



สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อสังเคราะห์งานวิจัยเชิงปริมาณด้วยวิธีการวิเคราะห์ห่อภิมาณคุณภาพของแบบสอบหลายตัวเลือกทั้งค่าความเที่ยง ความตรง ความยากง่ายและค่าอำนาจจำแนก ที่มีรูปแบบของแบบสอบ บริบทของแบบสอบ วิธีการตรวจให้คะแนนและประเภทของแบบสอบต่างกัน ตามแนวคิดของ Hunter

ผู้วิจัยดำเนินการวิจัยโดยการสืบค้นรายงานการวิจัยเชิงทดลองและรายงานการวิจัยเชิงสหสัมพันธ์ที่ทำการศึกษาเกี่ยวกับคุณภาพของแบบสอบหลายตัวเลือก ที่พิมพ์เผยแพร่ระหว่างปี พ.ศ. 2511 ถึง 2542 จากห้องสมุด 3 แห่ง คือ ศูนย์บรรณสารสนเทศทางการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย สำนักหอสมุดกลาง มหาวิทยาลัยศรีนครินทรวิโรฒและสำนักหอสมุดกลาง มหาวิทยาลัยเกษตรศาสตร์ ได้รายงานการวิจัยที่มีคุณสมบัติครบตามเกณฑ์ที่กำหนดจำนวน 96 เล่ม จำแนกเป็นงานวิจัยเชิงทดลอง ๗2 เล่ม และงานวิจัยเชิงสหสัมพันธ์จำนวน 4 เล่ม

การเก็บรวบรวมข้อมูลเพื่อการวิเคราะห์ห่อภิมาณ ผู้วิจัยใช้เครื่องมือในการเก็บรวบรวมข้อมูล 2 ชุด ประกอบด้วยแบบบันทึกคุณลักษณะงานวิจัยและแบบประเมินคุณภาพงานวิจัย ที่มีข้อกำหนดเกณฑ์ในการให้คะแนนหรือแบบ rubric ผลของการเก็บรวบรวมข้อมูลได้ผลการวิจัยวัดในรูปค่าดัชนีมาตรฐานได้ 4 ชุด จำนวน 1.837 ค่า คือ ค่าความเที่ยง 777 ค่า ค่าความตรงจำนวน 461 ค่า ค่าความยากง่ายจำนวน 283 ค่า และค่าอำนาจจำแนกจำนวน 316 ค่า

การวิเคราะห์ข้อมูลเพื่อการวิเคราะห์ห่อภิมาณแบ่งออกเป็น 2 ตอน ตอนแรกเป็นการวิเคราะห์ข้อมูลเบื้องต้น โดยนำเสนอค่าสถิติพื้นฐานของตัวแปรในการวิจัย และตอนที่สอง เป็นการวิเคราะห์ห่อภิมาณคุณภาพของแบบสอบหลายตัวเลือก แบ่งออกเป็น 4 ส่วน คือ ส่วนแรกเป็นการวิเคราะห์ห่อภิมาณค่าความเที่ยงของแบบสอบหลายตัวเลือก ส่วนที่สองเป็นการวิเคราะห์ห่อภิมาณค่าความตรงของแบบสอบหลายตัวเลือก ส่วนที่สามเป็นการวิเคราะห์ห่อภิมาณค่าความยากง่ายของแบบสอบหลายตัวเลือก และส่วนสุดท้ายเป็นการวิเคราะห์ห่อภิมาณค่าอำนาจจำแนกของแบบสอบหลายตัวเลือก

สรุปผลการวิจัย

ตอนที่ 1 ลักษณะของงานวิจัยที่นำมาสังเคราะห์

1.1 ลักษณะของงานวิจัย

งานวิจัยที่เป็นกลุ่มตัวอย่างในการสังเคราะห์งานวิจัยในครั้งนี้ ส่วนใหญ่เป็นงานวิจัยระดับมหาบัณฑิต (ร้อยละ 96.9) ในจำนวนงานวิจัย 96 เรื่อง พบว่าเป็นผลงานวิทยานิพนธ์ของมหาวิทยาลัยศรีนครินทรวิโรฒ มากกว่าครึ่ง (ร้อยละ 53.1) ส่วนที่เหลือเป็นผลงานวิทยานิพนธ์ของจุฬาลงกรณ์มหาวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์ และแหล่งอื่น (รายงานการวิจัย) ร้อยละ 35.4, 9.4 และ 2.1 ตามลำดับ นอกจากนี้ยังพบว่าเป็นงานวิจัยซึ่งเป็นวิทยานิพนธ์ที่ผลิตโดยผู้วิจัยจากภาควิชาวิจัยการศึกษาถึงร้อยละ 86.5 และร้อยละ 83.3 เป็นงานวิจัยที่ผลิตโดยผู้วิจัยจากสาขาวิชาการวัดและประเมินผลการศึกษา เมื่อจำแนกงานวิจัยตามช่วงปี พบว่างานวิจัยในช่วงระหว่างปีพุทธศักราช 2520-2524 มีมากที่สุดคือร้อยละ 25.0 รองลงมาในช่วงปีพุทธศักราช 2530-2534 (ร้อยละ 19.8) ในส่วนของจำนวนหน้าของงานวิจัยทั้งหมดมีค่าเฉลี่ยเท่ากับ 114.51 หน้า และจำนวนหน้าของงานวิจัยที่ไม่รวมภาคผนวกมีค่าเฉลี่ยเท่ากับ 75.39 หน้า

งานวิจัยส่วนมาก (ร้อยละ 71.9) มีวัตถุประสงค์เพื่อเปรียบเทียบตัวแปรอิสระที่กล่าวถึง มากกว่าเป็นการศึกษาค่าคุณภาพของแบบสอบหรือการศึกษาความสัมพันธ์ระหว่างแบบสอบ (ร้อยละ 26.0) โดยมีการกำหนดสมมติฐานการวิจัยเป็นแบบไม่มีทิศทางถึงร้อยละ 70.8 ในส่วนของกลุ่มตัวอย่างที่ใช้ในการวิจัยพบว่า ขนาดของกลุ่มตัวอย่างในการวิจัยโดยเฉลี่ยเท่ากับ 755.48 คน กลุ่มตัวอย่างที่ใช้เกือบครึ่งมีภูมิลำเนาอยู่ในกรุงเทพมหานครและเขตปริมณฑล (ร้อยละ 49.0) และมีการใช้กลุ่มตัวอย่างที่อยู่ในระดับชั้นมัธยมศึกษาสูงสุด (ร้อยละ 65.8) โดยได้มาจากการสุ่มอย่างง่ายมากที่สุดร้อยละ 43.8 นอกจากนี้ยังพบอีกว่ามีการแบ่งกลุ่มตัวอย่างเพื่อใช้ในการวิจัยโดยการสุ่มอย่างง่ายสูงสุด (ร้อยละ 37.5) สำหรับค่าเฉลี่ยคะแนนการประเมินคุณภาพงานวิจัยมีค่าเท่ากับ 85.57 คะแนน

1.2 ลักษณะการสร้างแบบสอบก่อนนำไปใช้ในการวิจัย

งานวิจัยที่เป็นกลุ่มตัวอย่างในการสังเคราะห์งานวิจัยครั้งนี้จำนวน 96 เรื่อง มีการสร้างแบบสอบก่อนนำไปใช้ในการวิจัยจำนวน 180 ฉบับ พบว่า แบบสอบที่สร้างขึ้นเป็นแบบสอบที่มีลักษณะตัวเลือกถูกตัวเดียวมากกว่าครึ่ง (ร้อยละ 55.6) รองลงมาเป็นแบบสอบตัวเลือกถูกที่สุด (ร้อยละ 23.3) วิชาคณิตศาสตร์เป็นวิชาที่ผู้วิจัยนิยมนำมาสร้างแบบสอบสูงสุด (ร้อยละ 42.2) รองลงมาเป็นวิชาวิทยาศาสตร์และวิชาภาษาไทย (ร้อยละ 24.4 และ 11.7 ตามลำดับ) โดยมีการสร้างแบบสอบที่เป็นแบบรายวิชามากกว่าแบบหัวข้อ/เรื่อง (ร้อยละ 61.7 และ 38.3 ตามลำดับ) นอกจากนี้ยังพบอีกว่าแบบสอบที่ใช้ในการวิจัยนั้นผู้วิจัยเป็นผู้สร้างขึ้นเองสูงถึงร้อยละ 92.2 ส่วนที่เหลือเป็นแบบสอบที่มีผู้อื่นสร้างไว้แล้ว

ต้นฉบับ หน้าขาดหาย

ของแบบสอบทั้งหมด ส่วนตัวแปรตามที่เป็นการตรวจสอบความตรง ความยากง่าย และค่าอำนาจจำแนกของแบบสอบ เป็นจำนวนร้อยละ 59.6, 37.3 และ 42.4 ตามลำดับ โดยมีค่าความเที่ยงเฉลี่ยเท่ากับ .715 ค่าความตรงเฉลี่ยเท่ากับ .524 ค่าความยากง่ายเฉลี่ยเท่ากับ .473 และค่าอำนาจจำแนกเฉลี่ยเท่ากับ .433 นอกจากนี้ยังพบว่าขนาดของกลุ่มตัวอย่างที่ใช้กับแบบสอบฉบับต่างๆ มีค่าเฉลี่ยเท่ากับ 209.28 คนต่อแบบสอบ 1 ฉบับ จำนวนข้อสอบเฉลี่ยเท่ากับ 38.07 ข้อ และเวลาที่ใช้ในการทำแบบสอบเฉลี่ยเท่ากับ 47.78 นาที

