

บทที่ 5

สรุปและอภิปรายผล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาอิทธิพลที่เกิดขึ้นจากการตัดข้อสอบทำหน้าที่ย่างกันออกจากแบบทดสอบ ที่ส่งผลต่อคุณภาพของแบบทดสอบวัดผลฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6 โดยเปรียบเทียบค่าความเที่ยง ความตรงเชิงโครงสร้าง ความคงที่ของโครงสร้างองค์ประกอบและค่าสหสัมพันธ์อันดับของผู้สอบ ฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ

ข้อมูลที่ใช้เป็นข้อมูลทุติยภูมิ ซึ่งเป็นผลการตอบข้อสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ปีการศึกษา 2546 วิชาภาษาไทย ชั้นประถมศึกษาปีที่ 6 โดยผู้วิจัยขอความร่วมมือจากเจ้าหน้าที่สำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ ในการเก็บรวบรวมข้อมูล

กลุ่มตัวอย่าง เป็นผลการตอบของนักเรียนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ในเขตพื้นที่การศึกษานครศรีธรรมราช ได้มาโดยการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) แบบไม่กำหนดสัดส่วน โดยใช้ระดับความสามารถของผู้สอบเป็นชั้น ซึ่งแบ่งเป็นระดับดี ระดับพอใช้ และระดับปรับปรุง สุ่มแยกตามเพศ โดยสุ่มมาชั้นละเท่า ๆ กัน ใช้กลุ่มตัวอย่างทั้งหมด จำนวน 2,000 คน แยกเป็นผลการตอบของผู้สอบเพศชาย จำนวน 1,000 คน และผลการตอบของผู้สอบเพศหญิง จำนวน 1,000 คน

วิเคราะห์ค่าสถิติพื้นฐานโดยใช้โปรแกรม SPSS วิเคราะห์คุณภาพของข้อสอบรายข้อ และค่าความเที่ยงแบบสัมประสิทธิ์แอลฟาของครอนบาคโดยใช้โปรแกรม LERTAP 5 ตรวจสอบความตรงเชิงโครงสร้างโดยการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง โดยใช้โปรแกรม LISREL 8.50 และตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้โปรแกรม SIBTEST และ SPSS

สรุปผลการวิจัย

1. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จำนวน 40 ข้อ ด้วยวิธีการตรวจสอบสองวิธี คือวิธีถดถอยโลจิสติก และวิธีชิปเทสที่ปรับใหม่พบ DIF ตรงกันและถูกระบุว่าข้อสอบทำหน้าที่ต่างกัน จำนวน 12 ข้อ คิดเป็นร้อยละ 30
2. ผลการเปรียบเทียบค่าความเที่ยงของแบบทดสอบฉบับก่อนตรวจสอบ DIF กับฉบับหลังตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ ปรากฏว่า แบบทดสอบฉบับก่อนตรวจสอบ DIF มีข้อสอบจำนวน 40 ข้อ ค่าความเที่ยงแบบสัมประสิทธิ์แอลฟาของครอนบาคเท่ากับ .86 และ

แบบทดสอบฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ มีข้อสอบจำนวน 28 ข้อ มีค่าความเที่ยงหลังปรับขยายสัดส่วนจำนวนข้อให้เท่ากับฉบับก่อนตัดข้อสอบ มีค่าความเที่ยงแบบ สัมประสิทธิ์แอลฟาของครอนบาคเท่ากับ .87 ผลการเปรียบเทียบปรากฏว่า ค่าความเที่ยงของ แบบทดสอบทั้งสองฉบับมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3. ความตรงเชิงโครงสร้างของแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ ต่างกันออกจากแบบทดสอบปรากฏว่า โมเดลแบบทดสอบภาษาไทยทั้งสองฉบับมีความสอดคล้อง กลมกลืนกับข้อมูลเชิงประจักษ์อยู่ในระดับดี คำนวณหาค่าประกอบด้านความรู้ของแบบทดสอบ ฉบับก่อนตัดข้อสอบที่พบ DIF มีค่าเท่ากับ .32 ฉบับหลังตัดข้อสอบที่พบ DIF มีค่าเท่ากับ .53 คำนวณหาค่าประกอบด้านกระบวนการฉบับก่อนตัดข้อสอบมีค่าเท่ากับ .42 ฉบับหลังตัดข้อสอบ ที่พบ DIF มีค่าเท่ากับ .43 คำนวณหาค่าประกอบทุกค่ามีนัยสำคัญ แสดงว่าแบบทดสอบทั้งสอง ฉบับประกอบด้วยสององค์ประกอบเช่นเดียวกันและโครงสร้างของแบบทดสอบไม่ได้เปลี่ยนแปลง ไปจากเดิม เมื่อพิจารณาเปรียบเทียบค่าสถิติไค-สแควร์สอดคล้องระหว่างโมเดลแบบทดสอบฉบับ ก่อนตัดข้อสอบที่พบ DIF กับ โมเดลแบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF มีค่าเท่ากับ 315.72 ที่องศาอิสระเท่ากับ 347 โดยค่าไค-สแควร์สอดคล้องมีนัยสำคัญทางสถิติ แสดงว่า โมเดล แบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับ โมเดลแบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF มีความสอดคล้องกับข้อมูลเชิงประจักษ์แตกต่างกันอย่างมีนัยสำคัญ โดยโมเดลแบบทดสอบ ฉบับ หลังตัดข้อสอบที่พบ DIF มีความสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่า แต่ทั้งสองโมเดลยัง ประกอบด้วยสององค์ประกอบเช่นเดียวกัน

