



รายงานการวิจัย

เรื่อง

ผลของการตัดข้อสอบที่มีอำนาจจำแนกติดลบออกต่อคะแนนการสอบ
และความเชื่อมั่นของข้อสอบทั้งฉบับ: รายวิชาระบบสุขภาพและการ
พยาบาล

The Effects on Test Scores and Test Reliability

after Removing Items with Negative Discrimination Power:

Health System and Nursing Subject

ดร.ศรีสุดา วงศ์วิเศษกุล

มหาวิทยาลัยราชภัฏสวนดุสิต

พ.ศ. 2558

ลิขสิทธิ์ของมหาวิทยาลัยราชภัฏสวนดุสิต



รายงานการวิจัย

เรื่อง

ผลของการตัดข้อสอบที่มีอำนาจจำแนกติดลบออกต่อคะแนนการสอบ
และความเชื่อมั่นของข้อสอบทั้งฉบับ: รายวิชาระบบสุขภาพและการพยาบาล

The Effects on Test Scores and Test Reliability

after Removing Items with Negative Discrimination Power:

Health System and Nursing Subject

ดร.ศรีสุดา วงศ์วิเศษกุล

(คณะพยาบาลศาสตร์)

มหาวิทยาลัยราชภัฏสวนดุสิต

พ.ศ. 2558

ลิขสิทธิ์ของมหาวิทยาลัยราชภัฏสวนดุสิต

(งานวิจัยนี้ได้รับทุนอุดหนุนจากมหาวิทยาลัยราชภัฏสวนดุสิต ปีงบประมาณ 2556)

หัวข้อวิจัย ผลของการตัดข้อสอบที่มีอำนาจจำแนกติดลบออกต่อคะแนนการสอบ
และความเชื่อมั่นของข้อสอบทั้งฉบับ: รายวิชาระบบสุขภาพและการพยาบาล

ผู้ดำเนินการวิจัย ศรีสุตา วงศ์วิเศษกุล

ที่ปรึกษา ดร. เบ็ญจา เตากล้า

หน่วยงาน คณะพยาบาลศาสตร์ มหาวิทยาลัยราชภัฏสวนดุสิต

ปี พ.ศ. 2558

ค่าความยากง่าย (P) และค่าอำนาจจำแนก (R) เป็นค่าพารามิเตอร์ที่ใช้ในการทดสอบคุณภาพของแบบสอบชนิดหลายตัวเลือกตามทฤษฎีการสอบแบบดั้งเดิม การวิจัยเชิงบรรยายนี้มีวัตถุประสงค์เพื่อประเมินคุณภาพของข้อสอบ ศึกษาการเปลี่ยนแปลงของคะแนนสอบและลำดับคะแนนรวมทั้งค่าความเชื่อมั่นของข้อสอบกลางภาคและปลายภาคเมื่อตัดข้อสอบที่มีค่าอำนาจจำแนกเป็นลบออก ประชากรการวิจัยประกอบด้วยข้อสอบกลางภาคและปลายภาครายวิชาระบบสุขภาพและการพยาบาลปีการศึกษา 2555 สำหรับนักศึกษาพยาบาลที่ลงทะเบียนเรียนจำนวน 92 คน ค่าความยากง่าย (P) ค่าอำนาจจำแนก (R) และค่าความเชื่อมั่นของข้อสอบคำนวณด้วยโปรแกรมสำเร็จรูปคอมพิวเตอร์ การเปลี่ยนแปลงของคะแนนสอบและลำดับคะแนนวิเคราะห์ด้วยสถิติ Paired t-test และ Sign Rank test

ผลการวิจัยพบว่า

1) ในการสอบกลางภาค ข้อสอบร้อยละ 64.29 มีค่าความยากง่ายในเกณฑ์เหมาะสม ($P = .20 - .80$) และร้อยละ 45.71 มีค่าอำนาจจำแนกที่ใช้ได้ ($R \geq .20$) ข้อสอบที่มีค่าอำนาจจำแนกเป็นลบร้อยละ 5.71 สำหรับการสอบปลายภาค ข้อสอบร้อยละ 40 มีค่าความยากง่ายในเกณฑ์เหมาะสม ($P = .20 - .80$) และร้อยละ 47.50 มีค่าอำนาจจำแนกที่ใช้ได้ ($R \geq .20$) ข้อสอบที่มีค่าอำนาจจำแนกเป็นลบร้อยละ 7.50

2) หลังจากตัดข้อสอบที่มีค่าอำนาจจำแนกเป็นลบออก ค่าเฉลี่ยคะแนนสอบกลางภาคและปลายภาคเพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ ($t_{(91)} = -10.225$ ที่ $p \text{ value} < .001$ และ $t_{(91)} = 4.972$ ที่ $p \text{ value} < .001$ ตามลำดับ) เช่นเดียวกับกับลำดับคะแนนที่มีการเปลี่ยนแปลงอย่างมีนัยสำคัญทางสถิติ ($p \text{ value} < .001$)

3) ภายหลังจากตัดข้อสอบที่มีค่าอำนาจจำแนกเป็นลบออก ค่าความเชื่อมั่นของข้อสอบกลางภาคและปลายภาค ทั้งฉบับเพิ่มขึ้น

สรุปได้ว่า การตัดข้อสอบที่มีค่า r ติดลบออกทำให้คะแนนสอบและค่าความเชื่อมั่นของข้อสอบเพิ่มขึ้นรวมทั้งส่งผลต่อการเปลี่ยนแปลงของลำดับคะแนนการสอบเพิ่มขึ้น

คำสำคัญ: คุณภาพข้อสอบ อำนาจจำแนกเป็นลบ ความเชื่อมั่นของแบบสอบชนิดหลายตัวเลือก

Research Title The effects on test scores and the reliability of the test when taking out its items with negative discrimination power: Health System and Nursing subject

Researcher Sresuda Wongwiseskul

Research Consultants Dr. Benja Taoklam

Organization Faculty of Nursing, Suandusit Rajabhat University

Year 2014

Based on classical test theory, difficulty index (P) and discrimination index (R) are the parameters used to evaluate the standard of multiple choices questions (MCQs) in examination. This descriptive research aimed to evaluate the quality of test item, explore changes of test score and students' ranking, and changes of mid-term and final test reliability after removing the test items with negative discrimination power. Populations were midterm and final tests for 92 nursing students enrolled in Health System and Nursing subject, academic year 2012. Difficulty Index, discrimination index, test reliability were calculated using computerized program. Paired t-test and Sign Rank test were applied to test the changes of test score and students' ranking.

The study revealed that:

1) In mid-term test, 64.29% of items had index of difficulty (P) within a proper range ($P = .20 - .80$) and 45.71% had proper index of discrimination ($R \geq .20$). 5.71% of test items had negative discrimination. In the final, 40% of items had index of difficulty (P) within a proper range ($P = .20 - .80$) and 47.50% had proper index of discrimination ($R \geq .20$). 7.50% of test items had negative discrimination.

2) After removing the test items with negative discrimination power, mean score of mid-term and final test were higher and had significantly changes ($t_{(91)} = -10.225$ at p value $< .001$ and $t_{(91)} = 4.972$ at p value $< .001$ respectively). Similarly, the Signed Rank Test showed that students' rankings were significantly changed.

3) The reliability of mid-term and final test was increased after excluding.

In conclusion, excluding test items with negative discrimination power increased test reliability, mean score of the test, and affected score ranking.

Keywords: quality of test items, negative discrimination power, reliability of MEQ examination

กิตติกรรมประกาศ

การวิจัย เรื่องผลของการตัดข้อสอบที่มีอำนาจจำแนกติดลบออกต่อคะแนนการสอบ และความเชื่อมั่นของข้อสอบทั้งฉบับ: รายวิชาระบบสุขภาพและการพยาบาล คือความมุ่งหวังค้นหาคำตอบที่เหมาะสมสำหรับการจัดการกับข้อสอบที่มีอำนาจจำแนกต่ำซึ่งข้อสอบดังกล่าวหากเป็นข้อสอบที่ตั้งตามวัตถุประสงค์การเรียนรู้การตัดข้อสอบอาจจะส่งผลกระทบต่อ การตัดสินผลการเรียนของนักศึกษาได้ นอกจากนี้ การศึกษาคุณภาพข้อสอบอย่างลึกซึ้งจะช่วยในการพัฒนาคลังข้อสอบที่เป็นมาตรฐานของคณะได้

การวิจัยนี้สำเร็จลงได้โดยสมบูรณ์ด้วยความกรุณาของ ดร.เบ็ญจา เตากล่ำ คณบดีคณะพยาบาลศาสตร์ที่ปรึกษาโครงการวิจัย อาจารย์วิมล เตากล่ำ และผู้ช่วยศาสตราจารย์ ดร.ประกาย จิโรจน์กุลที่ให้ข้อเสนอแนะอันทรงคุณค่า อาจารย์จิราภรณ์ ชนมาสุขและอาจารย์มนชยา สมจรีต ที่สนับสนุนด้านเวลาและภาระงานวิชาการ รวมทั้งผู้อำนวยการ (ดร.ชนะศึก นิชานนท์) และเจ้าหน้าที่สถาบันวิจัยและพัฒนา ผู้วิจัยขอขอบพระคุณมา ณ โอกาสนี้ และขอขอบคุณมหาวิทยาลัยราชภัฏสวนดุสิตที่ให้การสนับสนุนด้านทุนวิจัย

ศรีสุดา วงศ์วิเศษกุล

พ.ศ. 2558

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญตาราง	ฉ
สารบัญภาพ	ช
บทที่ 1 บทนำ	1
ความเป็นมาและความสำคัญ	1
วัตถุประสงค์ของการวิจัย	3
ขอบเขตการวิจัย	3
คำจำกัดความที่ใช้ในการวิจัย	4
ประโยชน์ที่คาดว่าจะได้รับ	4
บทที่ 2 แนวคิด ทฤษฎี เอกสารและงานวิจัยที่เกี่ยวข้อง	5
การวิเคราะห์คุณภาพข้อสอบรายข้อ	5
ความเชื่อมั่นของข้อสอบ	10
วิธาระบบสุขภาพและการพยาบาล	12
งานวิจัยที่เกี่ยวข้อง	14
กรอบแนวคิดในการวิจัย	19
บทที่ 3 วิธีดำเนินการวิจัย	20
ประชากรและกลุ่มตัวอย่าง	20
เครื่องมือในการวิจัยและการตรวจสอบคุณภาพเครื่องมือ	20
การเก็บรวบรวมข้อมูล	21
การวิเคราะห์ข้อมูล	21

	หน้า
บทที่ 4 ผลการวิจัย	22
คุณภาพของข้อสอบรายวิชาระบบสุขภาพและการพยาบาล	22
การเปลี่ยนแปลงภายหลังจากตัดข้อสอบที่มีค่าอำนาจจำแนกเป็นลบออก	25
บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ	28
สรุปผลการวิจัย	28
อภิปรายผล	28
ข้อเสนอแนะในการนำผลการวิจัยไปใช้	31
ข้อเสนอแนะในการทำวิจัยครั้งต่อไป	31
บรรณานุกรม	32
บรรณานุกรมภาษาไทย	32
บรรณานุกรมภาษาต่างประเทศ	33
ภาคผนวก	
ข้อมูลคะแนนการสอบกลางภาคและปลายภาค	40
ประวัติผู้วิจัย	44

สารบัญตาราง

ตารางที่		หน้า
2.1	แผนการประเมินผลลัพธ์การเรียนรู้รายวิชาระบบสุขภาพและการพยาบาล	13
2.2	การกระจายตัวของข้อสอบตามระดับการวัดผลการเรียนรู้	14
4.1	ค่าความยากง่ายของข้อสอบกลางภาควิชาระบบสุขภาพและการพยาบาลทั้งฉบับ	25
4.2	ค่าความยากง่ายของข้อสอบปลายภาควิชาระบบสุขภาพและการพยาบาลทั้งฉบับ	25
4.3	ค่าอำนาจจำแนกของข้อสอบกลางภาควิชาระบบสุขภาพและการพยาบาลทั้งฉบับ	26
4.4	ค่าอำนาจจำแนกของข้อสอบปลายภาควิชาระบบสุขภาพและการพยาบาลทั้งฉบับ	26
4.5	ผลการทดสอบการเปลี่ยนแปลงของคะแนนการสอบของนักศึกษาในวิชา ระบบสุขภาพและการพยาบาล หลังการตัดข้อสอบที่มีค่า r ตีลบออก	27
4.6	ผลการทดสอบการเปลี่ยนแปลงของลำดับคะแนนจากการสอบของนักศึกษา ในวิชาระบบสุขภาพและการพยาบาล หลังการตัดข้อสอบที่มีค่า r ตีลบออก	28
4.7	ค่าความเชื่อมั่นของข้อสอบวิชาระบบสุขภาพและการพยาบาลก่อน และหลัง การตัดข้อสอบที่มีค่า r ตีลบออก	29

สารบัญภาพ

		หน้า
ภาพที่		
2.1	กรอบแนวคิดในการวิจัย	21

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญ

การจัดการเรียนรู้ในสถาบันอุดมศึกษาได้ปรับเปลี่ยนจากการเน้นกระบวนการเรียนการสอนมาเป็นการมุ่งเน้นผลลัพธ์อย่างเป็นรูปธรรมเมื่อสำนักงานคณะกรรมการการอุดมศึกษาได้นำกรอบมาตรฐานคุณวุฒิระดับอุดมศึกษา (Thailand Qualification Framework) มาประกาศใช้อย่างเป็นทางการในปี 2552 และกรอบมาตรฐานดังกล่าวกำหนดผลลัพธ์การเรียนรู้ไว้ 5 ด้าน คือ ด้านคุณธรรมจริยธรรม ด้านความรู้ ด้านทักษะทางปัญญา ด้านทักษะความสัมพันธ์ระหว่างบุคคลและความรับผิดชอบ ด้านทักษะการวิเคราะห์เชิงตัวเลข การสื่อสารและการใช้เทคโนโลยีสารสนเทศ สำหรับสาขาวิชาที่จัดการเรียนการสอนควบคู่ไปกับการฝึกภาคปฏิบัติ เช่น สาขาพยาบาลศาสตร์ ในรายวิชาภาคปฏิบัติได้เพิ่มทักษะปฏิบัติการทางวิชาชีพเข้าไว้อีก 1 ด้าน เพื่อให้นักศึกษาได้รับการพัฒนาสมรรถนะของการปฏิบัติงานตามบทบาทหน้าที่เชิงวิชาชีพ เมื่อผลลัพธ์การเรียนรู้ ได้กำหนดไว้ชัดเจนขึ้นรูปแบบการสอนได้มีการเปลี่ยนแปลงไปเพื่อให้ให้นักศึกษาได้รับการพัฒนาอย่างเป็นองค์รวมทั้ง 5 หรือ 6 ด้าน ในทุกรายวิชา ซึ่งน้ำหนักของผลลัพธ์การเรียนรู้ในแต่ละด้านแตกต่างกันไปตามลักษณะเฉพาะของวิชานั้นๆ เช่น ผลลัพธ์การเรียนรู้ในรายวิชาภาคทฤษฎี สัดส่วนของน้ำหนักคะแนนในการประเมินผลการเรียนด้านความรู้สูงกว่ารายวิชาภาคปฏิบัติ ในขณะที่ด้านอื่นๆอาจมีน้ำหนักใกล้เคียงกันเป็นต้น

การประเมินผลการเรียนด้านความรู้ดำเนินการได้หลายวิธี เช่น สอบ บันทึกย่อย การแก้ปัญหาด้วยโครงการและโครงงาน การฝึกประสบการณ์วิชาชีพร่วมกัน เป็นต้น (องอาจ นัยวัฒน์, 2556) โดยเฉพาะการศึกษาในยุคศตวรรษที่ 21 ที่มีการประเมินผู้เรียนในหลากหลายมิติ จึงมีการกล่าวถึงการประเมินตามสภาพจริง (Authentic assessment) มากขึ้น อีกทั้งการวัดและประเมินผลเป็นองค์ประกอบที่สำคัญของการจัดการศึกษาที่มีคุณภาพ และจุดมุ่งหมายของการวัดผลทางการศึกษา เพื่อให้ได้มาซึ่งข้อมูลที่เที่ยงตรง และเชื่อถือได้ ที่แสดงถึงความสามารถที่แท้จริง (True ability) ของผู้เรียน (อุทุมพร จามรมาน: 2535) สำหรับการประเมินผลในมิติของความรู้ วิธีการที่ใช้กันมาต่อเนื่อง คือ การสอบ ข้อสอบที่ดีต้องมีความเที่ยงตรง (Content validity) ความเป็นปรนัย (Objectivity) ความเชื่อถือได้ (Reliability) ความยากง่าย (Difficulty Index)พอเหมาะ ค่าอำนาจจำแนก (Discrimination power) ดี สร้างง่าย ใช้ง่าย ใช้สะดวก