1.5 ลักษณะการทดสอบสมมติฐาน

งานวิจัยจำนวน 96 เรื่อง มีการทดสอบสมมติฐานในการวิจัยรวม 2758 สมมติฐาน โดยพบว่าเป็นการทดสอบความแตกต่างของค่าความเที่ยงเป็นรายคู่มากที่สุด ร้อยละ 24.8 และสถิติที่ใช้ในการทดสอบสมมติฐานมากที่สุดคือ สถิติทดสอบซี (Z-test) ร้อยละ 42.2 รองลงมาเป็นสถิติทดสอบไค-สแควร์ (Chi-square) ร้อยละ 38.0 นอกจากนี้ยังพบว่าผลการทดสอบความมีนัยสำคัญทางสถิติที่พบมากที่สุดคือ แตกต่างอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 (ร้อยละ 58.2)

ตอนที่ 2 การวิเคราะห์ห่อภิมาณคุณภาพของแบบสอบหลายตัวเลือก

2.1 ผลการวิเคราะห์ห่อภิมาณค่าความเที่ยงของแบบสอบหลายตัวเลือก

จากการตรวจสอบความแปรปรวนของประชากรของค่าดัชนีมาตรฐานค่าความเที่ยงของแบบสอบที่ทำการปรับแก้ของข้อมูลทั้ง 3 ระดับ คือ ข้อมูลระดับเล่มวิทยานิพนธ์ (.743) ข้อมูลระดับของแบบสอบที่สร้างก่อนนำไปใช้ในการวิจัย (.733) และข้อมูลระดับของค่าคุณภาพของแบบสอบจากการวิจัย (.713) พบว่า ข้อมูลทั้ง 3 ระดับมีค่าความแปรปรวนของความเที่ยงต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งแสดงว่าค่าความเที่ยงของแบบสอบแต่ละระดับ ทั้ง 3 ระดับมีความแตกต่างกัน จึงต้องทำการแยกกลุ่มงานวิจัยตามตัวแปรปรับก่อนการสังเคราะห์ผลการวิจัยต่อไป

ผลการวิเคราะห์ความแปรปรวนของค่าดัชนีมาตรฐานค่าความเที่ยงของแบบสอบ แสดงให้เห็นว่าตัวแปรที่ทำให้คุณภาพของแบบสอบด้านค่าความเที่ยงแตกต่างอย่างมีนัยสำคัญมีจำนวน 11 ตัวแปร จากตัวแปรที่วิเคราะห์ทั้งหมดทั้งหมด 35 ตัวแปร คือ 1) ตัวแปรสถาบันที่ผลิตงานวิจัย 2) วิชาของแบบสอบ 3) การศึกษาหลักสูตร/เนื้อหาวิชา 4) การกำหนดวัตถุประสงค์เชิงพฤติกรรม 5) การเรียงตัวเลือก 6) การเรียงแบบสอบ 7) วิธีการตอบและตรวจให้คะแนนแบบสอบ 8) ประเภทของแบบสอบ 9) ขนาดของกลุ่มตัวอย่างที่ใช้ในการทดสอบรายฉบับ 10) จำนวนข้อสอบที่ใช้ในการวิจัย และ 11) เวลาที่ใช้ในการทำแบบสอบในการวิจัย ซึ่งทุกตัวแปรมีค่า F มีนัยสำคัญที่ระดับ .01 ยกเว้นค่า F สำหรับตัวแปรสถาบันที่ผลิตงานวิจัยและการกำหนดวัตถุประสงค์เชิงพฤติกรรมมีนัยสำคัญที่ระดับ .05

โดยกลุ่มของแบบสอบที่มีค่าเฉลี่ยของค่าความเที่ยงสูงสุด ได้แก่ 1) แบบสอบที่ผลิตโดยมหาวิทยาลัยเกษตรศาสตร์ (.832) 2) แบบสอบวิชาคณิตศาสตร์ (.790) 3) แบบสอบที่มีการศึกษาหลักสูตร/เนื้อหาวิชาก่อนการสร้าง (.749) 4) แบบสอบที่มีการกำหนดวัตถุประสงค์เชิงพฤติกรรมก่อนการสร้างแบบสอบ (.767) 5) แบบสอบที่มีการเรียงตัวเลือกแบบค่ามากไปน้อย (.775) 6) แบบสอบที่มีการเรียงข้อสอบจากพฤติกรรมการเรียนรู้ขั้นต่ำไปขึ้นสูง (818) 7) แบบสอบที่มีวิธีการตอบและตรวจให้คะแนนแบบสอบแบบให้คะแนนรายข้อต่างกันตามความสามารถของผู้ตอบ (.951) 8) แบบสอบประเภทเติมคำ (.852) 9) แบบสอบที่มีการใช้ขนาดของกลุ่มตัวอย่างในการสอบที่มากกว่า 200 คนต่อฉบับ (.761) 10) แบบสอบที่มีจำนวนข้อมากกว่า 40 ข้อ (.788) และ 11) แบบสอบที่ใช้เวลาในการทำมากกว่า 40 นาที (.764)

2.2 ผลการวิเคราะห์ห่อภิมาณค่าความตรงของแบบสอบหลายตัวเลือก

จากการตรวจสอบความแปรปรวนของประชากรของค่าดัชนีมาตรฐานค่าความตรงของแบบสอบที่ทำการปรับแก้ของข้อมูลทั้ง 3 ระดับคือ ข้อมูลระดับเล่มวิทยานิพนธ์ (.578) ข้อมูลระดับของแบบสอบที่สร้างก่อนนำไปใช้ในการวิจัย (.566) และข้อมูลระดับของค่าคุณภาพของแบบสอบจากการวิจัย (.528) พบว่า ข้อมูลทั้ง 3 ระดับมีค่าความแปรปรวนของความตรงต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งแสดงว่าค่าความตรงของแบบสอบแต่ละระดับ ทั้ง 3 ระดับมีความแตกต่างกัน จึงต้องทำการแยกกลุ่มงานวิจัยตามตัวแปรปรับก่อนการสังเคราะห์ผลการวิจัยต่อไป

ผลการวิเคราะห์ความแปรปรวนของค่าดัชนีมาตรฐานค่าความตรงของแบบสอบตามตัวแปรปรับประเภทต่างๆ พบตัวแปรที่ทำให้คุณภาพของแบบสอบด้านค่าความตรงแตกต่างกันอย่างมีนัยสำคัญมีจำนวน 8 ตัวแปร จากทั้งหมด 33 ตัวแปร ได้แก่ 1) ตัวแปรการกำหนดสมมติฐานในการวิจัย 2) เนื้อหาของแบบสอบ 3) รูปแบบตัวลวง 4) การเรียงแบบสอบ 5) วิธีการตอบและตรวจให้คะแนนแบบสอบ 6) ตัวแปรอิสระประเภทที่ 2 7) จำนวนข้อสอบที่ใช้ในการวิจัย และ 8) เวลาที่ใช้ในการทำแบบสอบที่ใช้ซึ่งทุกค่ามีนัยสำคัญที่ระดับ .01 ยกเว้นค่า F สำหรับตัวแปรการกำหนดสมมติฐานในการวิจัย ตัวแปรเนื้อหาของแบบสอบและตัวแปรรูปแบบตัวลวงมีนัยสำคัญที่ระดับ .05

โดยกลุ่มของแบบสอบที่มีค่าเฉลี่ยค่าความตรงสูงสุด ได้แก่ แบบสอบตามลักษณะตัวแปรทั้ง 8 ตัวแปร ดังนี้ 1) แบบสอบของงานวิจัยที่ไม่มีสมมติฐานในการวิจัย (.815) 2) แบบสอบที่สร้างจากเนื้อหาที่เป็นหัวข้อ/เรื่อง (.605) 3) แบบสอบที่มีรูปแบบตัวลวงจากคำตอบผิดของนักเรียน (.691) 4) แบบสอบที่มีการเรียงข้อสอบแบบเรียงจากพฤติกรรมการเรียนรู้ขั้นต่ำไปสูง (.623) 5) แบบสอบที่มีวิธีการตอบและตรวจให้คะแนนแบบสอบด้วยวิธีของสำราญ มีแจ้ง (.756) 6) แบบสอบจากงานวิจัยที่มีตัวแปรอิสระที่ไม่มีการกำหนดเงื่อนไขเพิ่มเติม (.561) 7) แบบสอบที่มีจำนวนข้อมากกว่า 40 ข้อ (.590) และ 8) แบบสอบที่ใช้เวลาในการทำมากกว่า 40 นาที (.565)

2.3 ผลการวิเคราะห์ห่อภิมาณค่าความยากง่ายของแบบสอบหลายตัวเลือก

จากการตรวจสอบความแปรปรวนของประชากรของค่าดัชนีมาตรฐานค่าความยากง่ายของแบบสอบที่ทำการปรับแก้ของข้อมูลทั้ง 3 ระดับ คือ ข้อมูลระดับเล่มวิทยานิพนธ์ (.506) ข้อมูลระดับของแบบสอบที่สร้างก่อนนำไปใช้ในการวิจัย (.497) และข้อมูลระดับของค่าคุณภาพของแบบสอบจากการวิจัย (.474) พบว่า ข้อมูลทั้ง 3 ระดับมีค่าความแปรปรวนของค่าความยากง่ายต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งแสดงว่าค่าความยากง่ายของแบบสอบแต่ละระดับ ทั้ง 3 ระดับมีความแตกต่างกัน จึงต้องทำการแยกกลุ่มงานวิจัยตามตัวแปรปรับก่อนการสังเคราะห์ผลการวิจัยต่อไป