4. ความคงที่ของโครงสร้างองค์ประกอบ สำหรับเพศชายและเพศหญิง

4.1 โครงสร้างแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออก จากแบบทดสอบ สำหรับเพศชาย ผลปรากฏว่า โมเดลแบบทดสอบภาษาไทยทั้งสองฉบับมีความ สอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์อยู่ในระดับดี เมื่อพิจารณาคำนวณหาค่าประกอบด้าน ความรู้ ฉบับก่อนตัดข้อสอบมีค่าเท่ากับ .31 ฉบับหลังตัดข้อสอบที่พบ DIF มีค่าเท่ากับ .52 ค่า คำนวณหาค่าประกอบด้านกระบวนการฉบับก่อนตัดข้อสอบมีค่าเท่ากับ .40 ฉบับหลังตัดข้อสอบที่ พบ DIF มีค่าเท่ากับ .48 คำนวณหาค่าประกอบทุกค่ามีนัยสำคัญ แสดงว่าเมื่อตัดข้อสอบทำหน้าที่ ต่างกันออกจากแบบทดสอบไม่ได้ทำให้โครงสร้างของแบบทดสอบเปลี่ยนแปลงไปจากเดิม นั่น คือโครงสร้างของแบบทดสอบ สำหรับเพศชาย มีความคงที่ เมื่อพิจารณาเปรียบเทียบค่าสถิติ ไค-สแควร์สอดคล้องระหว่างโมเดลของแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับ โมเดล แบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF มีค่าเท่ากับ 299.91 ที่องศาอิสระเท่ากับ 373 ค่า ไค-สแควร์สอดคล้องมีนัยสำคัญทางสถิติ แสดงว่าโมเดลแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ

DIF กับ โมเดลแบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF มีความสอดคล้องกับข้อมูลเชิงประจักษ์แตกต่างกันอย่างมีนัยสำคัญ โดยโมเดลแบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF มีความสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่า แต่ทั้งสองโมเดลประกอบด้วยสององค์ประกอบเช่นเดียวกัน

4.2 โครงสร้างแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ สำหรับเพศหญิง ผลปรากฏว่า โมเดลแบบทดสอบภาษาไทยทั้งสองฉบับมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์อยู่ในระดับดี เมื่อพิจารณาค่าน้ำหนักองค์ประกอบด้านความรู้ ฉบับก่อนตัดข้อสอบมีค่าเท่ากับ .33 ฉบับหลังตัดข้อสอบที่พบ DIF มีค่าเท่ากับ .56 ค่าน้ำหนักองค์ประกอบด้านกระบวนการฉบับก่อนตัดข้อสอบมีค่าเท่ากับ .43 ฉบับหลังตัดข้อสอบที่พบ DIF มีค่าเท่ากับ .45 ค่าน้ำหนักองค์ประกอบทุกค่ามีนัยสำคัญ แสดงว่าเมื่อตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบไม่ได้ทำให้โครงสร้างของแบบทดสอบเปลี่ยนแปลงไปจากเดิมนั่นคือโครงสร้างของแบบทดสอบ สำหรับเพศหญิง มีความคงที่ เมื่อพิจารณาเปรียบเทียบค่าสถิติไค-สแควร์สอดแทรกระหว่างโมเดลของแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับโมเดลแบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF มีค่าเท่ากับ 184.29 ที่องศาอิสระเท่ากับ 297 ค่าไค-สแควร์สอดแทรกมีนัยสำคัญทางสถิติ แสดงว่า โมเดลแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับ โมเดลแบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF มีความสอดคล้องกับข้อมูลเชิงประจักษ์แตกต่างกัน โดยโมเดลของแบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF มีความสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่า แต่ทั้งสองโมเดลประกอบด้วยสององค์ประกอบเช่นเดียวกัน

