



รายงานการวิจัย
เรื่อง

ผลของการตัดข้อสอบที่มีอำนาจจำแนกติดลบออกต่อคะแนนการสอบ และ
ความเชื่อมั่นของข้อสอบทั้งฉบับ: รายวิชากฎหมาย และจริยธรรมวิชาชีพ
The effects on the Test scores and Reliability after excluding
items with Negative Discrimination power: The Laws and
Professional Ethics Subject

นางนิภา ลีสุคนธ์
นางสาวจิตราพร จันทรวงศ์

มหาวิทยาลัยราชภัฏสวนดุสิต
2558
ลิขสิทธิ์ของมหาวิทยาลัยราชภัฏสวนดุสิต



รายงานการวิจัย
เรื่อง

ผลของการตัดข้อสอบที่มีอำนาจจำแนกติดลบออกต่อคะแนนการสอบ และ
ความเชื่อมั่นของข้อสอบทั้งฉบับ: รายวิชากฎหมาย และจริยธรรมวิชาชีพ
The effects on the Test scores and Reliability after excluding
items with Negative Discrimination power: The Laws and
Professional Ethics Subject

นางนิภา ลีสุคนธ์
(คณะพยาบาลศาสตร์)
นางสาวจิตราพร จันทรวงศ์
(สำนักส่งเสริมวิชาการและงานทะเบียน)

มหาวิทยาลัยราชภัฏสวนดุสิต
2558

ลิขสิทธิ์ของมหาวิทยาลัยราชภัฏสวนดุสิต
(งานวิจัยนี้ได้รับทุนอุดหนุนจากมหาวิทยาลัยราชภัฏสวนดุสิต ปีงบประมาณ 2556)

หัวข้อวิจัย	ผลของการตัดข้อสอบที่มีอำนาจจำแนกติดลบออกต่อคะแนนการสอบและความเชื่อมั่นของข้อสอบทั้งฉบับ: รายวิชากฎหมาย และจริยธรรมวิชาชีพ
ผู้ดำเนินการวิจัย	นางนิภา ลีสุคนธ์ นางสาวจิตราพร จันทรวงศ์
ที่ปรึกษา	ผศ.ดร.ประกาย จิโรจน์กุล
หน่วยงาน	คณะพยาบาลศาสตร์ มหาวิทยาลัยราชภัฏสวนดุสิต
ปี พ.ศ.	2558

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาคุณภาพของข้อสอบ การเปลี่ยนแปลงของคะแนนเฉลี่ย และอันดับของการสอบของนักศึกษารายบุคคล ก่อน และหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก การเปลี่ยนแปลงของค่าความเชื่อมั่นของข้อสอบ ก่อน และหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก ในรายวิชากฎหมาย และจริยธรรมวิชาชีพ โดยประชากร และกลุ่มตัวอย่าง ได้แก่ ข้อสอบ และคะแนนผลการสอบรายวิชารายวิชากฎหมาย และจริยธรรมวิชาชีพ ของนักศึกษาพยาบาลศาสตร์ ชั้นปีที่ 4 จำนวน 60 คน ที่ลงทะเบียนเรียน รายวิชากฎหมาย และจริยธรรมวิชาชีพ ในภาคการศึกษาที่ 1 ปีการศึกษา 2555 โดยใช้โปรแกรมการวิเคราะห์ข้อสอบรายข้อ (Items Analysis) สถิติที่ใช้ในงานวิจัยประกอบด้วย Paired t-test (one-tailed) และ Sign Ranked test ผลการวิเคราะห์ข้อมูลจากชุดข้อสอบแบบเลือกตอบ (Multiple choice) 4 ตัวเลือก จำนวน 2 ฉบับ เป็นชุดข้อสอบกลางภาค จำนวน 40 ข้อและปลายภาค 40 ข้อ พบว่า

1) ด้านคุณภาพข้อสอบ เมื่อพิจารณาจากดัชนีความยากง่าย (P) พบว่า ชุดข้อสอบกลางภาค และชุดข้อสอบปลายภาค มีข้อสอบที่มีค่าความยากง่ายในเกณฑ์ที่เหมาะสม (P อยู่ระหว่าง 0.20 – 0.80) จำนวนรวม 30, 25 ข้อ คิดเป็นร้อยละ 75, 65 ตามลำดับ ด้านค่าอำนาจจำแนก (r) พบว่า ชุดข้อสอบกลางภาค และชุดข้อสอบปลายภาค มีข้อสอบที่มีค่าอำนาจจำแนกในเกณฑ์ที่เหมาะสม ($r \geq 0.20$) จำนวนรวม 23, 19 ข้อ คิดเป็นร้อยละ 57.5 และร้อยละ 47.5 ตามลำดับ ทั้งนี้ ข้อสอบที่มีค่าอำนาจจำแนกติดลบ ($r = -1.00 - 0.00$) ในชุดข้อสอบกลางภาค และชุดข้อสอบปลายภาค มีจำนวนชุดละ 1 ข้อ คิดเป็นร้อยละ 2.5 ส่วนคุณภาพของข้อสอบทั้งฉบับ พบว่า ข้อสอบกลางภาคมีค่าความเชื่อมั่น 0.48 และเมื่อตัดข้อสอบที่มีค่า r ติดลบออก ทำให้ค่าความเชื่อมั่นเพิ่มขึ้นเป็น 0.51 ส่วนข้อสอบปลายภาค มีค่าความเชื่อมั่น = 0.40 และเมื่อตัดข้อสอบที่มีค่า r ติดลบออก ทำให้ค่าความเชื่อมั่นเพิ่มขึ้นเป็น 0.43

2) ผลการเปรียบเทียบคะแนนเฉลี่ยที่นักศึกษาได้รับ ในรายวิชากฎหมาย และจริยธรรมวิชาชีพ จากข้อสอบทั้งชุดกลางภาค และชุดปลายภาค ก่อน และหลังการตัดข้อสอบที่มีค่า r ติดลบ

ออก โดยสถิติ Paired t-test (one-tailed) พบว่า นักศึกษามีคะแนนเฉลี่ยสูงขึ้นอย่างมีนัยสำคัญทางสถิติ ที่ $\alpha = 0.05$ ($t=0.00$, $df = 59$, $P \text{ value} < 0.005$)

3) ผลการทดสอบการเปลี่ยนแปลงของลำดับคะแนนที่นักศึกษาได้รับ ในรายวิชากฎหมาย และจริยธรรมวิชาชีพ ก่อน และหลังการตัดข้อสอบที่มีค่า r ติดลบออก โดยสถิติ Signed Rank Test พบว่า ลำดับคะแนนที่นักศึกษาได้รับมีการเปลี่ยนแปลงลำดับที่อย่างมีนัยสำคัญทางสถิติ ที่ $\alpha = 0.05$ ($P \text{ value} < 0.01$) ในข้อสอบทั้ง 2 ชุด

คำสำคัญ: คุณภาพข้อสอบ ค่าอำนาจจำแนกติดลบ คะแนนสอบ ค่าความเชื่อมั่นของข้อสอบทั้งฉบับ

Research Title	The Effects on the Test Scores and Reliability after excluding items with a Negative Discrimination Power: The Laws and Professional Ethics Subject
Researcher	Mrs.Nipa Leesukol Miss. Jittraporn Jandharakhul
Research Consultants	Ass.Prof. Dr. Pragai Jirojjanakul
Organization	Faculty of Nursing, Suan Dusit Rajabhat University
Year	2015

The objectives of this research were 1) to describe the quality of the test in terms of each item's index of difficulty (P), index of discrimination (r) and its reliability. 2) to compare the test scores and students rank before and after excluding the items with negative discrimination power. A 40 item multiple choice midterm and final test were used in this study. The participants were 60 fourth year nursing students who were registered in the Laws and Professional Ethics Subject during the first semester of the academic year 2012. The item analysis was used to analyze the test items and scores. The test items were analyzed for the index of difficulty, index of discrimination and reliability by a computer program. The statistics used in this research included the Paired t-test (one – tailed) and the Sign Ranked test.

The results found that:

1) Quality of the test: there were 30 midterm and 25 final test items with a proper difficulty index ($P = 0.20 - 0.80$). This accounted for 75% and 65 % respectively. Twenty-three midterm test items (57.5%) and 19 final test items (47.5%) had an index of discrimination (r) with a proper range of ($r \geq 0.2$), while 1 item (2.5%) of each test had a negative discrimination power. ($r = -1.00 - 0.00$) The reliability of the midterm test and the final test were 0.48 and 0.40 respectively and it was increased to 0.51 and 0.43 after excluding items with negative discrimination power.

2) Comparing the mean scores before and after excluding the item with negative discrimination power, using a paired t-test (one-tailed) statistic, found that students had obtained significant higher scores at $\alpha = 0.05$ ($t=0.00$, $df=59$, p value < 0.005 in both tests.

3)The Signed Rank Test showed that the student's ranks were significantly changed in both tests with $\alpha = 0.05$, p value < 0.01 after excluding the item with negative discrimination power.

Key words: quality of test, negative discrimination power, test score, reliability of the test.

กิตติกรรมประกาศ

คณะผู้วิจัยขอขอบคุณ ผศ.ดร.ประกาย จิโรจน์กุล รองคณบดีฝ่ายวิชาการ คณะพยาบาล
ศาสตร์ มหาวิทยาลัยราชภัฏสวนดุสิต ที่กรุณาให้คำปรึกษาในการดำเนินโครงการวิจัยนี้เป็นอย่างดี
และขอบคุณอาจารย์วิมล เตากล่ำ อดีตผู้อำนวยการสำนักงานเขตพื้นที่การศึกษา กรุงเทพมหานคร
เขต 2 ซึ่งเป็นผู้จุดประกายให้ผู้วิจัยสนใจศึกษาเรื่องนี้ ทั้งยังได้กรุณาให้คำแนะนำที่เป็นประโยชน์
ตลอดระยะเวลาที่ศึกษา

ผู้วิจัยขอขอบคุณ มหาวิทยาลัยราชภัฏสวนดุสิต ที่ให้ทุนอุดหนุนการวิจัยครั้งนี้

คณะผู้วิจัย

2558

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ค
กิตติกรรมประกาศ	จ
สารบัญ	ฉ
สารบัญตาราง	ฅ
สารบัญแผนภูมิ	ฎ
บทที่ 1 บทนำ	1
ความเป็นมาและความสำคัญ	1
วัตถุประสงค์ของการวิจัย	4
ขอบเขตการวิจัย	4
คำจำกัดความที่ใช้ในงานวิจัย	5
ประโยชน์ที่คาดว่าจะได้รับ	5
บทที่ 2 แนวคิด ทฤษฎี เอกสารและงานวิจัยที่เกี่ยวข้อง	6
มโนทัศน์เกี่ยวกับการศึกษา (Educational Concept)	6
- ความหมายของการศึกษาและการเรียนรู้	
- กรอบแนวคิดในการเรียนรู้	
มโนทัศน์เกี่ยวกับการประเมินการเรียนรู้	11
- ความหมายของศัพท์เกี่ยวกับการประเมิน	
- ลักษณะการประเมินทางการศึกษา	
- ความแตกต่างระหว่างการประเมิน และการวัดผล	
- การวัด และการประเมินผล	
การประเมินผลการเรียนรู้ในนักศึกษาพยาบาล กับความสำคัญของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน	14
แนวคิดเกี่ยวกับแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน	15
แบบทดสอบชนิดเลือกตอบ	21
- คุณลักษณะ จุดเด่น จุดด้อย ของแบบทดสอบชนิดเลือกตอบ	
- จุดอ่อนของแบบทดสอบชนิดเลือกตอบ	
- โครงสร้างของข้อสอบแบบเลือกตอบ	
- ตัวอย่างข้อสอบแบบเลือกตอบ	
- แนวทางการสร้างโจทย์และคำถามทางการพยาบาลแบบเลือกตอบ	
- แนวทางการสร้างตัวเลือก	

การวิเคราะห์คุณภาพแบบทดสอบชนิดเลือกตอบ	26
- ประโยชน์ของการวิเคราะห์ข้อสอบภายหลังการสอบ	
- ประเภทของการวิเคราะห์คุณภาพแบบทดสอบ	
- การวิเคราะห์ข้อสอบชนิดเลือกตอบแบบอิงกลุ่ม	
- การวิเคราะห์ข้อสอบชนิดเลือกตอบแบบอิงเกณฑ์	
- การวิเคราะห์หาระดับความยากง่าย	
- การวิเคราะห์หาอำนาจจำแนก	
- ความจำกัดของการวิเคราะห์ข้อสอบด้วย ค่า P และค่า r	
- ประสิทธิภาพของตัวลวง	
- การเลือกข้อสอบจากผลการวิเคราะห์ข้อสอบ	
- ความเชื่อมั่นของแบบทดสอบ	
- องค์ประกอบที่มีผลทำให้ค่าความเชื่อมั่นสูงหรือต่ำ	
งานวิจัยที่เกี่ยวข้องกับคุณภาพของข้อสอบและการสร้างข้อสอบที่มีคุณภาพ	34
สรุป	39
กรอบแนวคิดในการวิจัย	39
บทที่ 3 วิธีดำเนินการวิจัย	41
ประชากรและการเลือกกลุ่มตัวอย่าง	41
เครื่องมือในการวิจัยและการตรวจสอบคุณภาพเครื่องมือ	41
การเก็บรวบรวมข้อมูล	42
การวิเคราะห์ข้อมูล	42
บทที่ 4 ผลการวิจัย	43
ข้อมูลทั่วไป	43
คุณภาพข้อสอบรายวิชากฎหมาย และจริยธรรมวิชาชีพ	43
การเปลี่ยนแปลงของคะแนนสอบรายวิชาและอันดับ ก่อน และหลัง	46
การตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก	
การเปลี่ยนแปลงค่าความเชื่อมั่นของข้อสอบทั้งฉบับ ก่อน และหลัง	48
การตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก	

บทที่ 5	สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ	49
	สรุปผลการวิจัย	49
	อภิปรายผล	50
	ข้อเสนอแนะในการนำผลการวิจัยไปใช้	55
	ข้อเสนอแนะในการทำวิจัยครั้งต่อไป	56
บรรณานุกรม		57
	บรรณานุกรมภาษาไทย	57
	บรรณานุกรมภาษาไทยทางสื่ออิเล็กทรอนิกส์	58
	บรรณานุกรมภาษาต่างประเทศ	59
ภาคผนวก		61
	ภาคผนวก ก ตารางกำหนดข้อสอบ (Test Blueprint) รายวิชากฎหมาย และจริยธรรมวิชาชีพ	62
	ภาคผนวก ข สรุปผลการวิเคราะห์ข้อสอบชุดกลางภาค และชุดปลายภาค รายวิชากฎหมายและจริยธรรมวิชาชีพ	66
ประวัติผู้วิจัย		75

สารบัญตาราง

ตารางที่		หน้า
2.1	ความแตกต่างระหว่างระหว่าง การประเมิน (Assessment) กับการประเมินผล(Evaluation)	13
2.2	ความแตกต่างระหว่างโครงสร้างระดับการคิด (Cognitive Domain) ของบลูม (Bloom) เดิม และโครงสร้างปรับปรุงใหม่	16
2.3	ข้อเสนอแนะในการเลือกข้อสอบ / แบบทดสอบจากความยากง่าย	28
2.4	เกณฑ์การแปลความหมายค่าอำนาจจำแนก (r) ของข้อสอบ	29
2.5	แสดงผลการวิเคราะห์ข้อสอบจำนวน 1 ข้อ ที่ใช้ทดสอบกับนักเรียนประถมศึกษาชั้นปีที่ 6 จำนวน 5 โรงเรียน	30
2.6	แสดงผลการวิเคราะห์จำนวนผู้ตอบ กลุ่มสูงและกลุ่มต่ำที่เลือกตอบตัวเลือกและตัวลวง	31
2.7	เกณฑ์การแปลความหมายของค่าสัมประสิทธิ์ความเชื่อมั่น	33
4.1	ชุดข้อสอบ จำนวนนักศึกษาที่เข้าสอบ จำนวนชุดข้อสอบ จำนวนข้อคำถามในแต่ละชุด ที่ใช้ในการศึกษาวิจัย	43
4.2	จำนวน และร้อยละ ของข้อสอบชุดกลางภาค แต่ละข้อ จำแนกตามค่าความยากง่าย (P)	44
4.3	จำนวน และร้อยละ ของข้อสอบชุดปลายภาค แต่ละข้อ จำแนกตามค่าความยากง่าย (P)	44
4.4	จำนวน และร้อยละ ของข้อสอบชุดกลางภาค จำแนกตามค่าอำนาจจำแนก (r) ของข้อสอบแต่ละข้อ	45
4.5	จำนวน และร้อยละ ของข้อสอบชุดปลายภาค จำแนกตามค่าอำนาจจำแนก (r) ของข้อสอบแต่ละข้อ	45
4.6	ภาพรวมของคุณภาพข้อสอบ แสดงค่าความยากง่าย และค่าอำนาจจำแนกเฉลี่ยที่เหมาะสม รวมทั้งจำนวนข้อสอบที่มีค่าอำนาจจำแนก (r) ตีติล	46
4.7	ค่าสัมประสิทธิ์ความเชื่อมั่นของข้อสอบทั้งฉบับ ในข้อสอบชุดกลางภาค และชุดปลายภาค ก่อนตัดค่า r ตีติล	46
4.8	ผลการเปรียบเทียบคะแนนเฉลี่ย คิดเป็นร้อยละ ที่นักศึกษาได้รับในรายวิชา กฎหมาย และจริยธรรมวิชาชีพ ก่อน และหลังการตัดข้อสอบที่มีค่า r- ออก ทั้งจากชุดข้อสอบกลางภาค และชุดข้อสอบปลายภาค โดยใช้สถิติ Pair t – test (one-tailed) (n=60)	

ตารางที่		หน้า
4.9	ผลการทดสอบการเปลี่ยนแปลงของลำดับคะแนนที่นักศึกษาได้รับในรายวิชา กฎหมาย และจริยธรรมวิชาชีพ ก่อน และหลังการตัดข้อสอบที่มีค่า r - ออก โดยสถิติ Signed Rank Test (n=60)	48
4.10	การเปลี่ยนแปลงค่าความเชื่อมั่นของข้อสอบ ชุดกลางภาค และชุดปลายภาค ในรายวิชากฎหมายและจริยธรรมวิชาชีพ ภายหลังจากที่มีการตัดข้อสอบ ที่มีค่า r ต่ำลบออก	48

สารบัญแผนภูมิ

แผนภูมิที่		หน้า
2.1	กรอบแนวคิดในการเรียนรู้ของมนุษย์	9
2.2	ความเชื่อมโยงต่อเนื่องของความหมายการวัด และการประเมินผล	12
2.3	วงจรความสัมพันธ์ระหว่างวัตถุประสงค์ของการเรียนการสอน การจัด ประสบการณ์การเรียนรู้ และการวัดประเมินผล	14
2.4	ตัวอย่างของโครงสร้างข้อสอบแบบเลือกตอบ (multiple choice question)	23
2.5	ช่วงค่าความยากง่าย และค่าอำนาจจำแนกของข้อสอบที่ได้เกณฑ์มาตรฐาน และไม่ได้เกณฑ์มาตรฐาน นำไปใช้พลอตกราฟผลการวิเคราะห์ข้อสอบแต่ละ ข้อในชุดเดียวกัน	32
2.6	กรอบแนวคิดในการวิจัย	40

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญ

สำนักงานเลขาธิการสภาการศึกษาได้ดำเนินการศึกษาเปรียบเทียบสมรรถนะของประเทศไทยกับนานาชาติ โดยใช้ดัชนีของสถาบันเพื่อพัฒนาการจัดการ (International Institute for Management Development, IMD) เป็นกรอบหลักในการวิเคราะห์เป็นประจำทุกปีตั้งแต่ พ.ศ. 2540 โดยมีจุดมุ่งหมายสำคัญเพื่อเป็นข้อมูลพื้นฐานในการกำหนดนโยบาย และวางแผน ตลอดจนการปฏิรูปการศึกษา เพื่อพัฒนาการศึกษาไทยให้มีคุณภาพและได้ มาตรฐานระดับสากล จากผลการประเมินการจัดอันดับโดย IMD พบว่า ในปี พ.ศ.2554 การศึกษาไทยอยู่ในอันดับที่ 51 จาก 60 ประเทศ ขณะที่ผลการประเมินการทดสอบ PISA (ย่อมาจาก Programme for International Student Assessment) ซึ่งจัดโดยองค์การเพื่อความร่วมมือทางเศรษฐกิจและการพัฒนาของประเทศ กลุ่มยุโรป (Organization for Economic Cooperation and Development, OECD) มีวัตถุประสงค์ เพื่อประเมินว่าระบบการศึกษาของแต่ละประเทศสามารถเตรียมความพร้อมด้านการศึกษาให้กับเยาวชนในประเทศมากน้อยแค่ไหน โดยวัดจากทักษะสามด้าน คือ คณิตศาสตร์ วิทยาศาสตร์ และการอ่าน โปรแกรม PISA นี้ เป็นการประเมินสมรรถนะของนักเรียนอายุ 15 ปี ในการแก้ปัญหาและการประยุกต์ความรู้ซึ่งไม่ใช่การทบทวนความรู้ที่เรียนไปแล้ว ในปี พ.ศ. 2552 พบว่า เด็กไทยอยู่ในอันดับที่ 50 จาก 65 ประเทศ ส่วนผลการจัดอันดับ 400 มหาวิทยาลัยที่ดีที่สุดในโลก โดย Times Higher Education World Rankings ปี 2555-2556 มีมหาวิทยาลัยไทยเพียงแห่งเดียวที่ติดอยู่ในกลุ่ม 351- 400 จากผลการสำรวจเปรียบเทียบในระดับนานาชาติ โดยองค์การระหว่างประเทศ สำนักงานเลขาธิการสภาการศึกษาให้ภาพสรุปว่า สมรรถนะการศึกษาไทยในเวทีสากลยังอยู่ในระดับไม่เป็นที่น่าพอใจและยังล้าหลังกว่าหลายประเทศ ทั้งด้านโอกาส คุณภาพ และประสิทธิภาพการจัดการศึกษา โดยมีแนวโน้มการพัฒนาที่ตกต่ำลงเมื่อเปรียบเทียบกับนานาชาติ (สำนักงานเลขาธิการสภาการศึกษา, 2556) ทำให้ในปี พ.ศ.2556 นี้รัฐมนตรีว่าการกระทรวงศึกษาธิการ (นายจาตุรนต์ ฉายแสง) ประกาศนโยบายภายหลังที่เข้ารับตำแหน่งในปี พ.ศ. 2556 เป็นวาระแห่งชาติ ในประเด็น “ปีแห่งการรวมพลังยกระดับคุณภาพการศึกษา” เพื่อเร่งยกเครื่องการศึกษาไทยให้มีคุณภาพ มาตรฐานในระดับสากล และสอดคล้องกับสังคมโลกยุคใหม่(สำนักงานเลขาธิการสภาการศึกษา, 2556)

การจัดการศึกษาต้องการมาตรฐานคุณภาพตั้งแต่ระดับพื้นฐานและต่อยอดให้เพิ่มพูนองกวมไปในระดับที่สูงขึ้น เปรียบเสมือนการสร้างอาคารสูงที่ต้องการโครงสร้างที่แข็งแรงตั้งแต่ฐานราก และก่อฐานสร้างไปที่ละชั้นอย่างมั่นคงแข็งแรง หากรากฐานไม่มั่นคงแข็งแรงแล้วตัวอาคารอาจพังทลายลงมาได้เมื่อใช้งาน เช่นเดียวกับการศึกษา หากมุ่งหวังไปต่อยอดเอาคุณภาพเฉพาะส่วนอุดมศึกษาแล้ว กระบวนการเรียนการสอนจะเต็มไปด้วยอุปสรรคต่างๆ ต้องใช้ทรัพยากรและความทุ่มเทของผู้สอนที่สูงยิ่งเพราะการหยั่งรากความรู้ของผู้เรียนทำได้จำกัด จากปัญหาผู้เรียนที่ผ่านเข้ามาในสายวิทยาศาสตร์ หรือวิทยาศาสตร์ประยุกต์ มีทักษะการอ่าน และความรู้พื้นฐานทางด้านวิทยาศาสตร์ ต่ำ

กว่าเกณฑ์ที่ควรจะเป็นอาจได้ผลลัพธ์น้อยไม่คุ้มค่าการลงทุน แต่ถึงอย่างไรก็ตาม ถึงแม้ว่าพื้นฐานการศึกษาที่ผู้เรียนจะดำเนินมาอย่างไร เป็นหน้าที่ของผู้บริหารสถานศึกษา และอาจารย์ผู้สอนในระดับอุดมศึกษาของแต่ละวิชาชีพ ต้องอุดช่องว่างนั้นเพื่อช่วยให้ผู้เรียนได้พัฒนาต่อไปมิเช่นนั้นแล้วผู้สำเร็จการศึกษาเมื่อออกไปประกอบอาชีพหรือวิชาชีพแต่ละแขนง ก็จะทำให้การประกอบวิชาชีพฯ หลังสำเร็จการศึกษามีมาตรฐานตกต่ำลงไปด้วย

การพัฒนาคุณภาพการศึกษา องค์ประกอบสำคัญหนึ่ง ได้แก่ การพัฒนาคุณภาพการวัดและประเมินผลผู้เรียนซึ่งการวัดประเมินผลนี้เป็นองค์ประกอบที่สำคัญของการจัดการศึกษาที่มีคุณภาพและจุดมุ่งหมายของการวัดผลทางการศึกษาเพื่อให้ได้มาซึ่งข้อมูลที่เที่ยงตรง เชื่อถือได้ที่แสดงถึงความสามารถที่แท้จริง (true ability) ของผู้เรียน (อุทุมพร จามรมาน, 2535) เพื่อนำมาใช้ในการพัฒนาผู้เรียนให้มีความสามารถ คุณลักษณะ ทักษะ ตามที่กำหนดไว้ในหลักสูตรวิชาซึ่งเป็นประโยชน์โดยตรงต่อผู้เรียนเพื่อพัฒนาผลสัมฤทธิ์ทางการเรียน ขณะเดียวกันยังสะท้อนถึงประสิทธิภาพการการสอนของผู้สอน และเป็นข้อมูลสำคัญที่สะท้อนคุณภาพการดำเนินงานการจัดการศึกษาของสถานศึกษาด้วย (ทิวต์ถ์ มณีโชติ, 2549)

การวัดประเมินผลการเรียนรู้ด้วยแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนเป็นเครื่องมือที่ใช้วัดพฤติกรรมในด้านพุทธิพิสัยเป็นส่วนใหญ่ซึ่งได้แก่ พฤติกรรมด้านความรู้ ความจำ ความเข้าใจ การวิเคราะห์ การสังเคราะห์ และการประเมินค่าซึ่งแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนประเภทหนึ่งได้แก่ แบบทดสอบผลสัมฤทธิ์ที่ครูสร้างขึ้นมีจุดประสงค์เพื่อวัดผลผลิตของการศึกษาโดยตรง ได้แก่ วัดผลการเรียนรู้ด้านเนื้อหาวิชาและทักษะต่างๆของแต่ละสาขาวิชา แบบวัดเพื่อจุดประสงค์นี้มีอยู่ 2 ลักษณะ (เยาวดี รวงชัยกุล วิบูลย์ศรี, 2554, น. 16) ได้แก่

1. ลักษณะแบบสอบผลสัมฤทธิ์ที่เป็นข้อเขียน (Paper - Pencil Test)
2. ลักษณะแบบสอบผลสัมฤทธิ์ที่เป็นภาคปฏิบัติจริง (Performance Test)

แบบทดสอบผลสัมฤทธิ์ที่เป็นข้อเขียน ที่นิยมใช้ประเภทหนึ่ง ได้แก่ ข้อสอบแบบเลือกตอบ (multiple choice question หรือ MCQ) ซึ่งข้อสอบแบบเลือกตอบนี้ คุณสมบัติที่สำคัญของข้อสอบที่ดี ควรมีค่าความยากง่ายที่เหมาะสมและมีอำนาจจำแนกสูง ค่าเฉลี่ยความยากง่ายของข้อสอบทั้งฉบับควรมีค่าประมาณ 0.50 แต่อย่างไรก็ตามในการคัดเลือกข้อสอบจะต้องคำนึงถึงความสมดุลระหว่างข้อสอบที่มีสถิติเหมาะสม(อิงกลุ่ม) กับข้อสอบที่วัดครอบคลุมจุดประสงค์ และเนื้อหาที่ต้องการด้วย (อิงเกณฑ์) บางครั้งอาจมีความจำเป็นที่จะต้องใช้ข้อสอบที่มีอำนาจการจำแนกไม่สูงนักเพื่อให้มีข้อสอบที่วัดครอบคลุมเนื้อหาที่ต้องการ (ปวรส บุตะเขียว, 2555) ในการวิเคราะห์ข้อสอบ ถ้าพบว่า ข้อสอบไม่มีคุณภาพควรกลับมาพิจารณาว่าเกิดจากสาเหตุใด ข้อคำถามไม่มีความเป็นปรนัยหรือไม่ รวมทั้งต้องพิจารณาที่ตัวเลือกด้วยว่าเป็นไปตามหลักการเขียนข้อสอบที่ดีหรือไม่ และดำเนินการปรับปรุงข้อสอบ รวมทั้งนำไปทดสอบซ้ำ (try out) อีกครั้งเพื่อวิเคราะห์ประสิทธิภาพของข้อสอบต่อไป ในส่วนของตัวลวง ประสิทธิภาพของตัวลวง เมื่อสร้างข้อสอบอิงเกณฑ์แบบหลายตัวเลือก (Multiple choices) ผู้ที่เลือกตัวลวง ถือว่า ตอบผิด ซึ่งแสดงให้เห็นว่า ผู้นั้นยังไม่สัมฤทธิ์ผลตามเป้าหมายของการวัด ในการวิเคราะห์ประสิทธิภาพตัวลวงทำโดยการตรวจสอบความถี่ของผู้ที่เลือกตัวลวงนั้นๆ ถ้าตัวลวงใดมีผู้เลือกในสัดส่วนที่สูง ถือว่าเป็นตัวลวงที่ใช้ได้ แต่ถ้าตัวลวงใดมีผู้เลือกน้อย แสดงว่าตัวลวงนั้นขาดประสิทธิภาพ สมควรที่จะต้องมีการปรับปรุงแก้ไขให้ดีขึ้น (ศิริชัย กาญจน

วาลี, 2552, น.241) และตัวลวงที่จัดว่าเป็นตัวลวงที่ดีนั้นผู้ที่เรียนอ่อนควรจะต้องเลือกตอบมากกว่าผู้ที่มีการเรียนดี

การจัดการเรียนการสอนในหลักสูตรพยาบาลศาสตรบัณฑิตของมหาวิทยาลัยราชภัฏสวนดุสิต ได้ยึดหลักการเรียนรู้ที่เน้นผู้เรียนเป็นสำคัญ และใช้ระเบียบการวัดและประเมินผลของมหาวิทยาลัย ควบคู่กับแนวทางการวัดและประเมินผลที่คณะได้กำหนดขึ้น คือ ใช้การประเมินผลแบบอิงเกณฑ์ในรายวิชาภาคปฏิบัติ ส่วนรายวิชาภาคทฤษฎี ใช้การประเมินผลแบบอิงเกณฑ์ควบคู่กับอิงกลุ่ม กล่าวคือ นักศึกษาต้องมีคะแนนดิบถึงเกณฑ์ 60% จึงจะได้รับเกรด C นอกจากนั้น ยังได้มีการพัฒนามาตรการในการช่วยเหลือผู้เรียนที่เรียนอ่อน เพื่อพัฒนานักศึกษาให้มีความรู้ความสามารถตามเกณฑ์ที่หลักสูตรกำหนด และได้พัฒนาเต็มตามศักยภาพของตน

คณะพยาบาลศาสตร์ ตระหนักดีว่าข้อสอบที่จะนำมาใช้ประเมินผลการศึกษาของนักศึกษา นักศึกษาต้องเป็นข้อสอบที่มีคุณภาพ แต่ข้อสอบที่นำไปใช้จริงเป็นข้อสอบที่ผู้สอนเขียนขึ้นเอง (teacher-made test) แม้คณะพยาบาลศาสตร์จะพยายามเก็บรวบรวมข้อสอบที่มีคุณภาพไว้ใช้ในอนาคต แต่ก็ยังไม่มากพอ จึงเป็นความจำเป็นที่ผู้สอนจะต้องออกข้อสอบใหม่ทุกครั้ง อย่างไรก็ตาม คณะพยาบาลศาสตร์ได้แต่งตั้งผู้ทรงคุณวุฒิของคณะเพื่อทำหน้าที่วิพากษ์ข้อสอบร่วมกับผู้ออกข้อสอบ เพื่อปรับปรุงแก้ไขก่อนนำไปทดสอบจริง แต่ก็ไม่สามารถตรวจสอบคุณภาพกับนักศึกษา ก่อนนำไปใช้จริงได้ เพราะหากนำไปทดสอบก่อนก็จะเกิดปัญหาการจำข้อสอบได้ กว่าจะทราบระดับคุณภาพของข้อสอบก็ต่อเมื่อนำข้อสอบเหล่านี้ไปประเมินผลการศึกษาจริงแล้ว ซึ่งอาจจะพบว่ายังมีข้อสอบบางข้อที่ต่อคุณภาพอยู่บ้าง ประเด็นคุณภาพข้อสอบที่คณะพยาบาลศาสตร์ให้ความสำคัญเป็นอย่างมาก คือค่าอำนาจจำแนกของข้อสอบที่มีค่าติดลบ เพราะข้อสอบเหล่านี้เป็นข้อสอบที่ผู้ที่เรียนดีส่วนใหญ่มีโอกาสจะตอบผิดในขณะที่ผู้ที่เรียนอ่อนส่วนใหญ่มีโอกาสจะตอบถูก

ประสบการณ์ในการจัดการเรียนการสอนในปีการศึกษา 2554 ที่ผ่านมา พบว่า นักศึกษาบางคน สอบไม่ผ่านเกณฑ์ขั้นต่ำ 60% ตามที่หลักสูตรกำหนด ต้องมีการสอบซ่อมทั้งการสอบกลางภาคและปลายภาค จากผลการวิเคราะห์ข้อสอบแต่ละรายวิชา พบว่า มีจำนวนข้อสอบที่มีค่าอำนาจจำแนกติดลบ มากที่สุด คือร้อยละ 13 ซึ่งแสดงว่าคุณภาพของข้อสอบข้อนั้นๆไม่สามารถจำแนกผู้เรียนเก่งและอ่อนได้ ผู้เรียนอ่อนส่วนมากตอบข้อนั้นถูก แต่ผู้เรียนเก่งส่วนมากไม่ได้คะแนนจากข้อนั้น การนำคะแนนจากข้อสอบที่มีค่าอำนาจจำแนกติดลบไปประเมินผลการศึกษาของนักศึกษาย่อมส่งผลกระทบต่อนักศึกษาโดยตรง

คณะพยาบาลศาสตร์ เห็นความจำเป็นในการศึกษาคุณภาพของข้อสอบในแต่ละรายวิชาที่จัดโดยอาจารย์ของคณะพยาบาลศาสตร์ ทั้งวิชาในกลุ่มพื้นฐานวิชาชีพ หมวดวิชาเลือกเสรี และหมวดวิชาชีพ และเปรียบเทียบการเปลี่ยนแปลงของคะแนนการสอบของนักศึกษารายบุคคล และรายวิชา ก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก พร้อมทั้งผลที่เกิดขึ้นต่อการเปลี่ยนแปลงค่าความเชื่อมั่นของข้อสอบทั้งฉบับ หลังจากนั้นจึงนำผลที่ได้มาประมวลเพื่อให้ได้ข้อมูล และองค์ความรู้ที่เป็นหลักฐานเชิงประจักษ์ ที่นำมาสู่ข้อสรุป หรือข้อเสนอแนะเชิงนโยบายในด้านการวัดและประเมินผล การจัดการเรียนการสอนในแต่ละรายวิชา ในหลักสูตรพยาบาลศาสตรบัณฑิตต่อไป