คัมค่า (นภา หลิมรัตน์, ม.ป.ป.) ดังนั้น ในการสอบด้วยข้อสอบปรนัย สิ่งที่อาจารย์ผู้สอนต้องดำเนินการควบคู่กับการตรวจข้อสอบคือการวิเคราะห์ข้อสอบ กล่าวคือเมื่อส่งข้อสอบตรวจจะมีการวิเคราะห์ข้อสอบเพื่อคำนวณค่าความยากง่าย (Difficulty Index) และ ค่าอำนาจจำแนก (Discrimination power) ในเวลาเดียวกัน

ค่าความยากง่าย (Difficulty Index) และค่าอำนาจจำแนก (Discrimination power) สะท้อนผลสัมฤทธิ์ที่เกิดกับผู้เรียนผ่านการสอบ ค่าความยากง่าย (Difficulty Index) คำนวณภายหลังการนำข้อสอบไปใช้แล้ว เป็นการสะท้อนความสามารถของนักศึกษา โดยผู้เรียนเก่งจะสามารถตอบข้อสอบได้ถูก ในขณะที่ผู้เรียนอ่อนจะตอบผิด ดังนั้นค่าความยากง่ายจึงแบ่งผู้เรียนออกเป็นสองกลุ่ม การคำนวณค่าความยากง่ายที่แสดงด้วยค่า p และมีค่าระหว่าง $0 - 1$ (นภา หลิมรัตน์, ม.ป.ป.) ในขณะที่ค่าอำนาจจำแนก (Discrimination power) ใช้อธิบายคุณภาพของข้อสอบ มีค่าอยู่ระหว่าง -1.00 ถึง $+1.00$ ถ้ามีค่าเท่ากับ 1.00 ถือว่ามีค่าอำนาจจำแนกที่สมบูรณ์ ถ้ามีค่าเท่ากับ -1.00 ถือว่ามีอำนาจจำแนกในทางตรงกันข้ามสูงสุด ถ้ามีค่าเท่ากับ 0.00 นับว่าไม่มีอำนาจจำแนก ข้อสอบที่ดี ควรมีค่าอำนาจจำแนกเป็นบวกและมีค่าใกล้ 1 (บุญชม ศรีสะอาด. ม.ป.ป.) กล่าวได้ว่า ข้อสอบ เป็นเครื่องมือสำคัญ ในการประเมินว่าผู้เรียนมีความรู้ ตามที่หลักสูตรคาดหวังหรือไม่ ข้อสอบที่มีคุณภาพ จึงเป็นปัจจัยสำคัญของการจัดการเรียนการสอน

คณะพยาบาลศาสตร์ มหาวิทยาลัยราชภัฏสวนดุสิต ประเมินผลการเรียนด้านความรู้ด้วยวิธีการต่างๆ เช่น การสอบ การศึกษาเฉพาะกรณี การมอบหมายรายงานโครงการ ฯลฯ แต่ให้น้ำหนักกับการสอบมากกว่าวิธีอื่นๆ โดยเฉพาะในรายวิชาภาคทฤษฎี น้ำหนักของคะแนนการสอบคิดเป็นร้อยละ $65 - 70$ ของคะแนนรวมทั้งรายวิชาที่มีการประเมินมิติอื่นๆตามผลลัพธ์การศึกษาระดับอุดมศึกษา จะเห็นได้ว่าการประเมินผลลัพธ์การเรียนรู้ด้วยข้อสอบนั้น คุณภาพของข้อสอบมีความสำคัญอย่างยิ่งและส่งผลกระทบต่อความแม่นยำในการวัดและประเมินผล สำหรับรายวิชาระบบสุขภาพและการพยาบาลซึ่งเป็นขอบเขตในการวิจัยนี้เป็นรายวิชาภาคทฤษฎี กำหนดสัดส่วนให้คะแนนจากการสอบคิดเป็นร้อยละ 70 (คณะพยาบาลศาสตร์, 2554) รายงานการวิเคราะห์ข้อสอบของปีการศึกษา $2550 - 2554$ พบข้อสอบที่มีค่าอำนาจจำแนกเป็นลบโดยเฉลี่ยร้อยละ 5 ของข้อสอบทั้งหมดและในปีการศึกษา 2555 พบว่าข้อสอบกลางภาคและปลายภาคมีค่าอำนาจจำแนกเป็นลบร้อยละ 5.71 และ 7.50 ตามลำดับ ในการพิจารณาผลคะแนนเพื่อตัดเกรดยังไม่มีการตัดข้อสอบที่มีค่าอำนาจจำแนกเป็นลบออกเนื่องจากยังไม่มีการศึกษาอย่างเป็นระบบ ดังนั้นการวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาคุณภาพของข้อสอบในรายวิชาระบบสุขภาพ

และการพยาบาล และเปรียบเทียบการเปลี่ยนแปลงของคะแนนการสอบของนักศึกษาหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก และการเปลี่ยนแปลงค่าความเชื่อมั่นของข้อสอบทั้งฉบับเมื่อตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออกเช่นกัน โดยคาดว่าผลการวิจัยจะเป็นหลักฐานเชิงประจักษ์ (Evidence base) ในการจัดการความรู้เกี่ยวกับการพัฒนาคุณภาพข้อสอบรายวิชาระบบสุขภาพและการพยาบาล รวมทั้งนำไปสู่ข้อสรุป หรือข้อเสนอแนะเชิงนโยบายในด้านการวัดและประเมินผลการจัดการเรียนการสอนในรายวิชานี้ต่อไป

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาคุณภาพของข้อสอบรายข้อและข้อสอบทั้งฉบับรายวิชาระบบสุขภาพและการพยาบาล
2. เพื่อศึกษาการเปลี่ยนแปลงของคะแนนการสอบและลำดับคะแนนก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก
3. เพื่อศึกษาการเปลี่ยนแปลงค่าความเชื่อมั่นของข้อสอบรายวิชาระบบสุขภาพและการพยาบาล หลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก

ขอบเขตการวิจัย

1. ขอบเขตด้านประชากร

ประชากรในการวิจัยนี้ ประกอบด้วยข้อสอบกลางภาคและปลายภาครายวิชาระบบสุขภาพและการพยาบาลปีการศึกษา 2555 สำหรับนักศึกษาพยาบาลชั้นปีที่ 1 ที่ลงทะเบียนเรียนจำนวน 92 คน

2. ขอบเขตด้านเวลา

การวิจัยนี้ รวบรวมข้อมูลเกี่ยวกับการวิเคราะห์อำนาจจำแนกของข้อสอบวิชาระบบสุขภาพและการพยาบาลที่จัดสอบกลางภาคและปลายภาคในภาคการศึกษาที่ 1 ปีการศึกษา 2555

3. ขอบเขตด้านเนื้อหา

คุณภาพของข้อสอบในการวิจัยนี้ประกอบด้วย

3.1 คุณภาพของข้อสอบรายข้อพิจารณาจากค่าความยากง่ายและอำนาจ
จำแนก

3.2 คุณภาพของข้อสอบทั้งฉบับพิจารณาจากค่าความเชื่อมั่น

3.3 คุณภาพของข้อสอบด้านความตรง (Validity) คือความตรงเชิงโครงสร้าง (Construct validity) พิจารณาจากความสอดคล้องทั้งจำนวนข้อสอบและประเด็นเนื้อหาเกี่ยวกับตารางวิเคราะห์การออกแบบการสร้างแบบทดสอบ (Test Blueprint) ที่ได้ออกแบบและแสดงไว้ใน มคอ. 3 รายวิชาระบบสุขภาพและการพยาบาล รวมทั้งการวิพากษ์ข้อสอบโดยผู้เชี่ยวชาญก่อนนำข้อสอบไปใช้จริง

คำจำกัดความที่ใช้ในการวิจัย

คุณภาพของข้อสอบรายข้อ หมายถึง คุณภาพของข้อสอบในวิชาระบบสุขภาพและการพยาบาล ฉบับกลางภาคและปลายภาค ที่ได้มาจากการวิเคราะห์ค่าความยากง่ายและอำนาจจำแนกเป็นรายข้อ

คุณภาพของข้อสอบทั้งฉบับ หมายถึง คุณภาพของข้อสอบในวิชาระบบสุขภาพและการพยาบาล ฉบับกลางภาคและปลายภาค ที่ได้มาจากการวิเคราะห์ความเชื่อมั่น (KR - 20) ของข้อสอบทั้งฉบับ

ข้อสอบที่มีอำนาจจำแนกดี หมายถึง ข้อสอบรายข้อที่ได้วิเคราะห์ค่าอำนาจจำแนก (r) แล้ว ได้ค่าน้อยกว่า 0

การตัดข้อสอบที่มีค่าอำนาจจำแนกดี หมายถึง การนำข้อสอบที่มีอำนาจจำแนก (r) เป็นลบออกก่อนนำไปวิเคราะห์คุณภาพข้อสอบในครั้งที่สอง

คะแนนการสอบ หมายถึงคะแนนที่ได้จากการตอบแบบทดสอบชนิดเลือกตอบ (Multiple Choices Question) 4 ตัวเลือก และมีคำตอบที่ถูกที่สุดเพียงข้อเดียว ข้อที่ตอบถูกต้องคะแนน 1 ตอบผิดได้คะแนน 0

ประโยชน์ที่คาดว่าจะได้รับ

ผลการวิจัยสามารถนำไปใช้ในการปรับปรุงข้อสอบที่มีค่าอำนาจจำแนกเป็นลบ การคัดเลือกข้อสอบที่มีคุณภาพเพื่อจัดทำคลังข้อสอบ และเป็นแนวทางในการวิเคราะห์ข้อสอบในรายวิชาอื่นๆ ที่ใช้แบบทดสอบชนิดเลือกตอบ (Multiple Choices Question Examination)

บทที่ 2

แนวคิด ทฤษฎี เอกสารและงานวิจัยที่เกี่ยวข้อง

วรรณกรรมที่เกี่ยวข้องกับการวิจัยเรื่อง ผลของการตัดข้อสอบที่มีอำนาจจำแนกติดลบออกต่อ คะแนนการสอบ และความเชื่อมั่นของข้อสอบทั้งฉบับ: รายวิชาระบบสุขภาพและการพยาบาล ประกอบด้วย

1. การวิเคราะห์คุณภาพข้อสอบรายข้อ
 2. ความเชื่อมั่นของข้อสอบ
 3. วิชาระบบสุขภาพและการพยาบาล
 4. งานวิจัยที่เกี่ยวข้อง
- ดังรายละเอียดต่อไปนี้

การวิเคราะห์คุณภาพข้อสอบรายข้อ

การวิเคราะห์คุณภาพของข้อสอบเป็นรายข้อ (Item analyses) มีความจำเป็นอย่างยิ่งเพราะถึงแม้ว่าข้อสอบจะออกแบบตามหลักการและสอดคล้องกับวัตถุประสงค์การเรียนรู้ แต่การวิเคราะห์ข้อสอบรายข้อด้วยวิธีการทางสถิติจากผลการตอบของผู้เรียนยังมีความสำคัญและใช้เพื่อยืนยันคุณภาพของข้อสอบ การวิเคราะห์ข้อสอบจำแนกเป็น 2 แนวคิด คือ การวิเคราะห์คุณภาพรายข้อตามแนวคิดของการวัดอิงกลุ่ม และแนวคิดของการวัดอิงเกณฑ์ (เยาวตี วิบูลย์ศรี, 2551) ซึ่งทั้งสองแนวคิดมีความแตกต่างกันในวิธีการทางสถิติที่ใช้วิเคราะห์

การวิเคราะห์คุณภาพรายข้อตามแนวคิดของการวัดอิงกลุ่ม (norm referenced) วิธีการทางสถิติที่ใช้กันมากในการวิเคราะห์ข้อสอบตามแนวคิดนี้ คือ ระดับความยากง่าย (Level of difficulty) และอำนาจจำแนก (Power of discrimination) แบบทดสอบที่นำมาวิเคราะห์จะต้องเป็นแบบทดสอบประเภทเลือกตอบ เช่น แบบสอบถูก - ผิด และแบบสอบชนิดหลายตัวเลือก (Multiple choice test) กล่าวคือ

1. ระดับความยากง่าย (Level of difficulty) หมายถึง สัดส่วนของผู้ตอบในข้อนั้นๆ ได้ถูกต้องต่อจำนวนผู้ตอบข้อนั้นทั้งหมด ค่าสัดส่วนนี้เรียกว่า ดัชนีความยากง่าย (Index of difficulty,

สัญลักษณ์ P) มีค่าระหว่าง 0 ถึง 1 เช่น ข้อสอบข้อหนึ่ง มีผู้ตอบทั้งหมด 50 คน จำนวนผู้ที่ตอบถูก 40 คน ข้อสอบข้อนี้มีดัชนีความยากง่าย 0.80 เป็นต้น

สูตรการคำนวณดัชนีความยากง่าย (Index of difficulty) (พวงรัตน์ ทวีรัตน์, 2540)

$$P = R / N$$

เมื่อ P = ค่าความยากง่ายของคำถามข้อนั้น

R = จำนวนผู้ตอบถูกในข้อนั้น

N = จำนวนผู้ตอบทั้งหมด

ดัชนีความยากง่ายสามารถวิเคราะห์โดยใช้โปรแกรมวิเคราะห์ข้อสอบได้

ดัชนีความยากง่ายมีเกณฑ์การพิจารณาดังนี้

ค่าความยากง่าย (P)	ความหมาย	ข้อเสนอแนะ
0.81 - 1.00	ง่ายมาก	ควรปรับปรุง/ไม่นำมาใช้ใหม่
0.60 - 0.80	ค่อนข้างง่าย	ดี
0.40 - 0.59	ยากง่ายปานกลาง	ดีมาก
0.20 - 0.39	ค่อนข้างยาก	ดี
0.00 - 0.19	ยากมาก	ควรปรับปรุง/ไม่นำมาใช้ใหม่

โดยสรุป ข้อสอบที่ยาก มีค่า P ต่ำหรือใกล้ 0 ในขณะที่ ข้อสอบง่าย มีค่า P สูงหรือใกล้ 1 ข้อสอบที่ควรเก็บไว้เป็นคลังข้อสอบ คือ ข้อสอบที่มีค่าความยากง่าย ระหว่าง .20 - .80 (สุมาลี จันทร์ชะลอ, 2542; ล้วน สายยศและอังคณา สายยศ, 2543; เยาวดี วิบูลย์ศรี, 2551)

2. อำนาจจำแนก (Power of discrimination) หมายถึง ความสามารถของข้อสอบแต่ละข้อในการจำแนกผู้สอบได้ตามความสามารถ (เยาวดี วิบูลย์ศรี, 2551) เช่น แยกคนเก่งกับคนอ่อน คนที่มีความรู้ในข้อนั้นกับคนที่ไม่มีความรู้ หรือในแบบวัดทัศนคติจะเป็นการบอกความแตกต่างของความคิดเห็นของผู้ตอบ ดัชนีอำนาจจำแนก หรือ ค่าอำนาจจำแนก (Index of discrimination, สัญลักษณ์ r หรือ D) คือสัดส่วนของผลต่างระหว่างจำนวนผู้ตอบถูกในกลุ่มที่ได้คะแนนสูงกับกลุ่มที่ได้คะแนนน้อย มีค่าระหว่าง -1.0 ถึง +1.0 ข้อคำถามที่ดีควรมีค่าอำนาจจำแนกเป็นบวก เพราะถ้าเป็นลบ แสดงว่ามีอำนาจจำแนกในทางตรงกันข้าม กล่าวคือ คนเก่งตอบผิด แต่คนไม่เก่งตอบถูก ถ้าค่าอำนาจ

จำแนกเป็น 0 แสดงว่าผู้เรียนในกลุ่มเก่งและกลุ่มอ่อนทำข้อสอบข้อนั้นได้เท่าๆกัน จึงไม่สามารถจำแนกกลุ่มผู้เข้าสอบได้ ข้อคำถามที่ดีควรมีค่าอำนาจจำแนก 0.20 ขึ้นไป และหากมีค่าใกล้ +1 จัดว่าเป็นข้อสอบที่มีอำนาจจำแนกดี (Chase 1978:140)

หากพิจารณาถึงความสัมพันธ์ระหว่าง ความยากง่าย และ อำนาจจำแนกจะพบว่า ถ้าข้อสอบมีความยากง่าย .50 แสดงว่าผู้เรียนจำนวนครึ่งหนึ่งทำข้อสอบข้อนั้นถูก อำนาจจำแนกมีค่าสูงสุด แต่หากข้อสอบยากมาก หรือ ง่ายมาก ค่าอำนาจจำแนกจะลดลง (สุมาลี จันทร์ชโล, 2542)

สูตรการคำนวณค่าอำนาจจำแนก (สุมาลี จันทร์ชโล, 2542)

$$r = \frac{R_u - R_e}{N / 2}$$

เมื่อ r = ค่าอำนาจจำแนกของคำถามข้อนั้น
 R_u = จำนวนผู้ตอบถูกในข้อนั้น ในกลุ่มเก่ง
 R_e = จำนวนผู้ตอบถูกในข้อนั้น ในกลุ่มอ่อน
 N = จำนวนคนเข้าสอบ

