

การศึกษาผลการให้คะแนนรายวิชาปฏิบัติการพยาบาลมารดาทารกและการผดุงครรภ์ 2 ในอาจารย์ผู้สอนภาคปฏิบัติที่มีประสบการณ์การสอนแตกต่างกัน

เนตรชานี กมรัตน์นนท์^{1*} นีรามัย อูสาหะ²

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อ 1) ศึกษาความคลาดเคลื่อนแบบกต ปล่อยคะแนน 2) ศึกษาความสอดคล้องในการประเมินของอาจารย์ผู้สอนภาคปฏิบัติ ประชากรคืออาจารย์ผู้สอนภาคปฏิบัติ 6 คน ที่สอนนักศึกษาพยาบาล ชั้นปีที่ 3 รายวิชาปฏิบัติการพยาบาลมารดา ทารกและการผดุงครรภ์ 2 ภาคการศึกษาที่ 2 ปีการศึกษา 2555 เครื่องมือที่ใช้คือ 1) แบบประเมินภูมิหลังของผู้ประเมิน และ 2) รุบริคการประเมินผลการปฏิบัติการพยาบาล 5 ระดับ วิเคราะห์โดย ร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และประมาณค่าความคลาดเคลื่อนจากผู้ประเมินด้วยวิธีของกิลฟอร์ด วิเคราะห์ความสอดคล้องกันระหว่างผู้ประเมินด้วยค่า intraclass correlation coefficient: ICC (3.3) (95% CI) ผลการวิจัยคือ อาจารย์ที่หน่วยห้องคลอด 3 คน และหน่วยหลังคลอด 3 คนมีประสบการณ์การสอนภาคทฤษฎีแตกต่างกัน 3 กลุ่ม คือไม่เคยสอนภาคทฤษฎี มีประสบการณ์น้อยกว่า 3 ปี และมากกว่า 3 ปี แต่ทุกคนมีประสบการณ์การสอนภาคปฏิบัติ และมีทัศนคติต่อการสอนในคลินิกในระดับปานกลาง ผลการประเมินคะแนนการฝึกปฏิบัติงานพิจารณาจากค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ที่หน่วยห้องคลอด พบว่าอาจารย์คนที่ 1 และ 2 มีการกตคะแนนคนที่ 3 มีการปล่อยคะแนน ที่หน่วยหลังคลอด พบว่า อาจารย์คนที่ 1 มีการกตคะแนน คนที่ 2 และ 3 มีการปล่อยคะแนน เมื่อวิเคราะห์ด้วยวิธีของกิลฟอร์ดให้ผลการวิเคราะห์สอดคล้องกันทั้งสองหน่วยฝึก ผลการวิเคราะห์ความสอดคล้องระหว่างผู้ประเมิน ที่หน่วยห้องคลอด พบว่าสมรรถนะส่วนใหญ่ มีค่า ICC ดิตลบ ยกเว้นสมรรถนะด้านความสามารถในการปฏิบัติกิจกรรมพยาบาลมีความสัมพันธ์ระดับปานกลาง ICC = 0.44 (CI = -0.04 - 0.72) และหน่วยหลังคลอดทุกสมรรถนะมีค่าดิตลบ ยกเว้นจรรยาบรรณวิชาชีพ มีความสัมพันธ์ระดับต่ำมาก ICC = 0.04 (95.00% CI = -1.65-0.03) แสดงว่าในภาพรวมคะแนนการประเมินระหว่างผู้ประเมินมีความไม่สัมพันธ์กัน

คำสำคัญ : อาจารย์ผู้สอนภาคปฏิบัติ ความคลาดเคลื่อนจากผู้ประเมิน ความคลาดเคลื่อนแบบกต ปล่อยคะแนน

¹ สาขาวิชาการพยาบาลมารดา ทารกและการผดุงครรภ์ คณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต กรุงเทพมหานคร
e-mail: netrachanee_kam@dusit.ac.th

² สาขาวิชาการพยาบาลมารดา ทารกและการผดุงครรภ์ คณะพยาบาลศาสตร์ มหาวิทยาลัยสวนดุสิต กรุงเทพมหานคร
e-mail: niramaiu@yahoo.com