การจัดการเรียนการสอนในวิชากฎหมายและจริยธรรมวิชาชีพซึ่งเป็นวิชาในกลุ่มพื้นฐาน วิชาชีพ เนื่องจากผู้เรียนมีจำนวน 60 คน (ในชั้นปี 2556 เพิ่มเป็น 90 คน) และเวลามีจำกัดด้วยเหตุผลต่างๆ รวมทั้งเพื่อให้สอดคล้องกับการสอบขึ้นทะเบียนรับใบอนุญาตประกอบวิชาชีพของสภาการพยาบาลที่กำหนดให้ผู้สำเร็จการศึกษา ต้องสอบวิชากฎหมายและจริยธรรมวิชาชีพด้วยโดยวัดด้วยข้อสอบชนิดเลือกตอบ 4 ตัวเลือก ทำให้ผู้วิจัยในฐานะผู้ประสานงานวิชา (Course coordinator) และเป็นผู้สอนหลัก พิจารณาใช้วิธีวัดและประเมินความรู้ของผู้เรียนด้วยวิธีการสอบโดยใช้แบบทดสอบผลสัมฤทธิ์ หรือชุดข้อสอบ ชนิดเลือกตอบ (MCQ) 4 ตัวเลือก เป็นหลัก ร่วมกับการวัดและประเมินผลด้วยวิธีการอื่นๆ แต่ยังคงให้น้ำหนัก การวัดประเมินผลการเรียนจากการสอบด้วยชุดข้อสอบไว้สูงที่สุดกว่าการวัดและการประเมินแบบอื่นๆ ดังนั้นคุณภาพของข้อสอบจึงเป็นสิ่งสำคัญมากเพราะส่งผลกระทบต่อทั้งด้านการประเมินผู้เรียนระหว่างกระบวนการเรียนรู้ การตัดเกรดภายหลังสอน และการสอบขึ้นทะเบียนรับใบอนุญาตประกอบวิชาชีพจากสภาการพยาบาล จากเหตุผลที่กล่าวมาแล้วทำให้ผู้วิจัยสนใจที่จะทำการศึกษาวิจัยเกี่ยวกับการประเมินคุณภาพของข้อสอบในวิชานี้ โดยมีทิศทางของการศึกษาวิจัยเป็นส่วนหนึ่งที่คู่ขนานกับการวิจัยร่วมใหญ่ของคณะพยาบาลศาสตร์ซึ่งศึกษาวิจัยคุณภาพของข้อสอบในแต่ละรายวิชาอื่นๆ ไปพร้อมกัน

วัตถุประสงค์ของการวิจัย

เพื่อประมวลผลการศึกษาวิจัย จากโครงการวิจัยเกี่ยวกับผลของการตัดข้อสอบที่มีอำนาจจำแนกติดลบออกต่อคะแนนการสอบ และความเชื่อมั่นของข้อสอบทั้งฉบับ ในรายวิชากฎหมายและจริยธรรมวิชาชีพ จึงตั้งวัตถุประสงค์ไว้ดังนี้

- 1) คุณภาพของข้อสอบรายวิชา กฎหมายและจริยธรรมวิชาชีพ (Law and Ethics in Profession)
- 2) การเปลี่ยนแปลงของคะแนนการสอบของนักศึกษาและการเปลี่ยนแปลงลำดับที่ของการสอบรายบุคคล ก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก
- 3) ศึกษาการเปลี่ยนแปลงของค่าความเชื่อมั่นของข้อสอบ ก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก

ขอบเขตการวิจัย

ประชากร คือ ข้อสอบ และคะแนนการสอบวิชากฎหมายและจริยธรรมวิชาชีพ (Law and Ethics in Profession) ของนักศึกษาชั้นปีที่ 4 หลักสูตรพยาบาลศาสตรบัณฑิต

ประชากรทั้งหมดที่เข้าถึงได้ (Accessible population) คือ ข้อสอบ และคะแนนการสอบวิชากฎหมายและจริยธรรมวิชาชีพ (Law and Ethics in Profession) ของนักศึกษาชั้นปีที่ 4 หลักสูตรพยาบาลศาสตรบัณฑิต ที่ลงทะเบียนเรียน ในภาคการศึกษาที่ 1 ปีการศึกษา 2555

กลุ่มตัวอย่าง เลือกจากข้อสอบ และคะแนนการสอบของวิชากฎหมายและจริยธรรมวิชาชีพ (Law and Ethics in Profession) ที่มีการวัดและประเมินผลโดยใช้ข้อสอบแบบเลือกตอบ (Multiple choices) 4 ตัวเลือก

คำจำกัดความที่ใช้ในงานวิจัย

1. **คุณภาพของข้อสอบ/ แบบทดสอบ** หมายถึง คุณภาพของข้อสอบรายข้อ และค่าความเชื่อมั่นของแบบทดสอบทั้งฉบับ สำหรับคุณภาพของข้อสอบรายข้อพิจารณาจาก ค่าความยากง่าย และค่าอำนาจจำแนก ในการวิจัยนี้ใช้ข้อสอบแบบเลือกตอบที่ใช้ในการสอบกลางภาค และปลายภาค ของนักศึกษาในหลักสูตรพยาบาลศาสตรบัณฑิต โดยข้อสอบที่มีค่าความยากง่ายเหมาะสม (p) อยู่ระหว่าง 0.2-0.8 และข้อสอบที่ค่าอำนาจจำแนก (r) ที่เหมาะสม คือ ≥ 0.2 ส่วนข้อสอบที่ค่าอำนาจจำแนกติดลบ แสดงว่าเป็นข้อสอบที่ไม่ดี เนื่องจาก นักศึกษาที่เรียนเก่งส่วนใหญ่ตอบผิด แต่นักศึกษาที่เรียนอ่อนส่วนใหญ่ตอบถูก

2. **การตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก** หมายถึง การคิดคะแนนของข้อสอบทั้งฉบับใหม่ ภายหลังจากทราบผลการวิเคราะห์ข้อสอบ โดยการไม่ตรวจให้คะแนนข้อสอบที่มีค่าอำนาจจำแนกเป็นลบ และปรับคะแนนรวมของจำนวนข้อสอบที่เหลือให้เป็นร้อยละ 100 เหมือนเดิม

3. **ค่าความเชื่อมั่นของข้อสอบ** หมายถึง ค่าความเชื่อมั่น (Reliability) ของข้อสอบ/แบบทดสอบวัดผลสัมฤทธิ์ของนักศึกษาในรายวิชากฎหมายและจริยธรรมวิชาชีพทั้งฉบับ ทั้งข้อสอบกลางภาค และปลายภาค เป็นข้อสอบแบบเลือกตอบ (Multiple choices) 4 ตัวเลือก คำนวณค่าสัมประสิทธิ์ความเชื่อมั่น โดยใช้สูตรของ KR-20 (Kuder-Richardson formula 20, KR-20)

ประโยชน์ที่คาดว่าจะได้รับ

1. ได้แนวทางในพัฒนาข้อสอบแบบเลือกตอบ การประเมินผลการเรียนการสอนรายวิชา กฎหมายและจริยธรรมวิชาชีพ ที่มีความเที่ยงตรง และ เชื่อมั่นมากกว่าวิธีที่ปฏิบัติอยู่ในปัจจุบัน
2. เกิดการพัฒนาการจัดการเรียนการสอน และการออกแบบการประเมินผล รายวิชา กฎหมายและจริยธรรมวิชาชีพ หลักสูตรพยาบาลศาสตรบัณฑิต มหาวิทยาลัยราชภัฏสวนดุสิต ให้มีมาตรฐานยิ่งขึ้น

บทที่ 2

แนวคิด ทฤษฎี เอกสารและงานวิจัยที่เกี่ยวข้อง

มโนทัศน์เกี่ยวกับการศึกษา (Educational Concepts)

มโนทัศน์เกี่ยวกับการศึกษา (Educational Concepts) ตั้งแต่อดีตถึงปัจจุบัน มีมโนทัศน์เกี่ยวกับการศึกษาที่หลากหลายซึ่งปรับตัวเปลี่ยนแปลงไปตามพัฒนาการของสังคม และในปัจจุบันองค์ความรู้ต่างๆ เทคโนโลยี และปฏิสัมพันธ์ของบุคคล ชุมชน และสังคม ในการเรียนรู้ มีกระบวนการ (process) เปลี่ยนแปลงที่รวดเร็วตลอดเวลาอย่างทั่วถึงกันทั้งโลก (Globalization) ผ่านโลกออนไลน์ การศึกษาเกี่ยวกับมโนทัศน์ทางการศึกษามีบริบทเกี่ยวกับคำศัพท์พื้นฐานทางการศึกษา หากผู้ศึกษาทราบความหมายของคำศัพท์ต่างๆที่เกี่ยวข้องจะช่วยให้มองแนวคิดทางการศึกษา และเข้าใจความสำคัญของการวัดและประเมินผลมากขึ้น ความหมายของคำศัพท์ที่สำคัญต่างๆ มีดังนี้

ความหมายของคำว่า “การศึกษา” (education) พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 เป็นแผนแม่บทด้านการศึกษาของประเทศ และในมาตรา 4 ของ พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 ได้ให้ความหมายของการศึกษาไว้ดังนี้ “การศึกษา หมายความว่า กระบวนการเรียนรู้เพื่อความเจริญงอกงามของบุคคลและสังคมโดยการถ่ายทอดความรู้ การฝึก การอบรม การสืบสานทางวัฒนธรรม การสร้างสรรค์จรรโลงความก้าวหน้าทางวิชาการ การสร้างองค์ความรู้อันเกิดจากการจัดสภาพแวดล้อม สังคม การเรียนรู้และปัจจัยเกื้อหนุนให้บุคคลเรียนรู้อย่างต่อเนื่องตลอดชีวิต” (ราชกิจจานุเบกษา เล่มที่ 116 ตอนที่ 74 ก) ในการจัดการศึกษาจำเป็นต้องมีการประเมินการเรียนรู้ (Assessment) ให้ได้ถึงผลลัพธ์การเรียนรู้ (Learning Outcomes) ได้แก่ การประเมินสิ่งที่พัฒนาขึ้นในตัวผู้เรียนทั้งจากการศึกษาในห้องเรียน กิจกรรมในและนอกหลักสูตร ปฏิสัมพันธ์กับเพื่อนนักศึกษา อาจารย์ และประสบการณ์ต่างๆที่เกิดในมหาวิทยาลัย และผลลัพธ์การเรียนรู้นี้เป็นสิ่งที่ต้องวัดให้ครอบคลุมถึงสาระความรู้ ความเข้าใจ ในเนื้อหาวิชา ทักษะ หรือความสามารถในการนำความรู้ไปใช้ มีพัฒนาการของพฤติกรรม ทักษะคิด ความคิด ความเชื่ออุปนิสัย (สุคนธ์ชื่น ศรีงาม, มปป.)

พระพรหมคุณาภรณ์ (2552) ให้ความหมายว่า การศึกษา “การศึกษานั้น เป็นการเรียนรู้เพื่อให้พัฒนาให้อยู่ดียิ่งขึ้น”

จาก William – Webster encyclopedia of Britannica ให้ความหมายของคำว่า “การศึกษา”(education) ไว้ว่า “เป็นการกระทำ หรือ กระบวนการ ในการสอนบุคคล โดยเฉพาะในโรงเรียน วิทยาลัย หรือมหาวิทยาลัย : เป็นความรู้ ทักษะ และความเข้าใจต่างๆที่บุคคลได้รับจากการศึกษาในโรงเรียน วิทยาลัย หรือมหาวิทยาลัย เป็นขอบข่ายของการเรียนรู้อย่างมีหลักเกณฑ์ซึ่งเกี่ยวข้องกับวิธีการสอน และปัญหาการสอน

จากนิยามความหมายการศึกษาเป็นการเน้นการเรียนรู้ในระบบโรงเรียนหรือในสถาบันการศึกษาที่เป็นทางการ (formal education) รวมความถึงกระบวนการและผลลัพธ์ที่เกิดขึ้นในตัวผู้เรียน

ความหมายของคำว่า “การเรียนรู้” (Learning) ประดิษฐ์ อุปรมัย (2540, น.121) สรุปความหมายของการเรียนรู้ว่า “การเรียนรู้คือการเปลี่ยนแปลงของบุคคลอันมีผลเนื่องมาจากการได้รับประสบการณ์โดยการเปลี่ยนแปลงนั้น เป็นเหตุทำให้บุคคลเผชิญสถานการณ์เดิมแตกต่างไปจากเดิม ประสบการณ์ที่ก่อให้เกิดการเปลี่ยนแปลงพฤติกรรม หมายถึงทั้งประสบการณ์ทางตรงและประสบการณ์ทางอ้อม”

ประสบการณ์ทางตรง คือ ประสบการณ์ที่บุคคลได้รับหรือสัมผัสด้วยตนเองแล้วเกิดประสบการณ์ใหม่ เช่น การสัมผัสถูกความร้อน ทำให้ไม่กล้าเข้าใกล้อีก เป็นต้น พฤติกรรมการเรียนรู้ต้องเกิดขึ้นขณะบุคคลมีสติ สัมผัสัญญะ Reflex ต่างๆในร่างกายจากระบบประสาทอัตโนมัติ ไม่ถือว่าเป็นประสบการณ์ตรง

ประสบการณ์ทางอ้อม คือ ประสบการณ์ที่ผู้เรียนมิได้พบหรือสัมผัสด้วยตนเองโดยตรงแต่อาจได้รับประสบการณ์ทางอ้อมจากการอบรมสั่งสอนหรือการบอกเล่าการอ่านหนังสือต่างๆ และการรับรู้จากสื่อต่างๆ

จาก William – Webster encyclopedia of Britannica (online) ให้ความหมายของการเรียนรู้ว่า (learning) “เป็นกิจกรรม หรือกระบวนการที่บุคคลได้รับความรู้ ทักษะ จากการเรียน (study) จากการฝึกฝนปฏิบัติ (practice) จากการได้รับการสอน หรือการมีประสบการณ์ในบางสิ่งบางอย่างทำให้เกิดแนวโน้มที่จะปรับพฤติกรรมจากประสบการณ์ที่ได้รับ”

เคนดร้า เซอร์รี่ (Kendra Cherry) นักการศึกษาทางด้านจิตวิทยา สรุปความหมายของการเรียนรู้ว่า “การเรียนรู้ทำให้เกิดประสบการณ์ และประสบการณ์จะสัมพันธ์กับผลลัพธ์สุดท้ายที่ทำให้บุคคลมีการปรับพฤติกรรม” การเรียนรู้ได้รับความสนใจมากในการศึกษาทางจิตวิทยาในส่วนของพฤติกรรมศาสตร์ (behaviorism) ในช่วงต้นของศตวรรษที่ 20 และเป็นจุดเริ่มต้นของกลุ่มการศึกษาหลักทางด้านความคิด (major school of thought) ปัจจุบันการเรียนรู้แนวพฤติกรรมศาสตร์ยังคงเป็นแนวคิดสำคัญทางจิตวิทยาหลากหลายสาขา ทั้งทางการรู้คิด (cognitive) ด้านการศึกษา ด้านสังคม และจิตวิทยาพัฒนาการ”

สรุป ความหมายของการเรียนรู้เน้นไปที่ผลลัพธ์ของการเรียนรู้ทั้งในระบบโรงเรียนและนอกระบบโรงเรียนที่เป็นประสบการณ์ทำให้บุคคลเกิดการเปลี่ยนแปลงพฤติกรรม

วิเคราะห์ความแตกต่างระหว่างคำว่า การศึกษา (education) กับการเรียนรู้ (learning)

1. **การศึกษา** หมายถึงกระบวนการจัดการศึกษาในระบบสถาบันการศึกษา (formal learning) ที่เป็นโรงเรียน วิทยาลัย หรือมหาวิทยาลัย โดยการจัดการให้บุคคลเกิดประสบการณ์การเรียนรู้ โดยมีการกำหนดหลักสูตร (curriculum) การวางแผน กำหนดจุดมุ่งหมาย วิธีการ มีการสอน (teaching) และการวัดผลลัพธ์ที่เกิดขึ้นจากการเรียนรู้ (learning Outcome) และให้เอกสารรับรอง (certificate) เกี่ยวข้องกับการเรียนรู้ในชั้นเรียนที่มีบุคคลอื่นจัดมาตรฐานไว้แล้วเป็นสิ่งที่รัฐจัดสำหรับประชาชนในสังคม

2. **การเรียนรู้** เน้นกระบวนการที่เกิดภายในตัวบุคคล (inner self) ที่เกิดขึ้นตั้งแต่เกิดมีชีวิตจนกระทั่งตาย มีลักษณะเป็นกระบวนการอย่างไม่เป็นทางการ (informal process) แต่การศึกษาได้

รับมาจากเพียงบางช่วงของชีวิตเท่านั้นซึ่งมีแหล่งที่มาจากภายนอก (outside source) มีลักษณะเป็นกระบวนการที่เป็นทางการ (formal process)

การเรียนรู้ทางอิเล็กทรอนิกส์ หรือ E-learning หมายถึง รูปแบบการเรียนการสอนซึ่งใช้การถ่ายทอดเนื้อหา (delivery methods) ผ่านทางอุปกรณ์อิเล็กทรอนิกส์ไม่ว่าจะเป็น คอมพิวเตอร์ เครือข่ายอินเทอร์เน็ต อินเทอร์เน็ต เอ็กซ์ทราเน็ต หรือ ทางสัญญาณโทรทัศน์หรือ สัญญาณดาวเทียม และใช้รูปแบบการนำเสนอเนื้อหาสารสนเทศในรูปแบบต่างๆซึ่งอาจอยู่ในรูปแบบการเรียนที่เรา คำนึงกันมาพอสมควร เช่น คอมพิวเตอร์ช่วยสอน (Computer-Assisted Instruction) การสอนบนเว็บ (Web-Based Instruction) การเรียนออนไลน์ (Online Learning) การเรียนทางไกลผ่านดาวเทียม หรืออาจอยู่ในลักษณะที่ยังไม่ค่อยเป็นที่แพร่หลายนัก เช่น การเรียนจากวิดีโอทัศน์ตามออร์เดอ (Video On-Demand) และคำว่า e-Learning ไม่ได้จำกัดเฉพาะ นำเอกสารการสอนมาแปลงให้อยู่ในรูปดิจิทัลและนำไปวางไว้บนเว็บหรือระบบบริหารจัดการการเรียนรู้เท่านั้นแต่ครอบคลุมถึงกระบวนการ ในการเรียนการสอนหรือการอบรมที่ใช้เครื่องมือทางด้านเทคโนโลยีสารสนเทศเพื่อให้เกิดความยืดหยุ่นทางการเรียนรู้ (flexible learning) สนับสนุนการเรียนรู้ในลักษณะที่ผู้เรียนเป็นศูนย์กลาง (learner-centered) และการเรียนในลักษณะตลอดชีวิต (life-long learning) ซึ่งอาศัยการเปลี่ยนแปลงด้านกระบวนทัศน์ (paradigm shift) ของทั้งกระบวนการในการเรียนการสอนด้วย นอกจากนี้ e-Learning ไม่จำเป็นต้องเป็นการเรียนทางไกลเสมอคุณาจารย์สามารถนำไปใช้ในลักษณะการผสมผสาน (blended) กับการสอนในชั้นเรียนได้ (ศูนย์บริการสื่อสารสนเทศ มหาวิทยาลัยศรีประทุม วิทยาเขตชลบุรี)

การศึกษาผ่านระบบอิเล็กทรอนิกส์คอมพิวเตอร์ (E-education) ฌนอมพร เลาหจรัสแสง (มปป) (ออนไลน์) ให้นิยามว่า คือ กระบวนการจัดระบบการศึกษา (education system) ทางไกลผ่านระบบเครือข่ายอิเล็กทรอนิกส์คอมพิวเตอร์ ทั้งในแง่ข้อมูลสื่อการศึกษาและการบริการ ได้แก่ ตัวหลักสูตรและการสอน (course ware) ห้องสมุดอิเล็กทรอนิกส์ การชำระค่าลงทะเบียน ระบบช่วยเรียนรู้ การวัดและประเมินผล เป็นต้น

ดังนั้น คำว่า “การศึกษาผ่านระบบอิเล็กทรอนิกส์คอมพิวเตอร์” จึงมีความหมายกว้างกว่าคำว่า “การเรียนรู้ทางอิเล็กทรอนิกส์” โดยที่ การศึกษาผ่านระบบอิเล็กทรอนิกส์คอมพิวเตอร์ เป็นการศึกษาที่เน้นการสื่อสารทางไกล โดยอาศัย อิเล็กทรอนิกส์คอมพิวเตอร์ เป็นตัวกลางเชื่อมระบบการศึกษากับตัวผู้เรียน

กรอบแนวคิดในการเรียนรู้ (Learning Framework)

กรอบแนวคิดในการเรียนรู้ (Learning Framework) เป็นกรอบที่ทำให้เห็นภาพรวมทั้งระบบของโครงสร้างการเรียนรู้ซึ่งจะมีส่วนประกอบย่อยๆการเรียนรู้ที่ต่อเนื่องสัมพันธ์กันอยู่ภายในกรอบ เกี่ยวข้องกับกระบวนการจัดการเรียนรู้ ได้แก่ การกำหนดเนื้อหาหลักสูตร การจัดกระบวนการวิชา การสอน การประเมินผล และอื่นๆเป็นต้นเพื่อพัฒนาการเรียนรู้ กรอบการเรียนรู้ (Learning framework) ในสหรัฐอเมริกาหลายสถาบันการศึกษายึดกรอบที่มาจากรายงานการเรียนรู้ทางด้านวิทยาศาสตร์ของสภาวิจัยแห่งชาติของอเมริกา (guidance of the National Research Council's report on learning science titled How people learn: Brain, mind, experience, and

school) (National Academy Press, 2001) (online) ที่ศึกษา ว่ามนุษย์เรียนรู้ได้อย่างไร (A framework of how people learn, HPL framework)



แผนภูมิที่ 2.1 กรอบแนวคิดการเรียนรู้ของมนุษย์ (HPL framework)
ที่มา; The IRIS center, Peabody college, Vanderbilt University

อธิบายกรอบแนวคิดตามแผนภูมิที่ 2.1 การเรียนรู้ (learning) มีลักษณะเป็นกระบวนการ (process) ที่เกิดขึ้นจากการมีปฏิสัมพันธ์ (interact) ระหว่างองค์ประกอบหลักที่เกี่ยวข้องกัน องค์ประกอบหลักทั้ง 4 ด้าน ได้แก่

1. ศูนย์รวมในตัวผู้เรียน (learner centered)
2. ศูนย์รวมด้านองค์ความรู้ (knowledge centered)
3. ศูนย์รวมด้านการประเมินผลการเรียนรู้ (assessment centered)
4. ศูนย์รวมด้านชุมชนแห่งการเรียนรู้ (community centered).

การบริหารการเรียนรู้ให้มีประสิทธิภาพ มาจากการจัดการสิ่งแวดล้อมของบริบทต่างๆ ใน องค์ประกอบทั้ง 4 ให้ได้สมดุลกัน โดยมีรายละเอียดในส่วนของแต่ละองค์ประกอบ ดังนี้

1. ศูนย์รวมในตัวผู้เรียน (learner centered) ประกอบด้วย

- 1.1 ภูมิหลังเกี่ยวกับความรู้ในตัวผู้เรียน (student's background knowledge)
- 1.2 ความสนใจของผู้เรียน (interests)
- 1.3 ค่านิยมทางสังคมและวัฒนธรรม (social and culture values)

องค์ประกอบนี้ให้ความสำคัญกับตัวผู้เรียน เกี่ยวกับแบบแผนที่ผู้เรียนถนัดในการเรียนรู้ทัศนคติ และบุคลิกส่วนตัว ตระหนักถึงความรู้และทักษะเดิมของผู้เรียนที่นำมาใช้ในสิ่งแวดล้อมของการเรียนรู้ใหม่ การปรับตัว ฯลฯ

2. ศูนย์รวมด้านองค์ความรู้ (knowledge centered) ประกอบด้วย

- 2.1 สารความรู้ที่นำมาสอน ต้องถูกต้อง เพียงตรง
- 2.2 ช่วยให้ผู้เรียนได้เกิดความเข้าใจ เกิดความหมาย ในตัวผู้เรียนมากกว่าการท่องจำ

องค์ประกอบนี้ ให้ความสำคัญกับโอกาสที่ผู้เรียนได้มีปฏิสัมพันธ์กับบริบทต่างๆให้เกิดความเข้าใจในสิ่งที่เรียนมากกว่าการท่องจำความรู้เป็นส่วนๆ และสามารถนำความรู้ไปพัฒนาใช้ได้

3. ศูนย์รวมเกี่ยวกับการประเมินผลการเรียนรู้ (assessment centered) ประกอบด้วย การประเมินที่ได้มาตรฐานสูง สร้างโอกาสนำผลการประเมินป้อนย้อนกลับ (feed back) การสะท้อนคิด (reflection) การช่วยให้เกิดมุมมองใหม่ (revision) ผู้เรียนอย่างสม่ำเสมอเพื่อให้ผู้เรียนได้เกิดการพัฒนาคุณภาพของการเรียนรู้ องค์ประกอบนี้ให้ความสำคัญกับการแสวงหาวิธีที่จะเฝ้าติดตามความก้าวหน้าของการเรียนรู้อย่างต่อเนื่อง และให้ข้อมูลย้อนกลับ มากกว่าการทดสอบ เฉพาะให้เสร็จสิ้นเป็นครั้งๆเท่านั้น

4. ศูนย์รวมด้านชุมชนแห่งการเรียนรู้ (community centered) ประกอบด้วย

4.1 การบ่มเพาะ และสนับสนุน ให้เกิดบรรทัดฐานของชุมชนแห่งการเรียนรู้ โดยให้มีบรรยากาศที่ทุกคนในชุมชนตระหนักและมีความร่วมมือกันว่า “การเรียนรู้คือสิ่งที่สำคัญที่สุด”

4.2 บุคคลได้รับการเคารพ และได้รับการเชื่อมโยงให้เข้าถึงทรัพยากรแห่งการเรียนรู้ ต่างๆ

4.3 ยอมรับสิ่งผิดพลาดในกระบวนการเรียนรู้ของบุคคล

องค์ประกอบนี้ ให้ความสำคัญกับบริบทของสิ่งแวดล้อมทั้งที่เป็นตัวบุคคล ปัจจัยสนับสนุนต่างๆ และบริบทอื่นๆที่การเรียนรู้ดำเนินอยู่ ส่งเสริมให้เกิดลักษณะของชุมชนแห่งการเรียนรู้ มีกระบวนการแลกเปลี่ยนจุดมุ่งหมาย (Goal) และค่านิยม (value) ร่วมกัน

กรอบแนวคิดตามรูปแบบ HPL เกิดก่อนการพัฒนาปรัชญาการศึกษาแนว Existentialism และแนวมนุษยนิยม (Humanistic psychology) จึงได้รับการวิพากษ์วิจารณ์จากนักการศึกษาด้านมนุษยวิทยา และด้านสังคมวิทยาว่า เริ่มต้นมาจากการศึกษาวิจัยทางด้านวิทยาศาสตร์เกี่ยวกับการรู้คิด (cognitive science) ที่เน้นด้านจิตวิทยาด้านการรู้คิด (cognitive psychology) ยังขาดความชัดเจนในมิติ เกี่ยวกับความรู้สึกรู้จัก (sense making) การพัฒนาการ (development) การหยั่งเห็น (insight) อภิปัญญาการรู้คิด (metacognition) ความถนัดในการเรียน (learning style) หรือกล่าวให้เข้าใจง่ายๆว่า แนว HPL เน้นแต่ด้านสมองในการคิด (เน้นสมองซีกซ้าย) ขาดมิติด้านปัญญาญาณ ได้แก่ การหยั่งเห็น และอภิปัญญาการรู้คิดซึ่งเป็นปัญญาชั้นสูง (wisdom) (Bransford and et al., 2000, p.101-102 in Kuchi and et al., 2003, p.3) โดย นักการศึกษากลุ่มจิตวิทยาแนวมนุษยนิยม (Humanistic psychology) ให้ความสำคัญกับด้านในของมนุษย์ ได้แก่ การเคารพความเป็นปัจเจก และแนวการเรียนรู้ทางสังคม (social learning) ให้ความสำคัญเพิ่มขึ้นและขยายมิติการเรียนรู้ ในด้านการนำตนเองในการเรียนรู้ (self - directed Learning), ชุมชนแห่งการเรียนรู้ (Learning Communities) และวัฒนธรรมในการสอนที่เหมาะสมกับความถนัดในการเรียนรู้ของผู้เรียน (Grassian and Kaplowitz 2001, p.50 -55 ; Merriam and Caffarella 1999, p.252 - 261 in Kuchi and et al, 2003, p.4)

(หมายเหตุ อภิปัญญาการรู้คิด เกี่ยวกับแนวคิดทางการศึกษาที่ว่า บุคคลที่ได้รับการพัฒนา จะเกิดกระบวนการทางปัญญา มีความสามารถ ที่จะควบคุมและประเมินความคิดตนเอง มีความตระหนักรู้ในบุคคล และงานทั้งที่เป็นเป้าหมาย กระบวนการ และการประเมินผล และสามารถใช้ยุทธวิธีทำงานจนสำเร็จอย่างสมบูรณ์ องค์ประกอบที่สำคัญที่จะช่วยให้เกิดอภิปัญญาการรู้คิด ได้แก่ ความตระหนักรู้ (awareness) และความสามารถในการควบคุมตนเอง (self regulation) (วิจารณ์ พานิช, 2556; ทิศนา เขมมณี, 2551)

สรุป การวัดประเมินผลลัพธ์การเรียนรู้ (Learning outcome) เป็นส่วนประกอบหลักสำคัญตามกรอบการเรียนรู้ซึ่งต้องมีมาตรฐานเพราะใช้ผลของการประเมินเป็นข้อมูลย้อนกลับเพื่อตรวจสอบคุณภาพการเรียนรู้ของผู้เรียน ผู้สอน วัตถุประสงค์ และกระบวนการจัดประสบการณ์การเรียนรู้ในประเทศไทย กรอบแนวคิดการจัดการเรียนรู้มีพัฒนาการสอดคล้องกับกรอบ HPL ในอดีต แต่ปัจจุบันกรอบมาตรฐานการศึกษาอุดมศึกษา (TQF) (พ.ศ.2552) ได้ให้ความสำคัญกับผลลัพธ์การเรียนรู้ระดับปัญญา และคุณธรรมจริยธรรม มากขึ้น นอกจากนี้การทบทวนวรรณกรรมยังสะท้อนให้เห็นว่าหากองค์กรการศึกษา และผู้สอนมีความเชื่อแนวปรัชญาการศึกษาแบบใด การวัดประเมินผลจะออกแบบให้สอดคล้องและสะท้อนความเชื่อนั้นๆ การศึกษาสายวิทยาศาสตร์การแพทย์ตามแบบคลาสสิก (Classical Medical Education) ยังเน้นศาสตร์ที่เน้นเนื้อหาพื้นฐาน (content based) ซึ่งยังเป็นสิ่งจำเป็นเพราะเกี่ยวข้องกับความเป็นความตายของชีวิตจึงต้องมีความถูกต้องแม่นยำสูง ดังนั้นการวัดและประเมินผลผู้เรียน จึงให้ความสำคัญกับมิติการรู้คิดเป็นหลัก (cognitive domain) และให้ความสำคัญกับการวัดระดับการคิด ตามรูปแบบของบลูม (Bloom Taxonomy) (พวงพยอม แก้วพิลา, วราภรณ์ เชื้ออินทร์, นภา หลิมรัตน์, ชาญชัย พานทองวิริยะกุล, 2548) ซึ่งเป็นไปในลักษณะเดียวกับวิชาชีวการพยาบาล-ซึ่งยังคงวัด และประเมินผลผู้เรียนตามศาสตร์ที่เน้นเนื้อหาพื้นฐาน (content based) ซึ่งเป็นฐานการวิจัยครั้งนี้

มโนทัศน์เกี่ยวกับการประเมินการเรียนรู้ (learning assessment concepts)

ความหมายของศัพท์เกี่ยวกับการประเมิน

การทดสอบ (Test or Quiz) เป็นวิธีวัดความสามารถของผู้เรียน เมื่อเรียนรู้เสร็จสิ้นหรือตรวจสอบเกี่ยวกับเนื้อหาความรู้ ทักษะที่เกิดขึ้นในตัวผู้เรียน แบบทดสอบที่ใช้วัด เช่น แบบทดสอบชนิดเลือกตอบ เป็นต้น มักใช้คู่กับการประเมินผล (evaluation) หรือการประเมิน (assessment) หรืออีกแง่หนึ่งคือการทดสอบเป็นส่วนหนึ่งของการประเมินผล

การวัด (Measurement) ความหมายทั่วไป หมายถึงการวัด แต่ในทางการศึกษา แปลว่า **การวัดผล** หมายถึง รูปแบบ และวิธีการของการวัดผล (set of procedures) และหลักการเกี่ยวกับวิธีการทดสอบ (test) หรือเป็นกระบวนการที่ได้มาซึ่งข้อมูลเชิงปริมาณและเชิงคุณภาพเกี่ยวกับการเรียนรู้ หรือการวัดผล หมายถึง การใช้เครื่องมืออย่างใดอย่างหนึ่งที่ช่วยค้นหา หรือตรวจสอบเพื่อให้ได้ข้อมูลเชิงปริมาณ หรือเชิงคุณภาพ แล้วแปลงเป็นตัวเลขที่มีความหมายแทนพฤติกรรมหรือผลงานที่ผู้เรียนแต่ละคนแสดงออกมา (ภัทรา นิคมานนท์, 2522, น.1) การวัดผลเป็นการวัดทางอ้อมในสิ่งที่ เป็นนามธรรม ได้แก่ วัดความสามารถในการเรียนรู้ในวิชาต่างๆ เครื่องมือที่ใช้วัดมีหลายประเภท ทั้งนี้ต้องเลือกใช้ให้สอดคล้องกับคุณลักษณะของสิ่งที่ต้องการวัด ตัวอย่างคำศัพท์ที่ใช้ในการวัดผลเช่น คะแนนดิบ (raw scores) ลำดับที่เปอร์เซ็นต์ไทล์ (percentile ranks) คะแนนมาตรฐาน (standard scores) เป็นต้น

การประเมินผล (Evaluation) หมายถึง กระบวนการตัดสินใจ ให้ค่า จาก ข้อมูลที่ได้มา จากการวัดผล หรืออีกนัยหนึ่ง เป็นการตีค่าหรือกำหนดระดับคุณค่าของผลจากการวัดตามเกณฑ์หรือ จุดมุ่งหมายที่กำหนด การประเมินผลที่เที่ยงธรรมย่อมาจาก การวัดที่ดี

การประเมิน (Assessment) เป็นกระบวนการเก็บรวบรวมข้อมูล สารสนเทศ ด้วยวิธีการที่ หลากหลาย ตีความข้อมูล และนำผลการประเมินมาใช้เพื่อประเมินความก้าวหน้าและตัดสินใจ ทางการศึกษา จากนियามการทดสอบ (test) จึงเป็นส่วนหนึ่งของการประเมิน วิธีการประเมินอื่นๆ ได้แก่ การสัมภาษณ์ การสังเกต เป็นต้น (อนันต์ ศรีโสภา, 2534; ทิวต์ล มณีโชติ, 2549; Overton, 2009)

การประเมินทางการศึกษา (Educational Assessment) เป็นกระบวนการเก็บรวบรวม สารสนเทศด้วยวิธีการที่หลากหลายทางการศึกษา และนำข้อมูลมาใช้ในการพัฒนาระบบการศึกษา

ลักษณะการประเมินทางการศึกษา มีดังนี้ (ทิวต์ล มณีโชติ, 2549)

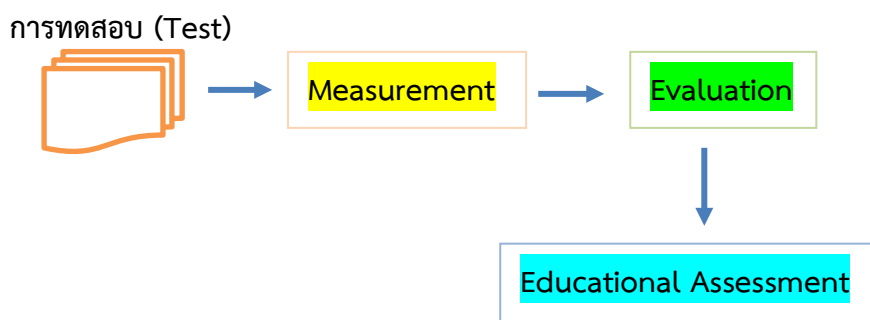
1. เป็นส่วนหนึ่งของกระบวนการเรียนการสอนหรือกระบวนการจัดการเรียนรู้ซึ่งควรทำการ ประเมินอย่างต่อเนื่อง เพื่อนำผลการประเมินไปปรับปรุงกระบวนการเรียนการสอนให้มีประสิทธิภาพ ยิ่งขึ้น

2. เป็นการประเมินคุณลักษณะหรือพัฒนาการการเรียนรู้ของผู้เรียนว่าบรรลุตามจุดประสงค์ หรือไม่

3. เป็นการประเมินในภาพรวมทั้งหมดของผู้เรียน โดยการรวบรวมข้อมูลและประมวลจากตัวเลข จาก การวัดหลายวิธีและหลายแหล่ง

4. เป็นกระบวนการเกี่ยวข้องกับบุคคลหลายกลุ่ม ทั้งครู นักเรียน ผู้ปกครองนักเรียน ผู้บริหาร โรงเรียน และอาจรวมถึงคณะกรรมการต่างๆ ของโรงเรียน

สรุป จากนิยามของคำที่ใช้เกี่ยวกับกระบวนการตรวจสอบคุณภาพการศึกษา เริ่มต้นจาก กระบวนการย่อยๆ เกิดขึ้นต่อเนื่องกันไป ได้แก่ การทดสอบ จนถึง กระบวนการสุดท้าย ได้แก่ การ ประเมินทางการศึกษา จากแผนภูมิด้านล่าง



แผนภูมิที่ 2.2 ความเชื่อมโยงต่อเนื่องของความหมายการวัดและการประเมินผล

งานวิจัยครั้งนี้เป็นการวิเคราะห์คุณภาพของแบบทดสอบในรายวิชากฎหมายและจริยธรรม เป็นเรื่องของคุณภาพเครื่องมือที่ใช้ในการทดสอบ(Test) และการพิจารณาตัดข้อสอบ r ติดลบ เพื่อตัดเกรด อยู่ในส่วนของ การวัดและประเมินผล (Measurement and Evaluation)