ค่าอำนาจจำแนกสามารถวิเคราะห์โดยใช้โปรแกรมวิเคราะห์ข้อสอบได้เช่นกัน (พวงรัตน์ ทวีรัตน์, 2540) เกณฑ์การแปลความหมายค่าอำนาจจำแนก ได้มีผู้เชี่ยวชาญด้านการสอบให้ความเห็นไว้หลากหลายเช่น รังสรรค์ มณีเล็ก และคณะ (2546) จำแนกระดับของค่าอำนาจจำแนกที่เป็นบวกมีเกณฑ์การแปลความหมาย ดังนี้

ค่าอำนาจจำแนก	ความหมาย
1.00	จำแนกดีเลิศ
0.80 - 0.99	จำแนกได้ดีมาก
0.60 - 0.79	จำแนกได้ดี
0.40 - 0.59	จำแนกได้ปานกลาง
0.20 - 0.39	จำแนกได้บ้าง
0.00 - 0.19	จำแนกไม่ค่อยได้

ในขณะที่ ปวรส บุตะเสียว (ม.ป.ป.) ได้จำแนกระดับของค่าอำนาจจำแนกไว้ดังนี้

ค่าอำนาจจำแนก	ความหมาย
0.60 - 1.00	จำแนกดีมาก
0.40 - 0.59	อำนาจจำแนกดี
0.20 - 0.39	อำนาจจำแนกพอใช้
0.10 - 0.19	อำนาจจำแนกต่ำ ควรปรับปรุงหรือตัดทิ้ง
- 1.00 - 0.09	อำนาจจำแนกต่ำมาก ควรปรับปรุงหรือตัดทิ้ง

นอกจากนี้ พบว่า เอเบลและฟรีสบี (Ebel, R. L. & Frisbie, D. A., 1986 อ้างถึงใน Backhoff, E., Larrazolo, N., & Rosas, M. (2000) ได้แบ่งระดับค่าอำนาจจำแนกและข้อเสนอในการพิจารณาข้อสอบไปใช้ดังนี้

ค่าอำนาจจำแนก	ความหมาย
> 0.39	จำแนกดีมาก ควรนำไปใช้ต่อไป
0.30 - 0.39	อำนาจจำแนกดี ควรพัฒนาให้ดีขึ้น
0.20 - 0.29	อำนาจจำแนกพอใช้ ควรทบทวน/ปรับแก้
0.00 - 0.20	อำนาจจำแนกต่ำ ควรปรับปรุงหรือตัดทิ้ง
< -0.01	อำนาจจำแนกต่ำมาก ควรปรับปรุงหรือตัดทิ้ง

สรุปได้ว่า อำนาจจำแนกควรมีค่าตั้งแต่ 0.20 ขึ้นไป ค่าอำนาจจำแนกมีความหมายดังนี้

ค่าอำนาจจำแนก (r)	ความหมาย
0.80 - 1.00	จำแนกได้ดีมาก
0.60 - 0.79	จำแนกได้ดี
0.40 - 0.59	จำแนกได้ปานกลาง
0.20 - 0.39	จำแนกได้น้อย
0.01 - 0.19	จำแนกได้น้อยมาก
-1.00 - 0.00	จำแนกไม่ได้

ข้อสอบที่มีอำนาจจำแนกเป็นลบ

ดังที่ได้กล่าวมาแล้วว่า ข้อสอบที่มีคุณภาพดี ควรสามารถจำแนกความสามารถผู้เรียนได้ตามความเป็นจริง มีค่าระหว่าง -1.0 ถึง $+1.0$ ซึ่งในกรณีที่มีค่าเป็นลบ ได้มีผู้ให้ความคิดเห็นไว้ดังนี้

ข้อสอบที่มีอำนาจจำแนกเป็นลบ เกิดจากข้อกระทงที่ใช้ภาษาซับซ้อนและทำให้ผู้สอบมีโอกาสเลือกคำตอบที่ถูกต้องได้โดยปราศจากความรู้ในเรื่องนั้นจริงๆ ผู้เรียนที่เรียนอ่อนอาจตอบถูกโดยใช้การเดา ในขณะที่ผู้เรียนที่เรียนเก่งอาจสับสนว่าข้อสอบง่ายเกินไปหรือไม่ ใช้กระบวนการคิดซับซ้อนเกินไป ดีความลึกซึ้งเกินไป จนตัดสินใจเลือกตัวลงในที่สุด จนทำให้ผลการวิเคราะห์อำนาจจำแนกพบว่า ผู้ที่เรียนเก่งทำข้อสอบได้ไม่ดีเท่าผู้ที่เรียนอ่อน หลักการโดยทั่วไปของค่าดัชนีอำนาจจำแนก ข้อที่มีค่า เท่ากับหรือมากกว่า $.40$ ถือว่าเป็นข้อสอบที่ดี ค่า $0.30 - 0.39$ เป็นข้อสอบที่ใช้ได้ แต่ต้องปรับแก้เล็กน้อย ค่า $0.20 - 0.29$ เป็นข้อสอบที่ยอมรับได้แต่ควรปรับแก้ และค่าต่ำกว่า 0.19 เป็นข้อสอบที่ไม่เหมาะสม ไม่ควรนำมาใช้ (Ebel & Frisbie, 1986 as cited in Susan Matlock-Hetzel (1997))

ค่าดัชนีอำนาจจำแนกสะท้อนคุณภาพของข้อสอบ ข้อสอบที่มีค่าอำนาจจำแนกต่ำเกิดจากสำนวนภาษาของข้อกระทง ที่ทำให้ผู้สอบสับสนหรือคิดวิเคราะห์มากเกินไป ซึ่งหากจะนำข้อดังกล่าวมาใช้สอบในครั้งต่อไป ผู้ออกข้อสอบควรปรับภาษาให้ชัดเจน ไม่กำกวม ข้อสอบที่มีค่าอำนาจจำแนกเป็นลบควรได้นำมาวิเคราะห์สาเหตุ เช่น ค่าเป็นลบอาจเกิดจากเฉลยผิดข้อ ทำให้ผู้สอบที่ทราบคำตอบที่ถูกต้องไม่เลือกข้อเฉลย เป็นต้น (Undergraduate Academic Affairs, 2005)

การประเมินคุณภาพของข้อสอบที่ต้องการวัดความรู้ของนักศึกษาควรวิเคราะห์ด้วยค่าความยากง่ายและอำนาจจำแนกเป็นวิธีที่ดี สาเหตุของข้อสอบที่มีอำนาจจำแนกไม่ดีหรือเป็นลบอาจเกิดจากข้อคำถามไม่ชัดเจน ตั้งคำถามในประเด็นที่ยังไม่มีข้อสรุปที่ชัดเจน เฉลยไม่ถูกต้องหรือเป็นคำถามเกี่ยวกับความคิดเห็นที่อาจเป็นไปได้ทั้งถูกและผิด (Bharti N Karella, Ajita Pillai, Bhavisha N Vegada (2013))

ในทำนองเดียวกัน Tavakol M, Dennick R (2011) เขียนบทความชื่อ การวิเคราะห์ผลการสอบวัดพฤติกรรมส่วนบุคคล โดยกล่าวว่าเป้าหมายที่สำคัญประการหนึ่งของการประเมินความรู้ความสามารถของนักศึกษา ในสาขาวิทยาศาสตร์สุขภาพ คือการลดความคลาดเคลื่อนในแบบทดสอบให้น้อยที่สุดเท่าที่จะทำได้ เพื่อที่จะได้ผลคะแนนสอบที่สามารถบ่งบอกระดับ ความรู้ ความสามารถที่แท้จริงของนักศึกษามากที่สุด ดังนั้นในการสร้างแบบทดสอบ ผู้สร้างแบบทดสอบจะต้องคำนึงทุกๆ

องค์ประกอบ ที่อาจก่อให้เกิดความลำเอียงในการประเมินความสามารถของนักศึกษา ซึ่งเริ่มตั้งแต่ขั้นตอนของการคิดสร้างข้อสอบขึ้นมา ไปจนถึงขั้นตอนการแปลผลคะแนนสอบ ในบทความนี้ได้กล่าวถึง การอธิบายถึงขั้นตอนในแนวทางการปฏิบัติเพื่อให้นักศึกษาสามารถเกิดการเรียนรู้ตามที่ผู้สอนต้องการได้นั้นมีองค์ประกอบที่สำคัญ 3 ประการคือ

1. ผู้สอนจะต้องทำการตัดสินใจเลือกประเด็นที่นักศึกษาจำเป็นต้องรู้ทั้งหมดออกมาให้ได้ก่อน
2. ผู้สอนจะต้องมีการปรับปรุง และออกแบบวิธีการจัดการเรียนการสอนทั้งในการสอนภาคทฤษฎีในชั้นเรียน การเรียนการสอนภาคปฏิบัติในศูนย์บริการสุขภาพต่างๆ ที่เหมาะสมเพื่อให้นักศึกษาเกิดการเรียนรู้ในเนื้อหา ทักษะที่ผู้สอนคาดหวังไว้ให้มากที่สุด
3. ผู้สอนจะต้องมีการประเมินระดับความรู้ ความสามารถ ในทักษะและเนื้อหาที่จำเป็นของนักศึกษาแต่ละราย ซึ่งโดยส่วนใหญ่จะใช้วิธีการทำข้อสอบ ซึ่งโดยส่วนใหญ่แบบทดสอบที่เหมาะสมและถูกเลือกใช้เพื่อประเมินความรู้ของนักศึกษา คือข้อสอบแบบเลือกตอบโดยมีตัวเลือกหลายข้อ (Multiple choice) ซึ่งโดยปกติข้อสอบที่ดี จะเป็นตัวที่สามารถแบ่งนักศึกษาออกเป็นกลุ่มเด็กเก่งและเด็กอ่อนได้ด้วย สามารถทำได้โดย ผู้สอนจะต้องมีการทดสอบความตรง ความเที่ยงของตัวข้อสอบ เพื่อให้เกิดความแน่ใจได้ว่าข้อสอบในชุดนั้นๆสามารถใช้วัดระดับความรู้ ความสามารถ ทักษะของนักศึกษาในเนื้อหาที่ผู้สอนต้องการได้อย่างแท้จริง และนอกจากการวัดความตรงและความเที่ยงของข้อสอบแล้ว ควรมีการคำนวณความยากง่ายของตัวเลือกที่ถูกต้องในข้อสอบแต่ละข้อ เพื่อเป็นการประเมินความยากง่ายของข้อสอบข้อนั้นๆ ซึ่งข้อสอบข้อที่ง่ายคือ ข้อสอบที่มีนักศึกษาส่วนใหญ่ (มากกว่า 60%) เลือกตอบตัวเลือกที่ถูกต้อง และการแปลผลข้อสอบที่ยากมาก คือข้อสอบที่มีนักศึกษาเลือกตอบในตัวเลือกที่ถูกต้องน้อยกว่า 30% ของจำนวนนักศึกษาที่ทำข้อสอบทั้งหมด ซึ่งในทางสถิติการจำแนกความยากง่ายของข้อสอบจะใช้สถิติ T-test และ Point bi-serial Correlation Coefficient ในการคำนวณ และสามารถเป็นตัวจำแนกข้อสอบดีซึ่งมักจะมีค่าบวกจำนวนมากจากการใช้สถิติทั้งสองในการคำนวณ และในทางกลับกันหากข้อสอบในข้อใดมีค่าบวกจากการคำนวณสถิติน้อย จะแปลได้ว่าข้อสอบข้อนั้นๆเป็นข้อสอบที่ไม่ดี ซึ่งหากข้อสอบข้อใดที่เมื่อใช้สถิติคำนวณแล้วพบว่าไม่มีความแตกต่างอย่างมีนัยสำคัญทางสถิติ ของจำนวนนักศึกษาเฉลี่ยที่เลือกคำตอบในตัวเลือกที่ถูกต้องและจำนวนนักศึกษาเฉลี่ยที่เลือกตัวเลือกอื่นที่ไม่ถูกต้อง แปลได้ว่าข้อสอบข้อนั้นไม่สามารถจำแนกกลุ่มเด็กเก่งและเด็กอ่อนออกจากกันได้ ซึ่งแนะนำให้ควรตัดข้อสอบข้อนั้นออกจากชุดข้อสอบ

กล่าวได้ว่า ข้อสอบที่มีอำนาจจำแนกเป็นลบ ผู้ออกข้อสอบควรได้ทบทวนว่า เฉลยผิดหรือไม่ ส่วนนวงษาของข้อสอบชัดเจน กระชับ ได้ใจความตามวัตถุประสงค์ของการวัดหรือไม่ และเมื่อได้มีการปรับปรุงแล้วยังมีดัชนีค่าอำนาจจำแนกเป็นลบ ข้อสอบนั้นไม่ควรนำมาใช้อีกไม่ควรตัดออกจากคลังข้อสอบ อย่างไรก็ตาม ควรพิจารณาค่าความยากง่ายร่วมด้วย

ความเชื่อมั่นของข้อสอบ

ความเชื่อมั่นของข้อสอบ หมายถึง ความคงที่ของผลการนำข้อสอบนั้นไปใช้ไม่ว่าจะสอบกี่ครั้ง ค่าความเชื่อมั่นที่ได้จะไม่แตกต่างกัน (มนต์ชัย เทียนทอง, 2547; ปรัชญนันท์ นิลสุข, ม.ป.ป.; เกียรติสุตา ศรีสุข, 2552) เทคนิคการหาค่าความเชื่อมั่นจำแนกได้เป็น 2 กลุ่มใหญ่ คือ อิงกลุ่มและอิงเกณฑ์ สถิติที่ใช้กันแพร่หลาย ได้แก่ Split-Half Method, KR - 20 ซึ่งเป็นชนิดอิงกลุ่ม และ ANOVA ที่เป็นชนิดอิงเกณฑ์ สำหรับการวิจัยนี้ ศึกษาความเชื่อมั่นด้วยสถิติ KR - 20

เทคนิคการวิเคราะห์ความเชื่อมั่นของข้อสอบด้วย KR - 20 เป็นการวิเคราะห์ความเชื่อมั่นจากความแปรปรวนของคะแนนที่สะท้อนความสอดคล้องภายใน (internal consistency) ของข้อสอบ สามารถวิเคราะห์ได้ในการสอบแต่ละครั้ง โดยที่คะแนนของข้อสอบเป็น 0/1

สูตรการคำนวณค่าความเชื่อมั่น KR-20 (เกียรติสุตา ศรีสุข, 2552)

$$r_t = [n/(n - 1)] \times [1 - (\sum pq)/\text{Var}]$$

เมื่อ r_t = ค่าสัมประสิทธิ์แห่งความเที่ยงของแบบสอบ

n = จำนวนข้อกระทงในแบบสอบ

Var = ความแปรปรวนของคะแนนสอบทั้งชุด

$\sum pq$ = ผลรวมของสัดส่วนของผู้ที่ตอบถูก คูณกับสัดส่วนของผู้ที่ตอบผิด

p = สัดส่วนของผู้ที่ตอบถูกในแต่ละข้อกระทง

q = สัดส่วนของผู้ที่ตอบผิดในแต่ละข้อกระทง หรือมีค่าเท่ากับ $1 - p$

เกณฑ์การแปลผลความเชื่อมั่น (เกียรตินิสิตา ศรีสุข, 2552)

ค่าความเชื่อมั่น	ความหมาย
0.00 – 0.20	ความเชื่อมั่นต่ำมากหรือไม่มีเลย
0.21 – 0.40	ความเชื่อมั่นต่ำ
0.41 - 0.70	ความเชื่อมั่นปานกลาง
0.71 – 1.00	ความเชื่อมั่นสูง

สรุปได้ว่า ค่าความเชื่อมั่นอยู่ระหว่าง 0.00 – 1.00 และค่าที่ใกล้ 1.00 แสดงว่าข้อสอบนั้นมีความเชื่อมั่นสูง ค่าความเชื่อมั่นที่ยอมรับได้คือ $R \geq .20$

ปัจจุบัน การวิเคราะห์ข้อสอบทำได้ง่ายขึ้นเนื่องจากมีโปรแกรมคอมพิวเตอร์สำเร็จรูป เช่น CTIA/ Grading, EVANA, IRT-BAY เป็นต้น ในการวิจัยนี้ ข้อสอบวิชาการระบบสุขภาพและการพยาบาล วิเคราะห์ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูปที่มหาวิทยาลัยราชภัฏสวนดุสิตจัดให้ ใช้เทคนิคการวิเคราะห์คุณภาพรายข้อตามแนวคิดของการวัดอิงกลุ่มซึ่งสอดคล้องกับทฤษฎีการทดสอบแบบประเพณีนิยม (Classical Test Theory)