* ผู้นิพนธ์หลัก e-mail: netrachanee_kam@dusit.ac.th

The study of rating results among nurse instructors with difference instruction experiences in Maternal and Newborn nursing and midwifery practicum²

Netrachanee Kamonranatanun^{1*} Niramai Usaha²

Abstract

The purposes of this study were: (1) to investigate rater effects, leniency, stringency error (2) to demonstrate the inter-rater agreement among the instructors. Population were six instructors with vary in experience of instruction the 3rd year student nurses who enrolled in practicum in maternal newborn and midwifery II course in 2nd semester of 2012. The Instruments were 1) the background of rater assessment questionnaire, 2) the 5 level of scoring rubrics for evaluating nursing outcome. Percent, mean, standard deviation, estimating of leniency/stringency error by means of Guilford (1954) Intraclass correlation coefficient: ICC (3.3) (95% CI) were employed. The results concluded that the clinical instructors' background were vary in 3 groups from no experience, less than 3 years and more than 3 years of experience in theoretical teaching, but all had experience in clinical teaching and had moderate attitude to clinical teaching. By means of mean and standard deviation rater 1 and rater2 at delivery room setting had stringency error, rater 1 was more stringency than rater2, but rater 3 was leniency. At postpartum care unit, rater 1 was stringency but rater2 and rater3 were leniency. Comparing to means of Guilford the results were presented accordingly at both setting. The inter-rater agreement among raters on six competency: competency of using nursing process, competency of nursing intervention, communication competency, nursing ethics and self- improvement at delivery room setting, most of them were demonstrated negative value of ICCs. Except for competency of nursing intervention was fair ICC =0.44 (95.00% CI=-0.04-0.72). At postpartum care setting, all competency were negative value of ICCs, but for nursing ethics was very poor ICC = 0.04 (95.00% CI=-1.65-0.03). In conclusion, the rating results were not in accord with each other.

Keywords : clinical instructor, rater effects, leniency/stringency error

¹ Bachelor of Nursing Science Program, Department of Maternal and Newborn Nursing and Midwifery, Faculty of Nursing, Suan Dusit University, e-mail: netrachanee_kam@dusit.ac.th

² Bachelor of Nursing Science Program, Department of Maternal and Newborn Nursing and Midwifery, Faculty of Nursing, Suan Dusit University, niramaiu@yahoo.com

* Corresponding author, email: netrachanee_kam@dusit.ac.th

บทนำ

คณะพยาบาลศาสตร์ มหาวิทยาลัยสวนราชภัฏสวนดุสิต มีการจัดการเรียนการสอนวิชาปฏิบัติการพยาบาลมารดาทารกและการผดุงครรภ์ 2 ซึ่งนักศึกษาพยาบาลจะต้องขึ้นฝึกปฏิบัติงานบนหอผู้ป่วย โดยอาศัยศาสตร์ด้านการพยาบาลมารดาทารก และการผดุงครรภ์เป็นฐานความรู้ในการปฏิบัติการพยาบาล ซึ่งมีความสำคัญเนื่องจากเป็น 2 ใน 8 วิชาที่จัดสอบเพื่อรับใบอนุญาตประกอบวิชาชีพการพยาบาล การประเมินผลการเรียนในวิชานี้จึงเน้นการประเมินภาคปฏิบัติของนักศึกษา และการประเมินที่มีคุณภาพมีปัจจัยสำคัญ คือ 1) คุณภาพของเครื่องมือ ซึ่งรายวิชานี้ใช้รูบริค 5 ระดับในการประเมินซึ่งผ่านการตรวจสอบความตรง และยังไม่เคยได้รับการตรวจสอบความเที่ยงมาก่อน และ 2) คุณภาพของผู้ประเมิน ซึ่งมีคู่มือและเกณฑ์การให้คะแนนแก่อาจารย์ผู้สอนภาคปฏิบัติทุกคนได้ศึกษาด้วยตนเอง คุณภาพผู้ประเมินที่สำคัญคือการลดความคลาดเคลื่อนจากผู้ประเมิน (Rater effect) (Reynolds, Livingston, & Willson, 2009) เช่น ความคลาดเคลื่อนจากการกดปล่อยคะแนน (Leniency/stringency error) (Wolfe, 2004) การตรวจสอบใช้สถิติดังนี้ 1) ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน หากค่าเฉลี่ยของผู้ประเมินคนหนึ่งสูงกว่าค่าเฉลี่ยกลุ่มมากแสดงถึงการปล่อยคะแนน ถ้าต่ำกว่าค่าเฉลี่ยกลุ่มมากแสดงถึงการกดคะแนน 2) ความถี่ของการให้คะแนน หากกระจุกตัวอยู่ตรงกลางมากแสดงถึงให้คะแนนแบบกลางๆ หากกระจุกตัวอยู่ที่คะแนนสูงแสดงถึงปล่อยคะแนน และถ้ากระจุกตัวอยู่ที่คะแนนต่ำแสดงถึงการกดคะแนน (Johnson, Penny, & Gordon, 2008) 3) การประมาณค่าความคลาดเคลื่อนด้วยวิธีวิเคราะห์ความแปรปรวนแบบสองทาง และการตรวจสอบความสอดคล้องระหว่าง ผู้ประเมิน เช่น (1) ค่าสัมประสิทธิ์สหสัมพันธ์แบบอันดับ (Spearman's rank-order correlation coefficient) (2) ค่าสหสัมพันธ์ของผลการประเมินของผู้ประเมิน (Correlation of scorers' results) หรือ ความสอดคล้องกันระหว่าง ผู้ประเมิน (Inter-rater reliability or intraclass correlation) คำนวณจากค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน และ (3) ค่าสัมประสิทธิ์ แคปปา (Kappa coefficient) (Reynolds, Livingston, & Willson, 2009) การเลือกใช้ค่าสถิติใดขึ้นอยู่กับระดับข้อมูล การวิจัยครั้งนี้เลือกวิเคราะห์ความสอดคล้องกันระหว่าง ผู้ประเมินด้วยค่า intraclass correlation coefficient (Hallgren, 2012)