ผลการวิเคราะห์ความแปรปรวนของค่าดัชนีมาตรฐานค่าความยากง่ายของแบบสอบ แสดงให้เห็นว่าตัวแปรที่ทำให้คุณภาพของแบบสอบด้านค่าความยากง่ายแตกต่างกันอย่างมีนัยสำคัญมีจำนวน 7 ตัวแปร จากตัวแปรที่วิเคราะห์ทั้งหมดทั้งหมด 34 ตัวแปร คือ 1) ตัวแปรภูมิถิ่นของ กลุ่มตัวอย่าง 2) การเรียงแบบสอบ 3) วิธีการตอบและตรวจให้คะแนนแบบสอบ 4) ประเภทของแบบสอบ 5) ตัวแปรอิสระประเภทที่ 2 6) ขนาดของกลุ่มตัวอย่างที่ใช้ในการทดสอบรายฉบับ และ 7) จำนวนข้อสอบที่ใช้ในการวิจัย โดยทุกค่ามีนัยสำคัญที่ระดับ .01 ยกเว้นค่า F สำหรับตัวแปรภูมิถิ่นของ กลุ่มตัวอย่างมีนัยสำคัญที่ระดับ .05

โดยที่กลุ่มแบบสอบที่มีค่าเฉลี่ยค่าความยากง่ายอยู่ในระดับที่เหมาะสม ได้แก่ แบบสอบตามลักษณะตัวแปรทั้ง 7 ตัวแปร ดังนี้ 1) แบบสอบที่มีการใช้กลุ่มตัวอย่างที่ไม่ได้อยู่ในเขตกรุงเทพมหานครและปริมณฑล (.472) 2) แบบสอบที่เรียงข้อสอบโดยการสุ่ม (.484) 3) แบบสอบที่มีวิธีการตอบและตรวจให้คะแนนด้วยวิธีแก้การเดา (.497) 4) แบบสอบประเภทหลายตัวเลือก (.538) 5) แบบสอบจากงานวิจัยที่มีตัวแปรอิสระเกี่ยวกับแบบสอบ (.523) 6) แบบสอบที่มีการใช้ขนาดของกลุ่มตัวอย่างที่ใช้ในสอบที่มากกว่า 200 คน (.517) และ 7) แบบสอบที่มีจำนวนข้อมากกว่า 40 ข้อ (.527)

2.4 ผลการวิเคราะห์ห่อภิมาณค่าอำนาจจำแนกของแบบสอบหลายตัวเลือก

จากการตรวจสอบความแปรปรวนของประชากรของค่าดัชนีมาตรฐานค่าอำนาจจำแนกของแบบสอบที่ทำการปรับแก้ของข้อมูลทั้ง 3 ระดับ ข้อมูลระดับเล่มวิทยานิพนธ์ (.419) ข้อมูลระดับของแบบสอบที่สร้างก่อนนำไปใช้ในการวิจัย (.430) และข้อมูลระดับของค่าคุณภาพของแบบสอบจากการวิจัย (.433) พบว่า ข้อมูลทั้ง 3 ระดับมีค่าความแปรปรวนของค่าอำนาจจำแนกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งแสดงว่าค่าอำนาจจำแนกของแบบสอบแต่ละระดับ ทั้ง 3 ระดับมีความแตกต่างกัน จึงต้องทำการแยกกลุ่มงานวิจัยตามตัวแปรปรับก่อนการสังเคราะห์ผลการวิจัยต่อไป

ผลการวิเคราะห์ความแปรปรวนของค่าดัชนีมาตรฐานค่าอำนาจจำแนกของแบบสอบตามตัวแปรปรับประเภทต่างๆ พบว่า ตัวแปรที่ค่า F มีนัยสำคัญทางสถิติมีจำนวน 14 ตัวแปร จากทั้ง

หมด 34 ตัวแปร ได้แก่ 1) ตัวแปรปีที่ทำวิจัย 2) สถาบันที่ผลิตงานวิจัย 3) ภาควิชาที่ผลิตงาน 4) สาขาวิชาที่ผลิตงานวิจัย 5) ระดับการศึกษาของกลุ่มตัวอย่าง 6) ขนาดของกลุ่มตัวอย่างในระดับเล่มวิทยานิพนธ์ 7) วิชาของแบบสอบ 8) เนื้อหาของแบบสอบ 9) การกำหนดวัตถุประสงค์เชิงพฤติกรรม 10) การสร้างตารางแผนผังข้อสอบ 11) รูปแบบตัวลวง 12) วิธีการตอบและตรวจให้คะแนนแบบสอบ 13) ตัวแปรอิสระประเภทที่ 2 และ 14) ขนาดของกลุ่มตัวอย่างที่ใช้ในการทดสอบรายฉบับ โดยทุกค่ามีนัยสำคัญที่ระดับ .01 ยกเว้นค่า F สำหรับตัวแปรวิชาของแบบสอบ รูปแบบตัวลวง มีนัยสำคัญที่ระดับ .05

โดยที่กลุ่มแบบสอบที่มีค่าเฉลี่ยค่าอำนาจจำแนกสูงที่สุด ได้แก่ แบบสอบตามลักษณะตัวแปรทั้ง 14 ตัวแปร ดังนี้ 1) แบบสอบจากงานวิจัยที่ผลิตในปี 2530-2534 (.520) 2) แบบสอบจากงานวิจัยที่ผลิตจากมหาวิทยาลัยศรีนครินทรวิโรฒ (.459) 3) แบบสอบจากงานวิจัยที่ผลิตจากภาควิชาวิจัยการศึกษา (.440) 4) แบบสอบจากงานวิจัยที่ผลิตจากสาขาวิชาการวัดและประเมินผลการศึกษา (.446) 5) แบบสอบที่มีการใช้กลุ่มตัวอย่างระดับชั้นประถมศึกษา (.500) 6) แบบสอบของงานวิจัยที่มีขนาดของกลุ่มตัวอย่างที่ใช้ในการวิจัยในระดับเล่มวิทยานิพนธ์ที่มากกว่า 700 คน (.503) 7) แบบสอบวิชาคณิตศาสตร์ (.463) 8) แบบสอบที่สร้างโดยเนื้อหาที่เป็นหัวข้อ/เรื่อง (.496) 9) แบบสอบที่มีการกำหนดวัตถุประสงค์เชิงพฤติกรรมก่อนการสร้างแบบสอบ (.501) 10) แบบสอบที่ไม่มีการสร้างตารางแผนผังข้อสอบก่อนการสร้างแบบสอบ (.463) 11) แบบสอบที่มีรูปแบบตัวลวงจากคำตอบผิดจากนักเรียน (.550) 12) แบบสอบที่มีวิธีการตอบและตรวจให้คะแนนแบบสอบด้วยวิธีของอนันต์ ศรีโสภา (.693) 13) แบบสอบจากงานวิจัยที่มีตัวแปรอิสระที่เกี่ยวกับแบบสอบ (.488) และ 14) แบบสอบที่มีการใช้ขนาดของกลุ่มตัวอย่างในการสอบที่มากกว่า 200 คน (.498)

2.5 ผลการวิเคราะห์ห่อภิมาณคุณภาพของแบบสอบหลายตัวเลือกตามตัวแปรปรับประเภทต่างๆ ของข้อมูลระดับเล่มวิทยานิพนธ์

2.5.1ปีที่ทำวิจัย สถาบันที่ผลิตงานวิจัย ภาควิชาที่ผลิตงานวิจัย สาขาวิชาที่ผลิตงานวิจัยและประเภทของการวิจัย

ผลการวิเคราะห์ห่อภิมาณพบว่า งานวิจัยที่ทำในช่วงปีที่ต่างกัน ไม่ทำให้ค่าความเที่ยง ค่าความตรงและค่าความยากง่ายแตกต่างกัน แต่มีผลทำให้ค่าเฉลี่ยอำนาจจำแนกของแบบสอบแตกต่างกัน กล่าวคือ งานวิจัยที่ผลิตในช่วงปี 2530-2534 จะให้ค่าอำนาจจำแนกสูงที่สุด และงานวิจัยที่ผลิตในช่วงปี 2525-2529 จะให้ค่าอำนาจจำแนกต่ำที่สุด เช่นเดียวกับภาควิชาและสาขาวิชาที่ผลิตงานวิจัยที่พบว่า ใ้ว่าจะเป็งานวิจัยที่ผลิตโดยภาควิชาวิจัยการศึกษาหรือภาควิชาอื่น หรือเป็งานวิจัยที่ผลิตโดยสาขาวิชาการวัดและประเมินผลการศึกษาหรือสาขาวิชาอื่น ก็ไม่ทำให้ค่าคุณภาพของแบบสอบทั้งค่าความเที่ยง ค่าความตรงและค่าความยากง่ายแตกต่างกัน แต่มีผลทำให้ค่าอำนาจจำแนกแตกต่างกัน คือ งานวิจัยที่ผลิตโดยภาควิชาวิจัยการศึกษาจะมีค่าอำนาจจำแนกสูงกว่างานวิจัย

ที่ผลิตโดยภาควิชาอื่น และงานวิจัยที่ผลิตโดยสาขาวิชาการวัดและประเมินผลการศึกษาให้ค่าอำนาจจำแนกที่สูงกว่างานวิจัยที่ผลิตโดยสาขาวิชาอื่นๆ

ในส่วนของสถาบันที่ผลิตงานวิจัย พบว่า งานวิจัยที่ผลิตจากสถาบันที่ต่างกันไม่ทำให้ค่าความตรงและค่าความยากง่ายแตกต่างกัน แต่ทำให้ค่าความเที่ยงและค่าอำนาจจำแนกแตกต่างกัน โดยที่งานวิจัยจากมหาวิทยาลัยเกษตรศาสตร์จะให้ค่าความเที่ยงสูงสุด และงานวิจัยของมหาวิทยาลัยศรีนครินทรวิโรฒจะให้ค่าอำนาจจำแนกสูงที่สุด นอกจากนี้ยังพบอีกว่า ไม่ว่าจะเป็งานวิจัยที่เป็นวิทยานิพนธ์ระดับมหาบัณฑิต (ปริญญาโท) ระดับคุณวุฒิปบัณฑิต (ปริญญาเอก) หรือรายงานการวิจัย ต่างก็ไม่ทำให้ค่าคุณภาพของแบบสอบทั้งค่าความเที่ยง ค่าความตรง ค่าความยากง่าย และค่าอำนาจจำแนกแตกต่างกัน