วิชาการระบบสุขภาพและการพยาบาล

รายวิชานี้จัดการเรียนการสอนในชั้นปีที่ 1 เป็นวิชาภาคทฤษฎีในหมวดวิชาพื้นฐานทางวิชาชีพและเป็นวิชาภาคบังคับ มีสาระเนื้อหาเกี่ยวกับประวัติ และวิวัฒนาการของวิชาชีพการพยาบาล และการผดุงครรภ์ แนวคิดระบบสุขภาพ ระบบบริการสุขภาพ และระบบบริการพยาบาล ครอบคลุมระบบบริการสุขภาพระดับปฐมภูมิ ทูติยภูมิ ตติยภูมิ และการสาธารณสุขมูลฐาน ระบบหลักประกันสุขภาพถ้วนหน้า การประกอบวิชาชีพ การพยาบาล และการผดุงครรภ์ภายใต้ขอบเขตของกฎหมาย และจริยธรรมแห่งวิชาชีพ กิจกรรมการเรียนการสอนประกอบด้วย 1) การบรรยายประกอบสไลด์ PowerPoint 2) การมอบหมายให้นักศึกษาจัดทำรายงาน 2 ฉบับคือรายงานการศึกษาชีวประวัติบุคคลสำคัญในวิชาชีพการพยาบาล และ รายงานการศึกษาระบบสุขภาพของประเทศ

ประชาคมอาเซียน และ 3) การจัดนิทรรศการวัน ASEAN WEEK แผนการประเมินผลแสดงดังตารางที่ 2.1

ตารางที่ 2.1 แผนการประเมินผลผลลัพธ์การเรียนรู้รายวิชาระบบสุขภาพและการพยาบาล

วิธีการประเมินผล/ ผลการเรียนรู้ที่คาดหวัง	สัปดาห์ที่กำหนด	สัดส่วนของการประเมินผล
1. สอบ	สัปดาห์ที่ 9 และ 15	ร้อยละ 70
2. ประเมินพฤติกรรมกรรมการเข้าเรียน/ การส่งรายงาน/พฤติกรรมกรรมการทำงาน เป็นกลุ่ม	ตลอดภาคการศึกษา	ร้อยละ 5
3. รายงาน 2 ฉบับ (ประกอบด้วย 3.1 รายงานการศึกษาชีวประวัติ บุคคลสำคัญในวิชาชีพการพยาบาล 3.2 รายงานการศึกษาระบบสุขภาพ ของประเทศประชาคมอาเซียน	สัปดาห์ที่ 3 สัปดาห์ที่ 10	ร้อยละ 5 ร้อยละ 10
4. งานกลุ่มจัดนิทรรศการวัน ASEAN WEEK	สัปดาห์ที่ 11	ร้อยละ 10
	รวม	ร้อยละ 100

สำหรับวิธีการสอบใช้ในการประเมินผลด้านความรู้ แบบสอบเป็นข้อสอบปรนัยชนิดหลายตัวเลือก (Multiple choice test) 4 ตัวเลือกและมีคำตอบที่ถูกต้องเพียงข้อเดียว เนื้อหาข้อสอบเป็นไปตามคำอธิบายรายวิชาที่กำหนดไว้ในหลักสูตร กำหนดวัตถุประสงค์การเรียนรู้เชิงพุทธิพิสัยไว้ 3 ระดับ คือ รู้-จำ เข้าใจ และนำไปใช้ การกระจายระดับของการวัดผลการเรียนรู้ขึ้นกับธรรมชาติของเนื้อหาและสัปดาห์ของการเรียน เช่น หัวข้อวิวัฒนาการวิชาชีพการพยาบาลเป็นเนื้อหาในหน่วยแรกสำหรับนักศึกษาที่เริ่มเรียนวิชาการพยาบาลเป็นครั้งแรกเนื่องจากวิชานี้จัดการเรียนการสอนในภาคการศึกษาที่ 1 ชั้นปีที่ 1 การวัดผลจึงวัดที่ระดับ รู้-จำและเข้าใจ ในขณะที่หัวข้ออื่นๆที่เหลือจะวัดทั้งระดับรู้-จำ เข้าใจและนำไปใช้ สำหรับการแบ่งเนื้อหาเพื่อจัดทำแบบสอบกลางภาคและปลายภาคนั้น

เนื่องจากในปีการศึกษาที่ดำเนินการวิจัยนี้ จัดการเรียนการสอนให้นักศึกษาค้นคว้าด้วยตนเองเพื่อนำเสนอในชั้นเรียน จัดนิทรรศการร่วมด้วยส่งผลให้สัดส่วนจำนวนข้อสอบกลางภาคต่อปลายภาคเป็น 70:40 หรือ 2:1 โดยประมาณ จำนวนข้อสอบต่อชั่วโมงการสอนเท่ากับ 5:1 ทุกครั้งที่สอบเสร็จจะนำข้อสอบไปวิเคราะห์คุณภาพด้วยโปรแกรมสำเร็จรูปคอมพิวเตอร์ที่มหาวิทยาลัยจัดให้

การกระจายตัวของข้อสอบตามระดับการวัดผลการเรียนรู้แสดงดังตารางที่ 2.2

ตารางที่ 2.2 การกระจายตัวของข้อสอบตามระดับการวัดผลการเรียนรู้

ขอบเขตเนื้อหา	จำนวนข้อสอบตามระดับการเรียนรู้			รวม
	รู้-จำ	เข้าใจ	นำไปใช้	
ข้อสอบกลางภาค				
วิวัฒนาการวิชาชีพการพยาบาล	3	2	-	5
ระบบสุขภาพ และดัชนีสุขภาพ	5	5	10	20
ระบบบริการสุขภาพ	12	12	21	45
รวม	20	19	31	70 ข้อ
ข้อสอบปลายภาค				
ระบบบริการพยาบาล	2	3	5	10
สิทธิผู้ป่วย	3	3	4	10
กฎหมายเบื้องต้น	3	3	4	10
จริยธรรมวิชาชีพ	3	3	4	10
รวม	11	12	17	40 ข้อ

โดยสรุป วิชาระบบสุขภาพและการพยาบาลประเมินผลด้านความรู้ด้วยข้อสอบปรนัย แบ่งเป็นข้อสอบกลางภาค 70 ข้อ ข้อสอบปลายภาค 40 ข้อ รวม 110 ข้อ และมีน้ำหนักคะแนน ร้อยละ 70 ของคะแนนทั้งหมด ทั้งนี้ ประเมินผลการเรียนรู้โดยใช้คะแนนอิงเกณฑ์และอิงกลุ่ม

งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการศึกษาครั้งนี้ ส่วนใหญ่เป็นการวิเคราะห์คุณภาพข้อสอบรายข้อด้วยวิธีการวิเคราะห์ค่าความยากง่ายและค่าอำนาจจำแนก อาทิเช่น

Backhoff, E., Larrazolo, N., & Rosas, M. (2000) ได้ทำการศึกษาระดับความยากง่ายและอำนาจจำแนกของข้อสอบประเมินความรู้และทักษะพื้นฐาน (EXHCOBA) ซึ่งเป็นข้อสอบที่นำไปใช้กันอย่างแพร่หลายในเม็กซิโก ข้อสอบ EXHCOBA เป็นข้อสอบชนิดหลายตัวเลือก ใช้ในการสอบเพื่อเลื่อนชั้นเรียน ข้อสอบประกอบด้วย 3 ส่วนย่อย คือ ส่วนที่ 1 ข้อสอบสำหรับประเมินความสามารถพื้นฐาน มีขอบเขตเนื้อหาครอบคลุม 2 ประเด็นหลักจำนวน 60 ข้อ ส่วนที่ 2 ข้อสอบสำหรับประเมินความรู้พื้นฐาน มีขอบเขตเนื้อหาครอบคลุม 4 ประเด็นหลัก จำนวน 70 ข้อ และส่วนที่ 3 ข้อสอบพื้นฐานสำหรับประเมินความรู้เฉพาะสาขา มีขอบเขตเนื้อหาครอบคลุม 9 ประเด็นหลักจำนวน 180 ข้อ ผู้สอบทุกคนจะต้องตอบคำถามใน 2 ส่วนแรก สำหรับส่วนที่ 3 ผู้สอบสามารถเลือกตอบใน 3 ประเด็นหลักที่ตนเองต้องการมีความเชี่ยวชาญเฉพาะ จำนวนข้อสอบ 60 ข้อ ข้อสอบ EXHCOBA มีข้อสอบคู่ขนานอีก 5 ชุด แต่การวิจัยนี้ ศึกษาในข้อสอบชุดที่สองเท่านั้น ผลการวิจัยพบว่าข้อสอบส่วนใหญ่มีความยากง่ายระดับกลาง ข้อสอบที่วัดความสามารถทางคณิตศาสตร์จะมีระดับของอำนาจจำแนกสูงกว่าข้อสอบที่มีเนื้อหาเชิงสังคม ซึ่งอธิบายได้ว่าหลักการทางคณิตศาสตร์เป็นหลักการที่ตายตัว และเรียงลำดับความยากง่ายไว้แล้ว เช่น ก่อนที่ผู้เรียนจะเข้าใจวิธีการได้ จะต้องผ่านความเข้าใจในการบวก ลบ คูณ มาก่อนในขณะที่หลักการเชิงสังคมไม่ได้มีการจัดลำดับการเรียนรู้เช่นคณิตศาสตร์

Ascalon, M., Evelina, Meyers, S., Lawrence, Davis, W., Bruce, & Smits, Niels (2007) วิจัยเรื่อง ความคล้ายคลึงของตัวลวงและโครงสร้างของข้อสอบ:ผลกระทบต่อความยากง่ายของข้อสอบ การวิจัยครั้งนี้ เป็นการทดสอบผลจากการวิจัย ของ Haladyna Downing และ Rodriguez (2002) ที่พบว่า ข้อสอบชนิดหลายตัวเลือก (multiple choice) ที่มีตัวลวงคล้ายคลึงกันทั้งชุดกับคำตอบที่ถูกต้อง จะทำให้ข้อสอบข้อนั้นเป็นข้อสอบที่ยากมากขึ้น และ จากการวิจัยของ Smith and Smith (1988) ที่พบว่า การใช้ตัวลวงที่มีคำตอบที่ถูกต้องแตกต่างกันอย่างสิ้นเชิง หรือตัวลวงที่มี

ความคล้ายคลึงกันกับคำตอบที่ถูกต้อง นั้นความยากง่ายของข้อสอบจะขึ้นอยู่กับการเล่นโยทย์ คำถามมากกว่า โดยบอกว่า โทย์คำถามที่มีลักษณะเป็นคำถามจะง่ายกว่า โทย์ที่มีลักษณะข้อคำถามปลายเปิดโดยเว้นไว้เติมคำตอบ การวิจัยครั้งนี้ถูกออกแบบให้มี 5 ขั้นตอนและตั้งสมมุติฐาน 2 ข้อคือ 1) ลักษณะของตัวลวงที่มีความคล้ายคลึงกันทั้งชุดกับคำตอบที่ถูกต้องจะทำให้ข้อสอบข้อนั้น ยากมากกว่าตัวลวงที่มีคำตอบที่ถูกต้องแตกต่างกันอย่างสิ้นเชิง และ 2) ลักษณะโยทย์คำถามที่มี ลักษณะคำถามง่ายกว่า ลักษณะโยทย์คำถามที่มีลักษณะปลายเปิด เว้นไว้ให้ตอบ ผลการวิจัยครั้งนี้ พบว่า ตัวลวงของข้อคำถามที่มีความสั้นยาวของตัวอักษรแต่ละข้อตัวลวง ใกล้เคียงกัน เนื้อหาของ คำตอบคล้ายคลึงกัน กับข้อคำถามที่ถูกต้อง จะทำให้ข้อสอบข้อนั้นเป็นข้อสอบที่ยากมากกว่าข้อสอบ ที่มีตัวลวงที่มีเนื้อหาของคำตอบแตกต่างกันกับคำตอบที่ถูกต้อง อย่างมีนัยสำคัญทางสถิติ ($p < 0.01$) เพราะตัวลวงที่มีเนื้อหาแตกต่างกัน จะง่ายต่อการคาดเดา เป็นไปตามสมมุติฐานที่ตั้งไว้ ข้อ 1 ส่วน สมมุติฐานที่ 2 ผลการวิจัย ไม่พบว่า ลักษณะการตั้งโยทย์คำถามที่เป็นข้อคำถามหรือโยทย์คำถามที่เป็น ข้อความปลายเปิด ไม่พบความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ในเรื่องความยากง่ายของ ข้อสอบ จากผลการวิจัยสรุปได้ว่า ข้อสอบที่ไม่ได้มาตรฐานจะมีความยาก ทำให้นักศึกษาได้คะแนน น้อยกว่าข้อสอบที่เป็นมาตรฐาน แต่ไม่สามารถแยกแยะกลุ่ม ถูก/ผิด (เก่ง/อ่อน) ได้ และนักศึกษาที่ เรียนอ่อนมักจะทำผิด ดังนั้นการใช้รูปแบบการสร้างข้อคำถามและตัวลวงควรมีการทดสอบเชิง โครงสร้างของคำถาม และความสอดคล้องระหว่างข้อถูกและผิด

Mitra, N. K., Nagaraja, H. S., Ponnudurai, G., Judson, J. P. (2009) ศึกษาระดับของ ความยากง่ายและดัชนีอำนาจจำแนกของข้อสอบชนิดหลายตัวเลือกชุด A ในการสอบรวบยอดสห สาขาวิชาปริ-คลินิกภาคการศึกษาที่ 1 การศึกษานี้เป็นการศึกษาแบบติดตามไปข้างหน้า (Cohort study) ในมหาวิทยาลัยแพทยศาสตรนานาชาติ เริ่มศึกษาตั้งแต่ปี 2003 – 2006 จากข้อสอบชุด A ซึ่งเป็นข้อสอบรวบยอด จำนวน 12 ชุด ลักษณะข้อสอบเป็นชนิด 5ตัวเลือก และมีคำตอบที่ถูกต้อง เพียงข้อเดียว ถ้าตอบถูกให้ 2 คะแนน ตอบ ผิด ให้ 0 คะแนน ไม่มีการให้คะแนนติดลบสำหรับข้อที่ ตอบผิด ค่าความยากง่ายคำนวณเป็นร้อยละ ค่าอำนาจจำแนกวิเคราะห์ด้วยโปรแกรม Excel และ วิเคราะห์ความสัมพันธ์ด้วยสถิติเพียร์สัน ผลการวิจัยพบว่า ข้อสอบที่มีอำนาจจำแนกดีคือข้อสอบที่ ยากปานกลาง ค่าเฉลี่ยความยากง่ายของแบบทดสอบรวบยอดแต่ละชุดอยู่ระหว่างร้อยละ 64 – 89 ข้อสอบจำนวน 1 ใน 3 มีค่าความยากง่ายร้อยละ 80 แสดงว่าข้อสอบง่าย ข้อสอบร้อยละ 67 มีดัชนี

อำนาจจำแนกที่ยอมรับได้ (มากกว่า 0.20) ข้อสอบจำนวน 45 ข้อจาก 120 ข้อมีดัชนีอำนาจจำแนกดี มาก และพบว่าดัชนีอำนาจจำแนกมีความสัมพันธ์กับดัชนีความยากง่าย ($r = -0.325$) ข้อสรุปจากการ วิจัยควรนำไปใช้เพื่อการเปลี่ยนแปลงแนวทางการเลือกข้อสอบสำหรับการสอบใดๆที่จัดขึ้น รวมทั้ง นำไปกำหนดกลวิธีการประเมินผลการเรียนรู้เมื่อมีการพัฒนาหลักสูตร ควรวิเคราะห์ข้อสอบรายข้อ หลังการสอบทุกครั้งเพื่อวิเคราะห์จุดอ่อนของคำตอบที่ยังใช้ได้ดีเพื่อยกระดับการประเมินผลการ เรียนรู้

