

บทที่ 5

สรุปและอภิปรายผล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและเปรียบเทียบค่าความเที่ยงของแบบทดสอบผลสัมฤทธิ์ทางการเรียน ที่มีจำนวนข้อสอบทำหน้าที่ต่างกันแตกต่างกัน ภายใต้เงื่อนไขจำนวนของข้อสอบทำหน้าที่ต่างกัน 7 เงื่อนไข คือ 0%, 5 %, 10%, 15%, 20 %, 25% และ 30 % โดยใช้วิธีชิปเทสต์ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ และเปรียบเทียบค่าความเที่ยงของแบบทดสอบ แบบค่าสัมประสิทธิ์แอลฟาของครอนบัค

ข้อมูลที่ใช้เป็นข้อมูลทุติยภูมิ ที่เป็นผลการตอบข้อสอบวัดผลสัมฤทธิ์ทางการเรียน ระดับชาติ วิชาภาษาไทย ของนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ปีการศึกษา 2546 ของวิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา ซึ่งเป็นข้อมูลที่ได้จากสำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ

กลุ่มตัวอย่างที่ใช้ครั้งนี้ เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ได้มาโดยการสุ่มแบบแบ่งชั้นแบบไม่กำหนดสัดส่วน (Stratified Random Sampling) ใช้ระดับความสามารถ (คะแนนรวม) เป็นชั้น แบ่งเป็น 3 ระดับ คือ ดี พอใช้ และปรับปรุง ใช้นักเรียนเป็นหน่วยในการสุ่ม สุ่มมาจำนวน 2,000 คน แบ่งเป็นผลการตอบของนักเรียนชาย 1,000 คน และผลการตอบของนักเรียนหญิง 1,000 คน

วิเคราะห์ค่าสถิติพื้นฐาน โดยใช้โปรแกรม SPSS ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยโปรแกรมชิปเทสต์ และทดสอบความแตกต่างของค่าความเที่ยงของแบบทดสอบ โดยใช้การทดสอบสถิติ Z

สรุปผลการวิจัย

1. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จากข้อมูลทุติยภูมิซึ่งเป็นผลการตอบข้อสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ วิชาภาษาไทย 40 ข้อ ของกลุ่มตัวอย่างจำนวน 2,000 คน โดยโปรแกรมชิปเทสต์ พบข้อสอบทำหน้าที่ต่างกัน จำนวน 13 ข้อ ได้แก่ ข้อที่ 2, 9, 13, 21, 23, 24, 27, 28, 34, 36, 38, 39 และ 40 และข้อสอบทำหน้าที่ไม่ต่างกัน 27 ข้อ เมื่อจำแนกกลุ่มผู้สอบตามตัวแปรเพศ

2. ผลการเปรียบเทียบค่าความเที่ยงของแบบทดสอบ ระหว่างแบบทดสอบฉบับที่ไม่มีข้อสอบทำหน้าที่ต่างกัน หรือมีข้อสอบทำหน้าที่ต่างกัน 0 % ซึ่งมีค่าความเที่ยงของแบบทดสอบ