5. สหสัมพันธ์อันดับของผู้สอบ ของแบบทดสอบฉบับก่อนตรวจสอบ DIF กับฉบับหลังตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ พบว่า ลำดับที่การสอบที่ได้จากแบบทดสอบฉบับก่อนตรวจสอบ DIF กับลำดับที่การสอบที่ได้จากแบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยมีค่าสัมประสิทธิ์สหสัมพันธ์อันดับแบบสเปียร์แมนแรงค์เท่ากับ .48 ซึ่งมีความสัมพันธ์อยู่ในระดับปานกลาง

อภิปรายผลการวิจัย

ผลการวิจัยพบว่า แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6 วิชาภาษาไทย มีข้อสอบทำหน้าที่ต่างกันจำนวน 12 ข้อ คิดเป็นร้อยละ 30 ค่าความเที่ยงของแบบทดสอบฉบับก่อนตัดข้อสอบกับฉบับหลังตัดข้อสอบมีค่าความเที่ยงแตกต่างกันอย่างมีนัยสำคัญทางสถิติ แบบทดสอบฉบับก่อนตัดข้อสอบกับฉบับหลังตัดข้อสอบมีความตรง

เชิงโครงสร้างตามทฤษฎีและประกอบด้วยสององค์ประกอบเช่นเดียวกัน โดยโมเดลแบบทดสอบฉบับหลังตัดข้อสอบมีความสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่า เมื่อจำแนกกลุ่มผู้สอบตามตัวแปรเพศพบว่า แบบทดสอบทั้งฉบับก่อนตัดข้อสอบกับฉบับหลังตัดข้อสอบมีความคงที่ของโครงสร้างองค์ประกอบและลำดับที่การสอบที่ได้จากแบบทดสอบทั้งสองฉบับมีความสัมพันธ์กันจากการวิจัยครั้งนี้ มีประเด็นสำคัญที่ควรนำมาอภิปราย ดังต่อไปนี้

1. ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธีตรวจสอบสองวิธี คือ วิธีถดถอยโลจิสติกกับวิธีชิปเทสที่ปรับใหม่ พบข้อสอบที่มี DIF จากการตรวจพบตรงกัน จำนวน 12 ข้อ คิดเป็นร้อยละ 30 เมื่อพิจารณาจากจำนวนข้อและร้อยละของข้อสอบทำหน้าที่ต่างกันแสดงให้เห็นว่าแบบทดสอบมีความผิดปกติ อาจเนื่องมาจากข้อสอบมีเนื้อหาที่ผู้สอบกลุ่มเพศชายไม่คุ้นเคย หรือข้อสอบวัดในสิ่งที่ไม่สอดคล้องกับสิ่งที่กำหนดไว้ในหลักสูตร หรือข้อสอบมีเนื้อหาทักษะ หรือสารสนเทศที่ไม่เป็นการทั่วไปกับผู้สอบทุกคน ส่งผลให้เกิดการได้เปรียบเสียเปรียบกันในการตอบข้อสอบโดยโดยกลุ่มผู้สอบเพศหญิงเป็นกลุ่มที่ได้เปรียบในการตอบข้อสอบส่วนกลุ่มผู้สอบเพศชายเป็นกลุ่มที่เสียเปรียบ และอีกตัวแปรหนึ่งที่จะทำให้พบข้อสอบทำหน้าที่ต่างกันของข้อสอบหลายข้อก็คือ วิธีการตรวจสอบ โดยปกติแต่ละวิธีจะตรวจสอบพบ DIF ไม่ตรงกัน ในการวิจัยครั้งนี้ใช้วิธีการตรวจสอบสองวิธีที่มีอำนาจการทดสอบสูง ทำให้พบข้อสอบ DIF จำนวนหลายข้อ โดยความผิดปกติของข้อสอบที่เกิดขึ้นสอดคล้องกับข้อค้นพบของคลอสเซอร์ที่ระบุว่า แบบทดสอบมาตรฐานที่ใช้วัดผลสัมฤทธิ์ทางการเรียน ถ้ามีสัดส่วนของข้อสอบทำหน้าที่ต่างกันร้อยละ 10 ถึง 15 ถือว่าไม่ผิดปกติ แต่ถ้ามีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 ถือว่าเป็นเรื่องผิดปกติอย่างมาก (Clauser, 1993 cited in Narayanan & Swaminathan, 1994 อ้างถึงใน วลีมาศ แซ่ฮึ้ง, 2543, หน้า 1)