ความแตกต่างระหว่างคำว่า การประเมิน(assessment) กับการประเมินผล (evaluation)

ตารางที่ 2.1 แสดงความแตกต่างระหว่าง การประเมิน (assessment) และการวัดผล (evaluation)

	การประเมิน (assessment)	การประเมินผล (evaluation)
ความมุ่งหมาย (purpose)	เพื่อพัฒนาผลการปฏิบัติงานในอนาคต	การตัดสินผิดถูก หรือ ตัดสินคุณค่าของผล การปฏิบัติงาน โดยเปรียบเทียบกับเกณฑ์ มาตรฐานที่กำหนดไว้ล่วงหน้า
การกำหนดข้อบ่งชี้ในการ ประเมิน (setting Criteria)	ผู้ประเมิน (assessor) และผู้รับการประเมิน เลือก กำหนด ข้อบ่งชี้ หรือ เกณฑ์ การประเมินร่วมกัน	ผู้ตัดสิน หรือผู้ประเมิน (evaluation) เป็น ผู้กำหนดเกณฑ์ในการประเมิน
การควบคุม (control)	ผู้รับการประเมิน เป็นผู้เลือก ที่จะพัฒนาหรือ ปรับปรุงแก้ไข หรือ จะไม่กระทำใดๆ หลังได้รับผล การประเมิน จากการที่ผู้ประเมินให้ข้อมูลย้อนกลับ (feedback)	ผู้ประเมินเป็นผู้ตัดสิน กระทำสิ่งต่างๆ เกี่ยวกับผลกระทบของการประเมิน
ความลึกของการวิเคราะห์ (depth of analysis)	การวิเคราะห์การประเมินจะกระทำอย่างต่อเนื่อง โดยตั้งคำถาม เช่น ทำไม อย่างไร ถึงจะปรับปรุงผลการปฏิบัติงานในอนาคต	กระทำการวัด และตัดสินผล ไปตามเกณฑ์
การตอบสนอง (Response)	มองการลงมือกระทำ และการปฏิบัติตามแผน ในเชิงบวก	สิ้นสุดการประเมิน ด้วยคำตอบ ถึง ความสำเร็จ หรือ ล้มเหลว

ที่มา; Peter E. Parker and et al., 2001

หลักการประเมินทางการศึกษาโดยทั่วไป (ทีวิตต์ มณีโชติ, 2549) มีดังนี้

1. ขอบเขตการประเมินต้องตรงและครอบคลุมหลักสูตร
2. ใช้ข้อมูลจากผลการวัดที่ครอบคลุมจากการวัดหลายแหล่ง หลายวิธี
3. เกณฑ์ที่ใช้ตัดสินผลการประเมินมีความชัดเจน เป็นไปได้ มีความยุติธรรม ตรงตาม

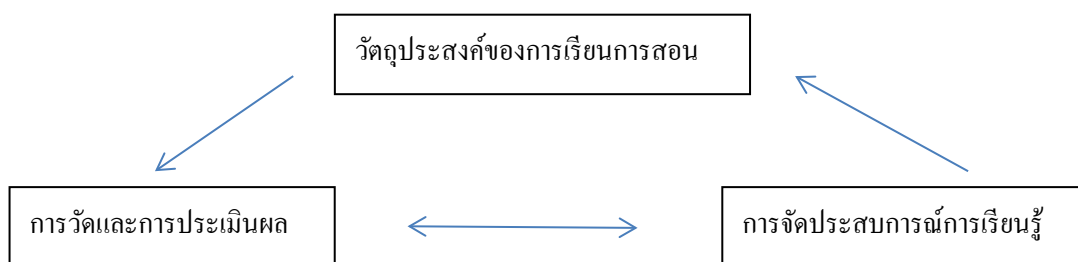
วัตถุประสงค์ของหลักสูตร

ขั้นตอนในการประเมินทางการศึกษา (ทีวิตต์ มณีโชติ, 2549)

การประเมินทางการศึกษามีขั้นตอนที่สำคัญ ดังนี้

1. กำหนดจุดประสงค์การประเมิน โดยให้สอดคล้องและครอบคลุมจุดประสงค์ของหลักสูตร
2. กำหนดเกณฑ์เพื่อตีค่าข้อมูลที่ได้จากการวัด
3. รวบรวมข้อมูลจากการวัดหลายๆ แหล่ง
4. ประมวลและผสมผสานข้อมูลต่างๆ ของทุกรายการที่วัดได้
5. วินิจฉัยชี้แจงและตัดสินโดยเทียบกับเกณฑ์ที่ตั้งไว้

การวัดและการประเมินผล (Measurement and Evaluation) ในทางการศึกษา มักใช้คำ ทั้ง 2 คำคู่กัน ทั้งวัดและประเมินผล เพื่อใช้ในการตัดสินผลการเรียนรู้ หมายความว่า กระบวนการ ส่วนหนึ่งของกระบวนการจัดการเรียนการสอนเพื่อตรวจสอบสิ่งที่เกิดขึ้นในการเรียนรู้โดยทำการ **วัดผล** เพื่อเก็บข้อมูลเกี่ยวกับผู้เรียนในรูปของคะแนนหรือคำบรรยายที่เกี่ยวกับผลการเรียนรู้ของผู้เรียนที่แสดงออกทั้งในระยะเวลา ก่อน ระหว่าง หรือหลังจากการเรียนการสอน ดังแผนภูมิที่ 2.3



แผนภูมิที่ 2.3 วงจรความสัมพันธ์ระหว่างวัตถุประสงค์ของการเรียนการสอน การจัด ประสบการณ์การเรียนรู้ และการวัดประเมินผล
 ที่มา; วสันต์ ทองไทย, มปป

ความสำคัญของการวัดและประเมินผลการเรียนรู้

การวัดและประเมินผลการเรียนรู้ เป็นองค์ประกอบสำคัญเพื่อ ทำให้ได้ข้อมูลสารสนเทศที่ จำเป็นในการพิจารณาว่าผู้เรียนเกิดคุณภาพการเรียนรู้ตามผลการเรียนรู้ที่คาดหวังและมาตรฐานการ เรียนรู้ การวัดและประเมินผลการเรียนนอกจากจะมีประโยชน์โดยตรงต่อผู้เรียนแล้ว ยังสะท้อนถึง ประสิทธิภาพการการสอนของครู และเป็นข้อมูลสำคัญที่สะท้อนคุณภาพการดำเนินงานการจัดการ การศึกษาของสถานศึกษาด้วย ดังนั้นครูและสถานศึกษาต้องมีข้อมูลผลการเรียนรู้ของผู้เรียน ทั้งจาก การประเมินในระดับชั้นเรียน ระดับภาพรวมทั้งหมด

การประเมินผลการเรียนรู้ในนักศึกษาพยาบาล (Nursing student learning outcome Assessment) กับความสำคัญของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน (Importance of Achievement Test)

กระบวนการจัดการเรียนการสอนทางวิทยาศาสตร์สุขภาพ ได้แก่ พยาบาลศาสตร์ศึกษา ต้องมีการออกแบบการจัดการเรียนรู้ (course designs) ให้เหมาะสมตามลักษณะวิชา และ วัตถุประสงค์รายวิชา รวมทั้งออกแบบการประเมินผลการเรียนรู้ ให้เหมาะสม สอดคล้อง กับการ ออกแบบการเรียนรู้ที่กำหนดไว้ เพื่อประเมินผลลัพธ์ว่าผู้เรียนได้เกิดการพัฒนาการตามวัตถุประสงค์ รายวิชา โดยการวัดและประเมินผล ความรู้ (knowledge) สติปัญญา (cognitive) ทักษะ (skills), อารมณ์ ความรู้สึก (emotion and perception) และทัศนคติ (Attitude) ของผู้เรียนด้วยเครื่องมือที่ ผสมผสานหลากหลาย เพราะโดยหลักการวัดประเมินผลไม่มีเครื่องมือชนิดใดที่ดีที่สุด และสามารถ ใช้ เก็บข้อมูลผู้เรียนได้ครอบคลุมด้วยเครื่องมือเพียงชนิดเดียว (รังสรรค์ มณีเล็ก, 2552, น.5) เครื่องมือที่

ใช้วัดจะสอดคล้องไปกับการออกแบบการเรียนรู้ ตัวอย่างของเครื่องมือ ที่ใช้วัดการเรียนรู้ในนักศึกษาพยาบาล ได้แก่ การประเมินการเรียนรู้ในสภาพการณ์จริง (Authentic Assessment) ในภาคปฏิบัติ (Performance-Based) จะใช้เครื่องมือ แฟ้มสะสมผลงาน (portfolio) (Marilyn H. Oermann, 2002) การสอบทักษะปฏิบัติแบบเสมือนจริง (Objective Structured Clinical Examination, OSCE) การประเมินการทำโครงการ (Project based learning evaluation) และ การใช้แบบทดสอบวัดผลสัมฤทธิ์ผู้เรียน (Achievement Test) การใช้การสังเกตพฤติกรรมการปฏิบัติงานโดยตรง โดยใช้แบบประเมินพฤติกรรม เป็นต้น (นภา หลิมรัตน์ และเชิดศักดิ์ ไอรณรัตน์, มปป.) ถึงจะมีการพัฒนาวิธีการอื่น ๆ มาใช้ประเมินผลการเรียนรู้ตามที่กล่าว แต่การวัดและประเมินผลด้วยแบบทดสอบผลสัมฤทธิ์ผู้เรียนที่ใช้กันมาตั้งแต่อดีตก็ยังนิยมใช้กันอยู่ในปัจจุบัน การพยาบาลศาสตร์ศึกษาก็เช่นเดียวกัน ดังที่กล่าวมาว่า ปรัชญาทางการศึกษาด้วยลักษณะของวิชาชีพ เมื่อวิเคราะห์เปรียบเทียบกับจัดการศึกษาของไทยแล้ว จะพบว่ามีลักษณะของปรัชญาการศึกษาแบบสารัตถนิยม (Essentialism) เป็นหลัก และการจัดการเรียนรู้มีส่วนของปรัชญาทางจิตวิทยา กลุ่มปัญญานิยม (Cognitivism) และปรัชญาอื่น ๆ เป็นส่วนประกอบซึ่งมีผู้วิพากษ์ไว้หลายท่าน (พระมหาทองมา อรุณรสี (แสงอรุณ), 2553, น.4, 20-25 ; ไพฑูรย์ สีนลารัตน์ และคณะ, 2548, น.20-30) เน้นการรู้คิดที่มีการจัดระดับความสามารถทางการคิดตาม แนวคิดของ เบนจามิน บลูม (Benjamin Bloom, 1956) การวัดผลแบบอื่น ๆ ก็เป็นไปตามหลักปรัชญาที่มาผสม โดยกำหนดน้ำหนักการวัดประเมินผลด้วยแบบทดสอบผลสัมฤทธิ์ผู้เรียนให้เป็นสัดส่วนเท่าใดของการประเมินแบบอื่น ๆ รวมกันทั้งหมด ความเหมาะสมขึ้นกับวัตถุประสงค์ของการเรียนรู้ในครั้งนั้น ๆ เน้นการวัดทักษะประเภทใด เป็นภาคทฤษฎีหรือปฏิบัติ และบริบทอื่น ๆ เช่น จำนวนผู้เรียน ศักยภาพของผู้เรียนและทีมผู้สอน ระยะเวลาที่กำหนดในวิชานั้น ๆ เป็นต้น แต่จากการออกแบบการเรียนรู้ วิชากลุ่มพื้นฐานวิชาชีพ และวิชากลุ่มวิชาชีพ ในภาคทฤษฎี ที่ผ่านมาจะให้น้ำหนักจากการสอบโดยใช้แบบทดสอบวัดผลสัมฤทธิ์ของผู้เรียนเป็นสัดส่วนสูงสุด ในวิชาที่กล่าวมาเป็นส่วนใหญ่

แนวคิดเกี่ยวกับแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน (Concepts of Achievement Tests)

การวัดประเมินผลการเรียนรู้ด้วยแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน เป็นเครื่องมือที่ใช้วัดผลลัพธ์การเรียนรู้ในมิติด้านพุทธิพิสัย (cognitive domain) หรือเป็นการวัดทางด้านสติปัญญา (intellectual) เป็นหลัก การพัฒนาทางด้านสติปัญญา หรือ ความสามารถทางการคิด มีฐานมาจากแนวคิดของเบนจามิน บลูม (Benjamin Bloom, 1956) ซึ่งแบ่งพัฒนาสติปัญญาของมนุษย์ หรือระดับของความคิด จากระดับล่าง ไปสู่ระดับสูง เป็น 6 ชั้น ได้แก่ 1) ด้าน ความรู้ (knowledge) 2) ความเข้าใจที่ครอบคลุม (comprehension) 3) การนำไปใช้ (application) 4) วิเคราะห์ แยกแยะ (analysis) 5) การสังเคราะห์ (synthesis) 6) การประเมินค่า (evaluation) ต่อมา ได้ปรับปรุง การแบ่งระดับแนวคิดใหม่โดยลูกศิษย์ของบลูม กลุ่มหนึ่ง ได้แก่ ลอริน แอนเดอร์สัน และคณะ (Lorin Anderson and et al, 2000) ได้แก่ 1) ความสามารถจำได้ (remembering) 2) ความเข้าใจในขอบเขตที่ขยายขึ้น (understanding) 3) การใช้ความรู้ (applying) 4) ความสามารถวิเคราะห์ (analyzing) 5) ความสามารถประเมินค่า (evaluating) 6) ความสามารถคิดสร้างสรรค์ (creating)

คำที่ใช้เปลี่ยนจากคำนามไปเป็นรูปของคำกริยา และจัดลำดับการสังเคราะห์ขึ้นเป็นระดับสติปัญญาชั้นสูงสุด ใช้คำใหม่เป็นการสร้างสรรค์แทน ดังตารางที่ 2.2

ตารางที่ 2.2 ความแตกต่างระหว่างโครงสร้างระดับการคิด (Cognitive Domain) ของบลูม (Bloom) เดิม และโครงสร้างปรับปรุงใหม่

โครงสร้างเดิมของบลูม (Original Domain)	โครงสร้างที่ปรับปรุงใหม่ (New Domain)
การประเมิน (evaluation)	ความสามารถคิดสร้างสรรค์ (creating)
การสังเคราะห์ (synthesis)	ความสามารถประเมิน (evaluating)
การวิเคราะห์ (analysis)	ความสามารถวิเคราะห์ (analyzing)
การนำไปใช้ (application)	ความสามารถใช้ความรู้ (applying)
ความเข้าใจ (comprehension) ซึ่งหมายถึง การสร้างความหมาย (construction of meaning)	ความเข้าใจ (understanding) :ซึ่งหมายถึง เป็นการนำความรู้เดิมที่ได้รับนำมาสร้างใหม่ขยายไปกว่าเดิม
ความรู้ (knowledge)	ความสามารถจำ (remembering)

ที่มา; Bloom's Revised Taxonomy

www.utar.edu.my/fegt/file/Revised_Blooms_Info.pdf

ชวาล แพร่สกุล, (ม.ป.ป.) อธิบายความหมายของคำที่ใช้ในโครงสร้างการวัดระดับการรู้คิด (Cognitive Domain) และตัวอย่างการเขียนคำถามเพื่อใช้วัดความคิดระดับต่างๆ จากระดับต่ำไปสูง ดังนี้

1. การจำ (remembering) ในการประเมินผลขั้นการจำ จะต้องประกอบไปด้วยการทำงานของจิต และสมอง 3 เรื่อง ได้แก่ ต้องได้รับความรู้ (knowledge) อาจจะใช้วิธีการต่างๆ เช่น ฟัง พูด อ่าน เขียน ลงมือกระทำ เป็นต้น ที่เรียกว่าเป็นประสบการณ์เรียนรู้ เกิดการจำ (memory) คือ การเก็บสิ่งที่ได้รู้ไว้ อาจจะเป็นชั่วคราว หรือถาวร และการระลึก (recall) คือความสามารถที่จะเรียกเอา ความรู้ที่เก็บไว้กลับคืนมาให้ปรากฏตัวอย่างคำถามเพื่อใช้ประเมินความคิดในระดับนี้ เช่น ให้บอกชื่อ คำศัพท์ นิยามต่างๆ ให้จับคู่ ให้เติมคำ

2. ความเข้าใจ (understanding) เป็นความสามารถในการผสมแล้วขยายความรู้ความจำให้ไกลออกไปจากเดิมอย่างสมเหตุสมผล ความเข้าใจเป็นสมรรถภาพขั้นต่อยอดจากความจำ เป็นการเสริมแต่งความรู้เดิมให้มีความหมายลึกกว่าขั้นจำได้ เป็นการเชื่อมโยงสิ่งที่รับเข้ามาใหม่ ให้ผสมผสานกับประสบการณ์เดิม ซึ่งต้องเกี่ยวข้องกับกระบวนการภายใน ดังนี้

- 2.1 ผู้เรียนต้องรู้ความหมาย และรายละเอียดปลีกย่อยต่างๆ ของเรื่องนั้นมาก่อนแล้ว
- 2.2 รู้ความเกี่ยวข้องสัมพันธ์ ระหว่างชั้นความรู้ย่อยๆ เหล่านั้น

2.3 ผู้เรียนผสมผสานความรู้ใหม่เข้ากับความรู้เดิมที่เก็บรักษาไว้ เชื่อมโยงจนเกิดเป็นความเข้าใจและเกิดมุมมองใหม่ที่ต่างจากความรู้ที่เก็บไว้เดิม จนสามารถอธิบายสิ่งนั้นด้วยภาษาของตนเอง และสามารถชี้แจงความรู้นั้นให้ผู้อื่นรู้ตาม ด้วยสำนวน ภาษาใหม่ของตนเองได้ ไม่ใช่พูดตามแบบครู หรือแบบอักษรในตำราที่เคยเรียนมา

2.4 เมื่อพบสิ่งอื่นใดที่มีสภาพทำนองเดียวกันกับที่เคยเรียนรู้มาแล้วสามารถตอบและอธิบายได้

2.5 สามารถฟังเรื่องเล่า อ่านข้อความที่เป็นตัวอักษร แล้วเกิดจินตภาพในใจ แล้วถ่ายทอดได้ การวัดพฤติกรรมการความเข้าใจ ด้วยการให้ผู้เรียนอธิบายความ แปลความหมาย ขยายความ ตีความ สรุปประเด็นและการคาดคะเน เหตุการณ์ ปัญญาประเภทนี้ และอื่นๆในขั้นสูง **หากวัดโดยนำเรื่องราว สารที่ครูเคย บอกมาถามกันโดยตรงไม่ได้ จะเป็นการวัดขั้นจำไปในทันที** ตัวปัญญาเกิดขึ้นโดยตัวของนักเรียนเองซึ่งต้องนำความรู้ต่างๆมาผสมผสานกัน แล้วขบคิด โดยมีครู ชี้นำให้คิด ให้ทำ เพราะความรู้ขั้นนี้ จะยึดยึด หรือหยิบยื่นให้แก่กัน โดยตรงเหมือนขั้นจำความรู้ไม่ได้ หากในการเรียนการสอน ผู้เรียนไม่ทำตาม ไม่คิด ไม่ทำด้วยตนเอง ก็ไม่เกิดปัญญา เช่น วิธีการเรียนที่ผิด เช่น การลอกการบ้าน การทำรายงานด้วยการลอกตัดปะ ไม่วิเคราะห์ สร้างสรรค์ด้วยตนเอง ตัวอย่างคำถามเพื่อใช้ประเมินความคิดในระดับนี้ เช่น แปลความหมายของคำและข้อความยกตัวอย่างที่แปลกใหม่กว่าที่ครูสอน ให้เปรียบเทียบ แปลความหมายของสัญลักษณ์ ตีความหมายของเรื่อง ของข้อเท็จจริง หากคำเปรียบเทียบกับเหตุการณ์ด้วยการหาคำ อุปมา อุปมัย เป็นต้น

3. การใช้ความรู้ (applying) เป็นความสามารถในการนำความรู้ ความเข้าใจในสถานการณ์เรื่องราวข้อเท็จจริง ไปแก้ปัญหาคำถามใหม่โดยการปฏิบัติหรือกระทำอย่างมีขั้นตอน ในสถานการณ์จริงหรือสถานการณ์ใหม่ที่คล้ายคลึงกัน หรือเป็นการนำความรู้ไปพลิกแพลงใช้ ตัวอย่าง ได้แก่ เมื่อผู้เรียนเรียนวิชากลศาสตร์เรื่องแรงและทิศทาง สามารถนำไปใช้เวลาโหม่งลูกฟุตบอลให้เข้าประตูได้ เรียกว่านำความรู้ไปใช้ ต่างกับ การที่ผู้สอนอธิบายวิธีเล่นฟุตบอลแล้วให้ผู้เรียนได้ทดลองปฏิบัติตามนั้น โดยการเล่นฟุตบอลจริง ไม่ถือว่า เป็นการนำเอาความรู้ไปใช้ เป็นแต่เพียงขั้น จำ และเข้าใจเท่านั้น ตัวอย่างคำถามเพื่อใช้ประเมินความคิดในระดับนี้ เช่น การแก้ปัญหาโจทย์ทางคณิตศาสตร์ที่ต่างจากที่ผู้สอนเคยฝึกให้ทำ ฝึกแก้ปัญหาคำถามใหม่ที่ต่างจากเดิม การให้อธิบายเหตุผลของการปฏิบัติ

4. การวิเคราะห์ (analyzing) เป็นความสามารถในการแยกสิ่งสำเร็จรูปตามหลักการและกฎเกณฑ์ที่กำหนดให้ เพื่อค้นคว้าหาความจริงต่างๆที่แฝงอยู่ภายในเรื่องราวนั้น สิ่งสำเร็จรูปนั้นอาจเป็นเรื่องราว บทความ รูปภาพ หรือองค์ประกอบของสสาร ในการแยกนี้ต้องอาศัยกฎเกณฑ์ต่างๆ เพื่อเป็นการค้นหาความจริงในเรื่องนั้นๆ โดยให้ทะลุผิวภายนอกที่อำพรางไว้ เพื่อลงไปสู่แก่นของเรื่องนั้นๆ ตัวอย่างคำถาม เพื่อใช้ประเมินความคิดในระดับนี้ เช่น คำถามวิเคราะห์ความสัมพันธ์ จุดเด่นจุดด้อย เหมาะสม ไม่เหมาะสม วิเคราะห์วัตถุประสงค์กับผลลัพธ์ เป็นต้น

5. การประเมินค่า (evaluating) เป็นการตีราคาสิ่งต่างๆโดยสรุปอย่างมีหลักเกณฑ์ ว่าสิ่งนั้น ดี เลว เหมาะควรอย่างไร เป็นความสามารถความสามารถในการวินิจฉัย สรุปผล หรือ ตัดสินคุณค่าของเรื่องราวความคิด การกระทำ และเหตุการณ์

6. การสร้างสรรค์ (creating) เป็นความสามารถในการรวบรวมสิ่งต่างๆตั้งสองชนิดเข้าด้วยกันเพื่อให้กลายเป็นสิ่งสำเร็จรูปชิ้นใหม่ มีคุณลักษณะแตกต่างไปจากคุณสมบัติของส่วนประกอบ

ย่อย การสร้างสรรค์ใช้ความสามารถที่สูงกว่าขั้นที่กล่าวมาแล้วทั้งหมด ต้องใช้ความรู้ และความสามารถหลายอย่างมาประกอบกัน ในทางความคิด อาจเป็นการเขียนบทความวิจารณ์ การแสดงความคิดเห็นในทัศนะของตนเอง การแต่งโคลง กลอน การเขียน บทความ ในทางปฏิบัติ เป็นการทำชิ้นงานประดิษฐ์ เช่น การออกแบบผลิตภัณฑ์”

สรุป ความสำคัญเรื่องโครงสร้างการวัดระดับการรู้คิด (Cognitive Domain) ของบลูม จะถูกนำมาใช้สร้างผังข้อสอบ (test blue print) และสร้างแบบทดสอบ (test) ให้สอดคล้องกับผังข้อสอบที่กำหนดไว้ โดยการเขียนข้อสอบให้เหมาะสมสอดคล้องกับการวัดในแต่ละระดับนั้นๆ เพื่อวัดผลการเรียนรู้ และประเมินผลว่าผู้เรียนมีความสามารถเรียนรู้ได้ถึงระดับใด

ประเภทของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน (Types of Achievement Test)

การแบ่งประเภทของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน สามารถแบ่งได้หลายลักษณะ ขึ้นกับจุดประสงค์ของการแบ่ง (ทิวต์ล มณีโชติ, 2549) ได้แก่

1. แบ่งตามกระบวนการออกข้อสอบ

1.1 แบบทดสอบที่ครูสร้างขึ้น (teacher made test) แบบทดสอบผลสัมฤทธิ์ที่ครูสร้างขึ้น โดยมีจุดประสงค์เพื่อวัดผลผลิตผลของการศึกษาโดยตรง ได้แก่ วัดผลการเรียนรู้ด้านเนื้อหาวิชาและทักษะต่างๆของแต่ละสาขาวิชา แบบวัดเพื่อจุดประสงค์นี้ มีอยู่ 2 ลักษณะ (เยาวดี ราชชัยกุล วิบูลย์ศรี, 2554, น.16) ได้แก่

1.1.1 ลักษณะแบบสอบผลสัมฤทธิ์ที่เป็นข้อเขียน (Paper - Pencil Tests)

1.1.2 ลักษณะแบบสอบผลสัมฤทธิ์ที่เป็นภาคปฏิบัติจริง (Performance Test)

1.2 แบบทดสอบมาตรฐาน (standardized test) เป็นข้อสอบที่มุ่งวัดผลสัมฤทธิ์ของผู้เรียน และได้มีการพัฒนาด้วยการวิเคราะห์ทางสถิติมาแล้วหลายครั้งจนมีคุณภาพสมบูรณ์ อีกทั้งยังมีเกณฑ์ปกติ (norm) ไว้สำหรับเปรียบเทียบคุณภาพต่าง ๆ ของนักเรียนต่างกลุ่มกันได้อีกด้วย (พิศเพลิน เขียวหวาน และสุจิตรา หังสพฤกษ์, 2552, น.174)

2. แบ่งตามลักษณะของแบบทดสอบ (ทิวต์ล มณีโชติ, 2549)

2.1 แบบทดสอบชนิดปรนัย ประกอบด้วย ข้อสอบแบบถูกผิด จับคู่ เติมคำหรือแบบตอบสั้น และแบบเลือกตอบ

2.2 แบบทดสอบชนิดอัตนัย แบบทดสอบอัตนัยเป็นแบบทดสอบที่ให้ผู้ตอบได้แสดงความคิดเห็นจึงเหมาะสำหรับวัดความรู้ขั้นสูงกว่าความจำและความเข้าใจ แบบทดสอบแบบอัตนัยแบ่งได้ 2 ลักษณะ คือ แบบจำกัดคำตอบ (ให้ผู้เรียนตอบตามประเด็นที่ระบุไว้) และแบบขยายความ (ให้ผู้เรียนแสดงความคิดเห็นอย่างเสรี)

จากที่กล่าว แบบทดสอบที่ใช้วัดทางด้านความรู้ เป็นการวัดทางสติปัญญา (Cognitive Domain) ที่มีการใช้ปัญญาระดับต่างๆ ได้แก่ ความจำ ความเข้าใจ การนำไปใช้ วิเคราะห์ ประเมินค่า และการสร้างสรรค์ ด้านต่างๆ ในการออกข้อสอบผู้ออกข้อสอบจะสร้างตารางวิเคราะห์หลักสูตร หรือ ตารางวิเคราะห์ข้อสอบ (Test Blue print) ซึ่งประกอบด้วยจุดประสงค์การเรียนรู้หรือเนื้อหา กับระดับปัญญาที่จะวัดโดยผู้ออกข้อสอบต้องวิเคราะห์ว่าในแต่ละจุดประสงค์การเรียนรู้หรือเนื้อหาต่าง ๆ นั้นจะวัดพฤติกรรมที่แสดงออกซึ่งสะท้อนระดับปัญญาใดบ้าง พฤติกรรมละก็ข้อ

ชนิดของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนแบบปรนัย (ทิวต์ล มณีโชติ, 2549)

1. แบบทดสอบแบบถูกผิด (true -false) แบบทดสอบประเภทนี้เหมาะสำหรับใช้วัดความจำที่เกี่ยวกับข้อเท็จจริงต่างๆ เช่น สถานที่ บุคคล เวลา เป็นต้น

2. แบบทดสอบแบบจับคู่ (matching) แบบทดสอบประเภทนี้เหมาะสำหรับการวัดความรู้ที่เป็นข้อเท็จจริง และความสัมพันธ์ระหว่างสิ่งต่างๆ โดยมีข้อความ 2 ส่วน คือ ส่วนที่เป็นคำถามและส่วนที่เป็นคำตอบ

3. แบบทดสอบแบบเติมคำหรือเติมความ (completion) เป็นแบบทดสอบที่ต้องการให้ผู้สอบเติมคำหรือข้อความสั้นๆ ข้อสอบประเภทนี้เหมาะสำหรับใช้วัดข้อเท็จจริง ข้อมูล และข้อสนเทศต่างๆ

4. แบบทดสอบแบบเลือกตอบ (multiple choice) เป็นแบบทดสอบที่นิยมใช้กันมากสำหรับแบบทดสอบแบบปรนัยเพราะสามารถวัดความรู้ได้ตั้งแต่ความจำ ความเข้าใจ การนำไปใช้ การวิเคราะห์ และการประเมินค่า

ลักษณะของเครื่องมือวัดที่ดี

เมื่อผู้สอนเลือกสร้างเครื่องมือที่มีความเหมาะสมแล้ว จะต้องหาคุณภาพของเครื่องมือ ซึ่งลักษณะของเครื่องมือที่มีคุณภาพต้องมีลักษณะดังนี้

1. เครื่องมือที่ดีต้องมีความตรง (Validity) หมายถึง ความสามารถวัดในสิ่งที่ต้องการจะวัด ความเที่ยงตรงที่สำคัญ แบ่งเป็น 3 แบบ

- 1.1 ความตรงตามเนื้อหา
- 1.2 ความตรงตามเกณฑ์ที่เกี่ยวข้อง
- 1.3 ความตรงตามโครงสร้าง

ความตรงตามเนื้อหา (content validity)

เนื่องจาก ในการศึกษาครั้งนี้เป็นการศึกษาคุณภาพของข้อสอบที่อาจารย์ผู้สอนสร้างขึ้น เพื่อวัดผลสัมฤทธิ์ของนักศึกษาพยาบาล จึงขอกล่าวถึงเฉพาะการหาความตรงตามเนื้อหาที่สอดคล้องกับที่ทางคณะพยาบาลศาสตร์ปฏิบัติอยู่ ซึ่งตามที่ บุญเชิด ภิญโญอนันตพงษ์ (2552, น.90, 94) ให้ความหมายไว้ว่า ความตรงตามเนื้อหา หมายถึง การลงความเห็นเกี่ยวกับความสอดคล้องระหว่างเนื้อหาของเครื่องมือวัดกับเนื้อหาของเกณฑ์ ถ้ามีความสอดคล้องมาก แสดงว่ามีหลักฐานแสดงความตรงตามเนื้อหามาก เพราะเนื้อหาของเครื่องมือวัดครอบคลุมเนื้อหาที่เป็นตัวเกณฑ์นั่นเอง ในการวิเคราะห์เพื่อลงความเห็นถึงความสอดคล้องระหว่างเนื้อหาของเครื่องมือกับเนื้อหาของเกณฑ์ จะใช้ดุลยพินิจของผู้เชี่ยวชาญตรวจสอบความตรง โดยการเทียบกับ ตารางกำหนดข้อสอบ (Test Blueprint) ซึ่งกำหนดตัวอย่างหัวข้อเนื้อหาวิชา และพฤติกรรมจากเนื้อหาวิชาทั้งหมด และถือว่าเป็นตัวแทนที่ดีแล้ว โดยที่หลักในการเทียบเคียงข้อสอบกับตารางวิเคราะห์ (บุญเชิด ภิญโญอนันตพงษ์ (2552, น.94-95) มีเกณฑ์ดังนี้

1. ข้อคำถามแต่ละข้อวัดเนื้อหาสอดคล้องกับตารางวิเคราะห์หลักสูตร หรือไม่
2. ข้อคำถามแต่ละข้อเป็นตัวแทนที่ดีของเนื้อหาวิชาหรือไม่
3. ข้อคำถามแต่ละข้อมีตัวแปรอื่นที่จะทำให้เกิดความคลาดเคลื่อนหรือไม่ เช่น คำถามกำกวม ใช้ศัพท์ยาก ความเหมาะสมกับวัย เป็นต้น
4. เครื่องมือวัดนั้นมีจำนวนข้อเหมาะสมเพียงพอที่จะเป็นตัวแทนของเนื้อหาวิชาหรือไม่

ขั้นตอนการหาความตรงตามเนื้อหาของข้อสอบ (content validity) ในการศึกษาวิจัยครั้งนี้

1. การจัดทำตารางกำหนดข้อสอบ (Test Blueprint) ที่มีมาตรฐาน

โดยหลักปฏิบัติตามมาตรฐานการจัดการเรียนการสอนของคณะพยาบาลศาสตร์ มหาวิทยาลัยราชภัฏสวนดุสิต ทุกวิชาต้องส่ง มคอ.3 และ มคอ. 4 เพื่อให้คณะกรรมการบริหารหลักสูตรที่ได้รับการแต่งตั้งจากผู้บริหารคณะพยาบาลศาสตร์ ตรวจสอบ และมีการแก้ไขปรับปรุง จนเชื่อได้ว่าการออกแบบจัดการเรียนการสอนมีมาตรฐาน เหมาะสมกับผู้เรียน ก่อนการดำเนินการจัดการเรียนการสอน โดยในแบบ มคอ.3 (ภาคทฤษฎี) ยังประกอบด้วยโครงสร้างตารางกำหนดข้อสอบ (Test Blueprint) ให้คณะกรรมการพิจารณาด้วย

หมายเหตุ (มคอ.ย่อมาจาก กรอบมาตรฐานคุณวุฒิระดับอุดมศึกษาแห่งชาติ” หรือภาษาอังกฤษใช้คำว่า TQF: HEd ย่อมาจาก. “Thai Qualifications Framework for Higher. Education”)

2. ข้อสอบที่อาจารย์สร้างขึ้นได้รับการตรวจสอบจากกรรมการวิพากษ์ข้อสอบ

ทุกครั้งที่อาจารย์ผู้สอนสร้างข้อสอบเสร็จทั้งฉบับ จะต้องนำไปให้กรรมการวิพากษ์ข้อสอบ ร่วมกันพิจารณา โดยองค์ประกอบของคณะกรรมการวิพากษ์ข้อสอบ มีอย่างน้อย 3 คน (โดยผู้เขียนข้อสอบเป็น 1 ในคณะกรรมการวิพากษ์ข้อสอบ) พิจารณาตรวจสอบว่าข้อสอบถูกต้องครบถ้วนตามโครงสร้างตารางกำหนดข้อสอบ และแก้ไขปรับปรุงจนเชื่อได้ว่าข้อสอบที่อาจารย์สร้างขึ้นถูกต้องสมบูรณ์ตามหลักเกณฑ์ของมาตรฐานที่กำหนด

รายละเอียดของโครงสร้างตารางกำหนดข้อสอบ รายวิชากฎหมายและจริยธรรมวิชาชีพอยู่ในภาคผนวก หน้า 66 - 73

3. เครื่องมือที่ดีต้องมีความเชื่อมั่น (Reliability) คือ เครื่องมือที่ใช้ประเมินผลหลาย ๆ ครั้ง ได้ผลคงเดิมหรือใกล้เคียง

4. เครื่องมือที่ดีต้องมีความเป็นปรนัย (Objectivity) พิจารณาจาก

- 4.1 คำถามชัดเจน ผู้เข้ารับการอบรมเข้าใจความหมายของคำถามตรงกัน
- 4.2 ภาษาที่ใช้รัดกุม
- 4.3 คำถามที่ใช้กำหนดเงื่อนไข มีความรัดกุม
- 4.4 คำถามอาจคิดในแง่มุมอื่นใดหรือไม่
- 4.5 ตัวเลือกในแบบเลือกตอบชัดเจนหรือไม่
- 4.6 ตัวเลือกของข้อมูลที่ใช้กับคำถามเป็นเรื่องเดียวกันหรือไม่
- 4.7 เกณฑ์ให้คะแนนชัดเจนหรือไม่

5. เครื่องมือที่ดีต้องมีอำนาจจำแนก (Discrimination) สามารถชี้เฉพาะให้เห็นลักษณะความแตกต่าง หรือความเหมือนของสิ่งที่ต้องการประเมิน

6. ความยากง่าย (difficulty)

7. เครื่องมือที่ดีต้องมีประสิทธิภาพ (Efficiency) ในการนำไปใช้ (บุญเชิด ภิญโญอนันตพงษ์, 2552, น.69; บุญศรี พรหมมาพันธ์ และนวลเสนห์ วงศ์เชิดธรรม, 2552, น.219; ทิวต์ล มณีโชติ, 2549)