0.799 กับแบบทดสอบฉบับที่มีข้อสอบทำหน้าที่แตกต่างกันระดับต่าง ๆ คือฉบับที่มีข้อสอบทำหน้าที่แตกต่างกัน 3.70 % (เป็นตัวแทนเงื่อนไข 5%) ซึ่งมีค่าความเที่ยงของแบบทดสอบ 0.8029 ฉบับที่มีข้อสอบทำหน้าที่แตกต่างกัน 11.11 % (เป็นตัวแทนเงื่อนไข 10 %) มีค่าความเที่ยงของแบบทดสอบ 0.8086 ฉบับที่มีข้อสอบทำหน้าที่แตกต่างกัน 14.81 (เป็นตัวแทนเงื่อนไข 15 %) มีค่าความเที่ยงของแบบทดสอบ 0.8086 ฉบับที่มีข้อสอบทำหน้าที่แตกต่างกัน 22.22 % (เป็นตัวแทนเงื่อนไข 20 %) มีค่าความเที่ยงของแบบทดสอบ 0.8041 ฉบับที่มีข้อสอบทำหน้าที่แตกต่างกัน 25.93 % (เป็นตัวแทนเงื่อนไข 25 %) มีค่าความเที่ยงของแบบทดสอบ 0.8019 และฉบับที่มีข้อสอบทำหน้าที่แตกต่างกัน 29.63 % (เป็นตัวแทนเงื่อนไข 30 %) มีค่าความเที่ยงของแบบทดสอบ 0.8018 ผลการเปรียบเทียบพบว่าค่าความเที่ยงไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ และเมื่อวิเคราะห์เปรียบเทียบค่าความเที่ยงของแบบทดสอบ ระหว่างฉบับที่ไม่มีข้อสอบทำหน้าที่แตกต่างกัน กับฉบับที่มีข้อสอบทำหน้าที่ต่างกันและถ้าเพียงเข้าหากลุ่มผู้สอบกลุ่มใดกลุ่มหนึ่ง ค่าความเที่ยงของแบบทดสอบไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

อภิปรายผลการวิจัย

ผลการวิจัยปรากฏว่า แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ วิชาภาษาไทย ของนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ปีการศึกษา 2546 มีข้อสอบทำหน้าที่ต่างกัน 13 ข้อ คิดเป็นร้อยละ 32.5 ค่าความเที่ยงของแบบทดสอบฉบับที่ตัดข้อสอบทำหน้าที่ต่างกันออก 13 ข้อ ซึ่งเหลือข้อสอบในแบบทดสอบ 27 ข้อ วิเคราะห์ค่าความเที่ยงของแบบทดสอบฉบับที่ไม่มีข้อสอบทำหน้าที่ต่างกัน เปรียบเทียบค่าความเที่ยงของแบบทดสอบกับฉบับที่มีข้อสอบทำหน้าที่ต่างกัน ระดับต่าง ๆ คือ 3.70 % , 11.11 , 14.81 % , 22.22 % , 25.93 % และ 29.63% ซึ่งผลการเปรียบเทียบพบว่าค่าความเที่ยงของแบบทดสอบไม่แตกต่างกัน จากผลการวิจัยครั้งนี้ มีประเด็นสำคัญที่นำมาอภิปรายดังต่อไปนี้

1. ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ วิชาภาษาไทยโดยโปรแกรมชิปเทสต์ พบข้อสอบทำหน้าที่ต่างกันจำนวน 13 ข้อ คิดเป็นร้อยละ 32.5 ซึ่งเป็นข้อสอบในกลุ่มวัดโครงสร้างความรู้ 3 ข้อ คือข้อ 2, 9 และ 13 และกลุ่มวัดกระบวนการ 10 ข้อ คือข้อ 21, 23, 24, 27, 28, 34, 36, 38, 39 และ 40 เมื่อพิจารณาโดยละเอียด พบว่า ในจำนวนข้อสอบ 13 ข้อ ซึ่งทำหน้าที่ต่างกัน มีค่า β_{umi} มากกว่า 0 จำนวน 9 ข้อ คือ 2, 9, 13, 27, 28, 34, 36, 38 และ 39 ซึ่งแสดงว่าข้อสอบทั้ง 9 ข้อ ถ้าเพียงเข้าข้างกลุ่มอ้างอิงหรือผู้สอบเพศหญิง และมีค่า β_{umi} น้อยกว่า 0 อยู่ 4 ข้อ คือ ข้อ 21, 23, 24 และ 40 ซึ่งเป็นข้อสอบที่