2. ค่าความเที่ยงของแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ พบว่า มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติ สอดคล้องกับสมมติฐานการวิจัยในข้อที่ 1 เนื่องจากตามปกติการลดจำนวนข้อสอบออกจากแบบทดสอบฉบับเดิมส่งผลให้ค่าความเที่ยงลดลง เหตุผลเนื่องมาจากการตัดข้อสอบออกจากแบบทดสอบนั้น อาจตัดข้อสอบที่มีผลทำให้การกระจายของคะแนนสูงออกไป ทำให้ความแปรปรวนของคะแนนที่สังเกตได้มีค่าลดลง จึงส่งผลให้ค่าความเที่ยงลดลง ดังนั้นเมื่อเปรียบเทียบโดยใช้สถิติทดสอบ ทำให้แบบทดสอบทั้งสองฉบับแตกต่างกันอย่างมีนัยสำคัญ แม้ว่าจะทำการปรับขยายสัดส่วนจำนวนข้อสอบฉบับหลังตัดข้อสอบที่พบ DIF ให้มีสัดส่วนข้อสอบเท่ากับก่อนตัด ก็ไม่ได้ส่งผลให้แบบทดสอบฉบับหลังมีความเที่ยงเทียบเท่ากับแบบทดสอบฉบับก่อนเพราะค่าความเที่ยงที่ปรับขยายนั้นเป็นเพียงค่าประมาณ สำหรับประเด็นด้านความยาวของแบบทดสอบที่เกี่ยวกับค่าความเที่ยง ศิริชัย

กาญจนวาลี (2545 ก) ระบุว่า การลดความยาวของแบบทดสอบมีผลต่อการลดลงของค่าความเที่ยง ดังนั้นการตัดข้อสอบออกจากแบบทดสอบจึงควรกระทำอย่างระมัดระวังและรอบคอบ โดยเฉพาะอย่างยิ่งกับแบบทดสอบที่มีจำนวนข้อสอบน้อยข้อ และเขาวดี วิบูลย์ศรี (2545) องค์ประกอบสำคัญที่มีผลต่อค่าความเที่ยงของแบบทดสอบก็คือความยาวของแบบทดสอบ สามารถสังเกตได้จากแบบทดสอบที่มีความยาวจะมีข้อสอบมาก จึงสามารถวัดคุณลักษณะที่ต้องการวัดได้ดี คือ ผู้สอบมีโอกาที่จะแสดงความสามารถที่แท้จริงได้มากขึ้น นอกจากนี้ผลการวิจัยยังสอดคล้องกับผลการศึกษาของ อารี วัชรโสติกุล (2543) ที่พบว่า ค่าความเที่ยงของแบบทดสอบฉบับหลังตัดข้อสอบทำหน้าที่ย่างกันออกจากแบบทดสอบ จากการปรับขยายค่าความเที่ยงโดยใช้สูตร สเปียร์แมน-บราวน์ พบว่า ค่าความเที่ยงที่ได้มีค่าแตกต่างจากแบบทดสอบฉบับก่อนตรวจสอบ DIF อย่างมีนัยสำคัญทางสถิติ รักชนก ยี่สุนศรี (2544) ที่ใช้กระบวนการดีเอฟไอทีในการตรวจสอบ พบว่าแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ย่างกันออกจากแบบทดสอบ ส่วนใหญ่มีค่าความเที่ยงลดลง