Hamzah, M.S.G. and Abdullah, S.K. (2011) วิจัยเรื่องการวิเคราะห์ข้อสอบ:แนวปฏิบัติ สำหรับนักการศึกษาวิชาชีพ ซึ่งกล่าวถึงกระบวนการประเมินผลอย่างเป็นระบบ ซึ่งประกอบด้วย การแบบทดสอบ วิธีการวัดผล และวิธีการประเมินผล ซึ่งในขั้นการทำแบบทดสอบนั้นได้ใช้มาตรวัด 4 แบบในการประเมินผล ได้แก่ มาตรวัดแบบเรียงอันดับ (Ordinal scale) มาตรวัดแบบนามบัญญัติ (Nominal scale) มาตรวัดแบบช่วง (Intermittent scale) and มาตรวัดแบบอัตราส่วน (Ratio scale) เลือกใช้ข้อสอบปรนัยเป็นภาษาอังกฤษจำนวน 40 ข้อ ซึ่งถูกเตรียมขึ้นโดย MSSPC (Malaysian Secondary School Principals Conference) ในรัฐกลันตัน ประเทศมาเลเซีย และ ใช้กลุ่มตัวอย่างจำนวน 40 คนโดยใช้การใช้เครื่องมือในการวัดได้แก่ 1) การหาค่าแนวโน้มสู่ส่วนกลาง พบว่ามีผู้สอบผ่าน 18 คนจาก 40 คน คิดเป็น 45% และได้คะแนนเฉลี่ย 40.62 แสดงว่านักเรียนมี คะแนนเกาะกลุ่มกัน 2) การหาค่าจำแนก (Discriminative Index Analysis) โดยนำคะแนนที่ได้มา จัดลำดับเป็นสามกลุ่มได้แก่ กลุ่มเก่ง (13คน) กลุ่มปานกลาง (14) และกลุ่มอ่อน (13คน) จากนั้นทำ การเปรียบเทียบโดยเลือกเปรียบเทียบในกลุ่มเก่งและกลุ่มอ่อน โดยการนำค่าการจำแนกมาวิเคราะห์ พบว่ามีข้อสอบจำนวน 11 ข้อควรได้รับการปรับปรุงหรือตัดออกเนื่องจากมีค่าจำแนก (Discrimination index= ต่ำกว่า 0.10) ซึ่งไม่สามารถใช้ในการจำแนกกลุ่มเก่งและกลุ่มอ่อนได้ 3) การหาค่าความยากง่าย (Difficulty Percentage Analysis) ซึ่งจากผลตรวจข้อสอบพบว่ามี ข้อสอบจำนวน 21 ข้อที่ควรพิจารณาเก็บไว้ เนื่องจากมีค่าความยาก อยู่ในเกณฑ์ที่เหมาะสมคือ 50%-74% ซึ่งใช้จำแนกกลุ่มเก่งและกลุ่มอ่อนได้ โดยข้อสอบ 9 ข้อพบว่าง่ายเกินไปใช้แยกกลุ่มเก่ง กับกลุ่มอ่อนไม่ได้ ซึ่งผลที่ได้รับจากการศึกษาครั้งนี้ ทำให้ผู้สอนนั้นสามารถเลือกเครื่องมือใน การประเมินผลเพื่อให้เกิดประสิทธิภาพสูงสุดได้

Adeyemo Emily Oluseyi และ Ajeigbe Taiwo Olufemi (2012) วิจัยเรื่องการวิเคราะห์ข้อสอบชนิดหลายตัวเลือกในวิชาเคมีพื้นฐานของมหาวิทยาลัยในไนจีเรีย โดยมีวัตถุประสงค์เพื่อวิเคราะห์ความยาก-ง่ายและอำนาจจำแนกของข้อสอบชนิดหลายตัวเลือกในวิชาเคมีพื้นฐาน รูปแบบการวิจัยเป็นแบบการวิเคราะห์ย้อนหลังเพื่อเปรียบเทียบสาเหตุ (Ex post facto, casual comparative design) กลุ่มตัวอย่างคือนักศึกษาจำนวน 800 คน ที่ได้มาจากการสุ่มตัวอย่างแบบง่ายจากจำนวนประชากรที่เป็นนักศึกษาที่เรียนวิชาเคมี 102 ในปีการศึกษา 2008 – 2009 ข้อสอบชนิด 4 ตัวเลือกจำนวน 40 ข้อสร้างขึ้นจากเนื้อหาในรายวิชาดังกล่าว โดยอาจารย์ผู้ร่วมสอนในแต่ละหัวข้อในรายวิชานั้น ผู้วิจัยคัดเลือกนักศึกษาจำนวน 442 คนที่ได้คะแนนในเปอร์เซ็นต์ที่ 30 ทั้งจากคะแนนที่น้อยที่สุดและมากที่สุดมาเป็นหน่วยในการวิเคราะห์ สถิติที่ใช้คือค่าความถี่ ค่าความยาก - ง่าย และค่าอำนาจจำแนกผลการศึกษาพบว่า ค่าความยาก - ง่ายรวมทั้งฉบับเท่ากับ 0.54 ค่าอำนาจจำแนกเท่ากับ 0.32 ตัวเลือกที่นักศึกษาเลือกตอบมากที่สุด คือ ตัวเลือกที่ 4 ซึ่งพบถึง 16 ข้อ ในจำนวนข้อสอบที่ 40 ข้อ พบข้อสอบที่มีความผิดพลาดในการวัด 3 ข้อ เนื่องจากอำนาจจำแนกเท่ากับ 0 ซึ่งผู้วิจัยอภิปรายว่าอาจจะเกิดจากการสร้างข้อสอบไม่ถูกต้องหรือผู้ตรวจเฉลยผิด นอกจากนี้ พบด้วยว่า ข้อสอบข้อที่ยากที่สุดจำนวน 1 ข้อ มีค่าความยาก-ง่ายเท่ากับ 0.245 และมีค่าอำนาจจำแนกเท่ากับ 0.23 ซึ่งมีนักเรียนจำนวน 76 จาก 442 คนที่ตอบได้ถูก การที่ค่าความยากง่ายของข้อสอบเท่ากับ 0.54 แสดงว่าข้อสอบไม่ยากหรือง่ายจนเกินไป ค่าอำนาจจำแนกเท่ากับ 0.32 แสดงว่า นักศึกษาที่ได้คะแนนสูงในวิชานี้เลือกคำตอบได้ถูกต้อง ผู้วิจัยสรุปในตอนท้ายว่าการสร้างข้อสอบมีความสำคัญอย่างยิ่งและควรทำตารางแสดงผลการวิเคราะห์ข้อสอบไว้ท้ายข้อเพื่อเลือกข้อที่ดีไว้ในธนาคารข้อสอบ

Bharti N Karella, Ajita Pillai, Bhavisha N Vegada (2013) ศึกษาความสัมพันธ์ของระดับความยาก-ง่ายและอำนาจจำแนกของข้อสอบชนิด 4 ตัวเลือกในรายวิชาเภสัชวิทยาที่จัดสอบในนักศึกษาชั้นปีที่ 2 โดยรวบรวมข้อมูลจากข้อสอบรบบยอด 10 ชุด ที่ใช้ในการสอบระหว่าง 2008 – 2012 จำนวน 200 ข้อ ข้อกระหนงเป็นคำถามชนิด 4 ตัวเลือก และมีคำตอบที่ถูกที่สุดเพียงข้อเดียว ผู้ที่ตอบถูก จะได้ 1 คะแนน ตอบผิด ได้ - 0.25 คะแนน และผู้ที่ไม่ตอบได้ 0 คะแนน วิเคราะห์ข้อมูลโดยการตรวจข้อสอบทั้งฉบับ เรียงลำดับคะแนนจากสูงที่สุดจนถึงต่ำที่สุด จากนั้นจึงแบ่งนักศึกษาเป็น 3 กลุ่ม กลุ่มแรกและกลุ่มสุดท้าย แต่ละกลุ่มมีจำนวน 1 ใน 3 ของนักศึกษาทั้งหมด กลุ่มแรกเริ่มต้น

นับจากนักศึกษาที่ได้คะแนนสูงที่สุด กลุ่มสุดท้ายเริ่มต้นนับจากนักศึกษาที่ได้คะแนนต่ำที่สุด กลุ่มที่สองเป็นนักศึกษาที่เหลือ คำนวณค่าความยากง่ายด้วยค่าร้อยละของนักศึกษาที่ตอบถูก กำหนดค่าความยากง่ายที่เหมาะสมอยู่ระหว่างร้อยละ 30 – 70 ค่าอำนาจจำแนกที่จัดว่าดีที่สุดคือ มากกว่า 0.35 ค่าระหว่าง 0.25 – 0.35 เป็นค่าอำนาจจำแนกที่ดี ค่าระหว่าง 0.20 – 0.24 เป็นค่าที่ยอมรับได้ และค่าน้อยกว่า 0.20 เป็นข้อสอบที่ไม่ดี ผลการวิจัยสรุปได้ว่า การประเมินคุณภาพของข้อสอบที่ต้องการวัดความรู้ของนักศึกษาควรวิเคราะห์ด้วยค่าความยากง่ายและอำนาจจำแนกเป็นวิธีที่ดีสาเหตุของข้อสอบที่มีอำนาจจำแนกไม่ดีอาจเกิดจากข้อคำถามไม่ชัดเจน ตั้งคำถามในประเด็นที่ยังไม่มีข้อสรุปที่ชัดเจน เฉลยไม่ถูกต้องหรือเป็นคำถามเกี่ยวกับความคิดเห็นที่อาจเป็นไปได้ทั้งถูกและผิด ข้อสอบที่มีอำนาจจำแนกดีคือข้อสอบที่มีค่าความยากง่ายปานกลาง ในขณะที่ข้อสอบที่มีอำนาจจำแนกไม่ดีคือข้อสอบที่ยากมากหรือง่ายมากแต่ไม่มีความสัมพันธ์กันเชิงสถิติ และเสนอแนะให้นำผลการวิเคราะห์ข้อสอบมาเป็นแนวทางในการพัฒนาวิธีการจัดสอบด้วยข้อสอบชนิดหลายตัวเลือกต่อไป

Sushma S. Pande, Santosh R. Pande, Vrushali R. Parate, Archana P. Nikam, Sushil H. Agrekar, (2013) ศึกษาเรื่องความสัมพันธ์ระหว่างดัชนีความยากง่ายและอำนาจจำแนกของข้อสอบชนิดเลือกตอบในการสอบย่อยวิชาสถิติวิทยา เพื่อนำผลการวิจัยมาใช้ในการคัดเลือกข้อสอบรวบรวม โดยศึกษาข้อสอบที่ใช้สอบมาแล้วกับนักศึกษา 100 คน ย้อนหลังไป 4 ปี จำนวน 240 ข้อที่เป็นข้อสอบชนิดเลือกตอบ ข้อสอบดังกล่าวนำมาวิเคราะห์ระดับความยากง่ายและดัชนีอำนาจจำแนก ศึกษาความสัมพันธ์ด้วยสถิติวิเคราะห์ความสัมพันธ์เพียร์สัน ผลการวิจัยพบว่า ดัชนีความยากง่ายมีการกระจายตัวกว้าง (8.57 to 95.71) ดัชนีอำนาจจำแนกอยู่ระหว่าง -0.54 ถึง 0.8 ดัชนีความยากง่ายเฉลี่ยเท่ากับ $52.53 + 20.59$ และดัชนีอำนาจจำแนกเฉลี่ยเท่ากับ $0.30 + 0.18$ ข้อสอบร้อยละ 23 เป็นข้อสอบง่าย ($P > 70\%$) ร้อยละ 15 เป็นข้อสอบยาก ($P < 30\%$) ส่วนที่เหลือ (ร้อยละ 62) เป็นข้อสอบที่มีค่าความยากง่ายในระดับที่ยอมรับได้ (30 ถึง 70%) ข้อสอบร้อยละ 4 มีอำนาจจำแนกเป็นลบ ร้อยละ 21 อำนาจจำแนกไม่ดี ร้อยละ 75 มีอำนาจจำแนกในระดับที่ยอมรับได้จนถึงระดับดีมาก ดัชนีอำนาจจำแนกมีความสัมพันธ์เชิงบวกกับดัชนีความยากง่าย สังเกตพบว่าข้อสอบที่มีอำนาจจำแนกสูง ($D=0.6-0.8$) จะมีความยากง่ายระดับกลาง ($P = 40\% - 60\%$) ผู้วิจัยสรุปว่า ข้อสอบส่วนใหญ่ (ร้อยละ 75)สามารถนำมาใช้ได้ ข้อสอบที่ยากหรือง่ายเกินไปจะมีอำนาจจำแนกไม่ดี ข้อสอบที่อำนาจจำแนกเป็นลบเกิดจากโจทย์ผิดหรือเฉลยผิด

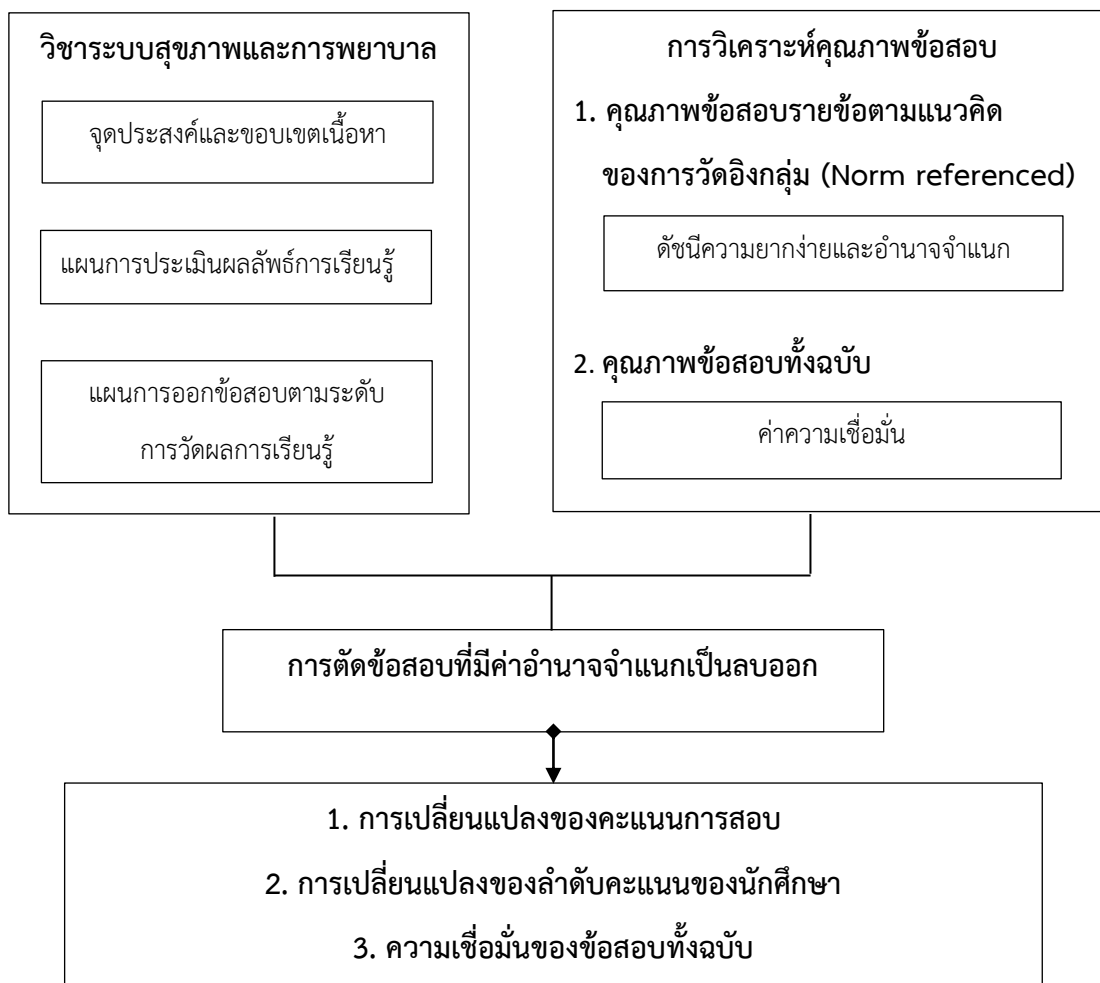
Suruchi and Surinder S. Rana. (2014) ศึกษาเรื่องความสัมพันธ์ระหว่างดัชนีความยากง่ายและดัชนีอำนาจจำแนกข้อสอบรายข้อในการสอบวัดผลรายวิชาชีววิทยา จำนวนข้อสอบที่วิเคราะห์ 120 ข้อ ใช้กับผู้เรียนจำนวน 500 คน ผู้วิจัยสร้างข้อสอบขึ้นจากเนื้อหาวิชาชีววิทยา 3 บท และผ่านการตรวจสอบจากผู้เชี่ยวชาญ กำหนดแผนการออกจำนวนข้อสอบคำนึงถึงปริมาณเนื้อหา และวัตถุประสงค์การเรียนรู้ ระดับของการทดสอบเป็นไปตามแนวคิดของบลูม (Bloom's revised taxonomy) ข้อสอบประกอบด้วยคำถามชนิดหลายตัวเลือก 80 ข้อ ข้อสอบให้ตอบ ถูก - ผิด 15 ข้อ ข้อสอบให้เติมประโยคให้สมบูรณ์ 18 ข้อ และข้อสอบจับคู่ 7 ข้อ ตรวจสอบการใช้ภาษาให้ง่ายต่อความเข้าใจ ความหมายชัดเจน ผลการวิเคราะห์ข้อสอบพบว่ามีความยากง่ายและอำนาจจำแนกในพิสัยที่ยอมรับได้ ข้อสอบจำนวน 9 ข้อถูกตัดออกเพราะมีค่าความยากง่ายสูงหรือต่ำเกินไป และอำนาจจำแนกไม่ดี ถึงแม้ว่าความสัมพันธ์ระหว่างดัชนีความยากง่ายกับอำนาจจำแนกมีลักษณะเป็นเส้นตรง แต่พบว่าข้อสอบมีอำนาจจำแนกเพิ่มขึ้นเมื่อค่าความยากง่ายเพิ่มขึ้น แต่จะลดลงในข้อสอบที่ยากหรือง่ายเกินไป