ลำเอียงเข้าข้างกลุ่มเปรียบเทียบหรือผู้สอบเพศชาย เมื่อพิจารณาจำนวนข้อสอบและร้อยละของข้อสอบที่ทำหน้าที่ต่างกันแสดงว่าแบบทดสอบมีความผิดปกติมาก เพราะมีจำนวนข้อสอบที่ทำหน้าที่ต่างกันมากกว่าร้อยละ 20 คือมีถึงร้อยละ 32.50 โดยความผิดปกติของข้อสอบที่เกิดขึ้นสอดคล้องกับข้อค้นพบของ คลอสเซอร์ ที่ระบุว่า แบบทดสอบมาตรฐานที่ใช้วัดผลสัมฤทธิ์ทางการเรียน ถ้ามีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 – 15 ถือว่าไม่ผิดปกติ แต่ถ้ามีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 ถือว่าผิดพลาดอย่างมาก (Clauser, 1993-citing Narayanan & Swaminathan, 1994 อ้างถึงใน วลีมาศ แซ่อึ้ง, 2543, หน้า 1) ทั้งนี้อาจเนื่องมาจากข้อสอบมีเนื้อหาที่ผู้สอบกลุ่มเพศชายซึ่งเป็นกลุ่มเปรียบเทียบ ไม่คุ้นเคย หรือข้อสอบวัดในสิ่งที่ไม่สอดคล้องกับสิ่งที่กำหนดไว้ในหลักสูตร หรือข้อสอบมีเนื้อหาทักษะ หรือสารสนเทศที่ไม่เป็นการทั่วไปกับผู้สอบทุกคน และเมื่อวิเคราะห์อย่างลึกซึ้งพบว่าข้อสอบในกลุ่มที่วัดทักษะกระบวนการมีข้อสอบทำหน้าที่ต่างกันถึง 10 ข้อ จาก 25 ข้อ ซึ่งอาจเนื่องมาจากผู้สอบเพศชายมีทักษะกระบวนการด้านภาษาน้อย หรือข้อสอบวัดไม่ตรงกับทักษะที่หลักสูตรกำหนด ส่งผลให้เกิดการได้เปรียบเสียเปรียบกันในการตอบข้อสอบ โดยกลุ่มผู้สอบเพศหญิงซึ่งเป็นกลุ่มอ้างอิงได้เปรียบในการตอบข้อสอบ ส่วนกลุ่มผู้สอบเพศชายเสียเปรียบในการตอบข้อสอบ

2. ผลการเปรียบเทียบค่าความเที่ยงของแบบทดสอบระหว่างฉบับที่ไม่มีข้อสอบทำหน้าที่ต่างกัน และฉบับที่มีข้อสอบทำหน้าที่ต่างกัน 6 ฉบับ ใจ คือ 3.70 %, 11.11%, 14.81 %, 22.22 %, 25.93 % และ 29.63% พบว่าค่าความเที่ยงของแบบทดสอบมีค่าไม่สูงมากและไม่แตกต่างกัน ซึ่งไม่เป็นไปตามสมมติฐานที่ตั้งไว้ ทั้งนี้อาจเนื่องจากแบบทดสอบฉบับที่ไม่มีข้อสอบทำหน้าที่ต่างกัน มีปริมาณข้อสอบน้อยเพียง 27 ข้อ และเมื่อจัดข้อสอบทำหน้าที่ต่างกันเข้าไปแทนข้อสอบที่ทำหน้าที่ไม่ต่างกัน เพื่อให้แบบทดสอบมีจำนวนข้อสอบ 27 ข้อเท่าเดิม จัดกระทำข้อมูลประมาณครั้งละ 5 เปอร์เซนต์ ซึ่งเมื่อคิดเป็นจำนวนข้อสอบ เป็นเพียง 1 – 2 ข้อ มีผลทำให้ค่าความแปรปรวนของคะแนนมีค่าแตกต่างจากเดิมเพียงเล็กน้อย เนื่องจากจำนวนข้อสอบทั้งหมดในแบบทดสอบมีจำนวนน้อย ส่งผลให้ค่าความเที่ยงของแบบทดสอบมีค่าแตกต่างจากเดิมเพียงเล็กน้อย ซึ่งแสดงให้เห็นว่าเมื่อจำนวนข้อสอบในแบบทดสอบมีจำนวนน้อย ข้อสอบซึ่งทำหน้าที่ต่างกันจะไม่มีผลต่อค่าความเที่ยงของแบบทดสอบ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. ในการเปรียบเทียบค่าความเที่ยงของแบบทดสอบระหว่างฉบับที่ไม่มีข้อสอบทำหน้าที่ต่างกัน กับฉบับที่มีข้อสอบทำหน้าที่ต่างกัน ปริมาณต่าง ๆ มีข้อสอบเพียง 27 ข้อ ซึ่งถือว่ามีจำนวนน้อย ทำให้ค่าความเที่ยงที่นำมาเปรียบเทียบมีค่าต่ำและมีค่าไม่แตกต่างกัน ดังนั้นถ้า