ในภาคการศึกษาที่ 1 ปีการศึกษา 2555 พบปัญหาการประเมินในรายวิชาปฏิบัติการพยาบาลมารดาทารกและการผดุงครรภ์ 2 คือ อาจารย์ผู้สอนภาคปฏิบัติในแต่ละกลุ่มมีการประเมินภาคปฏิบัติที่แตกต่างกัน อาจเนื่องจากอาจารย์มีความหลากหลายด้านประสบการณ์ซึ่งจะส่งผลต่อการกำหนดเกรด (Donaldson & Gray, 2012) ที่ผ่านมายังไม่มีการศึกษาถึงปัญหาการให้คะแนนที่มีรูปแบบชัดเจน ดังนั้นการศึกษาคความคลาดเคลื่อนจากผู้ประเมินในผู้สอนภาคปฏิบัติที่มีประสบการณ์การสอนแตกต่างกันจึงมีความจำเป็น เพื่อให้ได้ข้อมูลที่แสดงถึงปัญหาที่ชัดเจน และหาแนวทางแก้ไขเพื่อให้การประเมินผลการศึกษภาคปฏิบัติมีคุณภาพมากขึ้นต่อไป

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาความคลาดเคลื่อนจากการประเมินแบบกด ปล่อยคะแนน ของการให้คะแนนปฏิบัติการพยาบาลมารดาทารกและการผดุงครรภ์ 2 ของอาจารย์ผู้สอนภาคปฏิบัติที่มีประสบการณ์การสอนต่างกัน
2. เพื่อศึกษาความสอดคล้องกันในการให้คะแนนของอาจารย์ผู้สอนภาคปฏิบัติที่มีประสบการณ์การสอนต่างกัน

