกิดความคลาดเคลื่อนอันเนื่องมาจากการมีปฏิสัมพันธ์มาก่อน (Halo Error) ในระดับที่สูงกว่า

* ในที่นี้ได้ใช้ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ของคะแนนผลการประเมินรายการพฤติกรรมต่าง ๆ ของแต่ละบุคคลเป็นคะแนนความคลาดเคลื่อน (Halo Score) กรณีที่ Halo Score มีค่าน้อย แสดงว่าเกิด Halo Error มาก

โดยสรุป จากการตรวจสอบความคลาดเคลื่อน อันเนื่องมาจากการมีปฏิสัมพันธ์มาก่อน (Halo Error) พบว่า ในกรณีที่ใช้ค่าการกระจายของคะแนนรายพฤติกรรมของผู้ประเมินแต่ละคน หรือ ส่วนเบี่ยงเบนมาตรฐานของคะแนน เป็นดัชนีบ่งชี้การเกิดความคลาดเคลื่อนในการประเมิน (Halo Score) ผลการประเมินโดยใช้มาตรฐานค่าเชิงพฤติกรรม ปรากฏลักษณะความคลาดเคลื่อน น้อยกว่าการประเมินโดยใช้มาตรฐานค่าแบบกราฟิก ทั้งในสถานการณ์การประเมินพฤติกรรมตนเองของครูผู้สอน และการประเมินพฤติกรรมของครู โดยหัวหน้าหมวดวิชา