3. ความตรงเชิงโครงสร้างของแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ย่างกันออกจากแบบทดสอบ ผลปรากฏว่า โมเดลแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับโมเดลแบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF ยังประกอบด้วยสององค์ประกอบเช่นเดียวกัน โดยค่าน้ำหนักองค์ประกอบด้านความรู้ และด้านกระบวนการของแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับฉบับหลังตัดข้อสอบที่พบ DIF มีค่าใกล้เคียงกัน นั่นแสดงว่าเมื่อตัดข้อสอบทำหน้าที่ย่างกันออกจากแบบทดสอบ ไม่ได้ทำให้โครงสร้างของแบบทดสอบเปลี่ยนแปลงไปจากเดิม สอดคล้องกับสมมติฐานการวิจัยที่ตั้งไว้ในข้อที่ 2 เหตุผลเนื่องจากข้อสอบที่ตัดออกไปเป็นข้อสอบที่มีความคลาดเคลื่อนในการวัด ข้อสอบไม่ได้วัดความ สามารถเป้าหมายแต่อาจจะไปวัดความสามารถแทรกซ้อน การตัดข้อสอบที่มี DIF ดังกล่าวออกไปจึงไม่ได้ทำให้โครงสร้างของแบบทดสอบเปลี่ยนแปลงไป แต่กลับทำให้แบบทดสอบมีความตรงเชิงโครงสร้างเพิ่มมากขึ้น แบบทดสอบมีความสามารถในการวัดค่าคุณลักษณะที่แท้จริงของผู้สอบได้ตรงมากขึ้น โดยพิจารณาจากการเปรียบเทียบโมเดลของแบบทดสอบทั้งสองฉบับโดยใช้ค่าไค-สแควร์ทดสอบ พบว่า แบบทดสอบทั้งสองฉบับมีความสอดคล้องกับข้อมูลเชิงประจักษ์แตกต่างกันอย่างมีนัยสำคัญ โดยแบบทดสอบฉบับหลังตัดข้อสอบทำหน้าที่ย่างกันออกจากแบบทดสอบมีความสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่า ผลการวิจัยที่ได้สอดคล้องกับผลการศึกษาของ เกษร ห่วงจิตร (2539) ที่ศึกษาอิทธิพลของการทำหน้าที่ย่างกันของข้อสอบโดยใช้วิธีแมนเทิล-แฮนส์เชลในการตรวจสอบ พบว่า ความตรงเชิงโครงสร้างของแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบส่วนใหญ่ไม่แตกต่างกัน โรโซลสกี และไรส์

(Rozowski & Reith, 1999) ที่ศึกษาตรวจสอบโดยใช้วิธีแมนเทิล-แฮนส์เซล พบว่า แบบทดสอบที่มีข้อสอบทำหน้าที่ต่างกับแบบทดสอบที่ไม่มีข้อสอบทำหน้าที่ต่างกัน ไม่ได้ทำให้แบบทดสอบมีคุณภาพในการวัดหรือความตรงเชิงโครงสร้างต่ำลงมากนักและรักษนก ยี่สุนศรี (2544) พบว่าแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ มีค่าความตรงเชิงโครงสร้างของแบบทดสอบไม่แตกต่างกัน นอกจากนี้จากผลการวิจัยสามารถอธิบายเพิ่มเติมได้ว่า จากนิยามความตรงคือคุณสมบัติของแบบทดสอบที่แสดงถึงความสามารถในการวัดได้ถูกต้อง แม่นยำ ถ้าผลการวัดได้ค่าที่ใกล้เคียงกับค่าคุณลักษณะที่แท้จริงมากเพียงใด แสดงว่าแบบทดสอบมีความตรงมากขึ้นเท่านั้น ซึ่งสามารถประมาณค่าความตรงได้จากสัดส่วนความแปรปรวนของคะแนนจริงที่ตรงกับลักษณะที่มุ่งวัด ดังนั้นความแปรปรวนของคะแนนที่สังเกตได้ (σ_x^2) จึงประกอบด้วยความแปรปรวนของคะแนนที่เที่ยงตรงประเด็นตามคุณลักษณะที่มุ่งวัด (σ_v^2) ความแปรปรวนของคะแนนที่เที่ยงแต่ไม่ตรงประเด็นตามคุณลักษณะที่มุ่งวัด (σ_f^2) และความแปรปรวนของความคลาดเคลื่อน (σ_e^2) (ศิริชัย กาญจนวาสี, 2545 ก) การตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ อาจทำให้ความแปรปรวนของคะแนนที่สังเกตได้มีค่าลดลงหรือความแปรปรวนของคะแนนความคลาดเคลื่อนสูงขึ้น ซึ่งจะทำให้สัดส่วนความแปรปรวนที่เที่ยงต่ำลง (σ_v^2 / σ_x^2) จึงทำให้ค่าความเที่ยงของแบบทดสอบต่ำลง เนื่องจากความแปรปรวนที่เที่ยงมีทั้งความแปรปรวนของส่วนที่ตรงประเด็นตามคุณลักษณะที่มุ่งวัดกับส่วนที่ไม่ตรงประเด็น ดังนั้นสัดส่วนความแปรปรวนที่เที่ยงลดลง อาจมาจากการลดลงของความแปรปรวนในส่วนที่ไม่ตรงประเด็นตามคุณลักษณะที่มุ่งวัด จึงทำให้แบบทดสอบมีค่าความตรงสูงขึ้นในขณะที่ค่าความเที่ยงลดลง