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยเชิงบรรยาย

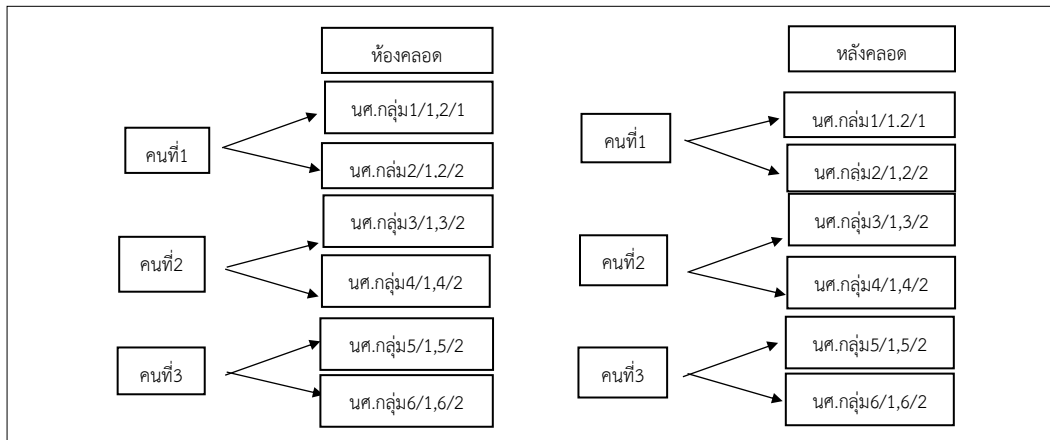
ประชากร อาจารย์ผู้สอนจำนวน 6 คน ในรายวิชาปฏิบัติการพยาบาลมารดาทารกและการผดุงครรภ์ 2 แก่นักศึกษาพยาบาลชั้นปีที่ 3 ภาคเรียนที่ 2 ปีการศึกษา 2555 คณะพยาบาลศาสตร์ มหาวิทยาลัยราชภัฏสวนดุสิต

เครื่องมือที่ใช้ในการวิจัย 1) แบบประเมินภูมิหลังของอาจารย์ผู้สอนภาคปฏิบัติ มีค่าสัมประสิทธิ์แอลฟาของครอนบาคเท่ากับ .92 2) แบบประเมินผลการปฏิบัติการพยาบาล ของคณะพยาบาลศาสตร์ มหาวิทยาลัยราชภัฏสวนดุสิต ที่ใช้เกณฑ์การประเมินแบบรูบริค 5 ระดับ 6 สมรรถนะคือ 1) ความสามารถในการใช้กระบวนการพยาบาล 2) ความสามารถในการปฏิบัติกิจกรรมพยาบาล 3) ความสามารถในการติดต่อสื่อสาร และ 4) จรรยาบรรณวิชาชีพ 5) การพัฒนาตนเอง 6) การประเมินผลการปฏิบัติการพยาบาล นำมาใช้โดยไม่ได้ดัดแปลงเนื่องจากต้องการศึกษาสภาพจริงจากการประเมิน

การเก็บรวบรวมข้อมูล

1. ส่งและเก็บรวบรวมแบบสอบถามภูมิหลังให้อาจารย์ผู้สอนภาคปฏิบัติทั้ง 6 คน

2. มอบหมายให้อาจารย์สอนภาคปฏิบัตินักศึกษา 2 ตอนเรียน กลุ่มย่อยละ 7-8 คน อาจารย์ 1 คนได้รับมอบหมายให้สอน นักศึกษา 4 กลุ่มย่อยรวม 28 ถึง 31 คน นักศึกษาแต่ละกลุ่มย่อยจะได้รับการคละระดับสมรรถนะโดยนำผลสัมฤทธิ์ทางการเรียนรายวิชาการพยาบาลมารดาทารกและการผดุงครรภ์ 2 และ รายวิชาปฏิบัติการพยาบาลมารดาทารกและการผดุงครรภ์ 1 มาเป็นเกณฑ์ในการจัดนักศึกษาเข้ากลุ่มย่อย จึงทำให้นักศึกษาแต่ละกลุ่มมีคุณลักษณะใกล้เคียงกัน เมื่อสิ้นสุดการฝึกปฏิบัติงานเก็บรวบรวมข้อมูลผลการประเมินการปฏิบัติจากอาจารย์ทั้ง 6 คน ดังแสดงในภาพที่ 1



ภาพที่ 1 แผนภาพการมอบหมายงานการสอนภาคปฏิบัติให้อาจารย์ผู้สอน

หมายเหตุ: 1. อาจารย์คนที่ 1-3 ที่หน่วยห้องคลอด และคนที่ 1-3 ของหน่วยหลังคลอดจะอยู่ประจำหน่วยการฝึกโดยไม่มีการหมุนเวียน