2.5.2 วัตถุประสงค์ของการวิจัยและการทดสอบสมมติฐาน

ผลการวิเคราะห์ห้อภิมาณพบว่า ไม่ว่าจะวัตถุประสงค์ของการวิจัยจะเป็นการเปรียบเทียบตัวแปรที่มีเงื่อนไขต่างกัน การศึกษาคุณภาพของแบบสอบหรือการพัฒนาแบบสอบ ไม่ทำให้ค่าคุณภาพของแบบสอบทั้งค่าความเที่ยง ค่าความตรง ค่าความยากง่ายและค่าอำนาจจำแนกแตกต่างกัน ในส่วนของการกำหนดสมมติฐานในการวิจัยพบว่า ไม่ว่าจะเป็นการกำหนดสมมติฐานแบบมีทิศทาง ไม่มีทิศทางหรือไม่มีการกำหนดสมมติฐานในการวิจัยก็ไม่ทำให้ค่าความเที่ยง ค่าความยากง่ายและค่าอำนาจจำแนกแตกต่างกัน แต่ส่งผลให้ค่าความตรงมีความแตกต่างกัน กล่าวคือ งานวิจัยที่ไม่มีการกำหนดสมมติฐานในการวิจัยจะให้ค่าความตรงที่สูงที่สุด

2.5.3 กลุ่มตัวอย่างที่ใช้ในการวิจัย

ผลการวิเคราะห์ห้อภิมาณพบว่า งานวิจัยที่มีระดับการศึกษาของกลุ่มตัวอย่าง ภูมิภาคเาของกุ่มตัวอย่าง วิธีการสุ่มกลุ่มตัวอย่าง วิธีการแบ่งกลุ่มตัวอย่างและขนาดของกลุ่มตัวอย่างที่ต่างกัน ไม่ทำให้ค่าคุณภาพของแบบสอบทั้งค่าความเที่ยง ค่าความตรง ค่าความยากง่ายและค่าอำนาจจำแนกแตกต่างกัน ยกเว้นงานวิจัยที่มีการใช้กลุ่มตัวอย่างในระดับชั้นประถมศึกษาและงานวิจัยที่มีการใช้ขนาดของกลุ่มตัวอย่างที่มากกว่า 700 คน จะทำให้มีค่าอำนาจจำแนกที่สูงที่สุด และงานวิจัยที่ใช้ขนาดของกลุ่มตัวอย่างที่มีภูมิภาคเาในกรุงเทพมหานครและปริมณฑลจะให้ค่าความยากง่ายสูงที่สุด

2.5.4 คะแนนการประเมิน

ผลการวิเคราะห์ห้อภิมาณพบว่า งานวิจัยที่มีคุณภาพการประเมินโดยผู้วิจัยที่ต่างกัน ไม่มีผลทำให้ค่าคุณภาพของแบบสอบทั้งค่าความเที่ยง ค่าความตรง ค่าความยากง่ายและค่าอำนาจจำแนกมีความแตกต่างกัน

ต้นฉบับ หน้าขาดหาย

ต้นฉบับ หน้าขาดหาย

2.7.5 การชี้แจงในการสอบ

ผลการวิเคราะห์ห่อภิมาณคุณภาพของแบบสอบเมื่อมีการชี้แจงในการสอบที่แตกต่างกัน พบว่า รูปแบบของการชี้แจงในการสอบ ไม่ว่าจะป็นแบบแนะนำให้เดา แบบไม่เสนอแนะให้เดา แบบห้ามเดา แบบให้รางวัล และแบบลงโทษ ไม่ทำให้ค่าความเที่ยงของแบบสอบนั้นแตกต่างกัน

2.7.6 การเรียงแบบสอบ

จากผลการวิเคราะห์ห่อภิมาณผลของการเรียงลำดับข้อสอบแบบต่างๆ ที่มีต่อคุณภาพของแบบสอบหลายตัวเลือก ผลปรากฏว่า การเรียงข้อสอบที่ต่างกันมีผลทำให้ค่าความเที่ยง ค่าความตรงและค่าความยากง่ายแตกต่างกัน โดยในส่วนของค่าความเที่ยงและค่าความตรงนั้น พบว่า การเรียงข้อสอบแบบเรียงตามพฤติกรรมการเรียนรู้จากขั้นต่ำไปขั้นสูงจะทำให้มีค่าคุณภาพสูงที่สุด รองลงมาเป็นการเรียงข้อสอบโดยการสุ่ม และการเรียงข้อสอบจากข้อที่ง่ายไปข้อที่ยากตามลำดับ แต่ในส่วนของค่าความยากง่ายนั้น การเรียงข้อสอบโดยการสุ่มจะทำให้ค่าความยากง่ายสูงที่สุด รองลงมาเป็นการเรียงข้อสอบจากข้อที่ง่ายไปยาก และการเรียงข้อสอบตามค่าความยากง่ายแบบหมุนเวียน ตามลำดับ

2.7.7 วิธีการตรวจให้คะแนน

ผลการวิเคราะห์ข้อมูลแสดงให้เห็นว่า วิธีการตรวจให้คะแนนที่ต่างกันมีผลทำให้ค่าคุณภาพของแบบสอบทั้งค่าความเที่ยง ค่าความตรง ค่าความยากง่าย และค่าอำนาจจำแนกแตกต่างกัน กล่าวคือ ในส่วนของค่าความเที่ยงนั้น พบว่า วิธีการตรวจให้คะแนนด้วยวิธีให้คะแนนรายข้อต่างกันตามความสามารถของผู้สอบจะทำให้มีค่าความเที่ยงสูงที่สุด รองลงมาป็นวิธีการตรวจให้คะแนนของเกริกชัย ชวบเจริญและวิธีใช้ค่าน้ำหนักคะแนนที่เหมาะสม ตามลำดับ สำหรับค่าความตรงพบว่า วิธีการตรวจให้คะแนนของสำราญ มีแจ้้ง ให้ค่าความตรงสูงที่สุด รองลงมาป็นวิธีการตรวจให้คะแนนของกิบบอนส์และคณะ และวิธีให้คะแนนรายข้อต่างกันตามความสามารถของผู้สอบ สำหรับค่าอำนาจจำแนกของแบบสอบพบว่า วิธีการตรวจให้คะแนนของอนันต์ ศรีโสภาก ให้ค่าสูงที่สุด รองลงมาคือวิธีของสำราญ มีแจ้้งและวิธีบอกระดับความมั่นใจ ตามลำดับ และในส่วนของค่าความยากง่ายของแบบสอบพบว่า วิธีการตรวจให้คะแนนทุกตัวเลือก ให้ค่าความยากง่ายสูงที่สุด รองลงมาป็นวิธีการตรวจให้คะแนนแบบนับ 2 และวิธีนับ 3 ตามลำดับ

2.7.8 ประเภทของแบบสอบ

ผลการวิเคราะห์ห่อภิมาณคุณภาพของแบบสอบในส่วนของประเภทของแบบสอบพบว่า ประเภทของแบบสอบที่ต่างกันมีผลทำให้ค่าความเที่ยง และค่าความยากง่ายของแบบสอบแตกต่างกัน แต่ไม่มีผลทำให้ค่าความตรงและค่าอำนาจจำแนกแตกต่างกัน กล่าวคือ แบบสอบแบบเดิมค่าจะให้ค่าความเที่ยงสูงที่สุด รองลงมาป็นแบบสอบอัตรนัย แบบสอบโคลซ และแบบสอบหลายตัวเลือก ตามลำดับ ในส่วนของค่าความยากง่าย พบว่า แบบสอบแบบถูกผิด จะให้ค่าความยาก

ง่ายที่สุด รองลงมาเป็นแบบสอบโคลซ แบบสอบหลายตัวเลือกและแบบสอบแบบเติมคำ ตามลำดับ

2.7.9 ตัวแปรอิสระประเภทที่ 2

ผลการวิเคราะห์ห่อภิมาณพบว่า งานวิจัยที่มีตัวแปรอิสระประเภทที่ 2 ที่ต่างกัน ไม่ทำให้ค่าความเที่ยงต่างกัน แต่มีผลทำให้ค่าความตรง ค่าความยากง่ายและค่าอำนาจจำแนกต่างกัน กล่าวคือ งานวิจัยที่มีตัวแปรอิสระประเภทที่ 2 ที่ไม่มีการกำหนดเงื่อนไขเพิ่มเติมจะให้ค่าความตรงสูงสุด และตัวแปรอิสระประเภทที่ 2 ที่เกี่ยวกับแบบสอบจะทำให้ค่าความยากง่ายและค่าอำนาจจำแนกสูงสุด และงานวิจัยที่มีตัวแปรอิสระประเภทที่ 2 ที่เป็นเรื่องเกี่ยวกับผู้สอบ จะให้ค่าความตรง ค่าความยากง่ายและค่าอำนาจจำแนกต่ำที่สุด

2.7.10 ขนาดของกลุ่มตัวอย่างที่ใช้ในการทดสอบรายฉบับ จำนวนข้อสอบและเวลาที่ใช้ในการทำแบบสอบ