4. ความคงที่ของโครงสร้างองค์ประกอบ ของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6 ที่จำแนกตามตัวแปรเพศ ผลปรากฏว่า ไม่ว่าจะจำแนกกลุ่มผู้สอบกลุ่มย่อยเป็นกลุ่มผู้สอบเพศชาย หรือกลุ่มผู้สอบเพศหญิง โมเดลแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับ โมเดลแบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF ยังมีองค์ประกอบคงที่ประกอบด้วยสององค์ประกอบเช่นเดียวกัน คำนวณน้ำหนักองค์ประกอบด้านความรู้ และด้านกระบวนการของแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับฉบับหลังตัดข้อสอบที่พบ DIF มีค่าใกล้เคียงกัน นั้นแสดงว่าเมื่อตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบไม่ได้ทำให้โครงสร้างของแบบทดสอบเปลี่ยนแปลงไปจากเดิม แบบทดสอบยังมีโครงสร้างองค์ประกอบคงที่สอดคล้องกับสมมติฐานการวิจัยในข้อที่ 3 เหตุผลเนื่องจากข้อสอบที่ตัดออกไปเป็นข้อสอบที่มีความคลาดเคลื่อนในการวัด ข้อสอบไม่ได้วัดความสามารถเป้าหมายแต่อาจจะไปวัดความสามารถแทรกซ้อน การตัดข้อสอบที่มี DIF ดังกล่าวออกไปจึงไม่ได้ทำให้โครงสร้างของแบบทดสอบ

เปลี่ยนแปลงไป แต่กลับทำให้แบบทดสอบมีความตรงเชิงโครงสร้างเพิ่มมากขึ้น แบบทดสอบมีความสามารถในการวัดค่าคุณลักษณะที่แท้จริงของผู้สอบได้ตรงมากขึ้น โดยพิจารณาจากการเปรียบเทียบโมเดลของแบบทดสอบทั้งสองฉบับโดยใช้ค่าไค-สแควร์ทดสอบไค-สแควร์ พบว่าแบบทดสอบทั้งสองฉบับมีความสอดคล้องกับข้อมูลเชิงประจักษ์แตกต่างกันอย่างมีนัยสำคัญ โดยแบบทดสอบฉบับหลังตัดข้อสอบทำหน้าที่ยกต่างจากแบบทดสอบที่มีความสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่า นอกจากนี้ตามนิยามความตรงคือคุณสมบัติของแบบทดสอบที่แสดงถึงความสามารถในการวัดได้ถูกต้อง แม่นยำ ถ้าผลการวัดได้ค่าที่ใกล้เคียงกับค่าคุณลักษณะที่แท้จริงมากเพียงใด แสดงว่าแบบทดสอบมีความตรงมากขึ้นเท่านั้น ซึ่งสามารถประมาณค่าความตรงได้จากสัดส่วนความแปรปรวนของคะแนนจริงที่ตรงกับลักษณะที่มุ่งวัด ดังนั้นความแปรปรวนของคะแนนที่สังเกตได้ (σ_x^2) จึงประกอบด้วยความแปรปรวนของคะแนนที่เที่ยงตรงประเด็นตามคุณลักษณะที่มุ่งวัด (σ_v^2) ความแปรปรวนของคะแนนที่เที่ยงแต่ไม่ตรงประเด็นตามคุณลักษณะที่มุ่งวัด (σ_r^2) และความแปรปรวนของความคลาดเคลื่อน (σ_e^2) (ศิริชัย กาญจนวาสี, 2545 ข) การตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ อาจทำให้ความแปรปรวนของคะแนนที่สังเกตได้มีค่าลดลงหรือความแปรปรวนของคะแนนความคลาดเคลื่อนสูงขึ้น ซึ่งจะทำให้สัดส่วนความแปรปรวนที่เที่ยงต่ำลง (σ_v^2 / σ_x^2) จึงทำให้ค่าความเที่ยงของแบบทดสอบต่ำลง เนื่องจากความแปรปรวนที่เที่ยงมีทั้งความแปรปรวนของส่วนที่ตรงประเด็นตามคุณลักษณะที่มุ่งวัดกับส่วนที่ไม่ตรงประเด็น ดังนั้นสัดส่วนความแปรปรวนที่เที่ยงลดลง อาจมาจากการลดลงของความแปรปรวนในส่วนที่ไม่ตรงประเด็นตามคุณลักษณะที่มุ่งวัด จึงทำให้แบบทดสอบมีค่าความตรงสูงขึ้นในขณะที่ค่าความเที่ยงลดลง การจำแนกกลุ่มผู้สอบตามตัวแปรเพศเป็นกลุ่มผู้สอบเพศชาย และเพศหญิง จึงยังมีโครงสร้างองค์ประกอบคงที่