ต้องการนำผลการเปรียบเทียบค่าความเที่ยงไปใช้ ไม่ควรตัดข้อสอบทำหน้าที่ต่างกันออกทั้ง 13 ข้อ เนื่องจากแบบทดสอบที่มีข้อสอบทำหน้าที่ต่างกัน 10 - 15 % ถือว่าไม่ผิดปกติ ดังนั้นควรตัดข้อสอบออกเพียงบางข้อ โดยดูเนื้อหาของข้อสอบเป็นเกณฑ์ว่ามีความเหมาะสมหรือไม่ เพื่อจะได้ค่าความเที่ยงที่ถูกต้องแม่นยำมากยิ่งขึ้น

2. จากผลการวิเคราะห์ข้อสอบ พบว่าในแบบทดสอบมีข้อสอบทำหน้าที่ต่างกันถึง 13 ข้อ เพื่อให้เกิดความยุติธรรมแก่ผู้เข้าสอบ ควรคิดคะแนนเฉพาะข้อสอบที่ทำหน้าที่ไม่ต่างกัน เพราะจะทำให้ได้ข้อมูลด้านความสามารถที่แท้จริงของผู้เข้าสอบ

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

1. จากการทดสอบการทำหน้าที่ต่างกันของข้อสอบ ผู้วิจัยใช้วิธีชิปทดสอบซึ่งเป็นวิธีที่เป็นมาตรฐานและเชื่อถือได้ รวมทั้งใช้โปรแกรมชิปทดสอบรุ่นล่าสุด ทำให้ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันหลายข้อ เพื่อให้ผลการตรวจสอบมีความน่าเชื่อถือมากขึ้นควรมีการตรวจสอบด้วยวิธีต่าง ๆ อีกหลาย ๆ วิธี แล้วพิจารณาเฉพาะข้อสอบทำหน้าที่ต่างกันที่ตรงกัน เพื่อลดความคลาดเคลื่อนประเภทที่ 1

2. ในการเปรียบเทียบค่าความเที่ยงของแบบทดสอบ แบบทดสอบที่เมื่อตัดข้อสอบทำหน้าที่ต่างกันออกแล้ว ควรมีจำนวนข้อสอบ 40 – 60 ข้อ เนื่องจากจำนวนข้อสอบที่ลดลงจะมีผลทำให้ค่าความเที่ยงของแบบทดสอบลดลง เมื่อตัดข้อสอบหรือเพิ่มข้อสอบเข้าเพียง 1 ข้อ หรือ 2 ข้อ จะไม่เห็นความแตกต่างของค่าความเที่ยง ทำให้ผลการวิจัยคลาดเคลื่อนจากความเป็นจริง

3. เนื่องจากงานวิจัยครั้งนี้ผู้วิจัยใช้ข้อมูลจริง จึงมีข้อจำกัดในด้านจำนวนข้อสอบ และลักษณะของแบบทดสอบ ที่ไม่สามารถระบุให้มีจำนวนข้อสอบทำหน้าที่ต่างกัน แตกต่างกันตามเงื่อนไขที่กำหนด คือ 0%, 5 %, 10%, 15%, 20 %, 25% และ 30 % ได้ ดังนั้น ในการวิจัยครั้งต่อไป ถ้าใช้วิธีการจำลองข้อมูลจะทำให้ได้ผลการวิจัยที่ถูกต้อง แม่นยำมากยิ่งขึ้น