ผลการวิเคราะห์ห่อภิมาณพบว่า งานวิจัยที่มีขนาดของกลุ่มตัวอย่างที่ใช้ในการทดสอบรายฉบับที่แตกต่างกัน มีผลทำให้ค่าความเที่ยง ค่าความยากง่ายและค่าอำนาจจำแนกต่างกัน แต่ไม่มีผลทำให้ค่าความตรงแตกต่างกัน กล่าวคือ งานวิจัยที่มีการใช้ขนาดของกลุ่มตัวอย่างที่มากกว่า 200 คน จะให้ค่าความเที่ยง ค่าความยากง่ายและค่าอำนาจจำแนกสูงกว่างานวิจัยที่มีการใช้ขนาดกลุ่มตัวอย่างที่น้อยกว่าหรือเท่ากับ 200 คน สำหรับจำนวนของข้อสอบที่ใช้ในการวิจัยพบว่า งานวิจัยที่มีจำนวนของข้อสอบที่ต่างกัน มีผลทำให้ค่าความเที่ยง ค่าความตรงและค่าความยากง่ายต่างกัน แต่ไม่มีผลทำให้ค่าอำนาจจำแนกต่างกัน กล่าวคือ งานวิจัยที่มีจำนวนข้อสอบที่มากกว่า 40 ข้อ จะให้ค่าความเที่ยง ค่าความตรง และค่าความยากง่ายที่สูงกว่างานวิจัยที่มีจำนวนข้อสอบน้อยกว่าหรือเท่ากับ 40 ข้อ นอกจากนี้ยังพบอีกว่า งานวิจัยที่มีการกำหนดเวลาในการทำแบบสอบที่ต่างกัน ก็มีผลให้ค่าความเที่ยงและค่าความตรงแตกต่างกัน แต่ไม่ส่งผลให้ค่าความยากง่ายและค่าอำนาจจำแนกแตกต่างกัน กล่าวคือ งานวิจัยที่มีการกำหนดเวลาในการทำแบบสอบที่มากกว่า 40 นาที จะให้ค่าความเที่ยง และค่าความตรงที่สูงกว่างานวิจัยที่มีการกำหนดเวลาที่น้อยกว่าหรือเท่ากับ 40 นาที

อภิปรายผลการวิจัย

การวิจัยครั้งนี้สามารถอภิปรายผลการวิจัยในประเด็นข้อค้นพบต่างๆ ได้ดังนี้

1. ลักษณะงานวิจัยที่นำมาสังเคราะห์

1.1 ช่วงเวลาในการทำวิจัย

จากผลการวิเคราะห์ข้อมูล พบว่า งานวิจัยที่ผลิตในปี 2520-2524 มีมากที่สุด ร้อยละ 25.0 และมีแนวโน้มที่จะลดลงเรื่อยๆ จนกระทั่งปี 2540-2542 มีเพียงร้อยละ 7.3 ซึ่งอาจจะเนื่องมาจากว่า มีการศึกษาวิจัยเกี่ยวกับแบบสอบหลายตัวเลือกมาก ตั้งแต่ยุคเริ่มแรกที่แบบสอบหลายตัวเลือกแพร่หลายและเป็นที่นิยมใช้กันอย่างมาก โดยจะเห็นได้อย่างชัดเจนว่าในช่วงระหว่างปี 2520-2534 มี

งานวิจัยเกี่ยวกับคุณภาพแบบสอบหลายตัวเลือกถึงร้อยละ 63.6 จนอาจจะทำให้ได้ข้อค้นพบจากการวิจัยมากเพียงพอจึงทำการศึกษาวิจัยในด้านอื่นๆ แทน ประกอบกับในปัจจุบันมีรูปแบบของการวัดและประเมินผลการศึกษาของผู้เรียนเพิ่มมากขึ้นหลายประเภท อาทิ การประเมินด้วยแฟ้มสะสมงาน แบบประเมินตนเอง เป็นต้น จึงทำให้การวัดผลการเรียนรู้ด้วยแบบสอบหลายตัวเลือกถูกลดบทบาทลงไปบ้าง แต่ถึงอย่างไร ในปัจจุบันแบบสอบหลายตัวเลือกก็ยังคงเป็นเครื่องมือที่สำคัญประเภทหนึ่งในกระบวนการวัดและประเมินผลผู้เรียน สาเหตุที่ทำให้แบบสอบหลายตัวเลือกเป็นที่นิยมใช้อย่างกว้างขวาง ก็เนื่องมาจากสามารถใช้วัดคุณลักษณะของมนุษย์ได้อย่างกว้างขวาง รูปแบบของข้อสอบเปิดโอกาสให้สามารถถ่มเนื้อเรื่องของสิ่งที่ต้องการวัดได้อย่างครอบคลุม ประกอบกับมีความเป็นปรนัยในการตรวจให้คะแนน ใช้เวลาตรวจน้อยและสามารถใช้เครื่องคอมพิวเตอร์ตรวจได้ (ศิริชัย กาญจนวาสี, 2542) ซึ่งจากจุดเด่นของแบบสอบหลายตัวเลือกข้างต้นนี้ ทำให้แบบสอบประเภทนี้ยังคงถูกใช้เป็นเครื่องมือในการวัดผลการเรียนรู้ที่ศึกษาค้นคว้าในปัจจุบัน

1.2 ลักษณะของกลุ่มตัวอย่างที่ใช้ในการวิจัย

เมื่อพิจารณาถึงระดับชั้นเรียนของกลุ่มตัวอย่าง พบว่า กลุ่มตัวอย่างที่อยู่ในระดับชั้นมัธยมศึกษาที่มีมากที่สุด ร้อยละ 65.8 และกลุ่มตัวอย่างที่อยู่ในระดับประถมศึกษาที่มีเพียงร้อยละ 17.7 เท่านั้น อาจเนื่องมาจากกลุ่มตัวอย่างที่อยู่ในระดับประถมศึกษานั้นยังมีความสามารถทางการอ่านที่ไม่ดีพอประกอบกับกลุ่มตัวอย่างในระดับชั้นมัธยมศึกษาเป็นเด็กที่อยู่ในวัยที่มีความพร้อมและมีความสามารถทั้งทางการอ่าน การเขียนหรือการสื่อสารมากกว่าเด็กในระดับชั้นประถมศึกษา ตลอดจนอาจเนื่องมาจากหน่วยงานต้นสังกัดของผู้วิจัยที่พบว่า ผู้วิจัยอยู่ในสังกัดกรมสามัญศึกษา มากที่สุด ร้อยละ 26.0 ซึ่งเป็นหน่วยงานที่เกี่ยวข้องกับการจัดการศึกษาในระดับมัธยมศึกษา จึงใช้กลุ่มตัวอย่างนี้เพื่อความสะดวกในการเก็บรวบรวมข้อมูล ดังนั้นจึงมีความสอดคล้องกับสภาพความเป็นจริงที่ว่าผู้วิจัยให้ความสนใจที่จะทำวิจัยเกี่ยวกับสายงานหรือหน่วยงานที่ตนสังกัดอยู่ นอกจากนี้ยังพบอีกว่ากลุ่มตัวอย่างที่อยู่ในระดับอุดมศึกษานั้นมีเพียงร้อยละ 12.5 เท่านั้น ก็เนื่องมาจากว่าในระดับอุดมศึกษาจะมีการใช้แบบสอบอัตนัยในการวัดผลการเรียนรู้มากขึ้น และใช้แบบสอบแบบหลายตัวเลือกลดลง ทั้งนี้เพื่อจะได้วัดระดับความรู้ของผู้เรียนได้ลึกซึ้งมากยิ่งขึ้น

1.3 จำนวนหน้าของงานวิจัย

เมื่อพิจารณาถึงจำนวนหน้าของงานวิจัยทั้งจำนวนหน้าทั้งหมด และจำนวนหน้าที่ไม่รวมกับภาคผนวก จะพบว่ามีจำนวนหน้าค่าสุดอยู่ที่ 14 หน้า เนื่องมาจากงานวิจัยที่มีจำนวน 14 หน้า ที่มีจำนวน 1 เล่มนั้นเป็นรายงานการวิจัยที่ตีพิมพ์ในวารสารจึงทำให้มีจำนวนหน้าที่น้อย เมื่อไม่รวมงานวิจัยเล่มนี้ จำนวนหน้าทั้งหมดและจำนวนหน้าที่ไม่รวมกับภาคผนวกค่าสุดอยู่ที่ 46 และ 30 หน้า มีค่าเฉลี่ยเท่ากับ 115.57 และ 76.03 หน้า ตามลำดับ

2. กระบวนการสร้างแบบสอบ

2.1 วิชาของแบบสอบที่ใช้ในการวิจัย

เมื่อพิจารณาถึงวิชาของแบบสอบที่ใช้ในการวิจัยมากที่สุด คือ วิชาคณิตศาสตร์ ร้อยละ 42.2 รองลงมาเป็นวิชาวิทยาศาสตร์ ร้อยละ 24.4 เหตุที่วิชาคณิตศาสตร์ได้รับความนิยมนำมาใช้มาก อาจเนื่องมาจากธรรมชาติของวิชาคณิตศาสตร์ที่อาจไม่มีปัญหาเรื่องความกำกวมของโจทย์และคำตอบ อีกทั้งวิชาคณิตศาสตร์ยังเป็นวิชาที่สำคัญ เป็นรากฐานของวิชาต่างๆ ทั้งทางด้านวิทยาศาสตร์และสังคมศาสตร์ (ยุพิน พิพิธกุล, 2523 อ้างถึงใน ประเสริฐ เศษะนาราเกียรติ, 2531) จึงเป็นที่นิยมใช้วัดผู้เรียนมากที่สุด ในส่วนของสาขาวิชาอื่นๆ พบว่า ยังคงมีจำนวนแบบสอบที่น้อย เพียงร้อยละ 20.7 เท่านั้น ส่วนใหญ่จะเป็นวิชาทางด้านสังคมศาสตร์ แสดงว่าการพัฒนาแบบสอบของไทยยังไม่ใคร่สมคูล ยังขาดการพัฒนาแบบสอบวิชาทางด้านสังคมศาสตร์ อย่างไรก็ตามข้อสรุปนี้ยังไม่สามารถยืนยันได้อย่างชัดเจน จึงควรได้มีการวิจัยเพื่อสังเคราะห์งานวิจัยเกี่ยวกับแบบสอบหลายตัวเลือกในภาพรวมทั้งหมดต่อไปอีก เพื่อจะได้ข้อสรุปที่ชัดเจนมากยิ่งขึ้น