งานวิจัยที่เกี่ยวข้องข้างต้น สรุปได้ว่า การวิเคราะห์ข้อสอบรายข้อส่วนใหญ่ใช้ทฤษฎีการสอบแบบดั้งเดิม (Classical Test Theory) ซึ่งประกอบด้วยการวิเคราะห์ความยากง่ายของข้อสอบและการวิเคราะห์อำนาจจำแนก นอกจากนี้ งานวิจัยบางส่วนได้วิเคราะห์ความสัมพันธ์ระหว่างความยากง่ายกับอำนาจจำแนกด้วยสถิติความสัมพันธ์ เพียร์สัน ผลการวิจัยในภาพรวมสรุปได้ว่า ความยากง่ายกับอำนาจจำแนกมีความสัมพันธ์กัน ข้อสอบที่มีอำนาจจำแนกดีจะมีความยากง่ายระดับกลางและเป็นข้อสอบที่ควรเลือกเก็บไว้ใช้ต่อไป

กรอบแนวคิดในการวิจัย

แสดงดังภาพที่ 2.1

ภาพที่ 2.1 กรอบแนวคิดในการวิจัย



บทที่ 3 วิธีดำเนินการวิจัย

การวิจัยเรื่อง ผลของการตัดข้อสอบที่มีอำนาจจำแนกติดลบออกต่อคะแนนการสอบ และความเชื่อมั่นของข้อสอบทั้งฉบับ: รายวิชาระบบสุขภาพและการพยาบาล เป็นการวิจัยเชิงบรรยาย มีวัตถุประสงค์ เพื่อคุณภาพของข้อสอบในรายวิชาระบบสุขภาพและการพยาบาลแบบรายข้อและทั้งฉบับ และศึกษาการเปลี่ยนแปลงของคะแนนการสอบและลำดับคะแนนก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออกโดยมีระเบียบวิธีวิจัยดังนี้

ประชากรและกลุ่มตัวอย่าง

การวิจัยนี้ศึกษาในกลุ่มประชากร คือ ข้อสอบกลางภาคและปลายภาครายวิชาระบบสุขภาพและการพยาบาลปีการศึกษา 2555 ที่ใช้กับนักศึกษาพยาบาลชั้นปีที่ 1 ที่ลงทะเบียนเรียนจำนวน 92 คน

เครื่องมือในการวิจัยและการตรวจสอบคุณภาพเครื่องมือ

เครื่องมือในการวิจัยนี้ คือ ข้อสอบวิชาระบบสุขภาพและการพยาบาลที่ผู้วิจัยสร้างขึ้น มีเนื้อหาสอดคล้องกับคำอธิบายรายวิชาที่กำหนดไว้ในหลักสูตรพยาบาลศาสตรบัณฑิต (หลักสูตรปรับปรุง พ.ศ. 2555) มหาวิทยาลัยราชภัฏสวนดุสิต และขอบเขตเนื้อหาพยาบาลที่กำหนดไว้ใน มคอ. 3 ปีการศึกษา 2555 ซึ่งผ่านการพิจารณารายละเอียดในการออกแบบรายวิชาจากกรรมการวิชาการ คณะพยาบาลศาสตร์ ข้อสอบจำนวน 2 ชุด ในการวิจัยนี้ได้แก่ ข้อสอบชุดที่ 1 ใช้ในการสอบกลางภาค จำนวน 70 ข้อ และข้อสอบชุดที่ 2 ใช้ในการสอบปลายภาค จำนวน 40 ข้อ เนื้อหาของข้อสอบแต่ละชุดเป็นอิสระต่อกัน ตามแผนการเรียนรู้ที่กำหนดไว้ใน มคอ. 3 จำนวนข้อของแต่ละชุดแปรตามจำนวนชั่วโมงสอน กล่าวคือ ชั่วโมงสอน 1 ชั่วโมงต่อข้อสอบ 5 ข้อ ลักษณะข้อสอบเป็นคำถามชนิดเลือกตอบสี่ตัวเลือก โดยมีคำตอบที่ถูกต้องที่สุดเพียงข้อเดียว เกณฑ์การให้คะแนน ในกรณีที่ตอบถูกให้ 1 คะแนน ตอบผิดให้ 0 คะแนน

การตรวจสอบคุณภาพข้อสอบกลางภาคและปลายภาคดำเนินการโดยการวิพากษ์ของอาจารย์ร่วมสอน ร่วมกับส่งข้อสอบให้ผู้ทรงคุณวุฒิภายนอกอ่านประเมิน ซึ่งพบว่าต้องปรับปรุงข้อคำถามบางข้อที่ใช้ภาษาไม่ชัดเจน จำนวนข้อสอบกลางภาคและปลายภาคที่ปรับการใช้ภาษา 4 ข้อ และ 2 ข้อ ตามลำดับ ในส่วนของเนื้อหาของข้อสอบครอบคลุมตามที่กำหนดไว้ในคำอธิบายรายวิชา จำนวนข้อสอบและ

ระดับของการวัดผลการเรียนรู้ตามแนวคิดของบลูม (Bloom's revised taxonomy) และสอดคล้องกับพิมพ์เขียวข้อสอบที่ระบุไว้ใน มคอ. 3

การเก็บรวบรวมข้อมูล

เมื่อสิ้นสุดการสอบกลางภาคและปลายภาค ภาคการศึกษาที่ 1 ปีการศึกษา 2555 แล้ว ผู้วิจัยส่งข้อสอบพร้อมเฉลยของแต่ละชุดให้สำนักส่งเสริมวิชาการและงานทะเบียน มหาวิทยาลัยราชภัฏสวนดุสิตนำไปคำนวณคะแนนรวม วิเคราะห์ความยากง่าย ค่าอำนาจจำแนกของข้อสอบรายข้อและวิเคราะห์ความเชื่อมั่นของข้อสอบทั้งฉบับด้วยโปรแกรมสำเร็จรูปคอมพิวเตอร์ หลักการวิเคราะห์ที่อยู่ภายใต้ทฤษฎีการสอบแบบดั้งเดิม (classical test theory) และใช้เทคนิคการวิเคราะห์คุณภาพรายข้อตามแนวคิดของการวัดอิงกลุ่ม (Brown, W. and Thomson, H.G.,1921; เยาวดี ราชชัยกุล, 2551)

การแบ่งกลุ่มผู้เรียน ในการวิจัยครั้งนี้จะใช้เทคนิค 27% วิธีนี้นิยมใช้เมื่อผู้สอบแบบทดสอบมีจำนวนมากหรือการกระจายของคะแนนสอบเป็นปกติ วิธีนี้เชื่อว่ากลุ่มขนาดตัวอย่าง 27% จากผู้สอบที่ได้คะแนนสูงหรือผู้ที่สอบได้คะแนนต่ำสามารถเป็นตัวแทนประชากรของผู้ที่สอบทั้งหมด ส่วนคะแนนที่เหลืออีก 46% ซึ่งเป็นของคนที่มีความสามารถปานกลางไม่ต้องนำมาวิเคราะห์ การใช้ 27% นี้จะได้ผลเหมือนกับการนำเอาผู้เรียนทั้งหมด (100%) มาวิเคราะห์

การวิเคราะห์ข้อมูล

วัตถุประสงค์การวิจัยครั้งนี้ คือเพื่อศึกษาคุณภาพของข้อสอบในรายวิชาระบบสุขภาพและการพยาบาล แบบรายข้อและทั้งฉบับ และศึกษาการเปลี่ยนแปลงของคะแนนการสอบและลำดับคะแนนก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออกดังนั้น สถิติที่ใช้ในการวิจัยประกอบด้วยสถิติต่อไปนี้

1. ศึกษาคุณภาพของข้อสอบรายข้อด้วยการคำนวณค่าความยากง่าย (P) และค่าอำนาจจำแนก (R) และศึกษาความเชื่อมั่นของข้อสอบทั้งฉบับด้วยการคำนวณค่าสัมประสิทธิ์ความเชื่อมั่นของคูเดอร์ - ริชาร์ดสัน (KR - 20)
2. การเปลี่ยนแปลงของคะแนนสอบรายวิชาก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออกด้วยสถิติ Paired t-test (one-tailed)
3. การเปลี่ยนแปลงของลำดับคะแนนก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออกด้วยสถิติ Sign Ranked test

บทที่ 4

ผลการวิจัย

การวิจัยเรื่อง ผลของการตัดข้อสอบที่มีอำนาจจำแนกติดลบออกต่อคะแนนการสอบ และความเชื่อมั่นของข้อสอบทั้งฉบับ: รายวิชาระบบสุขภาพและการพยาบาล เป็นการวิจัยเชิงบรรยาย มีวัตถุประสงค์ เพื่อคุณภาพของข้อสอบในรายวิชาระบบสุขภาพและการพยาบาลแบบรายข้อและทั้งฉบับ และศึกษาการเปลี่ยนแปลงของคะแนนการสอบและลำดับคะแนนก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก ประชากรที่ศึกษาประกอบด้วยข้อสอบวิชาระบบสุขภาพและการพยาบาลที่ใช้กับนักศึกษาจำนวน 92 คนในปีการศึกษา 2555 จำนวน 2 ชุด คือชุดที่ 1 ข้อสอบกลางภาค จำนวน 70 ข้อ และ ชุดที่ 2 ข้อสอบปลายภาค จำนวน 40 ข้อ วิเคราะห์ ความยากง่ายและอำนาจจำแนกของข้อสอบรายข้อและวิเคราะห์ความเชื่อมั่นของข้อสอบรวมทั้งฉบับด้วยโปรแกรมสำเร็จรูปคอมพิวเตอร์ หลักการวิเคราะห์ที่อยู่ภายใต้ทฤษฎีการสอบแบบดั้งเดิม (Classical Test Theory) และใช้เทคนิคการวิเคราะห์คุณภาพรายข้อตามแนวคิดของการวัดอิงกลุ่ม ผลการวิเคราะห์ข้อมูลแสดงดังต่อไปนี้

คุณภาพของข้อสอบรายวิชาระบบสุขภาพและการพยาบาล

ผลการวิเคราะห์คุณภาพของข้อสอบกลางภาคและปลายภาควิชาระบบสุขภาพและการพยาบาลในภาคการศึกษาที่ 1 ปีการศึกษา 2555 ประกอบด้วยผลการวิเคราะห์ค่าความยากง่าย และค่าอำนาจจำแนกสำหรับการวิเคราะห์รายข้อ และผลการวิเคราะห์ค่าความเชื่อมั่นสำหรับการวิเคราะห์คุณภาพข้อสอบทั้งฉบับผลการศึกษานำเสนอโดยลำดับดังนี้

1. ผลการวิเคราะห์คุณภาพข้อสอบรายข้อ

1.1 ผลการวิเคราะห์ความยากง่ายของข้อสอบ แสดงดังตารางที่ 4.1 และ 4.2

ตารางที่ 4.1 ค่าความยากง่ายของข้อสอบกลางภาควิชาระบบสุขภาพและการพยาบาลทั้งฉบับ
(N = 92 ชุด)

ค่าความยากง่าย (P)	ความหมาย	จำนวนข้อสอบ	ร้อยละ
0.80 – 1.00	ง่ายมาก	18	25.71
0.60 – 0.79	ค่อนข้างง่าย	23	32.86
0.40 – 0.59	ยากง่ายปานกลาง	13	18.57
0.20 – 0.39	ค่อนข้างยาก	9	12.86
0.00 – 0.19	ยากมาก	7	10.00

ตารางที่ 4.1 แสดงให้เห็นว่าข้อสอบกลางภาควิชาระบบสุขภาพและการพยาบาลร้อยละ 64.29 (12.86 + 18.57 + 32.86) เป็นข้อสอบที่มีค่าความยากง่าย (P) อยู่ระหว่าง .20 - .80

ตารางที่ 4.2 ค่าความยากง่ายของข้อสอบปลายภาควิชาระบบสุขภาพและการพยาบาลทั้งฉบับ
(N = 92 ชุด)

ค่าความยากง่าย (P)	ความหมาย	จำนวนข้อสอบ	ร้อยละ
0.80 – 1.00	ง่ายมาก	21	52.50
0.60 – 0.79	ค่อนข้างง่าย	8	20.00
0.40 – 0.59	ยากง่ายปานกลาง	5	12.50
0.20 – 0.39	ค่อนข้างยาก	3	7.50
0.00 – 0.19	ยากมาก	3	7.50

ตารางที่ 4.2 แสดงให้เห็นว่าข้อสอบปลายภาควิชาระบบสุขภาพและการพยาบาลเป็นข้อสอบร้อยละ 40 (7.5 + 12.5 + 20.0) เป็นข้อสอบที่มีค่าความยากง่าย(P) อยู่ระหว่าง .20 - .80

1.2 ผลการวิเคราะห์อำนาจจำแนกของข้อสอบ แสดงดังตารางที่ 4.3 และ 4.4

ตารางที่ 4.3 ค่าอำนาจจำแนกของข้อสอบกลางภาควิชาระบบสุขภาพและการพยาบาลทั้งฉบับ
(N = 92 ชุด)

ค่าอำนาจจำแนก (R)	ความหมาย	จำนวนข้อสอบ	ร้อยละ
0.80 – 1.00	จำแนกได้ดีมาก	-	-
0.60 – 0.79	จำแนกได้ดี	-	-
0.40 – 0.59	จำแนกได้ปานกลาง	6	8.57
0.20 – 0.39	จำแนกได้น้อย	26	37.14
0.00 – 0.19	จำแนกได้น้อยมาก	34	48.57
-1.00 – 0.00	จำแนกไม่ได้	4	5.71

ตารางที่ 4.3 แสดงให้เห็นว่าข้อสอบกลางภาควิชาระบบสุขภาพและการพยาบาลร้อยละ 45.71 (37.14 + 8.57) เป็นข้อสอบที่มีค่าอำนาจจำแนก (R) .20 ขึ้นไป

ตารางที่ 4.4 ค่าอำนาจจำแนกของข้อสอบปลายภาควิชาระบบสุขภาพและการพยาบาลทั้งฉบับ
(N = 92 ชุด)

ค่าอำนาจจำแนก (R)	ความหมาย	จำนวนข้อสอบ	ร้อยละ
0.80 – 1.00	จำแนกได้ดีมาก	-	-
0.60 – 0.79	จำแนกได้ดี	-	-
0.40 – 0.59	จำแนกได้ปานกลาง	2	5.00
0.20 – 0.39	จำแนกได้น้อย	17	42.50
0.01 – 0.19	จำแนกได้น้อยมาก	18	45.00
-1.00 – 0.00	จำแนกไม่ได้	3	7.50

ตารางที่ 4.4 แสดงให้เห็นว่าข้อสอบปลายภาควิชาระบบสุขภาพและการพยาบาลร้อยละ 47.50 ($42.50 + 5.00$) เป็นข้อสอบที่มีค่าอำนาจจำแนก (R) .20 ขึ้นไป

โดยสรุป คุณภาพข้อสอบวิชาระบบสุขภาพและการพยาบาลเมื่อวิเคราะห์รายข้อพบว่าเป็นข้อสอบที่มีค่าความยากง่ายและค่าอำนาจจำแนกอยู่ในเกณฑ์นำไปใช้ได้

2. ผลการวิเคราะห์คุณภาพข้อสอบทั้งฉบับ พบว่าคำนวณได้ค่าความเชื่อมั่น ($KR - 20$) ของข้อสอบกลางภาคและปลายภาค 0.505 และ 0.344 ตามลำดับ

การเปลี่ยนแปลงภายหลังจากตัดข้อสอบที่มีค่าอำนาจจำแนกเป็นลบออก

เมื่อตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออกส่งผลต่อการเปลี่ยนแปลงของคะแนนการสอบ ลำดับคะแนนและค่าความเชื่อมั่นของข้อสอบทั้งฉบับ ซึ่งนำเสนอโดยลำดับดังนี้

1. การเปลี่ยนแปลงของคะแนนการสอบและลำดับคะแนน แสดงดังตารางที่ 4.5 และ ตารางที่ 4.6

ตารางที่ 4.5 ผลการทดสอบการเปลี่ยนแปลงของคะแนนการสอบของนักศึกษาในวิชา ระบบสุขภาพและการพยาบาล หลังการตัดข้อสอบที่มีค่า r ติดลบออก ($N = 92$ ชุด)

ชุดข้อสอบ	คะแนนข้อสอบต้นฉบับ			คะแนนภายหลังตัดข้อที่มีค่า r ติดลบออก			Paired t-test
	พิสัย	Mean	s.d.	พิสัย	Mean	s.d.	
กลางภาค	29 - 51	41.12	4.932	30 - 53	42.23	5.351	-10.225**
ปลายภาค	21 - 36	28.70	2.843	21 - 36	28.43	3.011	4.972**