การวิจัยครั้งนี้ เป็นการศึกษาเกี่ยวกับแบบทดสอบที่ครูสร้างขึ้น (teacher made test) เป็นแบบทดสอบชนิดเลือกตอบ (multiple choice question) 4 ตัวเลือก ดังนั้น สารระสำคัญจะขอนำเสนอรายละเอียดที่เกี่ยวข้องกับแบบทดสอบที่ครูสร้างขึ้นชนิดเลือกตอบเป็นหลัก

แบบทดสอบชนิดเลือกตอบ (Multiple choice Question, MCQ)

แบบทดสอบชนิดเลือกตอบ (multiple choice question หรือ MCQ) เป็นเครื่องมือวัดผล ซึ่งเป็นที่รู้จักและใช้กันอย่างแพร่หลายเหมาะสำหรับการประเมินด้านพุทธิพิสัย (cognitive domain) ของผู้เรียนมีความเป็นปรนัย (objectivity) สูง กล่าวคือ สามารถให้คะแนนได้ใกล้เคียงหรือเหมือนกัน แม้ว่าผู้ประเมินจะต่างคนกันก็ตาม นอกจากนี้ยังสามารถบริหารจัดการสอบได้ง่าย เหมาะกับผู้เรียนจำนวนมาก แต่อย่างไรก็ตามยังมีปัญหาและข้อจำกัดในการสร้างข้อสอบทำให้ได้ข้อสอบที่ไม่ดี ส่งผลให้ไม่สามารถประเมินการเรียนรู้ของผู้เรียนได้อย่างแท้จริง การสร้างข้อสอบที่ดีจะต้องอาศัยทั้งความรู้ความเข้าใจเกี่ยวกับข้อสอบแบบเลือกตอบและประสบการณ์ของผู้ประเมิน จึงจำเป็นที่ผู้ประเมินควรใช้เวลาและความสำคัญกับการสร้างข้อสอบเพื่อจะได้ข้อสอบที่มีคุณภาพ วัดได้ตรงตามสิ่งที่ควรจะวัด และตรงตามวัตถุประสงค์ของเครื่องมือ (อานุกาพ เลขะกุล, 2556, น.2)

คุณลักษณะ จุดเด่น จุดด้อย ของแบบทดสอบชนิดเลือกตอบ มีดังนี้

1. เป็นเครื่องมือชนิดเลือกตอบ (selected response) โดยมีตัวเลือก 3-5 ตัวให้ผู้สอบเลือกตอบคำตอบที่ถูกต้องเพียงคำตอบเดียว

2. ใช้วัดด้านความรู้ ความคิดเป็นหลัก (cognitive) ตามระดับความสามารถ (ความจำ ความเข้าใจ การประยุกต์ใช้ การวิเคราะห์ การประเมินค่า การสร้างสรรค์) ข้อสอบประเภทนี้ สามารถประเมินเนื้อหาได้ครอบคลุมในระยะเวลาที่เหมาะสม วัดได้หลายระดับ แต่การวัดความคิดระดับสร้างสรรค์ MCQ สร้างไปถึงได้ยาก กว่าข้อสอบแบบอัตนัย หรือการวัดทักษะปฏิบัติจากการทำชิ้นงาน

3. มีความตรงเชิงเนื้อหา (content validity) สามารถสร้างให้ครอบคลุมเนื้อหาได้ดี โดยสร้างข้อสอบตามผังข้อสอบ (test blue print)

4. มีความเป็นปรนัยสูง สามารถตรวจให้คะแนนได้เหมือนกัน แม้ว่าผู้ประเมินจะต่างคนกัน (inter-rater reliability) เพราะมีการให้คำตอบที่ถูกต้องก่อนการสอบ การให้คะแนนมีความเที่ยงหรือเชื่อถือได้ (reliability) ลดปัญหาการให้คะแนนระหว่างผู้ประเมิน

5. สามารถจัดสอบได้ง่าย ไม่ยุ่งยาก ใช้กับผู้เรียนจำนวนมากได้ดี

6. สามารถตรวจข้อสอบและวิเคราะห์ข้อสอบได้ง่าย ปัจจุบันสามารถใช้ตรวจและวิเคราะห์ได้ด้วยระบบคอมพิวเตอร์

7. ไม่ถูกรบกวนจากทักษะการเขียนของผู้สอบที่ต้องใช้ทักษะการใช้ภาษาซึ่งจะเป็นตัวแปรแทรกซ้อน ทำให้ผลการวัดได้ไม่ตรงตามวัตถุประสงค์รายวิชา

8. สามารถพัฒนาคุณภาพข้อสอบให้เป็นข้อสอบมาตรฐานได้ (standard test) โดยเก็บข้อสอบที่มีคุณภาพเข้าคลังข้อสอบ ลดปัญหาการจัดหาข้อสอบในอนาคตได้ (อานุกาพ เลขะกุล, 2556, น.2)

จุดอ่อนของแบบทดสอบชนิดเลือกตอบ ได้แก่

1. การเขียนข้อสอบเพื่อวัดความสามารถในการจัดระบบความคิด หรือการตัดสินใจด้วยตนเอง ทำได้ยากเนื่องจากมีคำตอบให้เลือก

2. ผู้สอบไม่สามารถแสดงความคิดเห็นหรือเหตุผลในการเลือกตอบ

3. จำกัดความคิดสร้างสรรค์ และการแก้ปัญหาของผู้สอบ ให้คิดเฉพาะในส่วนของข้อสอบ

4. ผู้เขียนข้อสอบ ต้องใช้ศักยภาพความรู้ ความสามารถ ความตั้งใจเอาใจใส่พัฒนาการทักษะการออกข้อสอบซึ่งต้องใช้เวลา

5. ผู้ทำข้อสอบมีโอกาสเดาคำตอบได้ ถ้าข้อสอบไม่ดี ผู้สอบที่ไม่มีความรู้ อาจทำข้อสอบได้ถูกต้อง

6. ผู้ทำข้อสอบอาจจำข้อสอบได้ ดังนั้นข้อสอบต้องมีจำนวนมากพอเก็บไว้ในคลังข้อสอบ (อานุกาพ เลขะกุล, 2556, น.2-3)

โครงสร้างของข้อสอบแบบเลือกตอบ (Structure of multiple choice question)

ข้อสอบแบบเลือกตอบมีส่วนประกอบสำคัญ (อานุกาพ เลขะกุล, 2556, น.3) ได้แก่

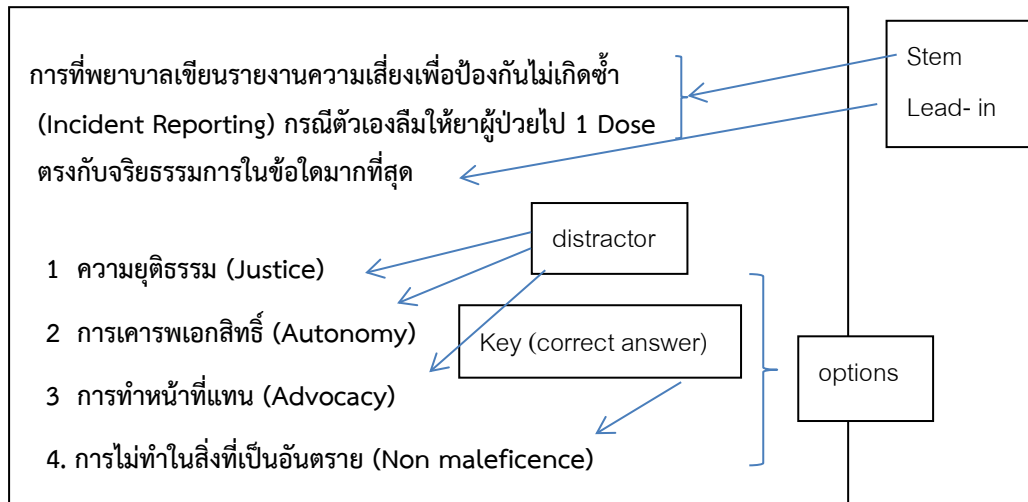
1. โจทย์ (stem) และคำถาม (lead-in) เป็นส่วนที่ให้ข้อมูลเพื่อใช้ในการตอบคำถาม อาจมีรูปภาพ กราฟ หรือแผนภูมิ ประกอบ

2. ตัวเลือก (option หรือ alternative) เป็นคำตอบที่ให้ผู้สอบเลือกตอบ โดยทั่วไป อาจมี 3-5 ตัวเลือก การมีจำนวนตัวเลือกน้อยเกินไป อาจมีโอกาสเดาถูกได้มากขึ้น ตัวเลือก 5 ข้อ จะมีความเที่ยง (reliability) มากที่สุด

3. คำตอบที่ถูกต้อง (correct answer หรือ key) เป็นตัวเลือกที่ถูกต้องที่สุด

4. ตัวลวง (distractor) เป็นตัวเลือกที่เป็นคำตอบที่ไม่ถูกต้อง เพื่อลวงให้ผู้สอบเลือกตอบ การสร้างตัวลวงที่ดี จะทำให้ความเที่ยงของข้อสอบข้อนั้นดีขึ้น

ตัวอย่างข้อสอบแบบเลือกตอบ



แผนภูมิที่ 2.4 ตัวอย่างของโครงสร้างข้อสอบแบบเลือกตอบ (multiple choice question)
ที่มา: ดัดแปลงจาก อานุกาพ เลขะกุล, 2556

แนวทางการสร้างโจทย์และคำถามทางการพยาบาล แบบเลือกตอบ ดัดแปลงโดยสรุปจาก
หลักการเขียนข้อสอบแบบเลือกตอบของ อานุกาพ เลขะกุล (2556, น.7-20)

1. เนื้อหาของโจทย์และคำถามควรสอดคล้องกับวัตถุประสงค์และตารางข้อสอบที่กำหนดไว้
2. ระดับความยากง่ายเหมาะสมกับผู้สอบ ตามระดับชั้นปี และเนื้อหาไม่สูงกว่าพื้นฐานความรู้ที่ผู้สอบได้เรียนรู้ผ่านมาแล้ว
3. โจทย์ควรจะกระชับ รัดกุม ไม่เยิ่นเย้อ เข้าใจได้ชัดเจน แต่มีข้อมูลเพียงพอสำหรับการตอบคำถามได้
4. ควรเน้นความรู้ที่นำไปใช้ (working knowledge) มากกว่าความรู้แบบท่องจำ (factual knowledge) คือเป็นการวัดในระดับ สูงกว่า ความจำ ได้แก่ ระดับความเข้าใจขึ้นไป
5. การสร้างโจทย์ปัญหาในคลินิก (case based) หากเป็นวิชาที่สอนก่อนนักศึกษาขึ้นฝึกปฏิบัติงานในคลินิกไม่ควรสลับซับซ้อน และควรโยงไปสู่ความรู้พื้นฐาน มากกว่าการประยุกต์ใช้ในการดูแลผู้ป่วย เพราะนักศึกษายังไม่เกิดประสบการณ์ในคลินิก โลกทัศน์ความลุ่มลึกเกี่ยวกับความเข้าใจที่เกิดจากการลงมือทำเหตุการณ์ และการดูแลผู้ป่วยยังไม่มี หากเป็นวิชาที่ผู้สอบผ่านการเรียนรู้ระดับคลินิก มาแล้ว โจทย์อาจซับซ้อนมากขึ้นและอาศัยทักษะการปฏิบัติตามกระบวนการพยาบาลมาออกได้ เนื้อหาโจทย์ที่นำมาออก ควรเป็นสาระสำคัญ เป็นปัญหาที่พบบ่อยในผู้ป่วย
6. Lead-in ควรอยู่ในรูปแบบของประโยคคำถาม เช่น อาการที่เกิดขึ้นในผู้ป่วยรายนี้เกิดจากยาตัวใดมากที่สุด การรักษาเบื้องต้นที่เหมาะสมที่สุดสำหรับผู้ป่วยรายนี้คือข้อใด พฤติกรรมของพยาบาลดังกล่าว สอดคล้องกับหลักจริยศาสตร์เรื่องใด
7. โจทย์ควรคล้ายคลึงกับสถานการณ์จริงที่ผู้สอบจะต้องประสบในการนำไปใช้ หรือในคลินิก

8. ข้อมูลที่ให้ควรเป็นข้อมูลดิบ หลีกเลี่ยงคำหรือข้อความที่ซ้ำเฉพาะ ซึ่งจะทำให้ผู้สอบตอบโดยใช้ความจำ (pattern recognition) มากกว่าการคิดวิเคราะห์

9. ข้อมูลที่จำเป็นทั้งหมดควรปรากฏอยู่ในโจทย์ ไม่ควรนำไปเพิ่มเติมในตัวเลือก มิฉะนั้นจะเหมือนสร้างโจทย์ใหม่ขึ้นมาในข้อสอบหลายข้อ ในข้อสอบข้อเดียว

10. โจทย์และคำถามต้องชัดเจน เมื่อผู้สอบอ่านแล้วควรจะสามารถตอบได้เลยโดยไม่ต้องดูว่าคำตอบในตัวเลือกเป็นอย่างไร

11. ผู้ออกข้อสอบควรตระหนักว่า ผู้สอบที่ทำข้อสอบอยู่กำลังใช้กระบวนการทางปัญญาพิจารณาข้อสอบผ่านการทำความเข้าใจกับภาษาเขียน ดังนั้นผู้ออกจำเป็นต้องมีทักษะการใช้ภาษาที่ดีในการสื่อความหมายในภาษาเขียน เขียนได้สละสลวย สั้นกระชับ ทำให้ผู้สอบอ่านโจทย์ และตัวเลือกได้เข้าใจง่ายชัดเจน ไม่สับสน

12. หลีกเลี่ยงโจทย์หรือคำถามที่บอกใบ้ หรือชี้แนะคำตอบซึ่งทำให้ผู้สอบอาจเดาคำตอบที่ถูกต้องได้

13. หลีกเลี่ยงโจทย์และคำถามที่เป็นเชิงลบหรือปฏิเสธ เช่น ไม่ (not), น้อยที่สุด (least), ยกเว้น (except), ไม่ถูกต้อง (incorrect) เป็นต้น เนื่องจากเป็นข้อสอบที่ไม่ตรงวัตถุประสงค์ และผู้สอบอาจตอบได้ว่าข้อใดไม่ถูก โดยที่ไม่รู้ว่าคำตอบที่ถูกคือข้อใด จะพบข้อสอบแบบนี้ได้บ่อยเนื่องจากเขียนข้อสอบง่ายกว่า สร้างคำตอบที่ผิดเพียงคำตอบเดียว

14. หลีกเลี่ยงโจทย์ที่เป็นคำบรรยายรูปหรือลักษณะที่ทำให้ผู้สอบตีความหมายหรือสร้างจินตนาการเองเพราะผู้สอบแต่ละคนอาจมีการรับรู้และตีความแตกต่างกัน หากต้องการให้แปลผลอาจใช้รูปภาพประกอบแทน

15. หลีกเลี่ยงการใช้ภาษาอังกฤษ เพราะจะทำให้การวัดไม่ตรงกับการพฤติกรรมการเรียนรู้ที่เป็นวัตถุประสงค์รายวิชา ยกเว้นเป็นคำศัพท์ทางเทคนิคที่ผู้เรียนจำเป็นต้องทราบ มิฉะนั้นจะเป็นการวัดผลที่มีวัตถุประสงค์ซ้อนกัน หากผู้เรียนทำข้อสอบผิด จะไม่ทราบว่าผู้เรียนไม่รู้ภาษาอังกฤษ หรือไม่รู้อาณาในวิชานั้น ยกเว้นผู้สอบที่เป็นชาวต่างชาติ

แนวทางการสร้างตัวเลือก สรุปรจาก อานุภาพ เลขะกุล (2556, น.10-20)

1. จำนวนตัวเลือก ควรมี 4 หรือ 5 ตัวเลือก ตัวเลือก 5 ข้อ จะมีความเที่ยงดีที่สุด การใช้ตัวเลือกมากกว่า 5 ข้อ ไม่มีความจำเป็นและไม่ทำให้ค่าอำนาจจำแนกสูงขึ้น

2. ต้องแน่ใจว่ามีคำตอบที่ถูกต้องเพียงคำตอบเดียว

3. ตัวเลือกแต่ละข้อควรเป็นอิสระต่อกัน

4. ตัวเลือกแต่ละข้อควรมีความเป็นไปได้และดึงดูดให้ผู้สอบเลือก

5. ความยาวของเนื้อหาของตัวเลือกแต่ละข้อควรใกล้เคียงกัน เนื่องจากคำตอบที่ถูกต้องมักมีแนวโน้มมีความยาวมากกว่าตัวเลือกที่ผิด

6. ความยาวของเนื้อหาของตัวเลือกแต่ละข้อไม่ควรยาวเกินไป ควรเป็นข้อความสั้นๆ และไม่ควรถูกเกินหนึ่งบรรทัด

7. เนื้อหาของตัวเลือกควรเป็นเรื่องในกลุ่มเดียวกัน ประเด็นที่คล้ายกัน เป็นเรื่องเดียวกัน หรือเป็นไปในทิศทางเดียวกัน (homogeneous)

8. หากตัวเลือกเป็นค่าตัวเลข ควรจะคงที่ (consistent) เช่น ค่าตัวเลขเป็นช่วง ก็ต้องเป็นช่วงทุกข้อ ไม่ใช่มีตัวเลือกบางข้อเป็นค่าเดี่ยว และควรเรียงลำดับกันไป เป็นต้น

9. หลีกเลี่ยงตัวเลือกที่ซับซ้อนหรือทำให้ผู้ตอบสับสน

10. หลีกเลี่ยงตัวเลือกที่ขัดแย้งกันเอง หรือตรงกันข้าม (mutually exclusive) เช่น ตัวเลือกหนึ่งเป็น “residual volumn เพิ่มขึ้น” อีกตัวเลือกหนึ่งเป็น “residual volumn ลดลง” เป็นต้น หากตัวคำตอบถูกเป็นข้อใดข้อหนึ่งในนี้ ตัวเลือกจะลดลงโดยปริยาย จะลดลงไป หนึ่งตัวเลือก

11. เนื้อหาในตัวเลือกแต่ละข้อไม่ควรทับซ้อนกัน (overlapping) เช่น คำตอบที่ถูกต้องเป็น “ให้ยาลดความดันโลหิต” ตัวลวงข้อหนึ่งเป็น “ให้ ACE inhibitor” ซึ่งเป็นกลุ่มยาลดความดันโลหิตพวกหนึ่ง เป็นต้น

12. หลีกเลี่ยงการใช้คำ “ทั้งหมด”, “เสมอ”, “ไม่เคย” เป็นต้น ตัวเลือกเหล่านี้มักจะผิด

13. หลีกเลี่ยงตัวเลือกที่เป็นปฏิเสธหรือปฏิเสธซ้อนปฏิเสธ เพราะทำให้ผู้อ่านงงได้

14. หลีกเลี่ยงตัวเลือก “ถูกหมดทุกข้อ” เพราะทำให้เดาได้ง่าย เพราะถ้ามีคำตอบถูก 2 ข้อไม่ทราบตัวเลือกที่สามก็จะเดาข้อถูกหมดทุกข้อหรือทราบตัวเลือกผิด 1 ข้อก็จะไม่เลือกข้อถูกหมดทุกข้อ

15. หลีกเลี่ยงตัวเลือกถูก “ผิดหมดทุกข้อ” เพราะผู้ตอบถูก อาจเลือกเพราะความไม่รู้ ซึ่งเป็นการเดา

16. หลีกเลี่ยงตัวเลือกที่ซ้ำกันหรือเหมือนกันในประโยคที่เป็นตัวเลือกเพราะจะทำให้ประโยคยาวเกินไป ควรจะนำไปเขียนไว้ในประโยคคำถาม หรือแยกเขียนไว้เป็นหัวข้อต่างหาก

17. ตำแหน่งของคำตอบที่ถูกไม่ควรซ้ำกัน หรือไม่ควรมีรูปแบบ ควรกระจายให้มีสัดส่วนที่ใกล้เคียงกัน

18. การจัดลำดับของตัวเลือก อาจทำได้หลายแบบ ได้แก่

18.1 หากตัวเลือกเป็นคำ อาจเรียงตามลำดับตัวอักษร

18.2 หากเป็นค่าตัวเลข ควรเรียงจากน้อยไปมาก

18.3 จัดเรียงตามลำดับเหตุการณ์ที่เกิดขึ้น

18.4 จัดเรียงตามกลุ่มที่มีแนวโน้มเหมือนกัน เช่น ลดลง เพิ่มขึ้น เป็นต้น

18.5 เรียงตามความยาวของข้อความจากสั้นไปยาว จะทำให้ผู้สอบอ่านได้ง่ายขึ้น

19. การสร้างตัวลวง ควรคำนึงถึงประเด็นต่อไปนี้

19.1 ตัวลวงแต่ละข้อควรมีความเป็นไปได้ และดึงดูดให้ผู้สอบเลือก

19.2 อาจใช้ข้อผิดพลาด หรือความเข้าใจผิดที่พบบ่อยของผู้เรียนเป็นตัวลวง

19.3 ใช้คำหรือข้อความที่ผู้ตอบคุ้นเคยแต่ไม่ถูกต้อง

19.4 หลีกเลี่ยงตัวลวงที่ผิดชัดเจนเกินไปหรือไม่มีเหตุผลจนไม่มีใครเลือก

19.5 หลีกเลี่ยงการสร้างตัวลวงที่เกิดจากความซับซ้อนหรือสับสนของคำตอบ

สรุป แบบทดสอบชนิดเลือกตอบเป็นที่นิยมใช้กันแพร่หลาย เพราะมีคุณสมบัติที่ดีหลายประการ ผู้สร้างข้อสอบต้องมีความรู้ ความเข้าใจเกี่ยวกับองค์ความรู้ในที่จะวัดผู้เรียนเป็นอย่างดี เข้าใจโครงสร้างระดับการรู้คิด เข้าใจการเขียนข้อคำถามที่สอดคล้องกับระดับการวัด รวมทั้งการเขียนตัวเลือก และตัวลวง ตามเกณฑ์คุณภาพ เพราะจากจุดนี้จะทำให้ผลการวิเคราะห์คุณภาพข้อสอบและแบบทดสอบภายหลังการสอบได้ตามเกณฑ์มาตรฐาน

การวิเคราะห์คุณภาพแบบทดสอบชนิดเลือกตอบ (Multiple choice test analysis)

ประโยชน์ของการวิเคราะห์คุณภาพแบบทดสอบภายหลังการสอบ เป็นกระบวนการในการพัฒนาคุณภาพเครื่องมือซึ่งเป็นส่วนหนึ่งของการวัดประเมินผล เพื่อให้ผลการประเมินมีคุณภาพเชื่อถือได้ กระบวนการนี้จะทำภายหลังการนำแบบทดสอบไปทดสอบ และนำแบบทดสอบนั้นมาวิเคราะห์หาค่าต่างๆ ซึ่งมีประโยชน์ดังนี้ (อนันต์ ศรีโสภา, 2525, น.185) ได้แก่

1. ประเมินคุณภาพของข้อสอบภายหลังการสอบ ก่อนใช้ตัดสินผลการสอบของผู้สอบ
2. เลือกเก็บข้อสอบที่มีคุณภาพไว้เป็นคลังข้อสอบ
3. ใช้เป็นแนวทางในการพัฒนาหรือปรับปรุงการสร้างข้อสอบในอนาคต
4. วินิจฉัยจุดแข็งและจุดอ่อนของผู้เรียน เพื่อใช้ในการให้ข้อมูลย้อนกลับ (feedback)
5. เป็นสิ่งที่สะท้อนถึงประสิทธิภาพของการจัดการเรียนการสอน ครอบคลุมเนื้อหาที่สำคัญตามวัตถุประสงค์หรือไม่ ผู้เรียนเข้าใจประเด็นใดยังไม่ถูกต้องหรือคลาดเคลื่อน สำหรับข้อสอบแบบเลือกตอบ จำนวนข้อสอบที่นำมาวิเคราะห์ จะต้องมีจำนวนมากกว่า 30 ข้อ และจำนวนผู้สอบควรมีจำนวนมากกว่า 30 คน ขึ้นไปเช่นกัน

ประเภทของการวิเคราะห์คุณภาพแบบทดสอบ แบ่งเป็น 2 ส่วน ได้แก่

1. การวิเคราะห์แบบทดสอบรายข้อ (Item analysis)
2. การวิเคราะห์แบบทดสอบทั้งฉบับ (Analysis of total test)

การวิเคราะห์ข้อสอบรายข้อ เป็นกระบวนการวิเคราะห์ย้อนรอยไปถึงคุณภาพที่เป็นผลจากการเขียนข้อสอบ โดยนำข้อสอบแต่ละข้อมาวิเคราะห์ในรายละเอียด การวิเคราะห์ข้อสอบรายข้อ แบ่งการวิเคราะห์ เป็น 2 ประเภท โดย พ.ท.หญิง ปวรส บุตะเขียว, ม.ป.ป. อธิบายไว้ ได้แก่

1. การวิเคราะห์ข้อสอบชนิดเลือกตอบแบบอิงกลุ่ม (Norm reference) เป็นการวิเคราะห์โดยนำผลคะแนนของผู้เรียนแต่ละคนไปเปรียบเทียบกับผลคะแนนของผู้เรียนคนอื่น ๆ ในกลุ่มเดียวกัน เพื่อจำแนกคนที่เรียนเก่งออกจากคนที่เรียนอ่อน ในกระบวนการประกอบด้วยการจัดการวิเคราะห์ ดังนี้ (ซวาล แพร์ตกุล, 2516, น.10-11; ปวรส บุตะเขียว, ม.ป.ป.; อนันต์ ศรีโสภา, 2525, น.185)

- 1.1 ระดับความยากง่าย (Level of difficulty)
- 1.2 อำนาจการจำแนก (Discriminative power)
- 1.3 ประสิทธิภาพของตัวลวง (Efficiency of distractors หรือ Distractability)

2. การวิเคราะห์ข้อสอบชนิดเลือกตอบแบบอิงเกณฑ์ (Criterion reference)

การวิเคราะห์ชนิดนี้ มีจุดมุ่งหมายเพื่อดูว่าผู้เรียนมีความรู้ ความสามารถในการเรียนซึ่งเป็นผลมาจากการเรียนการสอนตามมาตรฐานที่ตั้งไว้เพียงใด โดยใช้แบบทดสอบชุดเดียวกัน วัดก่อนเรียน (Pre Test) และหลังเรียน (Post Test) โดยการวิเคราะห์อยู่บนสมมติฐานที่ว่า เมื่อกระบวนการเรียนการสอนเสร็จสิ้นลง ผู้เรียนควรมีความรู้ในเรื่องนั้นๆดีขึ้น ดังนั้นควรมีผู้ตอบถูกมากหลังเรียน มากกว่าก่อนเรียน การวิเคราะห์จึงนำผลทดสอบก่อนเรียน และหลังเรียนมาคำนวณดัชนีความไวของผลการเรียนการสอน

ในที่นี้จะขอแนะนำรายละเอียดเฉพาะการวิเคราะห์ข้อสอบชนิดเลือกตอบแบบอิงกลุ่ม เนื่องจากเกี่ยวข้องกับการวิจัยครั้งนี้ ได้แก่

1. การวิเคราะห์หาระดับความยากง่าย (Level of difficulty)

ความยากง่าย (Difficulty) วัดได้จากดัชนีความยากง่าย (Difficulty index) โดยคำนวณจากจำนวนผู้ตอบถูกในข้อสอบข้อนั้นหาร ผู้ตอบข้อสอบทั้งหมด ซึ่งดัชนีความยากจะมีค่าอยู่ระหว่าง 0 ถึง 1 หรือ 0% - 100% ถ้าดัชนีความยากต่ำแสดงว่าข้อสอบข้อนั้นยากเพราะมีผู้ที่ทำข้อสอบตอบถูกน้อย และในทางกลับกัน ถ้าดัชนีความยากสูงแสดงว่าข้อสอบข้อนั้นง่ายเพราะมีผู้ที่ทำข้อสอบตอบถูกมาก (กรมวิชาการ, 2545, น.66) เขียนเป็นสูตรได้ดังนี้

จาก ค่า 0 หมายความว่า ไม่มีผู้ตอบข้อสอบนั้นได้ถูกเลย และค่า 1 หมายความว่า ทุกคนตอบได้ถูกหมด

จากสูตรในการคำนวณ

$$P = \frac{R}{N} \text{ เมื่อ}$$

P = ดัชนีความยากของข้อสอบ

R = จำนวนนักเรียนที่ตอบข้อสอบนั้นได้ถูกต้อง

N = จำนวนนักเรียนที่ตอบข้อสอบทั้งหมด

ยกตัวอย่างจากสูตร

สมมติว่า มีผู้ตอบ 60 คน ไม่มีใครตอบถูกเลย แทนค่าสูตร $P = 0/60 = 0$

หากตอบถูก ทั้ง 60 คน แทนค่าสูตร $P = 60/60 = 1$

เมื่อนำดัชนีความยากง่าย ระหว่าง 0 - 1 มาจัดกลุ่มแล้ว ประเมินคุณภาพของข้อสอบจากดัชนีความยากง่ายได้ดังนี้

ตารางที่ 2.3 ข้อเสนอแนะในการเลือกข้อสอบ/ แบบทดสอบจากค่าความยากง่าย

ค่าความยากง่าย (P)	ความหมาย	ข้อเสนอแนะ
.81 – 1.00	ง่ายมาก	ควรตัดทิ้ง
.61 - .80	ค่อนข้างง่าย	ดีพอใช้ ควรเก็บไว้ใช้
.41 - .60	ความยากง่ายปานกลาง	ดีมากเก็บไว้ใช้
.20 - .40	ค่อนข้างยาก	ดีพอใช้ ควรเก็บไว้ใช้
.00 - .19	ยากมาก	ควรตัดทิ้ง

ที่มา: ดัดแปลงจาก สุมาลี จันทรชลอ (2542, น.136)

ข้อสอบที่ดีควรมีระดับความยากง่าย เท่ากับ 0.5 ซึ่งจะทำให้เกิดค่าอำนาจจำแนกสูงสุด และมีความเที่ยงสูง อย่างไรก็ตามในการสอบวัดความรู้ผลการเรียนโดยทั่วไป มักนิยมให้มีข้อสอบที่มีระดับความยากง่ายในระดับต่าง ๆ ปะปนกันไป โดยจัดให้มีข้อสอบมีค่าความยากง่ายพอเหมาะ (p มีค่าใกล้เคียง 0.5) เป็นส่วนใหญ่ รวมทั้งให้มีข้อสอบที่ค่อนข้างยากและค่อนข้างง่ายอีกจำนวนหนึ่ง แต่ถ้าเป็นการสอบแข่งขันเพื่อคัดเลือกผู้ที่มีความรู้ความสามารถควรมีสัดส่วนของข้อสอบที่ยากสูงขึ้น ข้อสอบที่ดีควรมีค่าความยากง่ายระหว่าง 0.20 – 0.80 ในข้อสอบประเภท 4 ตัวเลือก ส่วนข้อสอบประเภทถูก – ผิด ค่าความยากง่าย ควรอยู่ระหว่าง 0.60 - 0.95 (เยาวดี รางชัยกุล วิบูลย์ศรี, 2552, น.144-145)

2. การวิเคราะห์หาอำนาจจำแนก (Discriminative power)

อำนาจจำแนก หมายถึงคุณสมบัติของข้อสอบ/แบบทดสอบ ในการจำแนกผู้สอบว่า ใครเก่งหรืออ่อนอย่างไร คำนวณจากหลักการ ที่ข้อสอบที่ผู้ได้คะแนนรวมส่วนมากตอบถูก และผู้ได้คะแนนรวมน้อยตอบผิด

การคำนวณ ค่าดัชนีอำนาจจำแนก (index of discrimination) ในอดีตที่ใช้การคำนวณด้วยมือคิดจากสัดส่วนผลต่างระหว่างจำนวนผู้ตอบถูกในกลุ่มที่ได้คะแนนสูงกับกลุ่มที่ได้คะแนนน้อย แต่ปัจจุบันที่ใช้การคำนวณด้วยคอมพิวเตอร์จะมีความแม่นยำมากกว่า เพราะใช้คะแนนของผู้สอบทุกคน โดยคำนวณจากสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson Product Moment Correlation) ระหว่างคะแนนที่ผู้สอบทำได้ในข้อนั้นๆ กับคะแนนรวมของข้อสอบข้ออื่นๆในฉบับนั้น (Office of Educational Assessment, University of Washington, 2005) ค่าดัชนีอำนาจจำแนกใช้แทนตัวอักษร r และมีค่าระหว่าง -1.0 ถึง +1.0 ข้อคำถามที่ดีจะต้องมีค่าอำนาจจำแนกเป็นบวกเพราะถ้าเป็นลบ แสดงว่ามีค่าอำนาจจำแนกกลับกัน คนเก่งตอบผิด แต่คนไม่เก่งตอบถูก ถ้าค่าอำนาจจำแนกเป็น 0 แสดงว่าผู้เรียนในกลุ่มเก่ง และกลุ่มอ่อนทำข้อสอบข้อนั้นได้เท่าๆกัน จึงไม่สามารถจำแนกกลุ่มผู้เข้าสอบได้ ข้อคำถามที่ดี ควรมีอำนาจจำแนก 0.2 ขึ้นไป และยังมีค่าเข้าใกล้ +1 ยิ่งดี (Chase 1978, p.140; เยาวดี รางชัยกุล วิบูลย์ศรี, น.154) จากหลักการที่กล่าวแล้ว ดังนั้นสูตรการคำนวณอำนาจจำแนก มีดังต่อไปนี้

$$\text{อำนาจจำแนก} = \frac{\text{จำนวนผู้เลือกตอบในกลุ่มสูง} - \text{จำนวนผู้เลือกตอบในกลุ่มต่ำ}}{\text{จำนวนคนในกลุ่มสูง หรือ จำนวนคนในกลุ่มต่ำ}}$$

$$\text{หรือ } r = \frac{\text{PH} - \text{PL}}{\text{NH or NL}}$$

ค่าดัชนีอำนาจจำแนก บ่งบอกว่าข้อสอบนั้น วัดความรู้ในเรื่องเดียวกัน หรือสอดคล้องกันกับข้อสอบข้ออื่นๆมากน้อยเพียงใด หากข้อสอบชุดนั้น วัดเนื้อหาที่มีความแตกต่างกันมาก ค่าสัมประสิทธิ์สหสัมพันธ์ก็จะต่ำกว่าข้อสอบที่วัดเนื้อหาที่เหมือนกัน ฉะนั้นในการพิจารณาค่าดัชนีอำนาจจำแนก ผู้สอนควรได้คำนึงถึงลักษณะของข้อสอบในวิชานั้นด้วย ข้อสอบที่ค่าดัชนีอำนาจจำแนกต่ำมักเกิดจากการใช้คำในข้อคำถามและตัวเลือกที่กำวม ไม่ชัดเจน ส่วนข้อที่มีค่าดัชนีอำนาจจำแนกติดลบ ผู้สอนควรได้นำมาพิจารณาว่าเกิดจากสาเหตุใด เช่น การเฉลยผิด ทำให้ผู้เรียนที่เก่งไม่ได้เลือกตัวเลือกนั้น เป็นต้น (office of Educational Assessment, University of Washington, 2005) รวมทั้งข้อสอบที่ยากมากสำหรับผู้เรียน ก็ทำให้ค่าอำนาจจำแนกติดลบได้เช่นกัน

ตารางที่ 2.4 เกณฑ์การแปลความหมายค่าอำนาจจำแนก (r) ของข้อสอบ

อำนาจจำแนกของข้อสอบ (r)	ความหมาย
0.80 - 1.00	อำนาจจำแนกดีมาก
0.60 - 0.79	อำนาจจำแนกดี
0.40 - 0.59	อำนาจจำแนกปานกลาง
0.20 - 0.39	อำนาจจำแนกพอใช้
0.01 - 0.19	อำนาจจำแนกต่ำ (ควรปรับปรุงหรือตัดทิ้ง)
-1.00 - 0.00	อำนาจจำแนกติดลบ (ควรตัดทิ้ง)

ที่มา: ดัดแปลงจาก Chase (1978, p.140)

ความจำกัดของการวิเคราะห์ข้อสอบด้วยค่า P และค่า r

ถึงแม้ว่าการวิเคราะห์ข้อสอบด้วย ด้วยค่า P และค่า r จะอาศัยหลักการทางสถิติที่มีระเบียบแบบแผน มีแนวทางในการแปลผลการวิเคราะห์ที่ชัดเจน และเป็นที่ยอมรับกันแพร่หลาย การวิเคราะห์ด้วยค่าสถิติดังกล่าว ยังมีข้อจำกัดบางประการที่อาจไม่เหมาะสมกับการวัดระดับชั้นเรียน ทั้งนี้ขึ้นอยู่กับความสามารถของผู้เรียนที่ทำข้อสอบนั้นๆ ดังตัวอย่าง ผลการวิเคราะห์หลังนำข้อสอบ 1 ข้อ ไปทดลองใช้กับนักเรียน ระดับชั้นประถมศึกษาปีที่ 6 จาก 5 โรงเรียน (ชยุตม์ ภิรมย์สมบัติ, 2555,น.12)