2.2 จำนวนผู้เชี่ยวชาญในการตรวจสอบความตรง

เมื่อพิจารณาถึงจำนวนผู้เชี่ยวชาญในการตรวจสอบความตรงในขั้นตอนการสร้างแบบสอบ พบว่า มีการใช้ผู้เชี่ยวชาญในการตรวจสอบความตรงโดยเฉลี่ยเท่ากับ 6 คน เป็นผู้เชี่ยวชาญทางด้าน การวัดผลการศึกษาโดยเฉลี่ย 2 คนและเป็นผู้เชี่ยวชาญเฉพาะสาขาวิชาโดยเฉลี่ย 4 คน ซึ่งก็เป็นไปในแนวทางที่เหมาะสมที่มีการใช้จำนวนผู้เชี่ยวชาญเฉพาะสาขาวิชามากกว่าเป็นจำนวน 2 เท่าของผู้เชี่ยวชาญด้านการวัดผลการศึกษา เพราะในการตรวจสอบความตรงของแบบสอบที่สร้างขึ้นนั้น ส่วนมากจะเป็นการตรวจสอบความตรงตามเนื้อหาว่า ได้ทำการสร้างข้อสอบได้ครอบคลุมเนื้อหาวิชาและวัดได้ตรงตามวัตถุประสงค์หรือไม่ รวมทั้งตรวจสอบความชัดเจนของข้อคำถามและคำตอบในแต่ละข้อซึ่ง ในส่วนนี้ผู้เชี่ยวชาญเฉพาะสาขาก็น่าจะมีส่วนสำคัญมากกว่าผู้เชี่ยวชาญด้านการวัดผลการศึกษา เนื่องมาจากวิสัยทัศน์ทางด้านเนื้อหาวิชามากกว่า แต่ถึงอย่างไรผู้เชี่ยวชาญทางด้าน การวัดผลการศึกษา ก็มีส่วนสำคัญในการพิจารณาตรวจสอบว่าข้อสอบที่สร้างขึ้นมีความครอบคลุมเนื้อหาวิชาและสามารถวัดได้ตรงตามวัตถุประสงค์การเรียนรู้หรือวัตถุประสงค์เชิงพฤติกรรมที่ตั้งไว้ก่อนการสอนหรือไม่ร่วมกับผู้เชี่ยวชาญเฉพาะสาขาวิชาด้วย

2.3 จำนวนข้อสอบ

เมื่อพิจารณาถึงจำนวนข้อสอบที่มีการสร้างขึ้นเบื้องต้นว่ามีจำนวนโดยเฉลี่ยเท่ากับ 60 ข้อ แต่เมื่อผ่านกระบวนการทดลองใช้แบบสอบ (try out) ในครั้งที่ 1 พบว่า มีแบบสอบที่มีคุณภาพเหมาะสมผ่านเกณฑ์การคัดเลือกโดยเฉลี่ยเท่ากับ 45 ข้อ หรือคิดเป็นร้อยละ 75 ของจำนวนข้อสอบที่สร้างไว้เบื้องต้น และยังพบอีกว่าหลังจากการทดลองใช้แบบสอบในครั้งที่ 2 แล้ว มีแบบสอบที่ผ่านการคัดเลือกโดยเฉลี่ยเท่ากับ 35 ข้อ หรือคิดเป็นร้อยละ 58.33 ของข้อสอบที่สร้างในเบื้องต้น จำนวนข้อสอบที่ผ่านการทดลองใช้แบบสอบในครั้งที่ 2 นี้ มีจำนวนข้อสอบที่ถูกตัดออกเนื่องจากมี

คุณภาพที่ไม่เหมาะสมจากจำนวนข้อสอบที่สร้างไว้ในเบื้องต้นประมาณร้อยละ 40 หรือเรียกได้ว่าเกือบครึ่งหนึ่งของข้อสอบที่ได้สร้างไว้ในครั้งแรก ซึ่งสอดคล้องกับข้อเสนอแนะของศิริชัย กาญจนวาสี (2542) ที่ได้กล่าวว่า ควรมีการเขียนข้อสอบเพื่อไว้ประมาณ 1-2 เท่าของจำนวนที่ต้องการใช้จริง เพื่อไว้สำหรับการตัด/ปรับปรุงข้อไม่ดีและเลือกข้อที่มีคุณภาพดีไว้ใช้ต่อไป ฉะนั้นในการที่จะสร้างแบบสอบ ควรที่จะมีการนำข้อเสนอแนะดังกล่าวข้างต้นไปพิจารณาใช้ เพื่อจะได้ข้อสอบที่มีคุณภาพดีครบถ้วนตามจำนวนที่ต้องการ

2.4 วิธีการที่ใช้ในการวิเคราะห์ข้อสอบ

จากผลการวิเคราะห์ข้อมูลพบว่า วิธีการที่ใช้ในการวิเคราะห์ข้อสอบในขั้นตอนของการทดลองใช้แบบสอบ (try out) ก่อนนำไปใช้ในการวิจัย เป็นวิธีการตามทฤษฎีการทดสอบแบบดั้งเดิม (CTT) มากถึงร้อยละ 86.1 และใช้วิธีตามทฤษฎีการตอบสนองข้อสอบ (IRT) เพียงร้อยละ .6 เท่านั้น ทั้งนี้อาจเนื่องมาจากความจำกัดของวิธีการตอบสนองข้อสอบที่ต้องใช้ขนาดของกลุ่มตัวอย่างที่มาก จึงทำให้มีผู้ใช้วิธีนี้ในการวิเคราะห์ข้อสอบในขั้นตอนของการทดลองใช้แบบสอบเป็นจำนวนน้อย แต่ถึงอย่างไรวิธีการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบก็เป็นวิธีการที่ให้สารสนเทศของข้อสอบมากกว่าวิธีการวิเคราะห์ด้วยทฤษฎีแบบดั้งเดิม เช่น ค่าโอกาสการตอบข้อสอบถูก การประมาณค่าความสามารถผู้สอบ หรือการประมาณสารสนเทศของข้อสอบ เป็นต้น จึงควรที่จะมีการนำวิธีการนี้มาใช้ในขั้นตอนของการทดลองใช้แบบสอบให้มากขึ้นกว่าเดิมที่ยังคงพบน้อยมาก

3. ค่าคุณภาพของแบบสอบหลายตัวเลือก

3.1 ค่าความเที่ยง

ค่าคุณภาพของแบบสอบจากการวิจัย ในส่วนของค่าความเที่ยงจำนวน 778 ค่า ในระดับค่าคุณภาพของแบบสอบพบว่า มีค่าเฉลี่ยเท่ากับ .715 จัดว่าแบบสอบโดยภาพรวมมีค่าความเที่ยงอยู่ในระดับปานกลาง แต่เมื่อพิจารณาตามหลักเกณฑ์ของการวัดและประเมินผลการศึกษาที่ระบุว่าค่าความเที่ยงที่จัดว่าอยู่ในระดับสูงควรมีค่าตั้งแต่ .80 ขึ้นไป เมื่อพิจารณาค่าคุณภาพจากแบบสอบพบว่า มีค่าความเที่ยงที่ผ่านหลักเกณฑ์ดังกล่าวจำนวน 287 ค่า (จากทั้งหมด 778 ค่า) หรือคิดเป็นร้อยละ 36.86 สามารถสรุปได้ว่าแบบสอบที่ได้จากการวิจัยมีคุณภาพอยู่ในระดับสูงประมาณ 1 ใน 3 ของแบบสอบทั้งหมด นอกจากนี้ยังพบอีกว่ามีค่าความเที่ยงที่ได้จากการวิจัยมีค่าเป็นลบจำนวน 1 ค่า ซึ่งไม่น่าจะเป็นไปได้ เพราะพิสัยของค่าความเที่ยงจะมีค่าอยู่ระหว่าง 0 ถึง 1 เท่านั้น

3.2 ค่าความตรง

สำหรับค่าความตรง จำนวน 464 ค่า ในระดับค่าคุณภาพของแบบสอบพบว่า มีค่าเฉลี่ยเท่ากับ .524 จัดว่าแบบสอบมีค่าความตรงอยู่ในระดับปานกลาง เมื่อพิจารณาถึงค่าความตรงตามที่มีค่ามากกว่า .80 ขึ้นไป ตามเกณฑ์แบบค่าความเที่ยง พบว่ามีจำนวน 32 ค่า (จากทั้งหมด 464 ค่า) หรือ

คิดเป็นร้อยละ 6.89 ที่มีค่ามากกว่า .80 สรุปได้ว่ามีแบบสอบที่มีค่าความตรงอยู่ในระดับคืออยู่ประมาณ 1 ใน 15 ของแบบสอบทั้งหมด ซึ่งถือว่ายังเป็นจำนวนที่น้อย