$df = 91$, ** p value $< .001$

ตารางที่ 4.5 แสดงให้เห็นว่าคะแนนเฉลี่ยทั้งฉบับของข้อสอบวิชาวิชาการระบบสุขภาพและการพยาบาล กลางภาคและปลายภาค ก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก มีการเปลี่ยนแปลงของคะแนนการสอบของนักศึกษา กล่าวคือ คะแนนเฉลี่ยของการสอบกลางภาคเพิ่มขึ้น จาก 41.12 คะแนน เป็น 42.23 คะแนน คะแนนเฉลี่ยของการสอบปลายภาคลดลงจาก 28.70 คะแนน เป็น 28.43 คะแนน การเปลี่ยนแปลงดังกล่าวมีนัยสำคัญทางสถิติ (p value < .001)

ตารางที่ 4.6 ผลการทดสอบการเปลี่ยนแปลงของลำดับคะแนนจากการสอบของนักศึกษาในวิชา ระบบสุขภาพและการพยาบาล หลังการตัดข้อสอบที่มีค่า r ติดลบออก (N= 92 ชุด)

ชุดข้อสอบ	จำนวนนักศึกษาที่มีลำดับคะแนนเปลี่ยนแปลง			Signed Rank Test
	เพิ่มขึ้น	ลดลง	คงเดิม	
กลางภาค	83	8	1	-7.339**
ปลายภาค	29	63	0	-4.884**

** p value < .001

ตารางที่ 4.6 แสดงให้เห็นว่าผลการทดสอบการเปลี่ยนแปลงของลำดับคะแนนจากการสอบของนักศึกษาในวิชา ระบบสุขภาพและการพยาบาล หลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก จำนวนนักศึกษาที่มีลำดับคะแนนจากการสอบกลางภาคเพิ่มขึ้น 83 คน ลดลง 8 คน และคงเดิม 1 คน จำนวนนักศึกษาที่มีลำดับคะแนนจากการสอบปลายภาคเพิ่มขึ้น 29 คน และลดลง 63 คน ไม่มีผู้ได้คะแนนคงเดิม การเปลี่ยนแปลงของลำดับคะแนนการสอบมีนัยสำคัญทางสถิติ (p -value < .001)

สรุป

เมื่อตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออกทำให้คะแนนการสอบแตกต่างจากคะแนนเดิมอย่างมีนัยสำคัญทางสถิติ (p value < .001) และมีการเปลี่ยนแปลงของลำดับคะแนนอย่างมีนัยสำคัญทางสถิติ (p -value < .001) เช่นกัน

2. การเปลี่ยนแปลงค่าความเชื่อมั่นของข้อสอบ

ผลการศึกษาคูณภาพข้อสอบทั้งฉบับด้วยค่าความเชื่อมั่นและการเปลี่ยนแปลงของค่าความเชื่อมั่นของข้อสอบเมื่อตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออกแสดงดังตารางที่ 4.7

ตารางที่ 4.7 ค่าความเชื่อมั่นของข้อสอบวิชาระบบสุขภาพและการพยาบาลก่อน และหลังการตัดข้อสอบที่มีค่า r ติดลบออก (N = 92 ชุด)

ชุดข้อสอบ	จำนวนและร้อยละของข้อสอบที่ตัดออก	ค่าความเชื่อมั่นต้นฉบับ	ค่าความเชื่อมั่นภายหลังตัดข้อสอบที่ r ติดลบออก
กลางภาค (70 ข้อ)	4 (5.71)	0.505	0.554
ปลายภาค (40 ข้อ)	3 (7.50)	0.344	0.388

ตารางที่ 4.7 อธิบายได้ว่า หลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออกจากชุดข้อสอบกลางภาค และข้อสอบปลายภาคพบว่าข้อสอบกลางภาคและข้อสอบปลายภาคมีค่าความเชื่อมั่นทั้งฉบับเพิ่มขึ้นจาก 0.505 เป็น 0.554 และจาก 0.344 เป็น 0.388 ตามลำดับ

สรุป

การตัดข้อสอบที่มีค่า r ติดลบออกทำให้ค่าความเชื่อมั่นของข้อสอบทั้งฉบับเปลี่ยนแปลงในทางเพิ่มขึ้น

ผลการวิจัยสรุปได้ว่าข้อสอบวิชาระบบสุขภาพและการพยาบาลส่วนใหญ่เป็นข้อสอบที่นำไปใช้ได้จำนวนข้อสอบกลางภาคและปลายภาคมีค่าอำนาจจำแนกเป็นลบร้อยละ 5.71 และ 7.50 ตามลำดับ เมื่อตัดข้อสอบที่มีค่าอำนาจจำแนกเป็นลบออกพบการเปลี่ยนแปลงของคะแนนการสอบและลำดับคะแนนจากการสอบอย่างมีนัยสำคัญทางสถิติและพบว่าค่าความเชื่อมั่นของข้อสอบวิชาระบบสุขภาพและการพยาบาลหลังการตัดข้อสอบที่มีค่า r ติดลบออกเพิ่มขึ้น

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

สรุปผลการวิจัย

การวิจัยเรื่อง ผลของการตัดข้อสอบที่มีอำนาจจำแนกติดลบออกต่อคะแนนการสอบและความเชื่อมั่นของข้อสอบทั้งฉบับเป็นการวิจัยเชิงบรรยาย มีวัตถุประสงค์เพื่อศึกษาคุณภาพของข้อสอบกลางภาคและปลายภาค การเปลี่ยนแปลงของคะแนนการสอบและลำดับคะแนน และการเปลี่ยนแปลงค่าความเชื่อมั่นของข้อสอบทั้งฉบับ ก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก ประชากรที่ศึกษาประกอบด้วยข้อสอบวิชาระบบสุขภาพและการพยาบาลในภาคการศึกษาที่ 1 ปีการศึกษา 2555 จำนวน 2 ชุด แบ่งเป็น ชุดที่ 1 ข้อสอบกลางภาค จำนวน 70 ข้อ และ ชุดที่ 2 ข้อสอบปลายภาค จำนวน 40 ข้อ ที่ใช้กับนักศึกษาจำนวน 92 คนทั้งสองชุด วิเคราะห์ความยากง่ายและอำนาจจำแนกของข้อสอบทั้งรายข้อและรวมทั้งฉบับ ด้วยโปรแกรมสำเร็จรูปคอมพิวเตอร์ หลักการวิเคราะห์ที่อยู่ภายใต้ทฤษฎีการสอบแบบดั้งเดิม (classical test theory) และใช้เทคนิคการวิเคราะห์คุณภาพรายข้อตามแนวคิดของการวัดอิงกลุ่ม ผลการวิจัยสรุปได้ว่าข้อสอบวิชาระบบสุขภาพและการพยาบาลส่วนใหญ่เป็นข้อสอบที่นำไปใช้ได้ จำนวนข้อสอบกลางภาคและปลายภาคมีค่าอำนาจจำแนกเป็นลบร้อยละ 5.71 และ 7.50 ตามลำดับ เมื่อตัดข้อสอบที่มีค่าอำนาจจำแนกเป็นลบออกพบการเปลี่ยนแปลงของคะแนนการสอบและลำดับคะแนนจากการสอบอย่างมีนัยสำคัญทางสถิติและพบว่าค่าความเชื่อมั่นของข้อสอบวิชาระบบสุขภาพและการพยาบาลหลังการตัดข้อสอบที่มีค่า r ติดลบออกเพิ่มขึ้น

อภิปรายผล

การศึกษาคุณภาพข้อสอบ

ข้อสอบวิชาระบบสุขภาพและการพยาบาลสร้างขึ้นโดยมีวัตถุประสงค์เพื่อประเมินผลสัมฤทธิ์การเรียนรู้ที่บังเกิดกับผู้เรียนผ่านแบบทดสอบ เป้าหมายของการสอบคือคะแนนที่แม่นยำและสามารถ

จำแนกความสามารถของผู้เรียนได้อย่างแท้จริง เนื้อหาของข้อสอบสร้างขึ้นจากเนื้อหาที่นักศึกษาได้รับจากกระบวนการเรียนการสอน ลักษณะข้อสอบเป็นคำถามชนิดหลายตัวเลือกโดยมีคำตอบที่ถูกต้องที่สุดเพียงข้อเดียวและมีตัวลวง 3 ตัว เมื่อสิ้นสุดการสอบทั้งกลางภาคและปลายภาค ได้มีการวิเคราะห์ความยากง่ายและอำนาจจำแนกของข้อสอบดังกล่าว การศึกษาคุณภาพข้อสอบเช่นนี้สอดคล้องกับแนวคิดของการวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการสอบแบบดั้งเดิม (Classical Test Theory) แบบอิงกลุ่ม (เยาวดี วิบูลย์ศรี, 2551) กล่าวได้ว่าการศึกษาคุณภาพข้อสอบโดยวิเคราะห์ในประเด็นของความยากง่ายและอำนาจจำแนกเป็นวิธีการที่เหมาะสมและเป็นสากล ซึ่งพบว่างานวิจัยเกี่ยวกับการศึกษาคุณภาพข้อสอบส่วนใหญ่วิเคราะห์ในประเด็นเช่นเดียวกันกับการศึกษาคั้งนี้ เช่นการวิจัยของ Backhoff, E., Larrazolo, N., & Rosas, M. (2000) Mitra, N. K., Nagaraja, H. S., Ponnud

ผลการวิเคราะห์ความยากง่ายของข้อสอบวิชาการระบบสุขภาพและการพยาบาลพบว่าข้อสอบส่วนใหญ่ค่อนข้างง่าย อภิปรายได้ว่า วิชานี้เป็นวิชาเป็นวิชาแรกของนักศึกษาที่มีเนื้อหาเกี่ยวข้องกับวิชาชีพการพยาบาลเพราะวิชาในชั้นปีที่ 1 ภาคการศึกษาที่ 1 เน้นวิชาในหมวดการศึกษาทั่วไปมากกว่า นักศึกษายังไม่มีความรู้เกี่ยวกับวิชาการพยาบาลมาก่อน พิมพ์เขียวข้อสอบที่กำหนดไว้ตามระดับของการเรียนรู้ส่วนใหญ่เน้นการรู้-จำและเข้าใจ ผลการวิเคราะห์ความยากง่ายจึงสะท้อนคุณลักษณะผู้เรียนในกลุ่มที่ศึกษาว่ามีเก่งในระดับใกล้เคียงกัน สามารถตอบข้อสอบได้ถูกต้อง เมื่อนำมาวิเคราะห์ความยากง่ายจึงพบว่าข้อสอบส่วนใหญ่ค่อนข้างง่าย ผลที่ตามมาของข้อสอบวิชานี้คืออำนาจจำแนกแตกต่างของผู้เรียนได้น้อย ดังจะพบจากผลการวิจัยว่าข้อสอบส่วนใหญ่มีค่าอำนาจจำแนกไม่เกิน 0.40 ซึ่งเมื่อเปรียบเทียบกับค่าอำนาจจำแนกของข้อสอบในการวิจัยนี้มีค่าใกล้เคียงกับงานวิจัยของ Adeyemo Emily Oluseyi และ AjeigbeTaiwoOlufemi (2012) ที่พบว่าข้อสอบที่ศึกษามีค่าอำนาจจำแนก 0.32 อย่างไรก็ตาม การวิจัยนี้ไม่ได้ศึกษาเชิงสถิติเพื่อวิเคราะห์ความสัมพันธ์ระหว่างความยากง่ายกับอำนาจจำแนกเพียงแต่พบความสัมพันธ์กันของผลการวิเคราะห์คุณภาพข้อสอบว่าส่วนใหญ่เป็นข้อสอบง่ายและมีอำนาจจำแนกต่ำ ซึ่งพบในการศึกษาของ Sushma S. Pande, Santosh R. Pande, Vrushali R. Parate, Archana P. Nikam, Sushil H. Agrekar, (2013) เช่นกัน แต่การศึกษาที่กล่าวอ้างนี้ได้วิเคราะห์นัยสำคัญของความสัมพันธ์ดังกล่าวและพบว่าดัชนีอำนาจจำแนกมีความสัมพันธ์เชิงบวกกับดัชนีความยากง่าย

การเปลี่ยนแปลงของคะแนนสอบและลำดับคะแนนการสอบภายหลังการตัดข้อสอบที่มีค่าอำนาจ จำแนกเป็นลบออก

ผลการวิจัยพบว่าเมื่อตัดข้อสอบที่มีค่าอำนาจจำแนกเป็นลบออกพบการเปลี่ยนแปลงของคะแนนการสอบอย่างมีนัยสำคัญทางสถิติ อภิปรายได้ว่า ภายหลังจากตัดข้อสอบที่ไม่สามารถจำแนกผู้เรียนที่เก่งออกจากผู้เรียนที่อ่อนได้แล้ว คะแนนสอบที่ได้จึงสอดคล้องกับความสามารถที่แท้จริงของนักศึกษามากขึ้น นอกจากนี้ เมื่อตัดข้อสอบที่มีค่าอำนาจจำแนกเป็นลบออกทำให้คะแนนที่ได้มีค่าเข้าใกล้คะแนนที่แท้จริง (true score) มากขึ้น สอดคล้องกับข้อเสนอแนะของ Tavakol & Dennick (2011) ที่กล่าวถึงความสำคัญและความจำเป็นของการวิเคราะห์คุณภาพข้อสอบ ทั้งนี้ ในการทดสอบควรลดความคลาดเคลื่อนของการวัดผลให้น้อยลงอย่างที่สุด

สำหรับการเปลี่ยนแปลงของลำดับคะแนนจากการสอบอย่างมีนัยสำคัญทางสถิติเมื่อตัดข้อสอบที่มีค่าอำนาจจำแนกเป็นลบออก อภิปรายได้ว่าการตัดข้อสอบดังกล่าว ทำให้คะแนนที่ได้ใหม่มีค่าเข้าใกล้คะแนนที่แท้จริง (true score) เช่นกัน อย่างไรก็ตาม ควรเพิ่มการศึกษาเกี่ยวกับการเปลี่ยนแปลงของลำดับคะแนนในทุกครั้งที่มีการวิเคราะห์คุณภาพข้อสอบเพื่อค้นหาข้อสรุปเชิงอภิมานถึงการเปลี่ยนแปลงดังกล่าวก่อนนำไปกำหนดนโยบายในการพิจารณาค่าคะแนน ทั้งนี้ จากงานวิจัยที่รวบรวมได้ ยังไม่ปรากฏผลการศึกษาที่สอดคล้อง

การเปลี่ยนแปลงค่าความเชื่อมั่นของข้อสอบทั้งฉบับ ก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจ จำแนกติดลบออก

ผลการวิจัยพบว่าเมื่อตัดข้อสอบที่มีค่าอำนาจจำแนกเป็นลบออกทำให้ค่าความเชื่อมั่นของข้อสอบวิชาระบบสุขภาพและการพยาบาลหลังการตัดข้อสอบที่มีค่า r ติดลบออกเพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ สามารถอภิปรายได้ว่าการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบมีผลให้ข้อสอบมีความคงที่ของผลการวัดเพิ่มขึ้น ซึ่งเยาเวดี วิบูลย์ศรี (2551) กล่าวว่า ปัจจัยที่มีอิทธิพลต่อความเชื่อมั่นของข้อสอบได้แก่ ระดับความยากง่าย กล่าวคือข้อสอบที่ยากหรือง่ายเกินไปจะส่งผลให้ค่าความเชื่อมั่นต่ำ ซึ่งในกรณีนี้ ผล

การวิเคราะห์คุณภาพข้อสอบวิชาระบบสุขภาพและการพยาบาลพบว่าเป็นข้อสอบที่ค่อนข้างง่าย จึงส่งผลให้ค่าความเชื่อมั่นปานกลางถึงต่ำ (0.554 และ 0.388) สำหรับการเปลี่ยนแปลงของค่าความเชื่อมั่นที่มีนัยสำคัญทางสถิติเป็นผลมาจากจำนวนแบบทดสอบที่มีมากเพียงพอที่จะลดความแปรปรวนของความสอดคล้องภายในของข้อสอบได้ (เยาวดี วิบูลย์ศรี (2551))

ข้อสังเกตจากการวิจัยนี้พบว่าการอภิปรายผลการวิจัยด้วยวรรณกรรมที่เกี่ยวข้องได้ในประเด็นของการศึกษาคุณภาพข้อสอบด้วยความยากง่ายเท่านั้น เนื่องจากในการทบทวนวรรณกรรมส่วนใหญ่พบการวิจัยเกี่ยวกับการศึกษาคุณภาพข้อสอบด้วยความยากง่ายและอาจจำแนก แต่ไม่พบรายงานการวิจัยผลกระทบของการตัดข้อสอบที่มีค่าอำนาจจำแนกเป็นลบ ดังนั้น ในทุกครั้งที่มีการสอบควรได้มีการวิเคราะห์คุณภาพข้อสอบทุกครั้งและควรเพิ่มการวิเคราะห์คุณภาพตัวลวง (distractor) เพื่อให้การปรับปรุงคุณภาพข้อสอบเป็นระบบมากขึ้น