ตารางที่ 2.5 แสดงผลการวิเคราะห์ข้อสอบจำนวน 1 ข้อ ที่ใช้ทดสอบกับนักเรียนประถมศึกษา
ชั้นปีที่ 6 จำนวน 5 โรงเรียน

โรงเรียน	ความยากง่าย P)		อำนาจจำแนก (r)	
	ค่าสถิติ	ความหมาย	ค่าสถิติ	ความหมาย
A	0.186	ยากมาก	0.062	จำแนกได้ต่ำ
B	0.285	ค่อนข้างยาก	0.235	จำแนกได้ปานกลาง
C	0.315	ค่อนข้างยาก	0.431	จำแนกได้ดี
D	0.483	ยากปานกลาง	0.640	จำแนกได้ดี
E	0.725	ค่อนข้างง่าย	0.496	จำแนกได้ดี

ที่มา: จาก ชยุตม์ ภิรมย์สมบัติ, 2555, น.12

จะเห็นว่าเมื่อเปลี่ยนกลุ่มตัวอย่างที่ทำแบบทดสอบ ค่า P และค่า r เปลี่ยนแปลงไปตามความสามารถของผู้เรียนในชั้นในแต่ละโรงเรียนที่มีความสามารถไม่เท่ากัน และการนำผลของค่า P และ r ที่ได้มาปรับปรุงเพื่อพัฒนาข้อสอบระดับชั้นเรียน ต้องพิจารณาให้รอบด้าน ปัญหาอาจมาจากกลุ่มมีขนาดเล็กเกินไป และกลุ่มอาจไม่เป็นตัวแทนที่ดีของประชากรผู้เรียนทั่วไป

3. ประสิทธิภาพของตัวลวง (Efficiency of distractors หรือ Distractability)

ประสิทธิภาพของตัวลวง (Distractor analysis) ประสิทธิภาพของตัวลวงเป็นอีกคุณลักษณะหนึ่งที่บ่งบอกถึงข้อสอบที่มีคุณภาพ ตัวลวงที่มีคุณภาพ พิจารณาได้จาก คนในกลุ่มอ่อน เลือกตอบมากกว่าคนในกลุ่มเก่ง โดยที่มีค่าสัดส่วนของผู้เลือกตัวลวง (P_w) ซึ่งหมายถึง สัดส่วนของผู้เลือกตัวลวง (P_w) ควรมีค่า ตั้งแต่ 0.05 (หรือร้อยละ 0.5) ขึ้นไป และมีอำนาจจำแนกของตัวลวง (r_w) หมายถึง ผลต่างระหว่างสัดส่วนของคนในกลุ่มอ่อนที่เลือกตัวลวงกับสัดส่วนของคนในกลุ่มเก่งที่เลือกตัวลวงนั้น ๆ ค่า r_w ควรมีค่า ตั้งแต่ 0.05 ขึ้นไป (Johnson, 1951; ชยุตม์ ภิรมย์สมบัติ, 2555, น.13)

ตัวอย่าง สัดส่วนของผู้ที่เลือกตัวลวง (P_w) เช่น ข้อสอบข้อหนึ่งมีคนสอบ 100 คน มีคนเลือกตัวลวง ก. 25 คน แสดงว่า สัดส่วนผู้เลือกตัวลวง $g. = 0.25$ หรือ 25 %

ดังนั้น หลังจากวิเคราะห์ข้อสอบแล้ว ควรพิจารณาต่อไปที่การวิเคราะห์ตัวลวง (Distractor analysis) วิธีง่ายๆ ทำได้โดยการคำนวณจำนวนร้อยละของผู้เรียนที่เลือกตัวลวงแต่ละตัว (P_w) โดยแบ่งกลุ่มผู้เรียนออกเป็นกลุ่มสูงและกลุ่มต่ำจากคะแนนรวมของกลุ่มสูงและกลุ่มต่ำของข้อสอบทั้งฉบับเป็นเกณฑ์ในการแบ่งกลุ่ม เพื่อวิเคราะห์ว่าผู้เรียนในแต่ละกลุ่มยังเข้าใจสาระคลาดเคลื่อนในแต่ละเนื้อหาเป็นจำนวนเท่าใด (จนทำให้ตอบข้อสอบผิด) ดังตัวอย่าง

ตารางที่ 2.6 ผลการวิเคราะห์จำนวนผู้ตอบ กลุ่มสูงและกลุ่มต่ำที่เลือกตอบตัวเลือกและตัวลวง

ตัวเลือก	1	2	3	4
กลุ่มคะแนนสูง	121	23	7	0
กลุ่มคะแนนต่ำ	24	67	41	17

ที่มา: จาก ชยุตม์ ภิรมย์สมบัติ, 2555, น.12

จากตารางที่ 2.6 เป็นการทำการวิเคราะห์สัดส่วนเลือก 1 ข้อ และตัวลวง 3 ข้อ ซึ่งปัจจุบันเครื่องคำนวณผลการวิเคราะห์ข้อสอบด้วยโปรแกรมคอมพิวเตอร์ จะให้ค่า P_w และ r_w มาพร้อมกันกับค่าการวิเคราะห์ข้อสอบอื่นๆ

ผลที่ได้ยังช่วยให้ผู้สอนได้ทราบเกี่ยวกับความเข้าใจคลาดเคลื่อนของผู้เรียนที่สะท้อนผ่านการเลือกตัวลวงแต่ละตัวในข้อสอบ สารสนเทศดังกล่าวจะช่วยให้การวางแผนการจัดการเรียนรู้เพื่อแก้ไขความเข้าใจคลาดเคลื่อนของผู้เรียนได้อย่างมีประสิทธิภาพต่อไป

การเลือกข้อสอบจากผลการวิเคราะห์ข้อสอบ

เมื่อวิเคราะห์ข้อสอบได้ค่าความยากง่ายและอำนาจการจำแนกแล้ว ควรนำมาจัดทำกราฟ ซึ่งจะแสดงการกระจายของข้อสอบและชี้ให้เห็นว่า ข้อสอบข้อใดบ้างที่อยู่ในเกณฑ์ใช้ได้หรือใช้ไม่ได้ จากเกณฑ์ที่ใช้ได้ นำไปไว้ในครั้งข้อสอบ ได้แก่ ข้อที่มีดัชนีความยากง่ายที่ 0.2–0.8 และมีดัชนีอำนาจจำแนก ตั้งแต่ 0.2 ขึ้นไปในข้อเดียวกัน

ตัวอย่าง การทำกรอบภาพกราฟเพื่อพิจารณาข้อสอบที่มีคุณภาพได้เกณฑ์มาตรฐานซึ่งสะท้อนจากการกระจายของข้อสอบตามผลการวิเคราะห์ข้อสอบ

ค่าสัมประสิทธิ์ความเชื่อมั่น มีค่าระหว่าง 0 - 1.00 สามารถคำนวณค่า โดยใช้ สัมประสิทธิ์ของครอนบัก (Cronbach's Alpha) ใช้กับแบบทดสอบเลือกตอบ ที่มีข้อที่ถูกเพียงข้อเดียว (ตอบถูกได้คะแนน 1 ตอบผิดได้คะแนน 0) และจะได้ค่าเท่ากับการใช้สูตร KR - 20

ตารางที่ 2.7 เกณฑ์การแปลความหมายของค่าสัมประสิทธิ์ความเชื่อมั่น

ค่าสัมประสิทธิ์ความเชื่อมั่น	การแปลความหมาย
> .90	ยอดเยี่ยม เทียบได้กับข้อสอบมาตรฐาน
.80 - .89	ดีมาก สำหรับใช้เป็นแบบทดสอบรายวิชาในชั้นเรียน
.70 - .79	ดี สำหรับใช้เป็นแบบทดสอบรายวิชาในชั้นเรียน อาจมีข้อสอบบางข้อจำเป็นต้องปรับปรุง
.60 - .69	ค่อนข้างต่ำ แบบทดสอบชุดนี้ ควรมีการวัดด้วยเครื่องมืออื่นเสริม ก่อนจะตัดสินเกรด
.50 - .59	ควรปรับปรุง จำเป็นต้องมีการวัดด้วยเครื่องมืออื่นเสริม ก่อนจะตัดสินเกรด
< .50	ไม่น่าเชื่อถือ จำเป็นต้องปรับปรุง และไม่ควรถูกใช้เป็นหลักในการตัดสินเกรด

ที่มา: ดัดแปลงจาก office of Educational Assessment, University of Washington, 2005

องค์ประกอบที่มีผลทำให้ค่าความเชื่อมั่น (reliability) สูงหรือต่ำ

เนื่องจากการวิจัยครั้งนี้คำนวณค่าสัมประสิทธิ์ความเชื่อมั่น โดยใช้สูตร KR-20 โดยใช้โปรแกรมคอมพิวเตอร์ วิเคราะห์หาความยากง่ายของแบบทดสอบ นำสัดส่วนของข้อที่ตอบถูกและผิดในแต่ละข้อคำถามและความแปรปรวน มาแทนค่าในสูตร KR-20 การใช้สูตร KR-20 ใช้เมื่อข้อคำถามแต่ละข้อ มีค่าความยากง่ายใกล้เคียงกัน และความสัมพันธ์ระหว่างข้อ (inter-correlation) เท่าๆ กัน ดังนั้นปัจจัยที่จะทำให้ค่า reliability ของข้อสอบทั้งฉบับมีค่าสูงหรือต่ำ มาจากปัจจัยดังนี้ (Crocker and Algina, 1986 อ้างอิงใน ศิริชัย กาญจนวาสิ, 2544, น.60-65)

1. ความเป็นเอกพันธ์ของกลุ่มผู้ให้ข้อมูล (Group Homogeneity) ในกลุ่มผู้ให้ข้อมูลที่มีลักษณะที่ใกล้เคียงกันเมื่อนำคะแนนมาหาค่าความเชื่อมั่นจะได้ค่าความเชื่อมั่นสูงกว่าที่ได้จากกลุ่มผู้ให้ข้อมูลที่มีลักษณะที่หลากหลายคละกัน (วิวิธพันธ์) และขนาดของของกลุ่มผู้ให้ข้อมูล ควรมีประมาณ 6-10 เท่าของ จำนวนข้อสอบจึงจะได้ความเชื่อมั่นที่เป็นจริง

2. ความยาวของแบบทดสอบ (test length) เมื่อเพิ่มจำนวนข้อสอบที่มีความคู่ขนานกับข้อสอบเดิมที่มีอยู่จะทำให้ค่าความเชื่อมั่นของแบบทดสอบฉบับนั้น ๆ มีค่าที่สูงขึ้น

3. ความสัมพันธ์ระหว่างข้อสอบ (Interitem Correlation) แบบทดสอบฉบับใดที่มีความเป็นเอกพันธ์ของคุณลักษณะหรือเนื้อหาแสดงว่าแบบทดสอบฉบับนั้นมีความสัมพันธ์ระหว่างข้อสอบสูง อันจะส่งผลต่อค่าสัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบฉบับนั้น

4. กำหนดเวลาที่ใช้ในการแบบทดสอบ (time limit) แบบทดสอบที่สร้างและพัฒนาเป็นอย่างดีและได้กำหนดเวลาที่ใช้ในการทดสอบที่เหมาะสมกับแบบทดสอบจะได้ค่าสัมประสิทธิ์ความเชื่อมั่นที่สูง แต่ถ้าให้เวลาที่จำกัดหรือมากเกินไปจะทำให้สัมประสิทธิ์ความเชื่อมั่นมีแนวโน้มลดลง

5. เกิดจากการสร้างข้อสอบที่ไม่ดี เช่น การใช้คำในคำถามที่ไม่รัดกุมพอ ทำให้เกิดการตีความต่างๆ กัน การสร้างตัวเลือก (options) ที่อาจถูกได้มากกว่า 1 ข้อ ในขณะที่เฉลยมีข้อถูกเพียงข้อเดียว หรืออาจมีการเฉลยผิด เป็นต้น (ล้วน สายยศ และอังคณา สายยศ, 2543; Gronlund NE., 1985)

แบบทดสอบรายวิชากฎหมายและจริยธรรม มีลักษณะบางอย่างที่สนับสนุนเหตุผลที่ทำให้ผลการวิเคราะห์มีค่าความเชื่อมั่นยังไม่สูงมาก ได้แก่ จำนวนข้อสอบ 40 ข้อ ซึ่งอาจยังไม่มากพอ และผู้ตอบมีจำนวนกลุ่มเล็ก 60 คน ข้อสอบไม่มีความเป็นเอกพันธ์ เพราะมี 2 วิชารวมกัน ได้แก่ วิชากฎหมาย และวิชาจริยธรรม ความสอดคล้องภายในต่างกันตามธรรมชาติ (เป็นสายกฎหมาย และสายปรัชญา ซึ่งผู้เรียนบางส่วนมีความถนัดในการเรียน 2 วิชานี้แตกต่างกันในตัวเอง ลักษณะคำถามและคำตอบในแบบทดสอบ ใช้ภาษาในการสื่อสารผ่านตัวข้อสอบคนละแบบแผน กฎหมายเป็นเรื่องของการรู้คิด (Cognitive) ความเป็นเหตุเป็นผล ส่วนจริยธรรม เป็นเรื่องของความรู้สึกถูกต้อง ดีงาม

งานวิจัยที่เกี่ยวข้องกับคุณภาพของข้อสอบ และการสร้างข้อสอบที่มีคุณภาพ

Oluseyi & Olufemi (2012) ได้ศึกษาวิจัยเกี่ยวกับการวิเคราะห์ข้อสอบชนิดหลายตัวเลือก ในวิชาเคมีพื้นฐานของมหาวิทยาลัยในจีเรีย โดยมีวัตถุประสงค์เพื่อวิเคราะห์ความยากง่าย และอำนาจจำแนกของข้อสอบชนิดหลายตัวเลือกในวิชาเคมีพื้นฐาน รูปแบบการวิจัยเป็นแบบการวิเคราะห์ย้อนหลังเพื่อเปรียบเทียบสาเหตุ (Ex post facto, casual comparative design) กลุ่มตัวอย่างคือนักศึกษาจำนวน 800 คน ที่ได้มาจากการสุ่มตัวอย่างแบบง่ายจากจำนวนประชากรที่เป็นนักศึกษาที่เรียนวิชาเคมี 102 ในปีการศึกษา 2008 – 2009 ข้อสอบชนิด 4 ตัวเลือก จำนวน 40 ข้อ สร้างขึ้นจากเนื้อหาวิชาดังกล่าว โดยอาจารย์ผู้ร่วมสอนในแต่ละหัวข้อในรายวิชานั้น ผู้วิจัยคัดเลือกนักศึกษาจำนวน 442 คน ที่ได้คะแนน เปอร์เซ็นไทล์ที่ 30 ทั้งจากคะแนนที่น้อยที่สุดและมากที่สุดมาเป็นหน่วยในการวิเคราะห์ สถิติที่ใช้ ได้แก่ ค่าความถี่ ความความยาก-ง่าย และค่าอำนาจจำแนก

ผลการศึกษาพบว่า ค่าความยาก-ง่าย ของข้อสอบทั้งฉบับ เท่ากับ 0.54 ค่าอำนาจจำแนก เท่ากับ 0.32 ตัวเลือกที่นักศึกษาเลือกตอบมากที่สุด คือ ตัวเลือกที่ 4 ซึ่งพบถึง 16 ข้อ จากจำนวนข้อสอบ 40 ข้อ พบข้อสอบที่มีความผิดพลาดในการวัด 3 ข้อ เนื่องจากอำนาจจำแนกเท่ากับ 0 ซึ่งผู้วิจัยอภิปรายว่า อาจเกิดจากการสร้างข้อสอบไม่ถูกต้อง หรือผู้ตรวจเฉลยผิด นอกจากนี้พบว่า ข้อสอบข้อที่ยากที่สุดจำนวน 1 ข้อ มีค่า ความยาก-ง่าย ของข้อสอบเท่ากับ 0. 24 มีค่าอำนาจจำแนก เท่ากับ 0.32 ซึ่งมีนักเรียนจำนวน 76 คน จาก 442 คน ที่ตอบได้ถูก การที่ค่าความยากง่ายของข้อสอบเท่ากับ 0.54 แสดงว่าข้อสอบ ไม่ยากหรือง่ายจนเกินไป ค่าอำนาจจำแนกเท่ากับ 0.32 แสดงว่า ค่าอำนาจจำแนก ได้พอใช้ ผู้วิจัยได้ให้ข้อเสนอแนะว่า การสร้างข้อสอบมีความสำคัญอย่างยิ่ง

และควรทำตารางแสดงผลการวิเคราะห์ข้อสอบไว้ท้ายข้อ เพื่อเลือกข้อที่ดี ไว้ในธนาคารข้อสอบ รวมทั้งการวิพากษ์ข้อสอบก่อนนำมาใช้กับนักศึกษาด้วย

Ascalon, Meyers, Davis & Smits (2007) ศึกษาผลของความคล้ายคลึงของตัวลวง (Distractor Similarity) และโครงสร้างของข้อคำถาม (Item-Stem Structure) ต่อความยากของข้อสอบ การวิจัยนี้มีวัตถุประสงค์เพื่อตรวจสอบแนวทางในการเขียนข้อสอบในการเขียนข้อสอบในสองประเด็น คือ รูปแบบของข้อคำถาม และความคล้ายคลึงของตัวเลือก ทั้งนี้เพื่อหาข้อมูลเชิงประจักษ์ในการตอบข้อเรียกร้องของ Haladyna Downing และ Rodriguez (2002) เกี่ยวกับแนวทางในการเขียนข้อสอบ และเพื่อขยายผลการศึกษาของ Smith and Smith (1988) เกี่ยวกับรูปแบบของโจทย์คำถาม โดยการวิจัยครั้งนี้ มีการตั้งโจทย์สมมุติฐาน 2 ข้อ ได้แก่ 1) ข้อสอบ Multiple choice ที่มีตัวเลือกซึ่งมีโครงสร้างคล้ายคลึงกัน ระหว่างตัวลวงกับคำตอบที่ถูกต้อง จะทำให้ข้อสอบข้อนั้นมีความยาก มากกว่าข้อสอบที่มีโครงสร้างของตัวลวง และคำตอบที่ถูกต้องแตกต่างกันอย่างสิ้นเชิง และ 2) รูปแบบของโจทย์คำถาม ที่มีลักษณะเป็นคำถามจะง่ายกว่าโจทย์ที่มีลักษณะข้อคำถามปลายเปิด โดยเว้นที่ไว้ให้เติมคำตอบ การวิจัยครั้งนี้ถูกออกแบบให้มี 5 ขั้นตอน โดยการสร้างและวิเคราะห์ข้อสอบในการสอบรับใบอนุญาตขับขี่ยานยนต์ ลักษณะของข้อสอบมีหลายแบบ แต่เป็นข้อสอบคู่ขนานกัน แต่ละขั้นตอนมีผู้เชี่ยวชาญวิเคราะห์และตรวจสอบข้อสอบแต่ละชุดอย่างละเอียด และนำข้อสอบแบบต่างๆ จำนวน 800 ชุด ไปใช้กับนักเรียนมัธยมศึกษา อายุระหว่าง 13-19 ปี จำนวน 3 โรงเรียน

ผลการวิจัย พบว่า ตัวลวงของข้อคำถามที่มีความสั้นยาวของประโยคแต่ละข้อตัวลวงใกล้เคียงกัน และเนื้อหาของข้อคำถามคล้ายคลึงกันกับข้อคำถามที่ถูกต้อง จะทำให้ข้อสอบข้อนั้นมีความยากมากกว่าข้อสอบที่มีตัวลวงที่มีเนื้อหาของคำตอบแตกต่างกันกับคำตอบถูก อย่างมีนัยสำคัญทางสถิติ ที่ระดับ 0.01 ($p < 0.01$) เพราะตัวลวงที่มีเนื้อหาแตกต่างกัน จะง่ายต่อการคาดเดา เป็นไปตามสมมุติฐานที่ 1) ส่วนสมมุติฐานที่ 2) ผลการวิจัยพบว่า ลักษณะการตั้งโจทย์คำถามที่เป็นข้อคำถามหรือโจทย์คำถามที่เป็นข้อความปลายเปิด ไม่พบความแตกต่างกันอย่างมีนัยสำคัญทางสถิติในเรื่องความยากง่ายของข้อสอบ

Caldwell D.J. & Pate A.N. (2013) ได้ศึกษารูปแบบของคำถามต่อคะแนนการสอบของนักศึกษาและคุณภาพของข้อสอบ โดยมีวัตถุประสงค์เพื่อเปรียบเทียบ การใช้รูปแบบคำถามของข้อสอบ ที่สร้างตามแนวปฏิบัติที่สร้างตามแนวปฏิบัติที่เป็นมาตรฐานและไม่มีมาตรฐาน ข้อสอบที่สร้างตามแนวปฏิบัติที่เป็นมาตรฐาน จะมีหลักการในการสร้างดังนี้ 1) การใช้คำถามเชิงบวก หลีกเลี่ยงคำถามเชิงลบ เช่น ยกเว้น ไม่ใช่ หากจำเป็นต้องใช้ ต้องขีดเส้นใต้ หรือใช้ตัวหนา เพื่อให้เห็นค่าที่เป็นเชิงลบอย่างชัดเจน 2) ในแต่ละคำถามสามารถใช้ตัวเลือกได้มากเท่าที่ต้องการ แต่ผลการวิจัยพบว่า การใช้ตัวเลือกเพียง 3 ตัวเลือก ก็เพียงพอแล้ว 3) การใช้ตัวเลือก “ไม่มีข้อใดถูกต้อง” ควรใช้ด้วยความระมัดระวัง

เครื่องมือการวิจัยเป็นข้อสอบปรนัยจำนวน 15 ข้อ จำนวน 2 ชุด สร้างโดยครูผู้สอนจำนวน 5 คน ชุดแรกสร้างตามแนวปฏิบัติที่เป็นมาตรฐาน เรียกว่า Standard scale ทดสอบกับนักศึกษาเภสัชศาสตร์ จำนวน 55 คน และชุดที่ 2 เป็นแบบสอบถามที่ไม่ได้สร้างตามแนวปฏิบัติที่เป็นมาตรฐาน เรียกว่า Nonstandard scale ทดสอบกับนักศึกษาจำนวน 54 คน สัดส่วนของคำถามเป็นการวัดด้านความรู้ – ความจำ ร้อยละ 60 เข้าใจ ร้อยละ 20 และนำไปใช้ ร้อยละ 20

ผลการวิจัยพบว่า ข้อสอบที่ไม่ได้สร้างตามแนวปฏิบัติที่เป็นมาตรฐาน (Nonstandard scale) มีความยากกว่า ข้อสอบ Standard scale (p value = 0.03) จากแนวปฏิบัติทั้ง 3 ข้อ มีเพียงข้อแนะนำที่ให้หลีกเลี่ยงการใช้ตัวลง “ไม่มีข้อใดถูกต้อง” เท่านั้น ที่ทำให้คะแนนของนักศึกษาทั้งสองกลุ่ม แตกต่างกันอย่างมีนัยสำคัญทางสถิติ โดยคะแนนเฉลี่ยของกลุ่มที่ทำข้อสอบ standard scale และกลุ่มที่มาข้อสอบ nonstandard scale เท่ากับร้อยละ 53.6 และ 41.3 ตามลำดับ (p value < 0.001)

จากผลการวิจัยสรุปได้ว่า ข้อสอบที่ไม่ได้มาตรฐานจะทำให้ยากต่อการที่การตอบของนักศึกษามากกว่าข้อสอบที่สร้างอย่างมาตรฐาน จึงไม่ช่วยในการแยกแยะกลุ่มที่เก่ง กับกลุ่มที่อ่อนได้ น่าจะเกิดจาก ทั้งสองกลุ่มเกิดความผิดพลาดจากการอ่านข้อสอบไม่ชัดเจน ทำให้ตอบผิดมากกว่าจากความบกพร่องในเรื่องความรู้ ทำให้ข้อสอบชุดที่กล่าวไว้เด็กเก่งและอ่อนไม่ได้

Phipps, Brackbill & Dunn (2009) ได้ศึกษาความสัมพันธ์ระหว่างข้อคำถามชนิดกรณีศึกษา (Case-based items) กับไม่ใช่กรณีศึกษา (Noncase-based items) ในแบบทดสอบชนิดเลือกตอบ (Multiple choice items) กับผลการวิเคราะห์ความยากง่าย และดัชนีค่าอำนาจจำแนก โดยแบบทดสอบที่นำมาวิเคราะห์ เป็นชุดข้อสอบชนิดเลือกตอบ 4-5 ตัวเลือก วิชาการบำบัดรักษา (Therapeutics) ที่ใช้สอบนักศึกษา หลักสูตรปริญญาเอกทางเภสัชศาสตร์ คณะเภสัชศาสตร์ (Shenandoah University) ชั้นปีที่ 2 และปีที่ 3 เป็นเวลา 4 ปี การศึกษา คือในช่วงปี ค.ศ. 2004 – 2005 และ 2007-2008 ชุดข้อสอบนี้เป็นข้อสอบกลางภาค และปลายภาค ให้เวลาทำข้อสอบ 2 ชั่วโมง จำนวน 1 ใน 4 ของข้อคำถาม คำถามจะถามเกี่ยวกับการสะสมของยาในร่างกายตามธรรมชาติ (Cumulative in nature) เนื้อหารายวิชาที่เรียน เป็นหัวข้อเกี่ยวกับระบบการไหลเวียนเลือด (Cardiovascular) โรคติดเชื้อ (infectious disease), วิทยาเนื้องอกและมะเร็งกับระบบเลือด (Hematology-oncology) ประสาทรับความรู้สึก (Neurosensory), และจิตเวชศาสตร์ (Psychiatry) ในชุดคำถามจะมีผลวิเคราะห์ ค่าความยากง่าย และดัชนีอำนาจจำแนกไว้ด้วย โดยการวิจัยจะปกปิดรายชื่อของผู้ทำข้อสอบ และเจ้าของข้อสอบไว้เป็นความลับ ชุดข้อสอบที่ไม่นำมาใช้วิจัยได้แก่ ข้อสอบที่มีข้อคำถามซึ่งถามต่อเนื่องมากกว่า 1 ข้อ (ข้อสอบที่ไม่เป็นเอกพันธ์ คือ ถามไม่เบ็ดเสร็จในข้อเดียว), เฉลยผิด, คำตอบถูกมีมากกว่า 1 ข้อ, ข้อสอบที่ให้เลือกตอบ ผิด-ถูก, และข้อสอบที่มีตัวเล็กลน้อยกว่า 4 ตัวเลือก หรือมากกว่า 5 ตัวเลือก การวิเคราะห์ใช้โปรแกรม SPSS (version 15) ใช้สูตร ค่า t -test independence เพื่อหาค่าความแตกต่างระหว่างกลุ่ม การวิจัยครั้งนี้ได้รับการตรวจสอบความถูกต้อง โดยคณะกรรมการจากสถาบันทวนสอบ The Shenandoah University Institutional Review Board

ผลการศึกษาพบว่า จากจำนวนข้อสอบทั้งหมด 1,575 ข้อ ร้อยละ 76 เป็นข้อสอบชนิดไม่ใช่ข้อคำถามกรณีศึกษา และพบว่า ค่าดัชนีอำนาจจำแนก ของข้อคำถามชนิดไม่ใช่กรณีศึกษา ($r=0.250$,

SD= 0.1) สูงกว่า คำถามชนิดกรณีศึกษา ($r=0.227$, $SD =0.1$) อย่างมีนัยสำคัญทางสถิติ (p value <0.01) ส่วนข้อคำถามที่เป็นกรณีศึกษาจะมีความยาวมากกว่า มีรายละเอียดมากกว่า แต่มีความยาก ($p=76.51$, $SD=19.2$) ไม่แตกต่างจากข้อสอบชนิด ไม่ใช่ข้อคำถามกรณีศึกษา ($P=76.86$, $SD=18.5$) อย่างมีนัยสำคัญทางสถิติ (p value = 0.75) นอกจากนั้นเมื่อวิเคราะห์ ข้อสอบชนิดที่มี 4 ตัวเลือกและ 5 ตัวเลือก เปรียบเทียบกัน พบว่า ข้อสอบชนิด 5 ตัวเลือก ยากกว่า ($P=75.04$, $SD=19.0$; $r=0.26$, $SD=0.1$) และค่าดัชนีอำนาจจำแนก สูงกว่า ชนิด 4 ตัวเลือก ($P=79.16$, $SD=17$; $r=0.22$, $SD=0.1$) อย่างมีนัยสำคัญทางสถิติ (p value < 0.001)

ผลการศึกษารูปได้ว่า ข้อคำถามชนิดกรณีศึกษา (cased-based items) กับไม่ใช่กรณีศึกษา(Noncase-based items) มีผลต่อคุณภาพของข้อสอบ

Lee HyeSun & Winke Paula (2012) ได้ศึกษาความแตกต่างของการใช้ข้อคำถามที่มีตัวเลือก 3, 4 และ 5 ตัวเลือก ในแบบทดสอบการฟังภาษาอังกฤษ College Scholastic Ability Test (CSAT) ซึ่งเป็นแบบทดสอบภาษาอังกฤษสำหรับผู้ใช้ภาษาอังกฤษเป็นภาษาที่สอง (High-stakes second language test) มีความสำคัญมากในประเทศเกาหลีเพราะเป็นการสอบที่ใช้การตัดสินว่าใครจะได้เข้าเรียนในมหาวิทยาลัย และได้เรียนที่ไหน ในปี 2011 มีผู้สมัครเข้าเรียนต่อมหาวิทยาลัยถึง 648,946 คนที่ได้ทำแบบทดสอบนี้ ผู้วิจัยต้องการศึกษาว่า ข้อสอบที่มีจำนวนตัวเลือกต่างกัน ในกลุ่มตัวเลือก 3,4,5 ตัว จะทำให้ความยากง่าย ค่าอำนาจจำแนก และเวลาในการทำข้อสอบแตกต่างกันหรือไม่ และสามารถวัดทักษะการฟังของผู้เข้าสอบได้คงเส้นคงวาหรือไม่ นอกจากนี้ยังมีการสอบถามความคิดเห็นของผู้เข้าสอบต่อข้อสอบที่มีจำนวนตัวเลือกต่างๆกัน

การศึกษาดำเนินการโดย ผู้วิจัยได้นำข้อสอบ CSAT ที่ใช้สำหรับเตรียมผู้เข้าสอบในปี 2007, 2008, 2009 รวม 3 ชุด ซึ่งมีจำนวนชุดละ 17 ข้อ ขอให้คนเกาหลีจำนวน 73 คน ประกอบไปด้วยผู้ที่จบปริญญาจากมหาวิทยาลัยในสหรัฐอเมริกา 13 คน, อาจารย์ที่สอนภาษาอังกฤษ 15 คน นักเรียนชั้นมัธยมศึกษา เกรด 9 จำนวน 45 คน เลือกตัวลงที่มีความเป็นไปได้น้อยที่สุดออกจากข้อสอบชนิด 5 ตัวเลือก จำนวน 2 รอบ จนเหลือ 4 ตัวเลือก และ 3 ตัวเลือก จนในที่สุดได้ข้อสอบทั้งหมด 3 ชุดๆ ละ 3 เวอร์ชัน รวมเป็น 9 ชุดย่อย หลังจากนั้นนำมาทดสอบกับนักเรียนชั้นมัธยมศึกษา (เกรด10) ที่เรียนพิเศษวิชาภาษาอังกฤษในกรุงโซล จำนวน 264 คน แบ่งเป็น 3 กลุ่ม มีการจัดตารางสอบแบบ Complete-block design ห่างกันครั้งละ 1 สัปดาห์ เพื่อให้แต่ละกลุ่มได้ทำแบบฝึกหัดชุดใดชุดหนึ่งใน 3 เวอร์ชัน คือ แบบทดสอบชนิดข้อคำถาม 5, 4 และ 3 ตัวเลือก เช่น กลุ่มที่ 1 ทำ ข้อสอบปี 2008 แบบ 4 ตัวเลือก สัปดาห์ต่อมาทำข้อสอบปี 2007 แบบ 5 ตัวเลือก และสัปดาห์ที่สาม ทำ ข้อสอบปี 2009 แบบ 3 ตัวเลือก เป็นต้น ทั้งนี้ คะแนนเฉลี่ยจากการสอบภาษาอังกฤษของทั้งสามกลุ่มก่อนการทดลองไม่แตกต่างกัน

ผลการศึกษาพบว่า แบบทดสอบชนิด 5 ตัวเลือก มีความยากที่สุด รองลงมาคือ ชนิด 4 ตัวเลือก และ 3 ตัวเลือก ตามลำดับ โดยคะแนนจากการทำแบบทดสอบ ชนิด 3 ตัวเลือก จะมีค่าคะแนนเฉลี่ยสูงกว่า แบบทดสอบที่มี 4 ตัวเลือก และ 5 ตัวเลือก อย่างมีนัยสำคัญทางสถิติ แต่ไม่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติของค่าอำนาจจำแนกโดยเฉลี่ยของข้อสอบทั้ง 3 แบบ ส่วนค่าความเชื่อมั่น ของแบบทดสอบ ไม่พบความคงเส้นคงวา ของค่าความเชื่อมั่นกับจำนวนตัวเลือก

กล่าวคือ แบบทดสอบ ปี 2007 ชนิด 4 ตัวเลือก มีค่าความเชื่อมั่นสูงสุด แบบทดสอบ ปี 2008 ชนิด 4 ตัวเลือก มีค่าความเชื่อมั่นต่ำสุด สำหรับเวลาในการทำแบบทดสอบ แบบ 3 ตัวเลือก ใช้เวลาน้อยที่สุด จากผลการศึกษานี้ ผู้วิจัยให้ข้อเสนอแนะว่า การใช้แบบทดสอบ 3 ตัวเลือก อาจไม่ใช่ทางเลือกที่ดีที่สุด ขึ้นอยู่กับมุมมองของผู้ออกแบบข้อสอบ หรือผู้มีส่วนได้ส่วนเสีย สำหรับการพัฒนาแบบทดสอบ ควรพิจารณาใช้สถิติในการทดสอบที่หลากหลาย และมองปัจจัยจากสภาพแวดล้อมรอบด้านในการตัดสินใจสร้างแบบทดสอบ ว่าควรมีจำนวนตัวเลือก 3, 4 และ 5 ตัวเลือก

Usova GM. (1997) ได้ศึกษาประสิทธิภาพของการใช้ข้อสอบเก่า ในคลังข้อสอบร่วมกับข้อสอบใหม่ในการจำแนกผู้สอบ โดยหน่วยงานภายใต้ Nuclear Regulator Commission (NRC), U.S. ที่รับผิดชอบเกี่ยวกับการออกใบอนุญาตผู้ควบคุมโรงไฟฟ้าพลังงานนิวเคลียร์ ได้จัดให้มีการสอบเรียกว่า NRC Generic Fundamentals Examination (GFE) สำหรับใช้สอบผู้ต้องการสอบใบอนุญาตเป็น Reactor Operator และ Senior Reactor Operator ตามโรงไฟฟ้าพลังงานนิวเคลียร์ทั่วประเทศ ข้อสอบ GFE เป็นข้อสอบแบบเลือกตอบ (Multiple choice items) 2 ชุดเกี่ยวกับ Boiling Water Reactors (BWR) และ Pressurized Water Reactors (PWR) อย่างละ 100 ข้อ ถามความรู้ของผู้สอบเกี่ยวกับ Reactor Theory, Plant Components, และ Thermodynamics ผู้เข้าสอบต้องทำคะแนนอย่างน้อย ร้อยละ 80 จึงจะสามารถผ่านเข้าไปสอบข้อเขียน และสอบปฏิบัติการต่อไป ข้อสอบ GRE ได้รับการพัฒนาให้สอดคล้องกับวัตถุประสงค์ที่ต้องการวัด ครอบคลุมความรู้และความสามารถในการปฏิบัติตามขอบเขตของงานที่ทำและมีความสอดคล้องกับการปฏิบัติงานจริง (Job-related knowledge) การคัดเลือกข้อสอบใช้ Bloom's Taxonomy เป็นเกณฑ์เพื่อจำแนกข้อสอบตามระดับความรู้เป็น 3 ระดับ คือ level 1) ความรู้ระดับพื้นฐาน (Fundamental Knowledge) 2) ความเข้าใจ (Comprehension) level 3) การวิเคราะห์/สังเคราะห์ และการนำไปใช้ หรือประยุกต์ ข้อสอบ GRE ได้รับการพิสูจน์ว่าเป็นข้อสอบที่ได้มาตรฐานระดับชาติ ทั้งขั้นตอนการออกข้อสอบ การควบคุมการสอบ และการทบทวนภายหลังการสอบ