และถึงแม้ว่าข้อมูลในระดับค่าคุณภาพของแบบสอบจะเป็นการตรวจสอบคุณภาพด้านค่าความตรงคิดเป็นร้อยละ 59.6 (464 ค่า จาก 778 ค่า) แต่ก็พบว่าในขั้นตอนของการสร้างแบบสอบก่อนนำมาใช้ในการวิจัยนั้น ผู้วิจัยส่วนใหญ่ได้ทำการตรวจสอบความตรงโดยใช้ผู้เชี่ยวชาญแล้ว ซึ่งถือว่าเป็นการตรวจสอบความตรงตามเนื้อหา จึงไม่ได้มีการรายงานค่าความตรงที่เป็นตัวเลข

3.3 ค่าความยากง่าย

ค่าคุณภาพของแบบสอบที่ได้จากผลการวิจัยที่เป็นค่าความยากง่ายจำนวน 284 ค่า ในระดับค่าคุณภาพของแบบสอบพบว่า มีค่าเฉลี่ยเท่ากับ .473 จัดว่าโดยภาพรวมแบบสอบมีค่าความยากง่ายอยู่ในระดับที่พอเหมาะ ไม่ยากหรือง่ายจนเกินไป เพราะมีค่าอยู่ในช่วงระหว่าง .40 ถึง .59 ซึ่งถือว่ามีระดับความยากง่ายอยู่ในระดับปานกลาง แต่เมื่อพิจารณาตามหลักเกณฑ์ของการวัดและประเมินผลการศึกษาที่ใช้โดยทั่วไปที่ระบุว่าค่าความยากง่ายควรมีค่าอยู่ในช่วงระหว่าง .20 ถึง .80 จะพบได้ว่าแบบสอบที่มีค่าความยากง่ายที่อยู่ในช่วงดังกล่าวจำนวน 270 ค่า (จากทั้งหมด 284 ค่า) หรือคิดเป็นร้อยละ 97.07 หรือเกือบทั้งหมดของแบบสอบที่มีค่าความยากง่ายที่เหมาะสม

3.4 ค่าอำนาจจำแนก

ในส่วนของคุณค่าอำนาจจำแนกของแบบสอบจำนวน 317 ค่า ในระดับค่าคุณภาพของแบบสอบพบว่า มีค่าเฉลี่ยเท่ากับ .433 จัดว่าแบบสอบที่ได้จากผลการวิจัยในภาพรวมมีค่าอำนาจจำแนกอยู่ในระดับดีมาก ตามเกณฑ์การแปลความหมายของ Ebel (1972) ที่กล่าวว่า ข้อสอบที่มีค่าอำนาจจำแนกตั้งแต่ .40 ขึ้นไป ถือว่าเป็นข้อสอบที่ดีมาก และเมื่อพิจารณาตามเกณฑ์ของค่าอำนาจจำแนกที่ใช้โดยทั่วไปที่ระบุว่าข้อสอบที่ดีควรมีค่าอำนาจจำแนกที่มากกว่า .20 ก็พบว่า มีแบบสอบที่อยู่ในเกณฑ์ดังกล่าวทั้งหมด 305 ค่า (จากทั้งหมด 317 ค่า) หรือคิดเป็นร้อยละ 96.21 ซึ่งกล่าวได้ว่าแบบสอบส่วนใหญ่มีค่าอำนาจจำแนกอยู่ในระดับที่พอใช้ได้

3.5 จากผลการวิเคราะห์ข้อมูลพบว่า ลักษณะการแจกแจงของค่าความเที่ยงและค่าความตรงมีลักษณะเบ้ซ้าย แสดงว่าค่าส่วนใหญ่กระจุกตัวไปทางค่าสูง แต่ลักษณะการแจกแจงค่าความยากง่ายและค่าอำนาจจำแนกมีลักษณะเบ้ขวา ค่าส่วนใหญ่กระจุกตัวไปทางค่าต่ำ ซึ่งโดยตามทฤษฎีแล้วข้อสอบที่มีค่าความยากง่ายในระดับที่พอเหมาะ และมีค่าอำนาจจำแนกที่สูงจะทำให้ค่าความเที่ยงและค่าความตรงสูงด้วย จากผลการวิจัยที่พบว่าค่าความยากง่ายและค่าอำนาจจำแนกส่วนใหญ่อยู่ในระดับที่ต่ำ แต่ค่าความเที่ยงและค่าความตรงส่วนใหญ่อยู่ในระดับที่สูง ที่เป็นเช่นนี้ก็เนื่องมาจากว่าค่าความยากง่ายและค่าอำนาจจำแนกนั้นเป็นค่าที่ได้จากการเฉลี่ยของค่าคุณภาพรายข้อให้เป็นรายฉบับ ค่าคุณภาพรายข้อที่มีค่าสูงหรือค่าที่ต่ำมากๆ จึงเกิดการชดเชยภายในซึ่งกันและกัน ส่งผลให้ค่าคุณภาพของแบบสอบทั้งค่าความยากง่ายและค่าอำนาจจำแนกมีค่าที่ต่ำ

4. ผลการวิเคราะห์ห่อภิมาณคุณภาพของแบบสอบหลายตัวเลือก

4.1 เมื่อพิจารณาค่าความแปรปรวนของประชากร (σ^2_p) ของค่าดัชนีมาตรฐานค่าอำนาจจำแนกที่พบว่า ส่วนใหญ่มีค่าที่ต่ำและคิดลบมากกว่าค่าความแปรปรวนของประชากรของค่าดัชนีมาตรฐานค่าความเที่ยง ค่าความตรง และค่าดัชนีมาตรฐานค่าความยากง่าย จึงเป็นสาเหตุให้ค่าความแปรปรวนของประชากรของค่าดัชนีมาตรฐานแตกต่างจากศูนย์อย่างไม่มีนัยสำคัญนั้นมีจำนวนมากที่สุด คือ ร้อยละ 39.64 (44 ตัวแปร จากจำนวนตัวแปรปรับทั้งหมด 111 ตัวแปร) ส่วนค่าความแปรปรวนของประชากรของค่าความเที่ยง ค่าความตรงและค่าความยากง่ายที่แตกต่างจากศูนย์อย่างไม่มีนัยสำคัญ มีจำนวนร้อยละ 25.90 (36 ตัวแปร จากจำนวนตัวแปรปรับทั้งหมด 139 ตัวแปร), ร้อยละ 17.39 (20 ตัวแปร จากจำนวนตัวแปรปรับทั้งหมด 115 ตัวแปร) และร้อยละ 24.76 (26 ตัวแปร จากจำนวนตัวแปรปรับทั้งหมด 105 ตัวแปร) ตามลำดับ ที่เป็นเช่นนี้อาจเนื่องมาจากพิสัยของค่าดัชนีมาตรฐานค่าอำนาจจำแนก (.713) ที่มีลักษณะที่แคบกว่าพิสัยของค่าความเที่ยง ค่าความตรงและค่าความยากง่าย (.979, .945 และ .771 ตามลำดับ) จึงเป็นผลให้ส่วนเบี่ยงเบนมาตรฐานมีค่าน้อย และส่งผลทำให้ค่าความแปรปรวนมีค่าน้อยตามไปด้วย จึงทำให้ค่าความแปรปรวนของประชากรของค่าดัชนีมาตรฐานค่าอำนาจจำแนกมีค่าแตกต่างจากศูนย์อย่างไม่มีนัยสำคัญเป็นจำนวนมากที่สุด

4.2 จากผลการวิเคราะห์ห่อภิมาณคุณภาพของแบบสอบหลายตัวเลือกเกี่ยวกับลักษณะของตัวเลือกที่พบว่า ลักษณะของตัวเลือกที่แตกต่างกัน เช่น ตัวเลือกปลายเปิด “ไม่มีข้อถูก” ตัวเลือกปลายปิด “ถูกทุกข้อ” หรือตัวเลือกธรรมดา เป็นต้น ไม่ทำให้ค่าคุณภาพของแบบสอบหลายตัวเลือกแตกต่างกัน ไม่ว่าจะเป็ค่าความเที่ยง ค่าความตรง ค่าความยากง่ายและค่าอำนาจจำแนก ซึ่งสอดคล้องกับผลการวิเคราะห์ห่อภิมาณของ Knowle และ Welch (1992) ที่พบว่า การใช้ตัวเลือกปลายเปิด “ไม่มีข้อถูก” ไม่ได้ทำให้ค่าความยากง่ายเพิ่มขึ้นและค่าอำนาจจำแนกลดลงกว่าการใช้แบบสอบตัวเลือกธรรมดา

4.3 ผลการวิเคราะห์ห่อภิมาณเกี่ยวกับจำนวนของตัวเลือกในแบบสอบหลายตัวเลือกที่พบว่า แบบสอบที่มี 2, 3, 4, 5 หรือ 6 ตัวเลือก ไม่ทำให้ค่าความเที่ยง ค่าความตรง ค่าความยากง่ายและค่าอำนาจจำแนกแตกต่างกัน ซึ่งต่างจากการวิเคราะห์ห่อภิมาณของ Aamod และ McShane (1992) ที่พบว่า แบบสอบที่มี 3 ตัวเลือกมีค่าความง่ายกว่าแบบสอบ 4 ตัวเลือก อาจเนื่องมาจากถ้ามีการสร้างตัวลวงที่มีคุณภาพดีเพียงพอที่สามารถดึงดูดให้ผู้ที่ไม่มีความรู้เลือกตัวเลือกนั้นๆ ได้ แบบสอบที่มี 2 ถึง 6 ตัวเลือก ก็ไม่น่าจะมีค่าคุณภาพที่แตกต่างกัน แต่การที่จะกำหนดจำนวนตัวเลือกที่เหมาะสมในการวัดผลการเรียนรู้ นั้น จะต้องทำการพิจารณาทั้งผู้สอบและแบบสอบ เพื่อให้ได้แบบสอบหลายตัวเลือกที่เหมาะสมกับผู้สอบในการวัดคุณลักษณะที่ต้องการได้อย่างน่าเชื่อถือ (ศิริชัย กาญจนวาสี, 2542)