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. ควรวิเคราะห์ข้อสอบทุกครั้ง ในกรณีที่พบข้อสอบที่มีค่าความยากง่ายและค่าอำนาจจำแนกต่ำ หรือมีค่าอำนาจจำแนกเป็นลบควรปรับปรุงคุณภาพข้อสอบก่อนนำไปใช้ใหม่
2. ควรคัดเลือกข้อสอบที่คุณภาพดีไว้ในคลังข้อสอบเพื่อใช้ในการสอบครั้งต่อไปและการพัฒนาข้อสอบมาตรฐาน
3. ควรพัฒนาคุณภาพข้อสอบด้วยกระบวนการวิเคราะห์ข้อสอบทั้งเชิงปริมาณและเชิงคุณภาพ เพื่อให้ได้ข้อสอบที่มีค่าความยากง่ายและค่าอำนาจจำแนกตามเกณฑ์มาตรฐาน มีความตรงตามโครงสร้าง (Content Validity) และสอดคล้องกับแผนการออกข้อสอบ (Test Blueprint) ซึ่งสะท้อนวัตถุประสงค์การเรียนรู้เชิงพฤติกรรมและเนื้อหาที่เป็นสาระสำคัญของรายวิชา

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

1. ควรศึกษาคุณภาพข้อสอบตามแนวทฤษฎีการตอบสนองข้อกระทง (Item Respond Theory) เพื่อการพัฒนาข้อสอบมาตรฐาน
2. ควรขยายขอบเขตการวิจัยให้ครอบคลุมข้อสอบที่มีค่าอำนาจจำแนกต่ำ (< 0.20) และการวิเคราะห์ตัวลวง
3. ควรศึกษาคุณภาพข้อสอบที่มีจำนวนข้อใกล้เคียงกันเพื่อลดความคลาดเคลื่อนในการวิเคราะห์

บรรณานุกรม

บรรณานุกรมภาษาไทย

- เกียรติสุดา ศรีสุข. (2552). *ระเบียบวิธีวิจัย*. พิมพ์ครั้งที่ 2. เชียงใหม่: โรงพิมพ์ครองช่าง.
- คณะพยาบาลศาสตร์ มหาวิทยาลัยราชภัฏสวนดุสิต. (2554). *คู่มือหลักสูตรพยาบาลศาสตร์ (หลักสูตรปรับปรุง 2554)*. กรุงเทพฯ: กราฟฟิคไซด์.
- เชิดศักดิ์ ไอรณณิ. (2552). การวิเคราะห์ข้อสอบปรนัย. *เวชบันทึกศิริราช*, 2(1), 31-37.
- ปรัชญนันท์ นิลสุข. (ม.ป.ป.). *การวิเคราะห์ข้อสอบ*. สืบค้น 10 กรกฎาคม 2555 จาก www.rtc.ac.th/download/item_analysis.ppt.
- ปวรส บุตะเชียว. (2555). *การวิเคราะห์ข้อสอบ (Item Analysis)*. สืบค้น 10 กรกฎาคม 2555 จาก www.rta.mi.th/630a0u/file/item_analysis.doc.
- นภา หลิมรัตน์. (2551). *การวัดผลและการประเมินผล*. สืบค้น 10 กรกฎาคม 2555 จาก <http://teachingresources.psu.ac.th/document/2551/limrat/9Item%20Analysis.pdf>.
- บุญชม ศรีสะอาด. (ม.ป.ป.). *การวิเคราะห์อำนาจจำแนก*. สืบค้น 10 กรกฎาคม 2555 จาก www.watpon.com/boonchom/04.doc.
- พวงรัตน์ ทวีรัตน์. (2540). *วิธีการวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์*. พิมพ์ครั้งที่ 7. กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.
- มนต์ชัย เทียนทอง. (2547). *การออกแบบและพัฒนาคอร์สแวร์สำหรับบทเรียนคอมพิวเตอร์ช่วยสอน*. กรุงเทพฯ: งานเอกสารและการพิมพ์ กองบริการการศึกษา สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- เยาวดี วิบูลย์ศรี. (2551). *การวัดผลและการสร้างแบบสอบผลสัมฤทธิ์*. พิมพ์ครั้งที่ 7. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- รังสรรค์ มณีเล็กและคณะ. (2546). *การพัฒนาเครื่องมือสำหรับการประเมินการศึกษา*. นนทบุรี: สำนักพิมพ์มหาวิทยาลัยสุโขทัยธรรมมาธิราช.

ล้วน สายยศและอังคณา สายยศ. (2543). *เทคนิคการวัดผลการเรียนรู้*. พิมพ์ครั้งที่ 2. กรุงเทพฯ: สุวีริยาสาส์น.

รังสรรค์ มณีเล็กและคณะ. (2546). *การพัฒนาเครื่องมือสำหรับการประเมินการศึกษา*. นนทบุรี: สำนักพิมพ์มหาวิทยาลัยสุโขทัยธรรมาธิราช.

สุมาลี จันทร์ชะลอ. (2542). *การวัดและประเมินผล*. กรุงเทพฯ: ศูนย์สื่อเสริมกรุงเทพ.

อุทุมพร จามรมาน. (2535). *ข้อสอบ: การสร้างและการพัฒนา*. กรุงเทพฯ: ฟีนนี่พับบลิชชิ่ง.

องอาจ นัยพัฒน์. (2556). *การวัดและประเมินผลการเรียนรู้ตามกรอบมาตรฐาน TQF*. สืบค้น

10 กรกฎาคม 2555 จาก

<http://www.thailandpod.net/documents/Ajarn08/02%20Associate%20Professor%20Dr.Ong-art%20Naiyapatana.pdf>.

บรรณานุกรมภาษาต่างประเทศ

Adeyemo Emily Oluseyi และ Ajeigbe Taiwo Olufemi. (2012). The Analysis of Multiple choices Item of the Test of an introductory course in Chemistry in a Nigerian University. *The International Journal of Learning*, 18#4:2012.

Ascalon, M., Evelina, Meyers, S., Lawrence, Davis, W., Bruce, & Smits, Niels. (2007). Distractor Similarity and Item-Stem Structure: Effects on Item Difficulty. *Applied Measurement in Education*, Volume 20, Issue 2.

Backhoff, E., Larrazolo, N., & Rosas, M. (2000). The Level of Difficulty and Discrimination Power of the Basic Knowledge and Skills Examination (EXHCOBA). *Revista Electrónica de Investigación Educativa*, 2 (1). Retrieved September 12, 2014 from: <http://redie.uabc.mx/vol2no1/contents-backhoff.html>.

- Bharti N Karelia, Ajita Pillai, Bhavisha N Vegada. (2013). The levels of difficulty and discrimination indices and relationship between them in four-response type multiple choice questions of pharmacology summative tests of Year II M.B.B.S students. *IeJSME* 2013 7(2): 41-46.
- Brown, W. and Thomson, H.G. (1921). *The essentials of mental measurement*. 2nd Edition. Cambridge, England: University Press.
- Chase, C. I. (1978). *Measurement for Educational Evaluation*. 2nd Edition. Reading; MA: Addison-Wesley Publishing.
- Hamzah, M.S.G. and Abdullah, S.K. (2011). Test Item Analysis: An Education Professionalism Approach. *US-China Education Review A* 3(2011), 307-322.
- Mitra, N. K., Nagaraja, H. S., Ponnudurai, G., Judson, J. P. (2009). The Levels Of Difficulty And Discrimination Indices In Type A Multiple Choice Questions Of Pre-clinical Semester 1 Multidisciplinary Summative Tests. *IeJSME* 2009: 3 (1): 2-7.
- Suruchi and Surinder S. Rana. (2014). Test Item Analysis and Relationship Between Difficulty Level and Discrimination Index of Test Items in an Achievement Test in Biology. *PARIPEX – INDIAN JOURNAL OF RESEARCH* Volume: 3 | Issue : 6 | June 2014.
- Susan Matlock-Hetzel. (1997). *Basic Concepts in Item and Test Analysis*. Texas A&M University, Retrieved September 12, 2014 from:
<http://ericae.net/ft/tamu/Espy.htm>.
- Sushma S. Pande, Santosh R. Pande, Vrushali R. Parate, Archana P. Nikam, Sushil H. Agrekar. (2013). Correlation between difficulty & discrimination indices of MCQs in formative exam in Physiology. *South -East Asian Journal of Medical Education*. Vol. 7 no. 1, 2013.
- Tavakol M, Dennick R. (2011). Post-examination analysis of objective tests. *Medical teacher*, 33: 447-458.

Undergraduate Academic Affairs. (2005). *Understanding Item Analysis Reports*

Understanding Item Analysis Reports. Retrieved September 12, 2014 from:

http://www.washington.edu/oea/services/scanning_scoring/scoring/item_analysis.html.

ภาคผนวก

ข้อมูลคะแนนการสอบกลางภาคและปลายภาค

วิชา ระบบสุขภาพและการพยาบาล ชั้นปีที่ 1

ลำดับที่	คะแนนการสอบกลางภาค			คะแนนการสอบปลายภาค		
	ก่อนตัดข้อ r ติดลบ ออก	หลังตัดข้อ r ติดลบออก	ค่าความแตกต่าง	ก่อนตัดข้อ r ติดลบ ออก	หลังตัดข้อ r ติดลบออก	ค่าความแตกต่าง
1.	40	41.36	1.36	28	27.37	-0.63
2.	29	29.70	0.70	24	23.16	-0.84
3.	46	47.73	1.73	27	27.37	0.37
4.	47	48.79	1.79	29	29.47	0.47
5.	37	37.12	0.12	26	25.26	-0.74
6.	36	37.12	1.12	26	26.32	0.32
7.	45	45.61	0.61	28	27.37	-0.63
8.	43	44.55	1.55	25	24.21	-0.79
9.	44	45.61	1.61	26	25.26	-0.74
10.	34	35.00	1.00	30	29.47	-0.53
11.	37	37.12	0.12	35	34.74	-0.26
12.	35	35.00	0.00	28	27.37	-0.63
13.	44	45.61	1.61	28	28.42	0.42
14.	38	38.18	0.18	26	26.32	0.32
15.	33	33.94	0.94	29	28.42	-0.58
16.	45	46.67	1.67	29	28.42	-0.58
17.	48	49.85	1.85	31	30.53	-0.47
18.	39	41.36	2.36	25	24.21	-0.79
19.	35	36.06	1.06	29	29.47	0.47
20.	31	31.82	0.82	28	27.37	-0.63
21.	33	32.88	-0.12	27	26.32	-0.68
22.	34	35.00	1.00	28	28.42	0.42
23.	32	32.88	0.88	25	24.21	-0.79
24.	31	32.88	1.88	27	26.32	-0.68

ลำดับที่	คะแนนการสอบกลางภาค			คะแนนการสอบปลายภาค		
	ก่อนตัดข้อ r ติดลบ ออก	หลังตัดข้อ r ติดลบออก	ค่าความ แตกต่าง	ก่อนตัดข้อ r ติดลบ ออก	หลังตัดข้อ r ติดลบออก	ค่าความ แตกต่าง
25.	37	38.18	1.18	26	25.26	-0.74
26.	41	40.30	-0.70	27	26.32	-0.68
27.	38	36.06	-1.94	31	31.58	0.58
28.	44	45.61	1.61	30	29.47	-0.53
29.	40	41.36	1.36	27	26.32	-0.68
30.	49	50.91	1.91	22	21.05	-0.95
31.	37	38.18	1.18	30	29.47	-0.53
32.	36	36.06	0.06	30	29.47	-0.53
33.	48	50.91	2.91	32	32.63	0.63
34.	34	33.94	-0.06	29	28.42	-0.58
35.	42	42.42	0.42	35	34.74	-0.26
36.	43	43.48	0.48	27	26.32	-0.68
37.	40	41.36	1.36	27	27.37	0.37
38.	40	38.18	-1.82	33	32.63	-0.37
39.	44	45.61	1.61	36	35.79	-0.21
40.	48	49.85	1.85	29	28.42	-0.58
41.	35	41.36	6.36	25	25.26	0.26
42.	46	47.73	1.73	26	25.26	-0.74
43.	45	45.61	0.61	33	32.63	-0.37
44.	43	44.55	1.55	32	31.58	-0.42
45.	41	42.42	1.42	28	27.37	-0.63
46.	34	35.00	1.00	28	27.37	-0.63
47.	43	44.55	1.55	29	28.42	-0.58
48.	34	33.94	-0.06	25	24.21	-0.79
49.	49	50.91	1.91	26	26.32	0.32
50.	38	39.24	1.24	28	27.37	-0.63

ลำดับที่	คะแนนการสอบกลางภาค			คะแนนการสอบปลายภาค		
	ก่อนตัดข้อ r ติดลบ ออก	หลังตัดข้อ r ติดลบออก	ค่าความ แตกต่าง	ก่อนตัดข้อ r ติดลบ ออก	หลังตัดข้อ r ติดลบออก	ค่าความ แตกต่าง
51.	41	41.36	0.36	28	27.37	-0.63
52.	43	44.55	1.55	30	29.47	-0.53
53.	42	43.48	1.48	32	31.58	-0.42
54.	49	50.91	1.91	29	29.47	0.47
55.	46	46.67	0.67	32	31.58	-0.42
56.	39	41.36	2.36	32	31.58	-0.42
57.	47	48.79	1.79	33	32.63	-0.37
58.	45	45.61	0.61	21	21.05	0.05
59.	37	38.18	1.18	29	29.47	0.47
60.	44	44.55	0.55	33	32.63	-0.37
61.	43	44.55	1.55	30	29.47	-0.53
62.	42	44.55	2.55	27	26.32	-0.68
63.	51	53.03	2.03	30	30.53	0.53
64.	42	42.42	0.42	29	28.42	-0.58
65.	44	44.55	0.55	30	29.47	-0.53
66.	39	39.24	0.24	25	24.21	-0.79
67.	45	45.61	0.61	33	32.63	-0.37
68.	40	40.30	0.30	30	29.47	-0.53
69.	46	46.67	0.67	27	27.37	0.37
70.	38	39.24	1.24	25	24.21	-0.79
71.	41	42.42	1.42	28	28.42	0.42
72.	42	43.48	1.48	33	33.68	0.68
73.	45	47.73	2.73	31	31.58	0.58
74.	50	51.97	1.97	30	29.47	-0.53
75.	38	39.24	1.24	24	23.16	-0.84
76.	41	40.30	-0.70	31	31.58	0.58

ลำดับที่	คะแนนการสอบกลางภาค			คะแนนการสอบปลายภาค		
	ก่อนตัดข้อ r ติดลบ ออก	หลังตัดข้อ r ติดลบออก	ค่าความ แตกต่าง	ก่อนตัดข้อ r ติดลบ ออก	หลังตัดข้อ r ติดลบออก	ค่าความ แตกต่าง
77.	39	41.36	2.36	23	22.11	-0.89
78.	49	51.97	2.97	28	27.37	-0.63
79.	51	53.03	2.03	31	31.58	0.58
80.	44	43.48	-0.52	29	29.47	0.47
81.	46	47.73	1.73	28	28.42	0.42
82.	44	45.61	1.61	29	28.42	-0.58
83.	40	40.30	0.30	32	31.58	-0.42
84.	42	43.48	1.48	29	28.42	-0.58
85.	45	45.61	0.61	27	27.37	0.37
86.	39	39.24	0.24	30	30.53	0.53
87.	42	43.48	1.48	29	28.42	-0.58
88.	38	38.18	0.18	31	31.58	0.58
89.	40	40.30	0.30	32	32.63	0.63
90.	42	44.55	2.55	30	29.47	-0.53
91.	41	42.42	1.42	26	26.32	0.32
92.	41	41.36	0.36	29	28.42	-0.58

ประวัติผู้วิจัย

นางศรีสุดา วงศ์วิเศษกุล

เกิดวันที่ 28 มกราคม 2500

ประวัติการศึกษา

- | | |
|------------|---|
| พ.ศ. 2520 | อนุปริญญาพยาบาลอนามัยและประกาศนียบัตรผดุงครรภ์
คณะพยาบาลศาสตร์ มหาวิทยาลัยมหิดล |
| พ.ศ. 2525 | ศิลปศาสตรบัณฑิต (ภาษาอังกฤษ)
คณะมนุษยศาสตร์ มหาวิทยาลัยรามคำแหง |
| พ.ศ. 2530 | พยาบาลศาสตรบัณฑิต (หลักสูตรต่อเนื่อง)
มหาวิทยาลัยสุโขทัยธรรมมาธิราช |
| พ.ศ. 2540 | สาธารณสุขศาสตรมหาบัณฑิต (การพยาบาลสาธารณสุข)
คณะสาธารณสุขศาสตร์ มหาวิทยาลัยมหิดล |
| พ.ศ. 2554 | ครุศาสตรดุษฎีบัณฑิต คณะครุศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย |
| ตำแหน่งงาน | หัวหน้าสาขาการพยาบาลชุมชน
คณะพยาบาลศาสตร์ มหาวิทยาลัยราชภัฏสวนดุสิต |