NRC ได้พัฒนาการเลือกข้อสอบ โดยให้ประกอบด้วย 3 กลุ่ม (1) test bank items เป็นข้อสอบที่เคยถูกนำมาใช้สอบมาก่อนและผ่านการพิจารณาความตรง และอำนาจจำแนกแล้ว (2) Modified items เป็นข้อสอบที่ปรับปรุงจากข้อสอบเดิม และ 3) new items ข้อสอบใหม่ ในสัดส่วน 50-40-10 เพื่อให้การสอบยังมีประสิทธิภาพในการจำแนกผู้สอบได้ แม้จะเป็นผู้ที่เคยสอบหลายครั้งแล้ว การศึกษาครั้งนี้ได้เก็บรวบรวมข้อมูลของการสอบเป็นระยะเวลา 5 ปี ตั้งแต่ ปี 1991-1996 จำนวน 28 ครั้ง มีผู้เข้าสอบทั้งหมด 2,064 คน พบว่าข้อสอบที่ได้จากการรวบรวมข้อสอบทั้ง 3 ลักษณะเข้าด้วยกัน ในสัดส่วนที่เหมาะสม สามารถจำแนกหรือแยกแยะผู้สอบได้ดี การใช้ข้อสอบเก่าจากคลังข้อสอบ เปิดโอกาสให้ผู้เข้าสอบได้ฝึกฝน ทบทวนตัวเอง เป็นการส่งเสริมการเรียนรู้ แต่ทำให้

ผู้เข้าสอบอาจคาดเดาได้ หากเคยเห็นข้อสอบข้อนั้นมาก่อน แต่การใช้ข้อสอบประเภท modified items และ new items ร่วมด้วย จะช่วยเพิ่มประสิทธิภาพ หรือแยกแยะผู้สอบได้

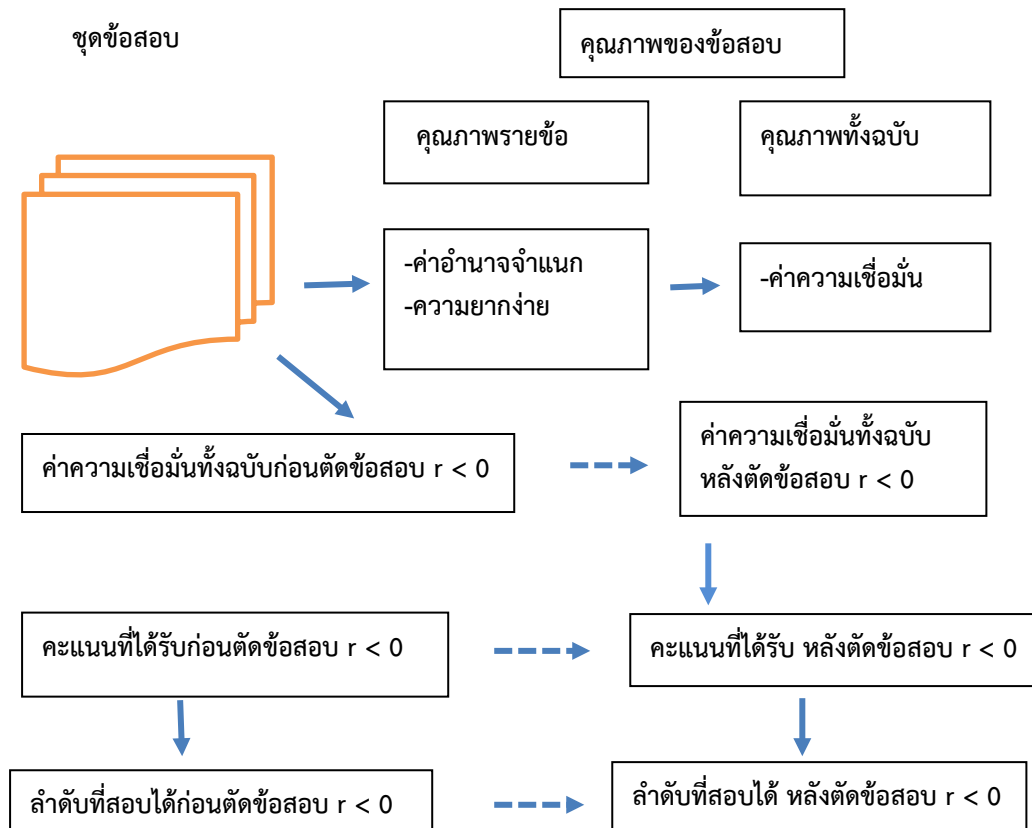
สรุป

จากการทบทวนวรรณกรรม พบว่า เรื่องการวัดและประเมินผลการศึกษาเป็นองค์ประกอบหลักที่สำคัญของการจัดการศึกษา ในปัจจุบันยึดแนวทางการวัดและประเมินผลหลากหลายมิติเพื่อพัฒนาผู้เรียนอย่างรอบด้าน การประเมินผลผู้เรียนแบบอิงกลุ่มด้วยการใช้ข้อสอบชนิดหลายตัวเลือก ยังเป็นที่นิยมใช้ เนื่องจากมีข้อดีหลายประการ จากผลการศึกษาวิจัยพบว่า คุณภาพของข้อสอบขึ้นกับโครงสร้างของระดับการวัด ยิ่งโครงสร้างของข้อสอบที่เขียนขึ้น วัดในระดับการนำไปใช้ การวิเคราะห์ สังเคราะห์ ตามหลักการวัดของ บลูม (Bloom, 1956; Lorin Anderson and et al., 2000; Usova GM.; 1997) สามารถจำแนกผู้เข้าสอบได้ดี เพราะตัวข้อสอบสะท้อนระดับการคิดของผู้ตอบขั้นสูง ทำให้สามารถแยกผู้เรียนอ่อนและเก่งออกจากกันได้ ส่วนรายละเอียดปลีกย่อย เกี่ยวกับ ข้อสอบชนิด 3, 4 และ 5 ตัวเลือก หรือ ลักษณะโจทย์คำถามที่เป็น ชนิดกรณีศึกษา และไม่ใช้กรณีศึกษา ถามแบบ ปลายเปิด หรือ คำถามปลายปิด ให้ความแตกต่างที่ยังไม่ชัดเจน นอกจากนี้ผลการศึกษา ยังสอดคล้องกับธรรมชาติของผู้เรียนในเรื่อง ตัวเลือกที่มีความสั้นยาวใกล้เคียงกัน เนื้อหาของคำตอบใกล้เคียงกัน และทั้งตัวคำตอบถูกและตัวหลง เนื้อความใกล้เคียงกัน จะทำให้ผู้ตอบที่อ่อนใช้การคาดเดายากขึ้น ทำให้ข้อสอบยากขึ้น และมีค่าอำนาจจำแนกสูงขึ้น นอกจากนี้ การลดความคาดเคลื่อนด้วยการเขียนข้อสอบให้มีความชัดเจน ทำเครื่องหมายให้ผู้ตอบสังเกตได้ง่าย ว่าเน้น คำถามปฏิเสธ เช่น ยกเว้น หรืออื่น ๆ จะช่วยป้องกันผู้ตอบที่มีความรู้ ทำข้อสอบผิด ที่เนื่องมาจากการอ่านไม่เข้าใจจากไม่มีความรู้จริงซึ่งจะทำให้การแปลความหมายของผลการทดสอบคลาดเคลื่อน วรรณกรรมและผลการวิจัยที่ผ่านมา สอดคล้องกัน อย่างไรก็ตาม ผู้วิจัยพบว่า ยังไม่มีการวิจัยเกี่ยวกับ การประเมินผลผู้เรียนเมื่อแบบทดสอบที่ใช้วัด ที่มีค่า r ตีลบ โดยเฉพาะจำนวนข้อ r ตีลบหลายๆข้อ แต่พบให้มีการปรับปรุงข้อสอบ ในครั้งต่อไป ผู้วิจัยจึงสนใจ การศึกษาผลของการตัดข้อสอบที่มีค่าอำนาจจำแนกตีลบออกต่อคะแนนของนักศึกษา

กรอบแนวคิดในการวิจัย

จากแผนภูมิ กรอบแนวคิดในการทำวิจัย แสดงให้เห็นถึง ตัวแปรที่ใช้ในการศึกษา คือ คุณภาพของข้อสอบ ประกอบไปด้วย คุณภาพของข้อสอบรายข้อซึ่งในการวิจัยครั้งนี้วัดจาก ค่าความยากง่าย

และค่าอำนาจจำแนก ส่วนคุณภาพของข้อสอบทั้งฉบับวัดจาก ค่าความเชื่อมั่น นอกจากนี้ กรอบแนวคิดยังแสดงถึงการเปลี่ยนแปลงของคะแนนสอบรายวิชา และลำดับที่ของนักศึกษาที่สอบได้ ก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก และการเปลี่ยนแปลงค่าความเชื่อมั่นของข้อสอบทั้งฉบับ ก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก



แผนภูมิที่ 2.6 กรอบแนวคิดในการวิจัย

บทที่ 3 วิธีดำเนินการวิจัย

การวิจัย เรื่องผลของการตัดข้อสอบที่มีอำนาจจำแนกติดลบออกจากคะแนนผลการสอบ และความเชื่อมั่นของข้อสอบทั้งฉบับ: รายวิชากฎหมายและจริยธรรมวิชาชีพ ฉบับนี้ เป็นส่วนหนึ่งของการวิจัยภายใต้ร่มใหญ่ ของ กรณีศึกษาคณะพยาบาลศาสตร์ มหาวิทยาลัยราชภัฏสวนดุสิต จำนวน 11 รายวิชา ที่เปิดสอนในหลักสูตรพยาบาลศาสตรบัณฑิต ปีการศึกษา 2555 ภาคการศึกษา ที่ 1 โดยวิธีดำเนินการวิจัยประกอบไปด้วย

- 1) ประชากร และการเลือกกลุ่มตัวอย่าง
- 2) เครื่องมือการวิจัย และการตรวจสอบคุณภาพของเครื่องมือ
- 3) การเก็บรวบรวมข้อมูล
- 4) การวิเคราะห์ข้อมูล

ประชากรและการเลือกกลุ่มตัวอย่าง

ประชากร คือ ข้อสอบ และคะแนนการสอบรายวิชากฎหมายและจริยธรรมวิชาชีพหลักสูตรพยาบาลศาสตรบัณฑิต

ประชากรทั้งหมดที่เข้าถึงได้ (Accessible population) คือ ข้อสอบ และคะแนนการสอบรายวิชากฎหมายและจริยธรรมวิชาชีพของนักศึกษา หลักสูตรพยาบาลศาสตรบัณฑิต ชั้นปีที่ 4 ที่ลงทะเบียนเรียน ในภาคการศึกษาที่ 1 ปีการศึกษา 2555 จำนวน 60 คน

เครื่องมือในการวิจัยและการตรวจสอบคุณภาพเครื่องมือ

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ ประกอบด้วย แบบบันทึกข้อมูลที่ผู้วิจัยใช้สำหรับเก็บข้อมูลคะแนนวิชากฎหมายและจริยธรรมวิชาชีพ และโปรแกรมการวิเคราะห์ข้อสอบรายข้อ (items Analysis) ที่พัฒนาโดยบริษัท Application Support & Development Department; ControlData (Thailand) Ltd. อาศัยหลักในการคำนวณและวิเคราะห์ข้อสอบโดยใช้ทฤษฎีการวิเคราะห์แบบทดสอบแบบประเพณีนิยม (Classical Test Theory) เพื่อศึกษาคุณภาพของข้อสอบทั้งรายข้อและรวมทั้งฉบับ ทฤษฎีการวิเคราะห์แบบทดสอบแบบประเพณีนิยมนี้ เหมาะสำหรับแบบทดสอบแบบเลือกตอบ (Multiple choice test) ของแบบทดสอบแบบอิงกลุ่ม (Norm reference test) โดยทฤษฎีการวิเคราะห์แบบทดสอบนั้น จะมีหลายเทคนิคที่ใช้กันอยู่ แต่ในการวิจัยครั้งนี้ใช้เทคนิค 27 % ซึ่งสำนักส่งเสริมวิชาการและงานทะเบียนของมหาวิทยาลัยเป็นผู้วิเคราะห์ให้

การเก็บรวบรวมข้อมูล

1) เข้าร่วมประชุมการออกแบบการวิจัย และการวางแผนการเก็บข้อมูลกับประธานคณะวิจัยและทีมผู้วิจัยทุกรายวิชา

2) ดำเนินการเรียนการสอนตามที่กำหนดไว้ ในการออกแบบการเรียนการสอนรายวิชา (มคอ.3) มีการออกข้อสอบตามตารางการวิเคราะห์หลักสูตร (Test blueprint) วิพากษ์ข้อสอบโดยทีมผู้สอน ที่เป็นอาจารย์อาวุโสและมีประสบการณ์การสอนในรายวิชามากกว่า 15 ปี อีก 2 ท่าน ดำเนินการสอบกลางภาค และปลายภาค รวมทั้งการตัดสินเกรดตามระเบียบการวัดและประเมินผลของมหาวิทยาลัย

3) ส่งข้อสอบไปวิเคราะห์ที่สำนักส่งเสริมวิชาการ และงานทะเบียนของมหาวิทยาลัยภายหลังการสอบ

4) รวบรวมผลการวิเคราะห์ข้อสอบรายวิชา ภาควิชา และจริยธรรมวิชาชีพ ทั้งการสอบกลางภาค และปลายภาค นำมาวิเคราะห์ข้อมูลตามวัตถุประสงค์ของการวิจัย

การวิเคราะห์ข้อมูล

1) การวิเคราะห์หาคุณภาพของข้อสอบรายวิชา ภาควิชา และจริยธรรมวิชาชีพ โดยการหาค่าความยากง่ายและอำนาจจำแนกเป็นรายข้อ ของข้อสอบกลางภาคและปลายภาคทุกชุด โดยใช้โปรแกรมการวิเคราะห์ข้อสอบรายข้อ (Items Analysis) ที่พัฒนาโดยบริษัท Application Support & Development Department; Control Data (Thailand) Ltd. ดำเนินการโดยสำนักส่งเสริมวิชาการและงานทะเบียนของมหาวิทยาลัย หลังจากนั้นนำผลการวิเคราะห์ มาแจกแจง จัดกลุ่ม และคำนวณหาค่าร้อยละ

2) การวิเคราะห์การเปลี่ยนแปลงของคะแนนสอบรายวิชา และอันดับ ก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก โดยมีการดำเนินการ ดังต่อไปนี้

(1) การเปรียบเทียบคะแนนเฉลี่ย ที่นักศึกษาได้รับในรายวิชา ภาควิชา และจริยธรรมวิชาชีพ ก่อนและหลังการตัดข้อสอบที่มีค่า r ติดลบออก โดยสถิติ Pair t-test (one-tailed)

(2) การทดสอบการเปลี่ยนแปลงของลำดับคะแนนที่นักศึกษาได้รับในรายวิชา ภาควิชา และจริยธรรมวิชาชีพ ก่อนและหลังการตัดข้อสอบที่มีค่า r ติดลบออก โดยใช้สถิติ Signed Rank Test

3) เปรียบเทียบค่าความเชื่อมั่นของข้อสอบทั้งฉบับ ก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก

บทที่ 4 ผลการวิจัย

การนำเสนอผลการวิเคราะห์ข้อมูลการวิจัย เรื่อง ผลของการตัดข้อสอบที่มีอำนาจจำแนกติดลบออกต่อคะแนนการสอบ และความเชื่อมั่นของข้อสอบทั้งฉบับ: รายวิชากฎหมาย และจริยธรรมวิชาชีพนี้ เป็นการประมวลผลการศึกษารายวิชา ในประเด็นต่างๆ ดังนี้

- 1) ข้อมูลทั่วไป
- 2) คุณภาพของข้อสอบวิชากฎหมาย และจริยธรรมวิชาชีพ
- 3) การเปลี่ยนแปลงของคะแนนสอบรายวิชา และอันดับ ก่อน และหลังการตัดข้อสอบที่มีอำนาจจำแนกติดลบออก
- 4) การเปลี่ยนแปลงค่าความเชื่อมั่นของข้อสอบทั้งฉบับ ก่อน และหลังการตัดข้อสอบที่มีอำนาจติดลบออก

ข้อมูลทั่วไป

การวิจัยครั้งนี้ เป็นการศึกษาย้อนหลังจากข้อสอบชุดกลางภาค และชุดปลายภาค รายวิชา กฎหมาย และจริยธรรมวิชาชีพ ของนักศึกษาพยาบาลศาสตร์ ชั้นปีที่ 4 ปีการศึกษา 2555 ภาค การศึกษาที่ 1

ตารางที่ 4.1 ชุดข้อสอบ จำนวนนักศึกษา และจำนวนข้อสอบในแต่ละชุด ที่ใช้ในการศึกษาวิจัย

ชุดข้อสอบ	จำนวนนักศึกษาที่เข้าสอบ (คน)	จำนวนชุดข้อสอบ (ชุด)	จำนวนข้อคำถามในแต่ละชุด (ข้อ)
ชุดข้อสอบกลางภาค	60	1	40
ชุดข้อสอบปลายภาค	60	1	40

จากตารางที่ 4.1 นักศึกษาพยาบาลศาสตร์ ชั้นปีที่ 4 มีจำนวน 60 คน เข้าสอบในวิชากฎหมาย และจริยธรรมวิชาชีพครบทุกคน จำนวน 60 คน จำนวนชุดข้อสอบกลางภาค และปลายภาคมีจำนวนอย่างละ 1 ชุด จำนวนข้อคำถาม ในแต่ละชุดมีจำนวน เท่ากัน คือ 40 ข้อ

คุณภาพข้อสอบรายวิชากฎหมาย และจริยธรรมวิชาชีพ

ผลการวิเคราะห์ข้อสอบกลางภาค และปลายภาค รายวิชากฎหมาย และจริยธรรมวิชาชีพ รวม 2 ชุด ในปีการศึกษา 2555 ภาคการศึกษาที่ 1 โดยใช้ค่าความยากง่าย และค่าอำนาจจำแนกของข้อสอบเป็นดัชนีแสดงถึงคุณภาพข้อสอบรายข้อ และใช้ค่าสัมประสิทธิ์ความเชื่อมั่น แสดงถึงคุณภาพของข้อสอบทั้งฉบับ โดยมีผลการวิเคราะห์ดังนี้

ตารางที่ 4.2 จำนวน และร้อยละ ของข้อสอบชุดกลางภาค แต่ละข้อ จำแนกตามค่าความยากง่าย (P)

ค่าความยากง่าย (P)	ความหมาย	จำนวนข้อสอบ	ร้อยละ
0.80 – 1.00	ง่ายมาก	8	20
0.60 – 0.79	ค่อนข้างง่าย	10	25
0.40 – 0.59	ความยากง่ายปานกลาง	12	30
0.20 – 0.39	ค่อนข้างยาก	8	20
0.00 – 0.19	ยากมาก	2	5
รวมข้อสอบ		40	100

จากตารางที่ 4.2 เมื่อพิจารณาคคุณภาพของข้อสอบชุดกลางภาค เป็นรายข้อ พบว่า ส่วนใหญ่เป็นข้อสอบที่มีความยากง่ายปานกลาง คิดเป็นร้อยละ 30 และหากพิจารณาข้อสอบที่มีค่าความยากง่ายในเกณฑ์ที่เหมาะสม (P อยู่ระหว่าง 0.20 - 0.79) จะมีจำนวนรวม 30 ข้อ คิดเป็นร้อยละ 75 และมีข้อสอบที่ง่ายมาก และยากมาก จำนวนรวมกัน 10 ข้อ คิดเป็นร้อยละ 25

ตารางที่ 4.3 จำนวนและร้อยละ ของข้อสอบชุดปลายภาคแต่ละข้อ จำแนกตามค่าความยากง่าย (P)

ค่าความ ยากง่าย (P)	ความหมาย	จำนวน ข้อสอบ	ร้อยละ
0.80 – 1.00	ง่ายมาก	11	27.5
0.60 – 0.79	ค่อนข้างง่าย	4	10
0.40 – 0.59	ความยากง่ายปานกลาง	14	35
0.20 – 0.39	ค่อนข้างยาก	8	20
0.00 – 0.19	ยากมาก	3	7.5
รวมข้อสอบ		40	100

จากตารางที่ 4.3 เมื่อพิจารณาคคุณภาพของข้อสอบชุดปลายภาคเป็นรายข้อ พบว่า ส่วนใหญ่เป็นข้อสอบที่มีความยากง่ายปานกลาง คิดเป็นร้อยละ 35 และหากพิจารณาข้อสอบที่มีค่าความยากง่ายในเกณฑ์ที่เหมาะสม (P อยู่ระหว่าง 0.20 - 0.79) จะมีจำนวนรวม 26 ข้อ คิดเป็นร้อยละ 65 มีข้อสอบที่ง่ายมากและยากมาก จำนวน 14 ข้อ คิดเป็นร้อยละ 35

ตารางที่ 4.4 จำนวน และร้อยละ ของข้อสอบชุดกลางภาค จำแนกตาม ค่าอำนาจจำแนก (r) ของข้อสอบแต่ละข้อ

ค่าอำนาจ จำแนก (r)	ความหมาย	จำนวนข้อสอบ	ร้อยละ
0.80-1.00	จำแนกได้ดีมาก	-	-
0.60-0.79	จำแนกได้ดี	-	-
0.40-0.59	จำแนกได้ปานกลาง	3	7.5
0.20-0.39	จำแนกได้พอใช้	20	50
0.01-0.19	จำแนกได้ต่ำ	12	30
-1.00-0.00	จำแนกไม่ดี	5	12.5
รวมข้อสอบ		40	100.00

จากตารางที่ 4.4 เมื่อพิจารณาคุณภาพของข้อสอบชุดกลางภาคเป็นรายข้อ พบว่า ส่วนใหญ่เป็นข้อสอบที่มีค่าอำนาจจำแนกได้พอใช้ จำนวน 20 ข้อ คิดเป็นร้อยละ 50 และหากพิจารณาข้อสอบที่มีค่าอำนาจจำแนกในเกณฑ์ที่เหมาะสม ($r \geq 0.2$) ขึ้นไปพบว่า มีจำนวนรวม 23 ข้อ คิดเป็นร้อยละ 57.5

ตารางที่ 4.5 จำนวน และร้อยละ ของข้อสอบชุดปลายภาค จำแนกตาม ค่าอำนาจจำแนก (r) ของข้อสอบแต่ละข้อ

ค่าอำนาจ จำแนก (r)	ความหมาย	จำนวนข้อสอบ	ร้อยละ
0.80-1.00	จำแนกได้ดีมาก	-	-
0.60-0.79	จำแนกได้ดี	-	-
0.40-0.59	จำแนกได้ปานกลาง	2	5
0.20-0.39	จำแนกได้พอใช้	17	42.5
0.01-0.19	จำแนกได้ต่ำ	19	47.5
-1.00-0.00	จำแนกไม่ดี	2	5
รวมข้อสอบ		40	100.00

จากตารางที่ 4.5 เมื่อพิจารณาคุณภาพของข้อสอบชุดปลายภาค เป็นรายข้อ พบว่า ส่วนใหญ่เป็นข้อสอบที่มีค่าอำนาจจำแนกได้พอใช้ จำนวน 17 ข้อ คิดเป็นร้อยละ 42.5 และมีข้อสอบที่มี

อำนาจจำแนกปานกลาง จำนวน 2 ข้อ คิดเป็นร้อยละ 5 และหากพิจารณาข้อสอบที่มีค่าอำนาจจำแนกในเกณฑ์ที่เหมาะสม ($r \geq 0.2$) ขึ้นไปพบว่า มีจำนวนรวม 19 ข้อ คิดเป็นร้อยละ 47.5

ตารางที่ 4.6 ภาพรวมของคุณภาพข้อสอบ แสดงค่าความยากง่าย และค่าอำนาจจำแนกเฉลี่ยที่เหมาะสม รวมทั้งจำนวนข้อสอบที่มีค่าอำนาจจำแนก (r) ติดลบ

ชุดข้อสอบ	จำนวนข้อ	จำนวนและร้อยละของข้อสอบที่มีค่าความยากง่าย (P) ระหว่าง 0.20 - 0.80	จำนวนและร้อยละของข้อสอบที่มีค่า $r \geq 2.0$	จำนวนและร้อยละของข้อสอบที่มีค่า r ติดลบ (r^-)
กลางภาค	40	30 (75)	23 (57.5)	1 (2.5)
ปลายภาค	40	26 (65)	19 (47.5)	1 (2.5)

จากตารางที่ 4.6 ภาพรวมของคุณภาพข้อสอบทั้งชุดกลางภาค และชุดปลายภาค พบว่า มีจำนวนข้อสอบที่มีค่าความยากง่าย (P) เหมาะสม $P = 0.20 - 0.80$ จำนวนทั้งหมด 30, 26 ข้อ คิดเป็นร้อยละ 75, 65 ตามลำดับ มีจำนวนข้อสอบที่มีค่า r เหมาะสม $r \geq 2.0$ ทั้งหมด 23, 19 ข้อ คิดเป็นร้อยละ 57.5, 47.5 ตามลำดับ โดยมีข้อสอบที่มีค่า r ติดลบ ชุดละ 1 ข้อ คิดเป็น ร้อยละ 2.5

ตารางที่ 4.7 ค่าสัมประสิทธิ์ความเชื่อมั่นของข้อสอบทั้งฉบับ ในข้อสอบชุดกลางภาคและชุดปลายภาคก่อนตัด ค่า r ติดลบ

ชุดข้อสอบ	ค่าสัมประสิทธิ์ความเชื่อมั่นของข้อสอบทั้งฉบับ	การแปลผล
กลางภาค	0.48	ไม่น่าเชื่อถือ จำเป็นต้องปรับปรุง และไม่ควรใช้เป็นหลักในการตัดสินเกรด
ปลายภาค	0.40	ไม่น่าเชื่อถือ จำเป็นต้องปรับปรุง และไม่ควรใช้เป็นหลักในการตัดสินเกรด

จากตารางที่ 4.7 พบว่าค่าสัมประสิทธิ์ความเชื่อมั่นของข้อสอบทั้งฉบับ ในข้อสอบชุดกลางภาค และชุดปลายภาคก่อนตัด ค่า r ติดลบ มีค่าใกล้เคียงกัน ที่ ระดับต่ำกว่า 0.50 ซึ่งจำเป็นต้องมีการปรับปรุง

การเปลี่ยนแปลงของคะแนนสอบรายวิชา และอันดับ ก่อน และหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก

ภายหลังจากได้รับผลการวิเคราะห์ข้อสอบ หากมีข้อสอบที่มีค่าอำนาจจำแนกติดลบ(r^-) แสดงว่า คุณภาพของข้อสอบข้อนั้นๆ ไม่สามารถจำแนกผู้เรียนเก่ง และอ่อนได้ โดยเป็นผลมาจาก

ผู้เรียนอ่อนตอบข้อสอบได้มากกว่าผู้เรียนเก่ง จะมากหรือน้อยเท่าใด จำนวนค่า r ติดลบ จะเป็นตัวบ่งชี้ ซึ่งจะแปรผันตาม กับจำนวนผู้เรียนอ่อนที่ตอบได้มากกว่าผู้เรียนเก่ง หากทั้งสองกลุ่มทำข้อสอบได้เท่ากัน ค่า r จะมีค่าเป็น 0 ดังนั้นในการศึกษาครั้งนี้ ผู้วิจัยจึงได้ตั้งข้อสอบที่มีค่า r ติดลบออก และส่งข้อสอบชุดเดียวกัน กลับไปคำนวณผล และวิเคราะห์ใหม่ หลังจากที่ได้ผลคืน นำคะแนนรวมจากการตรวจ ครั้งที่ 2 (ซึ่งนักศึกษาบางคนเมื่อตั้งข้อสอบซึ่ง คะแนน r ติดลบ ออก บางคนคะแนนเพิ่มขึ้น บางคนลดลง บางคนไม่เปลี่ยนแปลง ขึ้นกับว่าเดิมนักศึกษาทำข้อสอบที่มีค่า r ติดลบ ตอบผิดหรือถูก) มาคำนวณเทียบบัญญัติไตรยางค์กลับไปสู่ คะแนนเต็มส่วนในครั้งแรก เพื่อให้ได้คะแนนที่นักศึกษาแต่ละคน มีค่าคะแนนเต็มเท่าเดิม (ก่อนตัดข้อ r ติดลบออก) ดังตารางที่ 4.8 และทดสอบการเปลี่ยนแปลงของลำดับคะแนนที่นักศึกษาได้รับในรายวิชา กฎหมาย และจริยธรรมวิชาชีพ ก่อน และหลังการตัดข้อสอบที่มี ค่า r ติดลบ โดยสถิติ Signed Rank Test ดังตารางที่ 4.9

ตารางที่ 4.8 ผลการเปรียบเทียบคะแนนเฉลี่ย คิดเป็นร้อยละ ที่นักศึกษาได้รับในรายวิชากฎหมาย และจริยธรรมวิชาชีพ ก่อน และหลังการตัดข้อสอบที่มีค่า r ติดลบออกทั้งจากชุดข้อสอบกลางภาค และชุดข้อสอบปลายภาค โดยใช้สถิติ Pair t-test (one – tailed) (n=60)

ชุดข้อสอบ	จำนวนข้อ	ก่อนตัดข้อที่มีค่า r - ออก		หลังตัดข้อที่มีค่า r - ออก		คะแนนที่เพิ่มขึ้น	จำนวน, ร้อยละของข้อสอบ r -	Df = n-1	Paired t-test	P – Value
		Mean	S.D.	Mean	S.D.					
กลางภาค	40	22.53	3.86	22.71	4.06	0.18	1 (2.5)	59	.004	<0.01
ปลายภาค	40	22.52	3.44	22.78	3.53	0.26	1 (2.5)	59	.000	<0.01

จากตารางที่ 4.8 พบว่า คะแนนเฉลี่ยของข้อสอบชุดกลางภาค และชุดปลายภาค ของนักศึกษาทั้งกลุ่ม มีค่าใกล้เคียงกัน ที่ 22.53 คิดเป็นร้อยละ 56.33 และ 22.52 คิดเป็นร้อยละ 56.30 ตามลำดับ ซึ่งยังไม่ถึง ร้อยละ 60 ซึ่งเป็นเกณฑ์สอบผ่านของกลุ่มวิชาพื้นฐานวิชาชีพ และเมื่อเปรียบเทียบคะแนนการสอบกลางภาคของนักศึกษา ก่อนตัดข้อสอบที่มีค่า r ติดลบออก และภายหลังที่ตัดข้อสอบที่มีค่า r ติดลบออก โดยใช้สถิติ Paired t-test (one – tailed) พบว่า นักศึกษามีคะแนนเฉลี่ยสูงขึ้นอย่างมีนัยสำคัญทางสถิติ ($t = 0.00$, $df = 59$, $P \text{ value} < 0.005$)

ตารางที่ 4.9 ผลการทดสอบการเปลี่ยนแปลงของลำดับคะแนนที่นักศึกษาได้รับในรายวิชากฎหมายและจริยธรรมวิชาชีพ ก่อน และหลังการตัดข้อสอบที่มี ค่า r ติดลบออกโดยสถิติ Signed Rank Test (n=60)

ชุดข้อสอบ	จำนวนนักศึกษาทั้งหมด	จำนวนนักศึกษาที่คะแนนเพิ่มขึ้น		จำนวนนักศึกษาที่คะแนนคงเดิม		จำนวนนักศึกษาที่คะแนนลดลง		Signed Rank Test	P value
		จำนวน	%	จำนวน	%	จำนวน	%		
กลางภาค	60	37	61.67	-	-	23	38.33	000	<0.01
ปลายภาค	60	43	71.66	-	-	17	28.34	000	<0.01

จากตารางที่ 4.9 ผลการทดสอบการเปลี่ยนแปลงของลำดับคะแนนที่นักศึกษาได้รับในรายวิชากฎหมายและจริยธรรมวิชาชีพ ทั้งชุดข้อสอบกลางภาค และชุดข้อสอบปลายภาค ก่อน และหลังการตัดข้อสอบ ที่มีค่า r ติดลบออก โดยสถิติ Signed Rank Test (n=60) พบว่า ลำดับคะแนนที่นักศึกษาได้รับมีการเปลี่ยนแปลงลำดับที่อย่างมีนัยสำคัญทางสถิติ ที่ $\alpha = 0.05$ (p value <0.01)

การเปลี่ยนแปลงค่าความเชื่อมั่นของข้อสอบทั้งฉบับ ก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก

เพื่อศึกษาการเปลี่ยนแปลงค่าความเชื่อมั่นของข้อสอบทั้ง 2 ชุด ได้แก่ ชุดกลางภาค และชุดปลายภาค ในรายวิชากฎหมาย และจริยธรรมวิชาชีพ ก่อน และหลังการตัดข้อสอบที่มีค่า r- ออก ดังแสดงในตารางที่ 4.10

ตารางที่ 4.10 การเปลี่ยนแปลงค่าความเชื่อมั่นของข้อสอบ ชุดกลางภาค และชุดปลายภาค ในรายวิชากฎหมาย และจริยธรรมวิชาชีพ ภายหลังจากที่มีการตัดข้อสอบ ที่มีค่า r ติดลบออก

ชุดข้อสอบ	จำนวนข้อสอบ (ข้อ)	ร้อยละของข้อสอบที่ค่า r-	จำนวนนักศึกษาที่เข้าสอบ(คน)	ค่าความเชื่อมั่นของข้อสอบ	ค่าความเชื่อมั่นหลังการตัดข้อที่มีค่า r- ออก	ค่าความเชื่อมั่นที่เพิ่มขึ้นหรือลดลง
กลางภาค	40	2.50	60	0.48	0.51	0.03
ปลายภาค	40	2.50	60	0.40	0.43	0.03

จากตารางที่ 4.10 การเปลี่ยนแปลงของค่าความเชื่อมั่นของข้อสอบทั้งชุดกลางภาค และชุดปลายภาค ในรายวิชากฎหมาย และจริยธรรมวิชาชีพ ภายหลังจากที่มีการตัดข้อสอบ ที่มีค่า r ติดลบออก มีค่าเพิ่มขึ้นเล็กน้อย ทั้งสองชุด = 0.03

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

สรุปผลการวิจัย

การวิจัยครั้งนี้ เป็นการศึกษาย้อนหลังจากชุดข้อสอบกลางภาค และชุดข้อสอบปลายภาค จำนวน 2 ชุด ที่ใช้ในการสอบนักศึกษาพยาบาล ชั้นปีที่ 4 รายวิชากฎหมาย และจริยธรรมวิชาชีพ หลักสูตร พยาบาลศาสตร์บัณฑิต ภาคการศึกษาที่ 1 ปีการศึกษา พ.ศ.2555 จำนวน 60 คน โดยมีวัตถุประสงค์เพื่อศึกษา 1) คุณภาพข้อสอบรายวิชากฎหมาย และจริยธรรมวิชาชีพ (Laws and Professional Ethics Subject) 2) การเปลี่ยนแปลงของคะแนนการสอบของนักศึกษารายบุคคล ก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก 3) ศึกษาการเปลี่ยนแปลงของค่าความเชื่อมั่นของข้อสอบ ก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก

ผลการศึกษาวิเคราะห์ชุดข้อสอบกลางภาคและปลายภาค พบว่า

1) ด้านคุณภาพข้อสอบ เมื่อพิจารณาจากดัชนีความยากง่าย (P) พบว่า ชุดข้อสอบกลางภาค และชุดข้อสอบปลายภาค มีข้อสอบที่มีค่าความยากง่ายในเกณฑ์ที่เหมาะสม (P อยู่ระหว่าง 0.20 – 0.80) จำนวนรวม 30, 25 ข้อ คิดเป็นร้อยละ 75, 65 ตามลำดับ ด้านค่าอำนาจจำแนก (r) พบว่า ชุดข้อสอบกลางภาค และชุดข้อสอบปลายภาค มีข้อสอบที่มีค่าอำนาจจำแนกในเกณฑ์ที่เหมาะสม ($r \geq 0.20$) จำนวนรวม 23, 19 ข้อ คิดเป็นร้อยละ 57.5 และร้อยละ 47.5 ตามลำดับ ทั้งนี้ ข้อสอบที่มีค่าอำนาจจำแนกติดลบ ($r = -1.00 - 0.00$) ในชุดข้อสอบกลางภาค และชุดข้อสอบปลายภาค มีจำนวนชุดละ 1 ข้อ คิดเป็นร้อยละ 2.5 ส่วนคุณภาพของข้อสอบทั้งฉบับ พบว่า ข้อสอบกลางภาคมีค่าความเชื่อมั่น 0.48 และเมื่อตัดข้อสอบที่มีค่า r ติดลบออก ทำให้ค่าความเชื่อมั่นเพิ่มขึ้นเป็น 0.51 ส่วนข้อสอบปลายภาค มีค่าความเชื่อมั่น = 0.40 และเมื่อตัดข้อสอบที่มีค่า r ติดลบออก ทำให้ค่าความเชื่อมั่นเพิ่มขึ้นเป็น 0.43

2). ผลการเปรียบเทียบคะแนนเฉลี่ยที่นักศึกษาได้รับ ในรายวิชากฎหมาย และจริยธรรมวิชาชีพ จากข้อสอบทั้งชุดกลางภาค และชุดปลายภาค ก่อนและหลังการตัดข้อสอบที่มีค่า r ติดลบออก โดยสถิติ Paired t-test (one-tailed) พบว่า นักศึกษามีคะแนนเฉลี่ยสูงขึ้นอย่างมีนัยสำคัญทางสถิติ ที่ $\alpha = 0.05$ ($t=0.00$, $df = 59$, $P \text{ value} < 0.001$)