5. ค่าสหสัมพันธ์อันดับของผู้สอบ ฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ยกต่างจากแบบทดสอบวิชาภาษาไทย มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ สอดคล้องกับสมมติฐานการวิจัยที่ตั้งไว้ในข้อที่ 4 ซึ่งค่าสหสัมพันธ์อันดับ ที่ได้จากการวิจัยมีค่าปานกลาง แสดงว่าลำดับที่การสอบจากแบบทดสอบฉบับก่อนตรวจสอบ DIF กับฉบับหลังตัดข้อสอบที่พบ DIF แตกต่างกันในระดับปานกลาง โดยค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนแรงค์ที่ได้จากแบบทดสอบฉบับก่อนกับฉบับหลังตรวจสอบ DIF มีค่าสูง ปานกลาง หรือต่ำ ขึ้นอยู่กับจำนวนข้อสอบที่พบ DIF ที่ถูกตัดออกไป กล่าวคือถ้าตัดข้อสอบที่พบ DIF ออกไปมากเท่าใด ลำดับที่การสอบของผู้สอบจากแบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF ก็จะแตกต่างจากฉบับก่อนตรวจสอบ DIF มากขึ้นเท่านั้น ถ้าค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้มีค่าสูง แสดงว่าลำดับที่การสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับฉบับหลังตัดข้อสอบที่พบ DIF ไม่แตกต่างกันมากนัก ถ้าค่า

สัมประสิทธิ์สหสัมพันธ์ที่ได้มีค่าต่ำแสดงว่าลำดับที่การสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับ ฉบับหลังตัดข้อสอบที่พบ DIF แตกต่างกันมาก ซึ่งผลการวิจัยที่ได้ สอดคล้องกับโรโซลสกี และไรส์ (Rozowski & Reith, 1999) พบว่าค่าสหสัมพันธ์อันดับของผู้สอบ ระหว่างแบบทดสอบที่มีข้อสอบทำหน้าที่ย่อยต่างกันกับแบบทดสอบที่ไม่มีข้อสอบทำหน้าที่ย่อยต่างกันมีความสัมพันธ์กัน รักษนก ยี่สุนศรี (2544) ที่พบว่าสหสัมพันธ์อันดับ ในทุกกรณีมีความสัมพันธ์ในทางบวกซึ่งกันและกัน อย่างมีนัยสำคัญทางสถิติ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. จากผลการเปรียบเทียบคุณภาพของแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบพบว่า แบบทดสอบฉบับหลังตัดข้อสอบยังมีความตรงเชิงโครงสร้าง เช่นเดียวกับแบบทดสอบฉบับก่อนตรวจสอบ DIF แต่ค่าความเที่ยงของแบบทดสอบฉบับหลังมีค่าลดลงและแตกต่างจากแบบทดสอบฉบับก่อนตรวจสอบ DIF อย่างมีนัยสำคัญ ดังนั้นถ้าต้องการใช้ ผลคะแนนสำหรับตัดสินผลการสอบไม่ควรตัดข้อสอบทุกข้อที่พบ DIF ออกจากแบบทดสอบ แต่ควรเลือกตัดออกเพียงบางข้อโดยพิจารณาเนื้อหาของข้อสอบประกอบด้วย เพื่อแก้ปัญหา แบบทดสอบ มีคุณภาพในด้านความเที่ยงลดลง
2. จากผลการเปรียบเทียบคุณภาพของแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบโดยจำแนกกลุ่มผู้สอบตามตัวแปรเพศ ปรากฏว่า แบบทดสอบฉบับ หลังตัดข้อสอบยังมีโครงสร้างองค์ประกอบเช่นเดียวกับแบบทดสอบฉบับก่อนตรวจสอบ DIF แสดงให้เห็นว่าในการใช้ผลคะแนนสำหรับตัดสินผลการสอบในแบบทดสอบที่มี DIF ไม่จำเป็นต้องคิดคะแนนการสอบทุกข้อ ก็สามารถวัดองค์ประกอบด้านต่าง ๆ ของผู้เรียนได้ครบถ้วน ดังนั้น เพื่อให้เกิดความยุติธรรมต่อผู้เข้าสอบทุกกลุ่มอย่างเท่าเทียมกัน เราสามารถตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบได้โดยไม่ทำให้โครงสร้างขององค์ประกอบเปลี่ยนแปลงไปจากเดิม