4.4 ผลการวิเคราะห์ห่อภิมาณเกี่ยวกับการเรียงแบบสอบที่พบว่า การเรียงแบบสอบโดยการสุ่มจะให้ค่าความยากง่ายสูงสุด รองลงมาเป็นการเรียงแบบสอบจากข้อที่ง่ายไปยาก การเรียงแบบสอบตามค่าความยากง่ายแบบหมุนเวียน แบบเรียงตามพฤติกรรมกรเรียนรู้จากขั้นต่ำไปสูง แบบเรียงสลับพฤติกรรมกรเรียนรู้ และแบบเรียงตามพฤติกรรมกรเรียนรู้จากขั้นสูงมาต่ำ นั้นหมายความว่า การเรียงแบบสอบโดยการสุ่มจะทำให้แบบสอบมีความง่ายมากที่สุด รองลงมาเป็นการเรียงแบบสอบแบบเรียงจากข้อที่ง่ายไปยาก และการเรียงแบบสอบตามค่าความยากง่ายแบบหมุนเวียน ตามลำดับ ซึ่งต่างจากผลการวิเคราะห์ห่อภิมาณของ Aamod และ McShane (1992) ที่พบว่า การเรียงแบบเรียงจากข้อที่ง่ายไปยาก มีความง่ายมากกว่าการเรียงแบบสอบโดยการสุ่ม ผลการวิเคราะห์ห่อภิมาณที่ไม่สอดคล้องกันนั้น อาจเนื่องมาจากลักษณะของกลุ่มตัวอย่างที่ใช้ในการวิจัยมีระดับความสามารถที่ต่างกัน เพราะ งานวิจัยที่นำมาสังเคราะห์นั้นมีการใช้กลุ่มตัวอย่างจากหลายจังหวัด เช่น กรุงเทพมหานคร ลพบุรี ขอนแก่น นครสวรรค์ นครศรีธรรมราช เป็นต้น จึงทำให้กลุ่มตัวอย่างที่ใช้มีความสามารถที่แตกต่างกัน ประกอบกับผลของการวิเคราะห์ความแปรปรวนของค่าดัชนีมาตรฐานค่าความยากง่ายในตัวแปรปรับภูมิสำเนาของกลุ่มตัวอย่างที่พบว่า ค่าความยากง่ายของแบบสอบจากกลุ่มตัวอย่างที่มีภูมิสำเนาที่ต่างกัน คือ กลุ่มตัวอย่างที่มีภูมิสำเนาในเขตกรุงเทพมหานครและปริมณฑล และภาคอื่นๆ มีค่าที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งก็อาจจะสรุปได้ว่าผลการสังเคราะห์ที่ต่างกันอาจจะมาจากการใช้กลุ่มตัวอย่างที่มีความสามารถที่แตกต่างกัน

4.5 จากผลการสังเคราะห์พบว่า แบบสอบที่มีจำนวนข้อที่มากกว่า 40 ข้อจะให้ค่าความเที่ยงและค่าความตรงสูงสุด และให้ค่าความยากง่ายอยู่ในระดับปานกลาง และแบบสอบที่มีการใช้ขนาดของกลุ่มตัวอย่างในการทดสอบที่มากกว่า 200 คนต่อฉบับ จะให้ค่าความเที่ยงและค่าอำนาจจำแนกสูงสุด และให้ค่าความยากง่ายในระดับปานกลางนั้น ผลการวิจัยที่ปรากฏเช่นนี้เนื่องมาจากลักษณะของแบบสอบที่ได้จากงานวิจัยที่นำมาสังเคราะห์ ซึ่งในส่วนของกรนำผลการวิจัยในส่วนนี้ไปใช้เพื่อจะให้ได้แบบสอบที่มีคุณภาพสูงนั้น ควรที่จะต้องพิจารณาถึงเนื้อหาวิชาที่จะออกข้อสอบควรมีเนื้อหามากน้อยเพียงใดหรือเนื้อหาที่มีความซับซ้อนอย่างไร ซึ่งถ้าหากบางเนื้อหาวิชาที่มีขอบเขตที่มากหรือมีความซับซ้อนมากก็ควรที่จะออกข้อสอบให้มีจำนวนมากข้อเพื่อให้ครอบคลุมเนื้อหาวิชาโดยจะต้องคำนึงถึงความตรงตามเนื้อหาเป็นหลัก กล่าวโดยสรุปแล้วการที่จะกำหนดจำนวนข้อสอบหรือขนาดของกลุ่มตัวอย่างที่ใช้ในการสอบนั้นจะต้องคำนึงถึงบริบทของแต่ละเนื้อหาวิชาเป็นสำคัญ

5. การวิเคราะห์ห่อภิมาณกรณีข้อมูลหลายระดับ

ในการวิเคราะห์ข้อมูล ผู้วิจัยได้ทำการแบ่งข้อมูลออกเป็น 3 ระดับ แต่ไม่ได้มีการนำข้อมูลทั้ง 3 ระดับมาวิเคราะห์รวมกันด้วยเทคนิค HLM เนื่องจากไม่ใช่วัตถุประสงค์หลักของผู้วิจัย ประกอบกับตัวแปรต้นในแต่ละระดับไม่สอดคล้องกัน โดยจะมีระดับของข้อมูลที่สอดคล้องกัน

ต้นฉบับ หน้าขาดหาย

กรอบในการเขียนรายงานการวิจัย เพื่อให้การลำดับเนื้อหาในงานวิจัยเป็นไปในทิศทางเดียวกัน เชื่อว่าคงจะทำให้ผู้ที่ทำการศึกษางานวิจัยสามารถที่จะทำความเข้าใจในงานวิจัยนั้นๆ ได้ดีและง่ายมากยิ่งขึ้น

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

1. การวิจัยครั้งนี้เป็นการศึกษาที่มีตัวแปรต้นเป็นตัวแปรที่เกี่ยวกับรูปแบบของแบบสอบ บริบทของแบบสอบ วิธีการตรวจให้คะแนนแบบสอบ และประเภทของแบบสอบ ซึ่งยังมีงานวิจัยที่เกี่ยวข้องกับแบบสอบหลายตัวเลือกอีกเป็นจำนวนมาก จึงควรที่จะมีการศึกษาเพิ่มเติมโดยขยายขอบเขตของการศึกษาออกเป็นงานวิจัยที่มีการใช้แบบสอบหลายตัวเลือกเป็นเครื่องมือที่ใช้ในการวิจัย เพื่อจะได้เห็นถึงกระบวนการสร้างแบบสอบได้ชัดเจนมากขึ้น ตลอดจนเพื่อจะได้ผลการวิจัยที่กว้างขวางและเป็นประโยชน์เพิ่มขึ้น

2. ในการวิจัยครั้งต่อไป ควรที่จะมีการเก็บรวบรวมข้อมูลเกี่ยวกับตัวแปรภูมิหลังของผู้ควบคุมวิทยานิพนธ์ด้วย เช่น ประสบการณ์ด้านการสอน ประสบการณ์ด้านการควบคุมวิทยานิพนธ์ เป็นต้น เพราะตัวแปรเหล่านี้ น่าจะส่งผลต่อคุณภาพของงานวิจัยที่นำมาสังเคราะห์

3. ในการวิจัยครั้งต่อไป ควรที่จะมีการหาความสัมพันธ์ระหว่างคุณภาพของแบบสอบทั้งค่าความเที่ยง ค่าความตรง ค่าความยากง่ายและค่าอำนาจจำแนก กับคะแนนการประเมินคุณภาพงานวิจัย เพื่อที่จะได้ทำการตรวจสอบได้ว่าแบบสอบที่มีคุณภาพสูงจะมีคะแนนการประเมินคุณภาพงานวิจัยที่สูงด้วยหรือไม่

4. ในการวิจัยครั้งต่อไป ควรที่จะมีการวิเคราะห์ลักษณะการแจกแจงของค่าคุณภาพของแบบสอบทั้งค่าความเที่ยง ค่าความตรง ค่าความยากง่ายและค่าอำนาจจำแนก ในข้อมูลระดับเล่มวิทยานิพนธ์ ข้อมูลระดับแบบสอบที่สร้างก่อนนำไปใช้ในการวิจัยและข้อมูลระดับค่าคุณภาพของแบบสอบ เพื่อจะได้ตรวจสอบว่าค่าคุณภาพของแบบสอบทั้ง 3 ระดับนั้นมีความสอดคล้อง เป็นไปในทิศทางเดียวกันหรือแตกต่างกัน

5. เนื่องด้วยข้อมูลที่ได้จากขั้นตอนการเก็บรวบรวมข้อมูลมีลักษณะเป็นระดับชั้น 3 ระดับ แต่ผู้วิจัยไม่ได้นำข้อมูลทั้ง 3 ระดับมาทำการวิเคราะห์รวมกันเพราะไม่ใช่วัตถุประสงค์หลักของการวิจัยครั้งนี้ จึงควรที่จะมีการวิเคราะห์ข้อมูลด้วยโปรแกรม HLM เพื่ออธิบายความแปรปรวนของค่าดัชนีมาตรฐานจะได้ข้อค้นพบใหม่เพิ่มขึ้น

6. ควรทำการวิเคราะห์เพื่ออธิบายความแตกต่างของค่าดัชนีมาตรฐานต่อไป เมื่อพิจารณาใส่ตัวแปร 2 ตัวหรือมากกว่า เพื่อจะอธิบายให้ได้ว่าการที่แบบสอบมีคุณภาพสูงนั้น เนื่องมาจากตัวแปรตัวใด