3). ผลการทดสอบการเปลี่ยนแปลงของลำดับคะแนนที่นักศึกษาได้รับ ในรายวิชากฎหมาย และจริยธรรมวิชาชีพ ก่อนและหลังการตัดข้อสอบที่มีค่า r ติดลบออก โดยสถิติ Signed Rank Test พบว่า ลำดับคะแนนที่นักศึกษาได้รับมีการเปลี่ยนแปลงลำดับที่อย่างมีนัยสำคัญทางสถิติ ที่ $\alpha = 0.05$ ($P \text{ value} < 0.01$) ในข้อสอบทั้ง 2 ชุด

อภิปรายผล

1. ด้านคุณภาพข้อสอบรายข้อ

เมื่อพิจารณาค่าความยากง่าย (P) พบว่าข้อสอบรายข้อทั้งสองชุดมีความยากง่ายอยู่ในเกณฑ์เหมาะสมสูง คือ ข้อสอบส่วนใหญ่ มีจำนวนข้อสอบที่มีค่าความยากง่ายปานกลาง (P อยู่ระหว่าง 0.40 - 0.59) มากที่สุด เป็นไปตามทฤษฎีที่ว่าข้อสอบที่ดีควรมีค่า ระดับความยากง่าย เท่ากับ 0.5 (บุญเชิด ภิญโญนนตพงษ์, 2552, น.70) โดยให้เหตุผลว่า ข้อสอบที่มีคุณภาพควรมีผู้ตอบถูกประมาณร้อยละ 50 หรือครึ่งหนึ่งตอบถูก และครึ่งหนึ่งตอบผิด ข้อสอบที่ง่ายเกินไปไม่มีประโยชน์เพราะคนเก่งและคนไม่เก่งตอบถูกเหมือนกันหมด ข้อสอบที่ยากมากเกินไปถือว่าไม่มีประโยชน์เช่นกัน เพราะมีผู้ตอบถูกเพียงไม่กี่คน แสดงว่าข้อสอบไม่สามารถทำให้ผู้ตอบแสดงคุณลักษณะที่ต้องการวัดออกมาได้

เมื่อพิจารณาถึงโครงสร้างค่าความยากง่ายของข้อสอบทั้งฉบับ พบว่า 2 ชุด มีการกระจายของข้อสอบที่มีความยากง่ายค่อนข้างเหมาะสม (เยาวดี รางชัยกุล วิบูลย์ศรี, 2552, น.144-145) คือ มีข้อสอบง่ายปานกลางมากที่สุด จำนวนร้อยละ 30, 35 รองลงมา เป็นข้อสอบที่ค่อนข้างง่าย และค่อนข้างยาก ใกล้เคียงกัน ร้อยละ 25, 20 ในชุดกลางภาค ส่วนชุดปลายภาค มีข้อสอบที่ค่อนข้างง่ายมากกว่า ข้อสอบค่อนข้างยากเล็กน้อย

เมื่อพิจารณาค่าอำนาจจำแนก (r) พบว่า ข้อสอบทั้งสองชุด มีจำนวนข้อที่มีค่าอำนาจจำแนก อยู่ระหว่าง 0.20 - 0.39 สูงสุด ได้แก่ ร้อยละ 50 และร้อยละ 47.5 ตามลำดับ ซึ่งยังอยู่ในเกณฑ์มาตรฐานระดับพอใช้ ซึ่งควรมีการพัฒนาให้ค่าอำนาจจำแนกมีระดับสูงขึ้น

ผลที่วิเคราะห์ได้แตกต่างจากหลักการที่ว่า ข้อสอบที่ดีจะมีระดับความยากง่าย เท่ากับ 0.5 จะทำให้เกิดค่าอำนาจจำแนกสูงสุด และข้อสอบทั้งฉบับจะมีความสัมพันธ์กับความเชื่อมั่นสูง (เยาวดี รางชัยกุล วิบูลย์ศรี, 2552, น.144 -145) โดยครั้งนี้พบว่า ถึงแม้ค่าความยากง่ายของข้อสอบ 2 ชุด มีค่าความยากง่ายอยู่ในเกณฑ์มาตรฐานสูง แต่อำนาจจำแนกเฉลี่ยของข้อสอบทั้ง 2 ชุด มีค่าไม่สูงมาก อาจเนื่องมาจากสาเหตุดังนี้

1) จำนวนศึกษาพยาบาลที่ทำข้อสอบมีจำนวนน้อย (60 คน) ซึ่งใกล้เคียงกับเหตุผลที่ว่า เมื่อชั้นมีขนาดเล็ก เช่น ต่ำกว่า 50 คน จะทำให้อำนาจจำแนกลดลง เพราะจำนวนผู้เรียนเก่ง และผู้เรียนอ่อนจะแยกลักษณะเป็นกลุ่มก้อนไม่ชัดเจน (บุญชม ศรีสะอาด, 2514, น.58-66)

2) ธรรมชาติของวิชาทำให้ความสอดคล้องภายในของชุดข้อสอบมีน้อยกว่าปกติ รายวิชา ภูมิศาสตร์และจริยธรรมวิชาชีพ สำหรับนักศึกษาพยาบาลศาสตร์ชั้นปีที่ 4 มีลักษณะธรรมชาติของ 2 วิชา มารวมกัน จำนวน 2 หน่วยกิต ได้แก่ วิชาภูมิศาสตร์กับวิชาจริยธรรม วิชาภูมิศาสตร์ มีลักษณะของความเป็นรูปธรรมสูง ใช้กระบวนการคิดซึ่งต้องใช้ความจำเป็นฐานก่อนที่จะนำไปสู่ความเข้าใจ การประยุกต์ใช้ และการคิดวิเคราะห์เปรียบเทียบ ให้เข้าสถานการณ์ที่เป็นประเด็นภูมิศาสตร์ ส่วนวิชาจริยธรรมเป็นวิชาเกี่ยวกับนามธรรม ฝึกความรู้สึกขึ้นขมยินดีถึงความดี ความงาม ในคุณค่าของมนุษย์ นักศึกษาจำนวนบางส่วนมีความถนัดในส่วนของวิชาภูมิศาสตร์และจริยธรรมวิชาชีพไม่เท่ากัน นักศึกษาบางกลุ่มเรียนถนัดเรียนวิชาภูมิศาสตร์ นักศึกษาบางกลุ่มถนัดเรียนวิชาจริยธรรมทำให้ผลการสอบพบว่าความสอดคล้องภายในต่ำ เนื่องจากการคำนวณค่า r ติดลบ ด้วยคอมพิวเตอร์ซึ่งจะคำนวณโดยใช้คะแนนของผู้สอบทุกคน คำนวณจากสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน ระหว่างคะแนนที่

ผู้ทำข้อสอบได้ในข้อนั้นๆกับคะแนนรวมของข้อสอบข้ออื่นในฉบับนั้น (Office of Education Assessment, University of Washington, 2005) เมื่อความสอดคล้องภายในลดลง ค่าอำนาจจำแนกจึงลดลงด้วย รวมทั้งค่าสัมประสิทธิ์ความเชื่อมั่น (reliability) (บุญเชิด ภิญญอนันตพงษ์, 2552, น.84)

3) จำนวนข้อสอบมีน้อย ข้อสอบในแต่ละชุด มีจำนวนข้อสอบ 40 ข้อ (ตามเพื่อให้สอดคล้องกับชุดข้อสอบตาม Test Blue Print ของสภาการพยาบาลซึ่งเป็นการสอบขึ้นทะเบียนรับใบอนุญาตประกอบวิชาชีพ ซึ่งมี จำนวน 40 ข้อเช่นกัน) เพราะยิ่งจำนวนข้อสอบมากขึ้น ตั้งแต่ 80 ข้อขึ้นไปจะทำให้ค่าอำนาจจำแนกสูงขึ้นด้วย (อรวรรณ สุวรรณรัตน์, 2552, น.387) ในการทำชุดข้อสอบสำหรับการสอบครั้งต่อไป ควรเพิ่มจำนวนข้อสอบให้มากขึ้น เพื่อลดจุดอ่อนในเรื่องนี้

4) ความรู้ของนักศึกษาใกล้เคียงกัน จากกระบวนการเรียนการสอนในชั้นเรียน ทำให้ค่าอำนาจจำแนกต่ำลง (อรวรรณ สุวรรณรัตน์, 2552, น.387) รวมทั้งระดับความรู้ความสามารถของนักศึกษาในกลุ่มที่ทำแบบทดสอบว่า มีความสามารถเฉลี่ยทั้งกลุ่ม ในระดับ อ่อน – เก่งมาก มีลักษณะอย่างไร ซึ่งมีผลทำให้ค่าอำนาจจำแนกของข้อสอบชุดเดียวกัน เมื่อใช้ทดสอบผู้เรียนกลุ่มต่างกันได้ค่าไม่เท่ากัน (ชยุตม์ ภิรมย์สมบัติ, 2555, 12) เหตุผล เห็นได้จาก คะแนนเฉลี่ยของข้อสอบชุดกลางภาค และชุดปลายภาค ของนักศึกษาทั้งกลุ่ม มีค่าใกล้เคียงกัน ที่ 22.53 คิดเป็นร้อยละ 56.33 และ 22.52 คิดเป็นร้อยละ 56.30 ตามลำดับ ซึ่งยังไม่ถึง ร้อยละ 60 ซึ่งเป็นเกณฑ์สอบผ่านของกลุ่มวิชาพื้นฐานวิชาชีพ สะท้อนได้ว่า วิชานี้ยากสำหรับนักศึกษา นอกจากนี้ผู้วิจัยซึ่งเป็นผู้สอนหลักได้สัมผัสกับจากกระบวนการเรียนรู้ของผู้เรียน ลักษณะวิชากฎหมายซึ่งเป็นสายทางสังคมศาสตร์และวิชาจริยธรรมเป็นสายปรัชญา เป็นการเริ่มต้นเรียนครั้งแรกซึ่งแตกต่างไปจากวิชาทางการพยาบาลและวิชาพื้นฐานปรีคลินิกทางการแพทย์ที่นักศึกษาค้นเคยมากกว่า การเรียนให้เข้าถึงหัวใจของแต่ละวิชาแตกต่างกัน นอกจากนี้ จำนวนหน่วยกิตมีเพียง 2 หน่วยกิตซึ่งถือว่าน้อย เมื่อเทียบกับเนื้อหาสาระบางสถาบันกำหนดไว้ 3 หน่วยกิต ความไม่คุ้นเคย และเรียนภายในเวลาที่จำกัดทำให้คะแนนสอบ มีค่าเฉลี่ยค่อนข้างต่ำ ผลการสอบครั้งแรกนักศึกษาก่อนใหญ่สอบไม่ผ่านเกณฑ์คะแนนร้อยละ 60 ต้องสอนซ่อมเสริม และออกข้อสอบชุดใหม่เพื่อสอบซ่อม ในการสอบปลายภาคก็เป็นไปในทำนองเดียวกัน นอกจากนี้ถึงอาจารย์ผู้สอนจะเข้าแทรกแซงเพื่อพัฒนาการเรียนรู้ของผู้เรียน ยังมีนักศึกษาที่เรียนอ่อนจำนวนหนึ่ง สอบซ่อมไม่ผ่าน ทำให้อำนาจจำแนกของวิชานี้อยู่ในเกณฑ์พอใช้ และจำนวนผู้เรียนมี 60 คน มีความรู้ใกล้เคียงกัน ทำให้แยกกลุ่มเก่ง และกลุ่มอ่อนได้ไม่ชัดเจน ค่าอำนาจจำแนกจึงไม่สูงมาก

เหตุผลอีกด้านหนึ่งที่ทำให้ ความรู้ของนักศึกษาใกล้เคียงกัน ทำให้อำนาจจำแนกของวิชา นี้อยู่ในเกณฑ์พอใช้เพราะอาจารย์ผู้สอนไม่ได้ปล่อยให้นักศึกษาเรียนรู้ไปตามธรรมชาติของนักศึกษาแต่ละคนแต่เข้าไปแทรกแซงเพื่อพัฒนาผู้เรียนก่อนการสอบในแต่ละครั้ง เนื่องจาก รายวิชากฎหมายและจริยธรรมเป็นวิชาที่เรียนครั้งเดียว และนักศึกษาต้องนำความรู้ไปใช้สอบขึ้นทะเบียนรับใบอนุญาตประกอบวิชาชีพฯจากสภาการพยาบาล ผู้ประสานงานวิชาเป็นผู้สอนหลักและเป็นผู้เตรียมความรู้ นักศึกษาเพื่อสอบขึ้นทะเบียนรับใบอนุญาตประกอบวิชาชีพฯภายหลังนักศึกษาเรียนจบหลักสูตรแล้ว และเนื่องจากมีเวลาจำกัด ทำให้ผู้สอนวางแผนการเรียนรู้ของนักศึกษาในชั้นเรียนโดยพยายามให้นักศึกษาทุกคนเมื่อเรียนจบวิชานี้แล้ว มีศักยภาพที่จะสอบผ่านข้อสอบของสภาการพยาบาลให้ได้หมดทุกคน ที่คะแนนร้อยละ 60 ดังนั้นในกระบวนการเรียนการสอน ได้นำตัวอย่างข้อสอบจำนวน

หนึ่งให้นักศึกษาเรียนรู้ก่อนสอบกลางภาคและปลายภาค (แต่ไม่ใช่ชุดจริงที่สอบกลางภาค และปลายภาค) เพื่อทำความเข้าใจ และนอกจากนี้ ได้กำหนดไว้ใน Course design ให้นักศึกษาต้องสอบความรู้ซ่อมเสริมแต่ละครั้งเมื่อสอบไม่ผ่านเกณฑ์ ร้อยละ 60 โดยได้ทบทวนความรู้ก่อนสอบทุกครั้ง ทำให้นักศึกษามีความรู้ในวิชานี้ใกล้เคียงกัน

พบว่า คุณภาพของข้อสอบชุดกลางภาค มีข้อสอบที่มีค่าความยากง่ายและค่าอำนาจจำแนก ได้ค่าตามเกณฑ์ที่เหมาะสมในข้อเดียวกัน (P อยู่ระหว่าง 0.20 -0.80 และ $r \geq .20$) มีจำนวน 19 ข้อ คิดเป็นร้อยละ 47.5 ส่วนชุดกลางภาค ค่า P และ r ที่มีค่าตามเกณฑ์เหมาะสมในข้อเดียวกัน มีจำนวน 17 ข้อ คิดเป็นร้อยละ 42.5 ซึ่งถือว่าข้อสอบที่ได้มาตรฐานมากพอสมควร ทั้งหมดสามารถนำไปเก็บไว้ในคลังข้อสอบและสามารถนำกลับมาใช้ใหม่ (test bank items) และควรเพิ่มเติมจากการปรับปรุงข้อสอบที่ได้ผลจากการวิเคราะห์บางข้อ โดยไปพิจารณา คุณภาพของตัวลองให้ดีขึ้น รวมทั้งเขียนข้อสอบใหม่เพิ่มเติม ในสัดส่วน 50: 40: 10 ตามข้อสรุปของยูโซวา (Usova GM, 1997)

พบว่า ชุดข้อสอบกลางภาค มีข้อสอบที่ ได้ค่า $r = 0$ จำนวน 4 ข้อ คิดเป็นร้อยละ 10 และมีข้อที่ได้ ค่า r ติดลบ 1 ข้อ คิดเป็นร้อยละ 2.5 ส่วนข้อสอบปลายภาค มีข้อสอบที่ ได้ค่า $r = 0$ จำนวน 1 ข้อ คิดเป็นร้อยละ 2.5 และมีข้อที่ได้ ค่า r ติดลบ 1 ข้อ คิดเป็นร้อยละ 2.5 การวิเคราะห์ แผลผลได้ว่า การเขียนข้อสอบอาจมีปัญหาในเรื่องใดเรื่องหนึ่งที่ต้องปรับปรุงเหมือนการปรับปรุงข้อสอบที่ไม่ได้มาตรฐานทั่วไป แต่บางสถานการณ์ พบว่าข้อสอบที่วัดคนกลุ่มหนึ่ง ได้ผลวิเคราะห์คุณภาพข้อสอบไม่ได้มาตรฐาน ค่า P และ r แต่เมื่อนำข้อสอบชุดเดียวกันไปวัดคนอีกกลุ่มหนึ่ง เช่น นักศึกษาพยาบาลต่างสถาบัน พบว่า ค่าที่ได้อาจได้มาตรฐานก็ได้ เพราะเป็นการวัดแบบอิงกลุ่ม ผลลัพธ์มาจากความแตกต่างของมาตรฐานความรู้ ของนักศึกษาแต่ละกลุ่มซึ่งติดตัวนักศึกษารวมทั้งกระบวนการเรียนการสอนในชั้นเรียน เหตุนี้ อาจเป็นเพราะข้อสอบยากเกินระดับความสามารถของผู้เรียนทั้งชั้น ทำให้เกิดการคาดเดาในการตอบ และเมื่อความสามารถของผู้เรียนใกล้เคียงกันทั้งชั้น ทำให้ได้ผลการวิเคราะห์ดังที่พบในการวิจัยครั้งนี้ ทั้งนี้ผู้สอนสามารถนำผลการวิเคราะห์ที่ได้ไปใช้ให้เกิดประโยชน์ในการช่วยพัฒนาผู้เรียนว่ามีความเข้าใจคลาดเคลื่อนในประเด็นใดบ้าง และช่วยทำให้เกิดความกระจ่างขึ้น (Johnson,1951; ชยุตม์ ภิรมย์สมบัติ, 2555, น.13)

2. การเปลี่ยนแปลงของคะแนนรายวิชาและอันดับ ก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก

ภายหลังจากได้รับผลการวิเคราะห์ข้อสอบ หากมีข้อสอบที่มีค่าอำนาจจำแนกติดลบซึ่งแสดงว่าคุณภาพของข้อสอบข้อนั้นๆไม่สามารถจำแนกผู้เรียนเก่ง และอ่อนได้ ผู้เรียนอ่อนส่วนมากตอบข้อสอบนั้นถูกแต่ผู้เรียนเก่งส่วนมากไม่ได้คะแนนจากข้อนั้น ดังนั้นในการศึกษาครั้งนี้ ผู้วิจัยจึงได้ตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก และส่งตรวจเพื่อวิเคราะห์ข้อสอบใหม่ หลังจากนั้นนำมาคิดคำนวณเปรียบเทียบบัญญัติไตรยางศ์เพื่อให้ได้คะแนนที่นักศึกษาแต่ละคนได้รับจากคะแนนเต็มเท่าเดิม เพื่อให้สามารถเปรียบเทียบค่าเฉลี่ย และอันดับของคะแนนนักศึกษาเป็นรายบุคคลก่อน และหลัง

การตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก เมื่อนำมาวิเคราะห์เปรียบเทียบกันพบการเปลี่ยนแปลง ดังนี้

2.1 ผลการเปรียบเทียบค่าคะแนนเฉลี่ยที่นักศึกษาได้รับในวิชาทฤษฎีและจริยธรรม วิชาชีพก่อน และหลังจากตัดข้อสอบที่มีค่า r ติดลบออก ทั้งชุดข้อสอบกลางภาค และชุดข้อสอบปลายภาค โดยสถิติ Paired t - test (one-tailed) พบว่า นักศึกษามีคะแนนเฉลี่ยสูงขึ้นอย่างมีนัยสำคัญทางสถิติ ($t=0.00$, $df = 59$, $P \text{ value} < 0.005$) ทั้งสองชุด ทั้งนี้ อาจอธิบายได้ว่า ถึงแม้ว่าจำนวนข้อสอบทั้งสองฉบับ จะมีข้อสอบ r ติดลบอย่างละ 1 ข้อ คิดเป็นร้อยละ 2.5 แต่เนื่องจากจำนวนข้อสอบทั้งฉบับมีเพียง 40 ข้อต่อชุด การตัดข้อสอบ r ติดลบออก 1 ข้อ ทำให้มีการเปลี่ยนแปลงของคะแนนสูงขึ้นเป็นส่วนใหญ่ และมีส่วนน้อยที่คะแนนลดลง

ผลการวิจัยครั้งนี้ หากมีการตัดข้อสอบที่มีค่า r ติดลบออก แล้วนำมาคิดคะแนนใหม่ ในวิชานี้พบว่า มีค่าคะแนนเฉลี่ยสูงขึ้น อย่างมีนัยสำคัญทางสถิติ เนื่องจาก นักศึกษากลุ่มที่มีคะแนนเพิ่มขึ้น มีจำนวนมากกว่า นักศึกษาที่มีจำนวนคะแนนลดลง ทำให้ผลออกมาเช่นนี้ ตามแนวการวิจัยพบว่า หากดำเนินการปรับคะแนน จะมีนักศึกษากลุ่มหนึ่งได้ประโยชน์ และมีนักศึกษากลุ่มหนึ่งเสียประโยชน์ ทั้งที่ต้นเหตุของปัญหา อาจไม่ได้อยู่ที่ความบกพร่องของข้อสอบ แต่มาจากความเกี่ยวเนื่องตามธรรมชาติของกระบวนการจัดการเรียนการสอน ศักยภาพในตัวนักศึกษาเอง และจำนวนนักศึกษาในชั้นเรียน การตัดข้อสอบที่มีค่า r ติดลบ ควรพิจารณารายละเอียดในปัจจัยอื่นๆร่วมด้วย

2.2 ผลการทดสอบการเปลี่ยนแปลงของลำดับคะแนนที่นักศึกษาได้รับในรายวิชาทฤษฎีและจริยธรรมวิชาชีพ ก่อน และหลังการตัดข้อสอบที่มีค่า r ติดลบออก โดยสถิติ Signed Rank Test พบว่าลำดับคะแนนที่นักศึกษาได้รับมีการเปลี่ยนแปลงลำดับที่อย่างมีนัยสำคัญทางสถิติ ที่ $\alpha = 0.05$, $P \text{ value} < 0.01$ จากการทบทวนวรรณกรรมไม่พบงานวิจัยที่ศึกษาเรื่องนี้ แต่พบว่าลำดับที่ของนักศึกษาเปลี่ยนไปภายหลังการตัดข้อสอบที่มีค่า r ติดลบออก ถึงแม้ตัดออกเป็นจำนวนข้อต่ำสุดในวิชานี้ คือจำนวน 1 ข้อ จากข้อสอบทั้งหมด 40 ข้อ (ร้อยละ 2.5) ข้อค้นพบนี้ สอดคล้องกับข้อค้นพบในข้อ 2.1 การตัดข้อสอบ r ติดลบ ออก 1 ข้อ ทำให้มีการเปลี่ยนแปลงของคะแนนสูงขึ้นเป็นส่วนใหญ่ และมีส่วนน้อยที่คะแนนลดลง ผลจากข้อ 2.1 และ 2.2 หากพิจารณาเฉพาะการวัด และการประเมินผล การตัดข้อสอบที่ค่า r ติดลบ และกลับมาคำนวณคะแนนใหม่ ถือว่าเหมาะสม หากถือว่าข้อสอบไม่ได้เกณฑ์มาตรฐาน (แบบอิงกลุ่ม) แต่เมื่อพิจารณาถึงปัจจัยอื่นๆว่า ผลการวิเคราะห์ ค่า P และค่า r ยังสะท้อนการเรียนรู้ของนักศึกษา กระบวนการจัดการเรียน การสอน ในเนื้อหาที่วัดได้จากข้อสอบข้อนั้นด้วย (ชยุตม์ ภิรมย์สมบัติ, 2555, น.13) การตัดข้อสอบ r ติดลบออก แล้วคำนวณคะแนนใหม่จึงควรพิจารณาอย่างรอบคอบมากขึ้น

3. การเปลี่ยนแปลงค่าความเชื่อมั่นของข้อสอบทั้งฉบับ ก่อนและหลังการตัดข้อสอบที่มีค่าอำนาจจำแนกติดลบออก

การเปลี่ยนแปลงค่าความเชื่อมั่นของข้อสอบทั้งฉบับ ก่อนและหลังการตัดข้อสอบที่มีค่า r ติดลบออก พบว่า มีค่าเพิ่มขึ้นเล็กน้อยทั้งสองชุด = 0.03 (จากชุดข้อสอบกลางภาคเดิม 0.48 เพิ่มขึ้นเป็น 0.51 ชุดข้อสอบปลายภาค จาก 0.40 เพิ่มขึ้นเป็น 0.43) อาจเป็นเพราะจำนวนข้อสอบที่มีค่า r ติดลบ มีจำนวนเพียง ชุดละ 1 ข้อ ทำให้เกิดการเปลี่ยนแปลงเพิ่มขึ้นน้อย

หากพิจารณาความเหมาะสมของค่าความเชื่อมั่นของข้อสอบทั้งฉบับในรายวิชากฎหมายและจริยธรรม ตามเกณฑ์ข้อสอบมาตรฐาน ควรมีความเชื่อมั่น ตั้งแต่ 0.70 ขึ้นไป จึงถือว่า มีความเชื่อมั่นสูง แต่ถ้าเป็นข้อสอบที่ครูสร้างขึ้นเพื่อใช้ในห้องเรียน ค่าความเชื่อมั่นของข้อสอบ 0.50 อาจจะได้ถือว่ามีความเที่ยงสูง (บุญเชิด ภิญญอนันตพงษ์, 2552, น.117)

อภิปรายผล สาเหตุที่ค่าความเชื่อมั่นของข้อสอบทั้งฉบับมีค่าต่ำ

คุณภาพของข้อสอบทั้งฉบับที่พิจารณาจากค่าความเชื่อมั่นของข้อสอบทั้ง 2 ชุด มีค่า 0.48 และ 0.45 ตามลำดับ ซึ่งอยู่ในระดับที่ต้องปรับปรุง เพื่อให้ข้อสอบมีความเชื่อมั่นสูงขึ้น การปรับปรุงค่าความเชื่อมั่น ใช้หลักการคล้ายกับการปรับปรุงค่าอำนาจจำแนก (Crocker and Algina, 1986 อ้างอิงใน ศิริชัย กาญจนวาสี, 2544, น.60-65) โดยในการวิจัยครั้งนี้ การคำนวณคุณภาพของข้อสอบทั้งฉบับใช้วิธีของ คูเดอร์ – ริชาร์ดสัน (Kuder – Richardson -20, KR-20) ที่เป็นการหาสัมประสิทธิ์ความคงเส้นคงวาภายในแบบอิงกลุ่ม (coefficient of internal consistency) (เพราะใช้กับการวิเคราะห์ผลการตอบคำถามที่เป็น 0,1 คือ ตอบถูกได้ 1 คะแนน ตอบผิด ได้ 0 คะแนน) การได้ค่าสัมประสิทธิ์ ต่ำมาจากสาเหตุดังนี้ (บุญเชิด ภิญญอนันตพงษ์, 2552, น.148-149)

1. ผู้ทำข้อสอบที่มีความสามารถใกล้เคียงกันหรือเป็นกลุ่มเอกพันธ์ จะได้ค่าความเชื่อมั่นของแบบทดสอบต่ำกว่ากลุ่มที่มีความสามารถแตกต่างกันมากหรือกลุ่มวิวิธพันธ์
2. ระดับความสามารถของผู้ทำข้อสอบในกลุ่ม ถ้าผู้ตอบในกลุ่ม มีความสามารถเฉลี่ยสูงทำข้อสอบได้ถูกต้องแน่นอนทุกครั้ง แต่ถ้าผู้เรียนมีความสามารถเฉลี่ยต่ำ มักจะตอบโดยการเดาเป็นส่วนใหญ่ จึงทำให้คะแนนจากการสอบไม่แน่นอน ทำให้ค่าความเชื่อมั่นต่ำ
3. ระดับความยากรายข้อ การกระจายของคะแนนสูงจะทำให้ค่าความเชื่อมั่นสูงกว่าคะแนนที่มีการกระจายน้อย เกิดจากชุดข้อสอบมีข้อสอบที่ง่ายมากหรือยากมาก คะแนนที่ได้จะไม่ค่อยแตกต่างกัน เป็นเหตุให้ค่าความเชื่อมั่นของข้อสอบทั้งฉบับต่ำลง
4. ชุดข้อสอบที่มีจำนวนข้อสอบมาก จะทำให้ค่าความเชื่อมั่นของข้อสอบสูงขึ้น
5. ความคล้ายคลึงกันของเนื้อหาที่ออกข้อสอบ ข้อสอบที่วัดลักษณะเดียวกันทั้งฉบับจะมีค่าความเชื่อมั่นของข้อสอบสูงกว่าข้อสอบที่มีเนื้อหาแตกต่างกันมากๆ
6. ผู้ทำข้อสอบที่มีจำนวนน้อย และมีปัญหาเรื่องสภาพอารมณ์ ความวิตกกังวล จะส่งผลกระทบทำให้ค่าความเชื่อมั่นของข้อสอบต่ำลง
7. ความเป็นปรนัย หมายถึง ความชัดเจน แม่นยำของแบบวัดหรือชุดข้อสอบ หากผู้สอบทุก

คนเข้าใจคำถามตรงกันและตรงกับความมุ่งหมายที่จะถาม การใช้ภาษาในคำถามชัดเจน ไม่กำกวม ไม่วกวน ไม่พลิกแพลงเล่นภาษา จะทำให้ค่า ค่าความเชื่อมั่นของข้อสอบสูงขึ้น

สาเหตุที่กล่าว เป็นแนวทางนำไปศึกษาวิเคราะห์ปัจจัยที่ทำให้ ความเชื่อมั่นของชุดข้อสอบทั้งฉบับในวิชากฎหมาย และจริยธรรมวิชาชีพต่ำซึ่งได้อภิปรายบางส่วนไปแล้วในเรื่อง ค่าอำนาจจำแนก

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. ควรมีระบบการสร้างข้อสอบที่ได้มาตรฐาน ตั้งแต่

1.1 กำหนดผังข้อสอบ (Test blueprint) ให้สอดคล้องกับสาระตามวัตถุประสงค์การเรียนรู้ และจำแนกระดับการวัดให้สอดคล้องกับข้อกำหนดตามกรอบมาตรฐานคุณวุฒิระดับอุดมศึกษา แห่งชาติ ปี 2552 (Thailand Qualifications Framework for Higher Education; TQF: HED)

1.2 การตรวจสอบคุณภาพข้อสอบที่สร้างก่อนนำมาใช้โดยการวิพากษ์ข้อสอบโดยผู้ที่มีความรู้

1.3 มีการวิเคราะห์ข้อสอบหลังการสอบ และนำข้อสอบไปปรับปรุงหรือเก็บข้อสอบที่ได้มาตรฐานเข้าในคลังข้อสอบ

1.4 ควรมีการพัฒนาคณาจารย์ผู้สอนให้มีความรู้เรื่องหลักการเขียนข้อสอบคุณภาพ

หมายเหตุ ผู้รับผิดชอบรายวิชากฎหมาย และจริยธรรมวิชาชีพ ได้ดำเนินการ ตามข้อ 1 (1.1-1.4) ตามมาตรฐานการดำเนินงานของคณะพยาบาลศาสตร์ มหาวิทยาลัยราชภัฏสวนดุสิต มาตั้งเปิดสอน ในวิชานี้

2. การพิจารณาตัดข้อสอบที่มีค่า r ตีตก และนำผลคะแนนมาคำนวณใหม่ ควรพิจารณาด้วยความระมัดระวัง เพราะมีนักศึกษาที่ได้ประโยชน์ และเสียประโยชน์จำนวนหนึ่งเสมอ และปัจจัยที่มีผลทำให้เกิดค่า r ตีตก มาจากหลายปัจจัยอาจไม่ใช่เนื่องมาจากความบกพร่องของข้อสอบด้านเดียว ผู้ประเมินผล ตัดเกรด ควรเป็นผู้ที่มีส่วนร่วมในกระบวนการจัดการเรียนการสอนได้มีโอกาสปฏิสัมพันธ์กับผู้เรียนในชั้น ซึ่งจะช่วยให้สามารถมองตัวเลขที่ได้จากการวิเคราะห์ข้อสอบสะท้อนข้อเท็จจริงที่เกิดขึ้นในชั้นได้ ซึ่งแต่ละรายวิชา อาจมีสาเหตุที่ทำให้เกิดค่า r ตีตก ไม่เหมือนกัน

3. การพัฒนารายวิชากฎหมาย และจริยธรรมวิชาชีพที่ได้จากผลการศึกษา

ผลจากการศึกษา ทำให้ได้แนวทางการปรับปรุงข้อสอบให้มีค่าอำนาจจำแนก ค่าความเชื่อมั่นของข้อสอบทั้งฉบับให้สูงขึ้น รวมทั้งนำไปใช้พัฒนาผู้เรียน ดังนี้

3.1 ควรเพิ่มจำนวนข้อสอบให้มากขึ้น

3.2 ควรวิเคราะห์ตัวลงในแต่ละข้อ (และปรับปรุงตัวลง ให้มีความใกล้เคียงตัวเลือกมากขึ้น) ที่ทำให้นักศึกษาไม่สามารถใช้วิธีการเดาได้ (Ascalon, Meyers, Davis & Smits, 2007)

3.3 ควรเขียนข้อสอบให้มีระดับของการวัดสูงขึ้น เช่น ระดับการวิเคราะห์ และการให้ค่า เพราะเมื่อข้อสอบมีระดับการวัดสูงขึ้น ค่าอำนาจจำแนกจะสูงขึ้น เพราะคนเก่งจะตอบได้ มากกว่าคนอ่อน และจะทำให้ข้อสอบมีค่าความยากสูงขึ้นด้วย (Ascalon, Meyers, Davis & Smits, 2007) แต่ทั้งนี้ขึ้นกับว่า ผู้สอนเคยวิเคราะห์ให้นักศึกษาในประเด็นที่ถามมาก่อนหรือไม่ ถ้านักศึกษาเคยทราบผลการวิเคราะห์เปรียบเทียบมาก่อน ถึงแม้ว่าแบบแผนของการเขียนข้อคำถามจะเป็นการวิเคราะห์ แต่ระดับการวัดจริงยังเป็นระดับความจำหากนักศึกษายังไม่สามารถวิเคราะห์ด้วยตนเอง นอกจากนี้ มี

ข้อพิจารณาว่า การเขียนวัดระดับความรู้ในเชิงการวิเคราะห์สำหรับข้อสอบกลางภาคที่นักศึกษาเริ่มเรียนยังไม่สิ้นสุดเนื้อหาทั้งหมด ความสามารถของนักศึกษาทั้งชั้นยังไม่ถึง ทำให้มีการคาดเดาคำตอบสูง ผลที่ได้อาจทำให้เกิด ค่าอำนาจจำแนกต่ำ หรือติดลบ และค่าความยากง่ายต่ำ

3.4 การพัฒนาข้อคำถาม โดย ปรับปรุง ตัวคำถามในข้อสอบว่าควรเขียนข้อคำถามเป็น โจทย์สถานการณ์ดีหรือกรณีศึกษา (Case-based items) หรือไม่ใช่กรณีศึกษา (Noncase-based items) ตามผลการวิจัยของ Phipps, Brackbill & Dunn (2009) ได้ศึกษาความสัมพันธ์ระหว่างข้อคำถามชนิดกรณีศึกษา (Case-based items) กับไม่ใช่กรณีศึกษา (Noncase-based items) ในแบบทดสอบชนิดเลือกตอบ (Multiple choice items) กับผลการวิเคราะห์ความยากง่าย และดัชนีค่าอำนาจจำแนก พบว่าชนิดโจทย์คำถามไม่ใช่กรณีศึกษา ได้ค่าอำนาจจำแนกสูงกว่าโจทย์คำถามชนิดกรณีศึกษา ประเด็นนี้ผู้วิจัยเห็นว่า การเขียนโจทย์ให้มีข้อคำถามเป็นสถานการณ์ให้ตัดสินใจยังมีความจำเป็นในวิชากฎหมายและจริยธรรมวิชาชีพเพราะเป็นวัดในระดับการวิเคราะห์ การให้ค่า (ชวาล แพรัตกุล, ม.ป.ป.) ซึ่งเป็นการวัดขั้นสูง เหมาะสมสำหรับวัดตามธรรมชาติของการเรียนรู้ในวิชานี้ นอกจากนี้ ตามข้อกำหนด TQF ของ สมศ. (สำนักงานรับรองมาตรฐานและประเมินคุณภาพการศึกษา) การวัดทางด้านปัญญา ให้ได้มาตรฐาน นักศึกษาระดับอุดมศึกษาต้องสามารถวิเคราะห์ ให้คุณค่าแยกแยะความผิด ชอบ ชั่วดี ความดีความงาม ซึ่งคำถามที่เป็นโจทย์ประเภทสถานการณ์หรือกรณีศึกษาจะสนองตอบโจทย์ได้ดีระดับหนึ่ง

3.5 ควรนำผลจากการวิเคราะห์ตัวลงไปใช้ประโยชน์ในการทบทวนความรู้ให้กับนักศึกษา ภายหลังการสอบเพื่อทำความเข้าใจในส่วนที่คลาดเคลื่อนเกี่ยวกับสาระของการเรียนรู้ในประเด็นต่างๆที่สะท้อนผ่านการเลือกตัวลงแต่ละตัวของข้อสอบ สารสนเทศดังกล่าวจะช่วยในการวางแผนการจัดการเรียนรู้เพื่อแก้ไขความเข้าใจคลาดเคลื่อนของผู้เรียนได้อย่างมีประสิทธิภาพต่อไปทำให้การทบทวนเพื่อเตรียมสอบขึ้นทะเบียนรับใบอนุญาตประกอบวิชาชีพมีประสิทธิภาพยิ่งขึ้น