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. จากการศึกษาผู้วิจัยใช้วิธีวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธีลดข้อสอบ โลจิสติก และวิธีชิปเทสท์ปรับใหม่ ซึ่งเป็นวิธีที่เป็นมาตรฐานและเชื่อถือได้ โดยจะถือว่าข้อสอบ ข้อนั้นมี DIF ก็ต่อเมื่อตรงกันทั้งสองวิธี ดังนั้นควรใช้วิธีการตรวจสอบวิธีอื่น ๆ หลาย ๆ วิธีเพื่อเปรียบเทียบว่ามีการพบ DIF ตรงกันกับสองวิธีดังกล่าวข้างต้นหรือไม่ เพื่อที่จะลดความคลาดเคลื่อนประเภทที่ 1 ในการระบุผิดพลาด และเพื่อให้มีความแม่นยำถูกต้อง ควรใช้วิธีการตรวจสอบในกลุ่มที่ใช้ทฤษฎีการตอบสนองข้อสอบ

2. ควรมีการศึกษาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบวิชาอื่น ๆ และเมื่อจำแนกกลุ่มผู้สอบตามตัวแปรต่าง ๆ เช่น ภูมิภาค เน้นศาสนา เชื้อชาติ ภาษา อายุ เป็นต้นว่าจะส่งผลให้แบบทดสอบพบ DIF แตกต่างกันหรือไม่ และถ้าตัดข้อสอบที่พบ DIF ตามตัวแปรดังกล่าวออกทุกข้อแล้วจะส่งผลต่อค่าความเที่ยง ความตรงเชิงโครงสร้าง และความคงที่ของโครงสร้างองค์ประกอบเช่นเดียวกันหรือไม่

3. ในการศึกษาวิจัยเกี่ยวกับความคงที่ของโครงสร้างองค์ประกอบ ผู้วิจัยศึกษาเปรียบเทียบ ความคงที่ของโครงสร้างองค์ประกอบ โดยเปรียบเทียบเฉพาะกลุ่มเพศชายฉบับก่อนตัดข้อสอบที่พบ DIF กับฉบับหลังตัดข้อสอบที่พบ DIF และเพศหญิงฉบับก่อนตัดข้อสอบที่พบ DIF กับฉบับหลังตัดข้อสอบที่พบ DIF ซึ่งผู้วิจัยวิเคราะห์เปรียบเทียบแบบ Single Group จึงควรมีการศึกษาวิจัยเปรียบเทียบหลาย ๆ กลุ่ม แบบ Multiple Group เช่น เพศชายฉบับก่อนตัดข้อสอบที่พบ DIF กับเพศหญิงฉบับก่อนตัดข้อสอบที่พบ DIF หรือเพศชายฉบับหลังตัดข้อสอบที่พบ DIF กับเพศหญิงฉบับหลังตัดข้อสอบที่พบ DIF เป็นต้น โดยวิเคราะห์ในคราวเดียว ด้วยวิธี การวิเคราะห์กลุ่มพหุ