ข้อเสนอแนะในการทำการวิจัยครั้งต่อไป

1. นำข้อสอบในชุดที่ทำการทดสอบนี้ ไปทำการทดสอบกับนักศึกษาพยาบาลชั้นปีเดียวกัน ในสถาบันการศึกษาพยาบาลแห่งอื่นจำนวนหนึ่ง และนำข้อสอบมาวิเคราะห์หาค่า P และ r ตามวัตถุประสงค์ที่เหมือนกับการวิจัยครั้งนี้ ก่อนพิจารณาในเชิงนโยบาย ให้ตัดข้อสอบที่มีค่า r ติดลบ ออก

2. ทำการวิจัยว่า ในวิชาที่มีผู้สอนจำนวนมากร่วมกันออกข้อสอบ มีความแตกต่าง กับวิชาที่มีผู้สอนจำนวนน้อย เกี่ยวกับผลการวิเคราะห์ข้อสอบตามเกณฑ์มาตรฐานอย่างไร

บรรณานุกรม

บรรณานุกรมภาษาไทย

- เกษม สาหร่ายทิพย์.(2539) *การวัดผลสัมฤทธิ์*. พิษณุโลก: มหาวิทยาลัยนเรศวร
- กังวล เทียนกัณฑ์เทศน์. (2540). *การวัด การวิเคราะห์ การประเมิน*. กรุงเทพมหานคร: แวงค็อก ซอฟแวร์เทคโนโลยีเฮาส์
- ทีศนา แคมณี.(2551) ศาสตร์การสอน: องค์ความรู้เพื่อการจัดกระบวนการเรียนรู้ที่มีประสิทธิภาพ. กรุงเทพฯ สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย
- ประดิษฐ์ อูปรมย์ (2540) *ชุดวิชาพื้นฐานการศึกษา (มนุษย์กับการเรียนรู้)* พิมพ์ครั้งที่ 15 นนทบุรี: สุโขทัยธรรมมาธิราช
- พระมหาทองมา อรุณรสี (แสงอรุณ) (๒๕๕๓) การศึกษาวิเคราะห์ปรัชญาการศึกษาของสำนัก อรรถิภาวนิยม: A CRITICAL STUDY OF THE EDUCATIONAL PHILOSOPHY OF EXISTENTIALIS (วิทยานิพนธ์ปริญญาพุทธศาสตรมหาบัณฑิต) มหาวิทยาลัยมหาจุฬาลงกรณราชวิทยาลัย, บัณฑิตวิทยาลัย, สาขาวิชาปรัชญา
- พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ.2545 (2545, 19 ธันวาคม). *ราชกิจจานุเบกษา*. หน้า 6-21
- พิศเพลิน เขียวหวาน และสุจิตรา หังสพฤกษ์ (2552) การวัดและประเมินศักยภาพของบุคคล ในการพัฒนาเครื่องมือสำหรับการประเมินการศึกษา ใน วรดิ แสงประทีปทอง และคณะ (คณะกรรมการผลิต) การพัฒนาเครื่องมือสำหรับการประเมินทางการศึกษา (เล่ม 1 หน่วยที่ 1-7) พิมพ์ครั้งที่ 4 (หน้า 1-37) นนทบุรี: สุโขทัยธรรมมาธิราช
- พวงรัตน์ ทวีรัตน์. (2540). *วิธีการวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์*. พิมพ์ครั้งที่ 7. กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒประสานมิตร.
- ไพฑูริย์สินลารัตน์ และคณะ (2548) *รายงานการวิจัยเรื่องกรอบมาตรฐานคุณวุฒิระดับอุดมศึกษา ของประเทศไทย* กรุงเทพฯ: สำนักงานคณะกรรมการการอุดมศึกษา
- ภัทรา นิคมานนท์.(2543). *การประเมินผลการเรียน*. กรุงเทพมหานคร: อักษราพิพัฒน์.
- มนสิข สีทธิสมบุรณ์.(2537) *การวัดและการประเมินผลการเรียน*. พิษณุโลก: คณะศึกษาศาสตร์ มหาวิทยาลัยนเรศวร
- รังสรรค์ มณีเล็ก (2552) แนวคิดพื้นฐานเกี่ยวกับการพัฒนาเครื่องมือ สำหรับการประเมินทางการศึกษา ใน วรดิ แสงประทีปทอง และคณะ(คณะกรรมการผลิต)*การพัฒนาเครื่องมือ สำหรับการประเมินทางการศึกษา (เล่ม 1 หน่วยที่ 1-7)* พิมพ์ครั้งที่ 4 (หน้า 1-37) นนทบุรี: สุโขทัยธรรมมาธิราช
- ล้วน สายยศ และ อังคณา สายยศ (2543) *การวัดด้านจิตพิสัย* กรุงเทพฯ: สุวีริยาสาส์น.
- บุญเชิด ภิญโญอนันตพงษ์ (2552) ความรู้พื้นฐานเกี่ยวกับคุณภาพของเครื่องมือวัด ใน วรดิ แสงประทีปทอง และคณะ(คณะกรรมการผลิต) *การพัฒนาเครื่องมือ สำหรับการประเมินทางการศึกษา (เล่ม 1 หน่วยที่ 1-7)* พิมพ์ครั้งที่ 4 (หน้า 66-154) นนทบุรี: สุโขทัยธรรมมาธิราช

- บุญศรี พรหมมาพันธ์ และนวลเสนต์ วงศ์เชิดธรรม (2552) ความหมายและประเภทและ
แบบทดสอบวัดผลทางการเรียน ในวรรตี แสงประทีปทอง และคณะ(คณะกรรมการผลิต)
การพัฒนาเครื่องมือสำหรับการประเมินทางการศึกษา (เล่ม 1 หน่วยที่ 1-7) พิมพ์ครั้งที่ 4
(หน้า 215-256) นนทบุรี: สุโขทัยธรรมาธิราช
- ศิริชัย กาญจนวาสี. (2552). *ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory)*
พิมพ์ครั้งที่ 6. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สุมาลี จันทรชลอ. (2542). *การวัดและประเมินผล*. กรุงเทพฯ: ศูนย์สื่อเสริมกรุงเทพ.
- สมนึก ภัททิยธนี.(2537). *การวัดผลการศึกษา*. กทม. : ประสานการพิมพ์.
- สำเร็จ บุญเรืองรัตน์.(2527). *ทฤษฎีการวัดและประเมินผลการศึกษา*. กรุงเทพฯ: สำนักทดสอบทาง
การศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.
- อนันต์ ศรีโสภณ.(2534) *การวัดและการประเมินผลการศึกษา*. พิมพ์ครั้งที่ 2. กรุงเทพฯ:
ไทยวัฒนาพานิช
- อุทุมพร จามรมาน. (2535). *ข้อสอบ: การสร้างและการพัฒนา*. กรุงเทพฯ: ฟันนี้พับบลิชซิง.

บรรณานุกรมภาษาไทยทางสื่ออิเล็กทรอนิกส์

- ทิวต์ถ์ มณีโชติ. (2549) การวัดและประเมินผลการเรียนรู้ตามหลักสูตรการศึกษาขั้นพื้นฐาน. [ข้อมูล
อิเล็กทรอนิกส์]. กรุงเทพฯ : สำนักพิมพ์ศูนย์ส่งเสริมวิชาการ. ค้นจาก
<http://www.sotroiet.com/sotroiet/mainfile/files60.doc>
- ชยุตม์ ภิรมย์สมบัติ (2555,กันยายน – ตุลาคม) *วิเคราะห์ข้อสอบเลือกตอบอย่างไรสำหรับการวัด
ประเมินผลระดับชั้นเรียน* [ข้อมูลอิเล็กทรอนิกส์]. *นิตยสาร สสวท*. 40 (179), 11-13
- เชิดศักดิ์ ไอรรมณีรัตน์ (ม.ป.ป.) *ข้อแนะนำในการจัดสอบ OSCE (Objective Structured Clinical
Examination) (ตอนที่ 1)* ค้นจาก
http://teachingresources.psu.ac.th/document/2548/Ai_Rommani_Rat/2.pdf
- นภา หลิมรัตน์ (ม.ป.ป.) *Objective Structured Clinical Examination* ค้นจาก
[http://teachingresources.psu.ac.th//document/2551/rat /8Objective
%20Structured%20Clinical%20 Examination.pdf](http://teachingresources.psu.ac.th//document/2551/rat /8Objective %20Structured%20Clinical%20 Examination.pdf)
- บุญชม ศรีสะอาด (2514) *วิเคราะห์อำนาจจำแนก* เข้าถึงได้จาก [http:// www.watpon.com/
boonchom/02discrim.doc](http://www.watpon.com/boonchom/02discrim.doc)
- บุญชม ศรีสะอาด, นิภา ศรีไพโรจน์ และนุชชานา ทองทวี. (2528). *การวิเคราะห์ข้อสอบตามแนวคิด
อิงเกณฑ์* เข้าถึงได้จาก [http:// www.geocities.com/nincoo/mainb7.2.htm](http://www.geocities.com/nincoo/mainb7.2.htm).
- ปวรส บุตะเขี้ยว. (ม.ป.ป.). *การวิเคราะห์ข้อสอบ (Item analysis)*. ค้นจาก
www.rta.mi.th/630a0u/file/item_analysis.doc
- เพียว เนตรประชา (มปป.) *การวัดและประเมินผลการศึกษา* ค้นจาก
<http://www.ipesp.ac.th/learning/websatiti /manipulator.html>

- พระพรหมคุณาภรณ์ (2552) เอกสารสรุปผลการประชุมสัมมนา เรื่อง ปรัชญาการอุดมศึกษาไทย โดยสำนักงานคณะกรรมการการอุดมศึกษา เข้าถึงได้จาก <http://www.mua.go.th/users/the-commission/book/philosophyTHE.pdf>
- พวงพยอม แก้วพิลา วราภรณ์ เชื้ออินทร์ นภา หลิมรัตน์ ชาญชัย พานทองวิริยะกุล (2548) การศึกษาผลวิเคราะห์การสอบรวบยอดวิทยาศาสตร์การแพทย์คลินิก (สอบรวม 4 สถาบัน) ปีการศึกษา 2548 ของนักศึกษาแพทย์ชั้นปีที่ 6 คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น *ศรีนครินทร์เวชสาร* 21 (3), 1-7
- วิจารณ์ พานิช (2556) การเรียนรู้เกิดขึ้นอย่างไร : ๑๔. ผู้กำกับการเรียนรู้ของตนเอง : ทฤษฎี ค้นจาก <http://www.gotoknow.org/posts/531606>
- สุคนธ์ชื่น ศรีงาม (ม.ป.ป.) มาตรฐานการอุดมศึกษา เข้าถึงได้จาก http://www.mua.go.th/users/tqf-hed/news/FilesNews/FilesNews5/2_ProgSpac.pdf
- สมศักดิ์ สินธุระเวชชัย. (2522). *การวิเคราะห์ข้อสอบตามแนวคิดอิงเกณฑ์* . [http:// www.geocities.com/nincoo/mainb7.2.htm](http://www.geocities.com/nincoo/mainb7.2.htm). ค้นจาก .
- สำนักงานเลขาธิการสภาการศึกษา (2011) สรุปความสามารถในการแข่งขันของประเทศไทย ใน World Competitiveness Yearbook (IMD) 2011 ค้นจาก http://www.onec.go.th/onec_administrator/uploaded/Category/DepOpCenter/2011-Oct-11-SumIMD2011.pdf
- สำนักงานเลขาธิการสภาการศึกษา.(มปป.) รมว.ศธ. ประกาศนโยบาย ปีแห่งการรวมพลังยกระดับคุณภาพการศึกษา *ข่าว สกศ* ค้นจาก http://www.onec.go.th/onec_web/page.php?mod=Newseducation &file=view&itemId=964
- อานุกาพ เลชะกุล (2556) การสร้างข้อสอบแบบเลือกตอบ (Multiple choice question) ค้นจาก http://www.med.cmu.ac.th/secret/meded/AOMJAI/เอกสารดาวโหลด_module 4
- อรรวรรณ สุวรรณรัตน์ (2552) *การหาคุณภาพข้อสอบเพื่อบริหารจัดการด้านการเรียนการสอน สงขลานครินทร์เวชสาร* 27(5) (กันยายน- ตุลาคม), 381-388
- Johnson (1951) ใน ศิริชัย กาญจนวาสี การวิเคราะห์ข้อสอบ *เอกสารประกอบการอบรมเรื่อง การวิเคราะห์ข้อสอบเพื่อมาตรฐานการศึกษา* 26 ธันวาคม 2546 คณะสัตวแพทยศาสตร์ มหาวิทยาลัยขอนแก่น ค้นจาก <http://www.vet.kku.ac.th/ed/research11.doc>

บรรณานุกรมภาษาต่างประเทศ

- Ascaron M.E., Meyers L.S., Davis B.W. & Smits N. (2007). Distractor Similarity and Item - Stem Structure: Effect on item Difficulty. *Applied Measurement in Education*. 20(2): 153-170.
- Caldwell D.J. & Pate A.N. (2003). Effects on Question Formats on Student and Item Performance. *American Journal of Pharmaceutical Education*. 77(4). Article 71: 1-5

- Chase, C. I. (1978). *Measurement for Educational Evaluation*. 2nd ed. Reading, MA: Addison-Wesley.
- Crocker, Linda and Algina, James (1986) *Introduction to classical and Modern Test* California: Holt, Rinehart, and Winston อ้างถึงใน ศิริชัย กาญจนวาสี (2537), หน้า 60-65 *ทฤษฎีการประเมิน*. กรุงเทพมหานคร: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- Gronlund, NE. (1985) *Measurement and evaluation in teaching*. 5th ed. London: Collier Macmillan.
- Lee H. & Winke P. (2012). The differences among three-, four-, and five- option- item Formats in the context of a high – stakes English – language listening test. *Language Testing*. 30(1): 99-123.
- Merriam –Webster an encyclopedia of Britannica Company. *meaning of education* Retrived from <http://www.merriam-webster.com/dictionary/education>
- Oermann, Marilyn H. (2002) Developing a Professional Portfolio in Nursing [Electronic version] *Orthopaedic Nursing* 21(2), p 73-79
- Oluseyi A.E. & Olufemi A.T. (2012). The Analysis of Multiple Choice Item of the Test of an Introduction Course in Chemistry in a Nigerian University. *The international Journal of Learning* 18(4): 237-246
- Office of Education Assessment, University of Washington.(2005) *SCOREPAK @:Item Analysis*. Retrived from <http://www.washington.edu/oea/score1.htm>
- Overton, Terry.(2009). *Assessing Learners with special Needs*. 6 th edition Retrived from <http://www.slideshare.net/norazmi3003/testing-assessment-measurement-and-evaluation-definition>
- Peter E. Parker and et al. (2001) Effective and sustained proactivity in the workplace A self-determination Theory perspective Retrived from <https://books.google.co.th/books?isbn=019979491X>
- Revised Bloom’s Taxonomy (RBT) (no date) Retrieved from: <http://www.kurwongbss.qld.edu.au/thinking/Bloom/blooms.htm>
- The Iris Center, Vanderbilt (no date) Overview of the HPL Framework retrieved from http://iris.peabody.vanderbilt.edu/module/hpl/cresource/how-can-faculty-present-important-content-to-be-learned-in-ways-that-improve-student-learning/hpl_01

ภาคผนวก

ภาคผนวก ก

ตารางกำหนดข้อสอบ (Test Blueprint) รายวิชากฎหมาย และจริยธรรมวิชาชีพ

ภาคผนวก ข

สรุปผลการวิเคราะห์ข้อสอบ ชุดกลางภาค และชุดปลายภาค
รายวิชากฎหมาย และจริยธรรมวิชาชีพ

สรุปผลการวิเคราะห์ข้อสอบกลางภาค

สรุปผลการวิเคราะห์ข้อสอบกลางภาค (น.66-69)

ชื่อสถาบัน มหาวิทยาลัยราชภัฏสวนสุนันทา

ชื่อวิชา กฎหมายและจริยธรรม (6083101)

ชื่อผู้สอน

วันที่พิมพ์ 4 กันยายน 2555

ข้อที่	สรุป ค่า p ค่า d รายข้อ		กลุ่มสูง กลุ่มต่ำ 27 %	
	p	ค่าอำนาจจำแนก (d)	Delta	ค่าอำนาจจำแนก (r-bis)
1	0.969	0.063	6.200	0.087
2	0.563	0.125	12.400	0.158
3	0.407	0.188	14.000	0.199
4	0.125	0.000	17.700	0.045
5	0.344	-0.188	0.000	-0.091
6	0.532	0.438	12.700	0.400
7	0.313	0.250	15.100	0.185
8	0.844	0.313	8.400	0.303
9	0.313	0.375	15.200	0.381
10	0.813	0.000	9.500	0.043
11	0.532	0.563	12.600	0.591
12	0.376	0.375	14.400	0.353
13	0.500	0.500	13.000	0.568
14	0.844	0.313	8.400	0.298
15	0.688	0.125	11.100	0.236
16	0.719	0.188	10.700	0.178
17	0.719	0.313	10.600	0.421
18	0.625	0.250	11.700	0.279
19	0.907	0.188	7.300	0.193
20	0.282	0.188	15.400	0.220
21	0.750	0.250	10.300	0.241
22	0.719	0.188	10.700	0.207
23	0.594	0.438	12.000	0.338
24	0.126	0.125	17.800	0.039
25	0.844	0.188	8.900	0.223
26	0.688	0.375	10.900	0.299
27	0.782	0.313	9.700	0.244
28	0.563	0.000	12.400	0.159
29	0.532	0.063	12.700	0.268
30	0.344	0.188	14.700	0.251
31	0.438	0.375	13.700	0.377
32	0.250	0.250	15.900	0.277
33	0.875	0.250	7.900	0.197
34	0.813	0.375	8.800	0.318
35	0.219	0.313	16.500	0.545
36	0.657	0.313	11.300	0.334
37	0.532	0.313	12.700	0.322

งานตรวจกระดาษคำตอบและวิเคราะห์ข้อสอบ

มหาวิทยาลัยราชภัฏสวนสุนันทา

ชื่อสถาบัน มหาวิทยาลัยราชภัฏสวนคูสิต
 ชื่อวิชา กฎหมายและจริยธรรม (6083101)
 ชื่อผู้สอน

วันที่พิมพ์ 4 กันยายน 2555

ข้อที่	สรุป ค่า p ค่า d รายข้อ p ค่าอำนาจจำแนก (d)	กลุ่มสูง กลุ่มต่ำ 27 % Delta ค่าอำนาจจำแนก (r-bis)
38	0.500 0.000	13.000 0.128
39	0.626 0.375	11.700 0.252
40	0.407 0.313	14.000 0.222
เฉลี่ย	p d 0.566 0.239	Delta r-BIS 11.800 0.257

แบบทดสอบทั้งฉบับมีคุณภาพดีมากน้อยเพียงใดให้พิจารณาค่าตั้งต่อไปนี้
 ค่าเฉลี่ยของคะแนนคือ X ความถี่ประมาณ 50 % ของคะแนนเต็ม
 ค่าของ p ความถี่ใกล้เคียงกับ 0.50
 ค่าของ Delta ความถี่ใกล้เคียงกับ 13.0
 ค่าของ r-BIS และ d ความถี่ขนาดมากกว่า 0.20 ขึ้นไป

22.53 / 40

ชื่อสถาบัน มหาวิทยาลัยราชภัฏสวนดุสิต
 ชื่อวิชา กฎหมายและจริยธรรม (6083101)
 ชื่อผู้สอน

วันที่พิมพ์ 4 กันยายน 2555

จำนวนผู้สอบ	60	คะแนนต่ำสุด(Min)	14	ค่ามัธยฐาน(Median)	23.00
กลุ่มสูง	16	คะแนนสูงสุด(Max)	30	ค่าเบี่ยงเบนมาตรฐาน(S.D)	3.862
กลุ่มต่ำ	16	คะแนนเฉลี่ย(Mean)	22.533	ค่าความเชื่อมั่น KR-20	0.477
จำนวนข้อสอบ	40	ฐานนิยม (Mode)	24	ค่าความเชื่อมั่น KR-21	0.349

สรุปคุณภาพของข้อสอบ

ข้อสอบที่ดี ควรเก็บไว้ ได้แก่

1. ข้อที่ค่อนข้างง่าย จำนวนจำแนกดี มี 7 ข้อ ได้แก่ข้อ
17 18 21 26 27 36 39

2. ข้อที่ยากปานกลาง จำนวนจำแนกดี มี 7 ข้อ ได้แก่ข้อ
6 11 13 23 31 37 40

3. ข้อที่ค่อนข้างยาก จำนวนจำแนกดี มี 5 ข้อ ได้แก่ข้อ
7 9 12 32 35

ข้อสอบที่ควรปรับปรุง ได้แก่

1. ข้อที่ง่ายมาก จำนวนจำแนกดี มี 4 ข้อ ได้แก่ข้อ
8 14 33 34

2. ข้อที่ยากมาก จำนวนจำแนกดี ไม่มี

3. ข้อที่ความยากปานกลาง จำนวนจำแนกไม่ดี มี 8 ข้อ ได้แก่ข้อ
2 3 15 16 20 22 29 30

ข้อสอบที่ควรตัดทิ้ง ได้แก่

1. ข้อที่ยากมากหรือง่ายมาก จำนวนจำแนกไม่ดี มี 4 ข้อ ได้แก่ข้อ
1 19 24 25

2. ข้อที่ไม่มีจำนวนจำแนก (0) มี 4 ข้อ ได้แก่ข้อ
4 10 28 38

3. ข้อที่จำนวนจำแนกเป็นลบ (-) มี 1 ข้อ ได้แก่ข้อ
5

สรุปผลการวิเคราะห์ข้อสอบปลายภาค

ชื่อสถาบัน มหาวิทยาลัยราชภัฏสวนดุสิต

ชื่อวิชา กฎหมายและจริยธรรม (6083101)

ชื่อผู้สอน

วันที่พิมพ์ 4 ตุลาคม 2555

ข้อที่	สรุป ค่า p ค่า d รายข้อ		กลุ่มสูง กลุ่มต่ำ 27 %	
	p	ค่าอำนาจจำแนก (d)	Delta	ค่าอำนาจจำแนก (r-bis)
1	0.344	0.063	14.700	0.170
2	0.969	0.063	6.200	0.019
3	0.407	0.313	14.000	0.343
4	0.282	0.188	15.400	0.288
5	0.844	0.063	9.000	0.032
6	0.625	0.250	-11.700	0.295
7	0.032	0.063	0.000	0.000
8	0.500	0.250	13.100	0.284
9	0.500	0.500	13.000	0.411
10	0.407	0.313	14.000	0.319
11	0.282	0.313	15.500	0.381
12	0.125	0.000	17.700	-0.038
13	0.500	0.250	13.100	0.298
14	0.094	0.188	0.000	0.130
15	0.844	0.188	8.900	0.069
16	0.938	0.125	6.700	0.067
17	0.282	0.313	15.500	0.425
18	0.438	0.250	13.700	0.259
19	0.344	0.313	14.700	0.257
20	0.407	0.063	14.000	0.154
21	0.594	0.188	12.100	0.201
22	0.751	0.125	10.300	0.145
23	0.438	0.125	13.700	0.123
24	0.875	0.250	7.900	0.296
25	0.907	0.188	7.300	0.221
26	0.719	0.313	10.600	0.357
27	0.563	0.250	12.400	0.367
28	0.907	0.063	7.800	0.165
29	0.594	0.313	12.000	0.282
30	0.907	0.188	7.300	0.215
31	0.907	0.188	7.300	0.196
32	0.344	-0.063	0.000	-0.022
33	0.907	0.063	7.800	0.092
34	0.813	0.250	9.400	0.286
35	0.594	0.313	12.000	0.420
36	0.500	0.500	13.000	0.382
37	0.376	0.125	14.300	0.186

งานตรวจกระดาษคำตอบและวิเคราะห์ข้อสอบ

มหาวิทยาลัยราชภัฏสวนดุสิต

ชื่อสถาบัน มหาวิทยาลัยราชภัฏสวนคูสิต

ชื่อวิชา กฎหมายและจริยธรรม (6083101)

ชื่อผู้สอน

วันที่พิมพ์ 4 ตุลาคม 2555

ข้อที่	สรุป ค่า p ค่า d รายข้อ p ค่าอำนาจจำแนก (d)	กลุ่มสูง กลุ่มต่ำ 27 % Delta ค่าอำนาจจำแนก (r-bis)
38	0.750 0.250	10.300 0.316
39	0.407 0.313	14.000 0.300
40	0.376 0.125	14.300 0.202
เฉลี่ย	p d 0.559 0.203	Delta r-BIS 10.867 0.222

แบบทดสอบทั้งฉบับมีคุณภาพดีมากน้อยเพียงใดให้พิจารณาค่าดังต่อไปนี้
 ค่าเฉลี่ยของคะแนนดิบ X ความน่าจะเป็น 50 % ของคะแนนเต็ม
 ค่าของ p ความน่าจะเป็นที่ใกล้เคียงกับ 0.50
 ค่าของ Delta ความน่าจะเป็นที่ใกล้เคียงกับ 13.0
 ค่าของ r-BIS และ d ความน่าจะเป็นมากกว่า 0.20 ขึ้นไป

22.51 / 40

ชื่อสถาบัน มหาวิทยาลัยราชภัฏสวนดุสิต
 ชื่อวิชา กฎหมายและจริยธรรม (6083101)
 ชื่อผู้สอน

วันที่พิมพ์ 4 ตุลาคม 2555

จำนวนผู้สอบ	60	คะแนนต่ำสุด(Min)	11	ค่ามัธยฐาน(Median)	23.00
กลุ่มสูง	16	คะแนนสูงสุด(Max)	33	ค่าเบี่ยงเบนมาตรฐาน(S.D)	3.437
กลุ่มต่ำ	16	คะแนนเฉลี่ย(Mean)	22.517	ค่าความเชื่อมั่น KR-20	0.399
จำนวนข้อสอบ	40	ฐานนิยม (Mode)	24	ค่าความเชื่อมั่น KR-21	0.171

สรุปคุณภาพของข้อสอบ

ข้อสอบที่ดี ควรเก็บไว้ ได้แก่

1. ข้อที่ค่อนข้างง่าย อำนาจจำแนกดี
6 26 38 มี 3 ข้อ ได้แก่ข้อ
2. ข้อที่ยากง่ายปานกลาง อำนาจจำแนกดี
3 8 9 10 13 18 27 29 35 36 39 มี 11 ข้อ ได้แก่ข้อ
3. ข้อที่ค่อนข้างยาก อำนาจจำแนกดี
11 17 19 มี 3 ข้อ ได้แก่ข้อ

ข้อสอบที่ควรปรับปรุง ได้แก่

1. ข้อที่ง่ายมาก อำนาจจำแนกดี
24 34 มี 2 ข้อ ได้แก่ข้อ
2. ข้อที่ยากมาก อำนาจจำแนกดี
ไม่มี
3. ข้อที่ความยากปานกลาง อำนาจจำแนกไม่ดี
1 4 20 21 22 23 37 40 มี 8 ข้อ ได้แก่ข้อ

ข้อสอบที่ควรตัดทิ้ง ได้แก่

1. ข้อที่ยากมากหรือง่ายมาก อำนาจจำแนกไม่ดี
2 5 7 14 15 16 25 28 30 31 33 มี 11 ข้อ ได้แก่ข้อ
2. ข้อที่ไม่มีอำนาจจำแนก (0)
12 มี 1 ข้อ ได้แก่ข้อ
3. ข้อที่อำนาจจำแนกเป็นลบ (-)
32 มี 1 ข้อ ได้แก่ข้อ

มหาวิทยาลัยราชภัฏสวนดุสิต

สรุปผลการวิเคราะห์ข้อสอบ

วิชา กฎหมายและจริยธรรม (6083101)

รหัสวิชา ทฤษฎี หน่วยกิต ปฏิบัติ หน่วยกิต

ภาคการศึกษา

ปีการศึกษา

นักศึกษาชั้นปีที่

จำนวน 60 คน

จำนวนข้อสอบ	Max	Min	\bar{X}	S.D.	Median	Mode	Sk	Ku
40	33.000	11.0000	22.517	3.437	23.000	23.966	0.019	4.516
Normal Curve : Skewness (ความเบ้) = 0 Sk = (เบ้ขวา) คือ นศ. > 50% ได้คะแนน < \bar{X} Sk = (เบ้ซ้าย) คือ นศ. > 50% ได้คะแนน > \bar{X} Kurtosis (ความโค้ง) = 3 Ku > 3 คือ คะแนนกระจายแบบโค้งสูง (Leptokurtic) Ku < 3 คือ คะแนนกระจายแบบลาด								

ความเที่ยงของข้อสอบ				Standard Error of Measurement			
KR20	KR21	Alpha	Split-half(Rtt)	SEM 20	SEM 21	SEM -Alp	SEMTT
0.399	0.171	0.399	0.378	2.665	3.129	2.665	2.712
KR20 เหมาะสำหรับข้อสอบ Heterogeneous				Cronbach เหมาะสำหรับข้อสอบที่มีคะแนน 1,0			
KR21 ,Rtt เหมาะสำหรับข้อสอบ Homogeneous				(ค่า KR21 จะน้อยกว่า KR20 เสมอ และ Rtt จะได้ค่าความเที่ยงสูงสุด)			

ความยากง่ายและอำนาจจำแนกของข้อสอบทั้งฉบับ				
ความยากง่ายของข้อสอบ		อำนาจจำแนกของข้อสอบ (r)		
Difficulty Index (P)	Delta	Discrimination Index	Biserial RBIS	Point-biserial RBIS
0.560	10.868	0.203	0.222	0.207
Delta 6-20 คือยากง่ายพอเหมาะ		ค่าอำนาจจำแนก ส.ป.ศ.สหสัมพันธ์ Biserial RBIS ใช้กับ Normal Curve		
Delta >= 21 คือ ยากมาก		ค่าอำนาจจำแนก ส.ป.ศ.สหสัมพันธ์ Point-biserial RBIS ใช้กับ Curve เบ้		
				Delta <= 5 คือง่ายมาก

ข้อสอบรายข้อที่มีความยากง่าย (P) ตั้งแต่ 0.20 - 0.80 และมีค่าอำนาจจำแนก (r) ตั้งแต่ 0.2 ขึ้นไป มีจำนวน 17 ข้อ

จากจำนวนข้อสอบทั้งหมด 40 ข้อ คิดเป็นร้อยละ 42.50

3 6 8 9 10 11 13 17 18 19 26 27 29 35 36 38 39

สอบวันที่

วิเคราะห์เสร็จวันที่ 4 ตุลาคม 2555

ผู้วิเคราะห์ข้อสอบ

งานตรวจกระดาษคำตอบและวิเคราะห์ข้อสอบ

มหาวิทยาลัยราชภัฏสวนดุสิต

ประวัติผู้วิจัย

นางนิภา ลีสุคนธ์

วัน เดือน ปี เกิด 7 ธันวาคม 2497 อายุ 60 ปี

ตำแหน่งปัจจุบัน อาจารย์ประจำ สาขาการพยาบาลชุมชน คณะพยาบาลศาสตร์
มหาวิทยาลัยราชภัฏสวนดุสิต

ประวัติการศึกษา

พ.ศ. 2527 ครุศาสตร์มหาบัณฑิต (การบริหารการพยาบาล)

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ.2520 วิทยาศาสตร์บัณฑิต (พยาบาล)

คณะพยาบาลศาสตร์ มหาวิทยาลัยเชียงใหม่

พ.ศ.2520 ประกาศนียบัตรผดุงครรภ์ ชั้น 1

คณะพยาบาลศาสตร์ มหาวิทยาลัยเชียงใหม่

พ.ศ. 2545 วุฒิบัตรเฉพาะทางการพยาบาลเวชปฏิบัติทั่วไป(การรักษาโรคเบื้องต้น)

ภาควิชาพยาบาลศาสตร์ คณะแพทยศาสตร์ โรงพยาบาลรามาริบัติ

และขึ้นทะเบียนรับใบอนุญาตประกอบวิชาชีพ การรักษาโรค

เบื้องต้น จากสภาการพยาบาล

ผลงานวิจัยล่าสุด

หัวหน้าทีมวิจัย เรื่อง “ถอดบทเรียน การดำเนินงานระบบหลักประกันสุขภาพ ในระดับท้องถิ่น หรือพื้นที่ (กองทุนหลักประกันสุขภาพ อบต.และเทศบาล) ในพื้นที่ต้นแบบทั่วประเทศ: กรณีศึกษา เทศบาลตำบลบึงยี่โถ อ.ธัญบุรี จ.ปทุมธานี” พ.ศ.2551 และได้รับการคัดเลือกให้นำเสนอในการ ประชุมนำเสนอผลงานวิจัย รอบสุดท้าย เป็น 1 ใน 4 เรื่อง ต่อผู้บริหารคณะกรรมการหลักประกัน สุขภาพแห่งชาติ และผู้บริหารในชมรมผู้บริหารท้องถิ่น ผู้สนใจ และทีมผู้ร่วมทำวิจัยทั่วประเทศ ณ ห้องประชุม โรงแรม TK Palace ถนนแจ้งวัฒนะ

ตีพิมพ์ บทความวิจัยเชิงคุณภาพ (จากฐานข้อมูลวิจัยเดิม) ใน สำนักงานหลักประกันสุขภาพ แห่งชาติ (สปสช) ถอดบทเรียน การดำเนินงานระบบหลักประกันสุขภาพในระดับท้องถิ่นหรือพื้นที่ (กองทุนหลักประกันสุขภาพ อบต. และเทศบาล) ในพื้นที่ต้นแบบทั่วประเทศ พิมพ์ครั้งที่ 1 กรุงเทพฯ: สหมิตรพรินติ้งแอนด์พับลิชชิ่งจำกัด

ประวัติการปฏิบัติงาน

1) อาจารย์พยาบาล ระดับ ผู้ชำนาญการพิเศษ วิทยาลัยพยาบาลบรมราชชนนี กรุงเทพมหานคร สำนักงานปลัดกระทรวงสาธารณสุข พ.ศ. 2527 –2532, พ.ศ. 2541 – 2553

2) ผู้ช่วยเลขานุการ คณะกรรมการพัฒนาหลักสูตรการอบรมพยาบาลเวชปฏิบัติทั่วไป (การรักษาโรคเบื้องต้น) ณ วิทยาลัยพยาบาลบรมราชชนนี กรุงเทพมหานคร สำนักงานปลัดกระทรวงสาธารณสุข ได้รับการอนุมัติหลักสูตรจากสภาการพยาบาล พ.ศ. 2546 (วงรอบที่ 1)

3) หัวหน้าคณะกรรมการพัฒนาหลักสูตรการอบรมพยาบาลเวชปฏิบัติทั่วไป (การรักษาโรคเบื้องต้น) ณ วิทยาลัยพยาบาลบรมราชชนนี กรุงเทพมหานคร สำนักงานปลัดกระทรวงสาธารณสุข ได้รับการอนุมัติหลักสูตรจากสภาการพยาบาล พ.ศ. 2552 (วงรอบที่ 2)

4) กรรมการในคณะกรรมการพัฒนาหลักสูตรการอบรมพยาบาลเวชศาสตร์ฉุกเฉิน ณ วิทยาลัยพยาบาลบรมราชชนนี กรุงเทพมหานคร สำนักงานปลัดกระทรวงสาธารณสุข และได้รับการอนุมัติหลักสูตรจากสภาการพยาบาล พ.ศ. 2553

5) หัวหน้างานโครงการจัดการอบรมพยาบาลเวชปฏิบัติทั่วไป (การรักษาโรคเบื้องต้น) 2 รุ่น ระหว่าง พ.ศ. 2549-2550 (จำนวนรุ่นละ 50 คน) ณ วิทยาลัยพยาบาลบรมราชชนนี กรุงเทพมหานคร สำนักงานปลัดกระทรวงสาธารณสุข

6) วิทยากรรับเชิญสอนวิชากฎหมายและจริยธรรมวิชาชีพให้กับนักศึกษาพยาบาล วิทยาลัยพยาบาลพระมงกุฎ วิทยาลัยพยาบาลบรมราชชนนี ชลบุรี คณะพยาบาลศาสตร์ มหาวิทยาลัยราชภัฏสุราษฎร์ธานี วิทยาลัยพยาบาลบรมราชชนนี กรุงเทพมหานคร

7) วิทยากรรับเชิญเข้าร่วมประชุมเชิงปฏิบัติการพัฒนาอาจารย์ด้านจัดทำข้อสอบรวบยอดหลักสูตรพยาบาลศาสตร์บัณฑิต ประจำปีการศึกษา 2557 วิชากฎหมายและจริยธรรมวิชาชีพ จากกลุ่มงานพัฒนาการศึกษา สถาบันพระบรมราชชนก สำนักงานปลัดกระทรวงสาธารณสุข

8) วิทยากรรับเชิญเข้าร่วมประชุมเชิงปฏิบัติการวิพากษ์ข้อสอบรวบยอด หลักสูตรพยาบาลศาสตร์บัณฑิต ประจำปีการศึกษา 2555 วิชากฎหมายและจริยธรรมวิชาชีพ กลุ่มวิทยาลัยพยาบาลเครือข่ายภาคกลาง 1 ณ วิทยาลัยพยาบาลพระปกเกล้า จันทบุรี