2. ตัวเลขหน้าเครื่องหมาย / แทนนักศึกษาตอนเรียนนั้น ตัวเลขหลังเครื่องหมาย/ แทนนักศึกษาในกลุ่มย่อยนั้น เช่น ตอนเรียนที่ 1 กลุ่มย่อยที่ 1 เขียนสัญลักษณ์คือ นศ.กลุ่ม 1/1 นักศึกษาที่เขียนสัญลักษณ์ตอนเรียน และกลุ่มย่อยเหมือนกันคือนักศึกษากลุ่มเดียวกัน

การวิเคราะห์ข้อมูล การวิเคราะห์ความคลาดเคลื่อนแบบกต ปล่อยคะแนน ด้วยค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน และ Cell mean, mean all raters, mean all ratees, grand mean, standard deviation ด้วยวิธีของกิลฟอร์ด (Guilford, 1954) ค่าความสอดคล้องระหว่างผู้ประเมินของอาจารย์ผู้สอนภาคปฏิบัติที่หน่วยห้องคลอด และหน่วยหลังคลอด วิเคราะห์โดยค่า Intraclass correlation coefficient: ICC ใช้โมเดล 3 จากข้อตกลงเบื้องต้นคือผู้ประเมินทุกคนทำการประเมิน ผู้ถูกประเมินแต่ละคน (Fully –crossed designs) โดยไม่ได้มีการสุ่มผู้ประเมิน (Fixed effect) ทำการประเมินงานการประเมินเดียวกัน คือนักศึกษาที่ขึ้นฝึกปฏิบัติงานที่หน่วยห้องคลอด และหลังคลอดที่มีคุณลักษณะคล้ายคลึงกัน ประโยชน์คุณลักษณะสำหรับการประมาณค่า คือ ICC (3, 3) และใช้ค่าการประเมินเฉลี่ย (Average measures) ในการประมาณค่า (Hallgren, 2012)

ผลการวิจัย

ผลการวิจัยแบ่งเป็น 3 ส่วนดังนี้

1. ภูมิหลังของอาจารย์ผู้สอนภาคปฏิบัติ ณ หน่วยห้องคลอด อาจารย์คนที่ 1 มีประสบการณ์ทั้งภาคทฤษฎีและปฏิบัติมากกว่า 5 ปี คนที่ 2 มีประสบการณ์การสอนทั้งภาคทฤษฎีและปฏิบัติน้อยกว่า 3 ปี คนที่ 3 ไม่มีประสบการณ์การสอนภาคทฤษฎีเลยแต่มีประสบการณ์การสอนภาคปฏิบัติมากกว่า 3 ปี ณ หน่วยหลังคลอด มีประสบการณ์การสอนภาคทฤษฎีและปฏิบัติทุกคน คนที่ 1 มีประสบการณ์การสอนภาคทฤษฎีและปฏิบัติ 30 ปี คนที่ 3 มีประสบการณ์การสอนภาคทฤษฎีและปฏิบัติ 6 ปี และคนที่ 2 มีประสบการณ์การสอนภาคทฤษฎีและปฏิบัติ 1 ปี และและอาจารย์ทุกท่านมีทัศนคติต่อการสอนในคลินิกในระดับปานกลางดังแสดงในตารางที่ 1

ตารางที่ 1 ผลการวิเคราะห์ข้อมูลภูมิหลังของอาจารย์ผู้สอนภาคปฏิบัติ

อาจารย์	อายุ (ปี)	ประสบการณ์ การสอนในคลินิก	ประสบการณ์การสอนทฤษฎี	ทัศนคติต่อการสอนในคลินิก
ห้องคลอด				
คนที่ 1	61	6	6	ระดับปานกลาง
คนที่ 2	30	1	1	ระดับปานกลาง
คนที่ 3	39	4	0	ระดับปานกลาง
หลังคลอด				
คนที่ 1	58	30	30	ระดับปานกลาง
คนที่ 2	46	1	1	ระดับปานกลาง
คนที่ 3	44	6	6	ระดับปานกลาง

2. ผลการประมาณความคลาดเคลื่อนแบบกต ปล่อยคะแนน

2.1 หน่วยห้องคลอดพบว่า ค่าเฉลี่ยคะแนนของกลุ่ม = 4.47 คนที่ 1 และคนที่ 2 มีค่าเฉลี่ยน้อยกว่าค่าเฉลี่ยของกลุ่ม = -0.14 และ -0.08 ตามลำดับ คนที่ 3 มีค่าเฉลี่ยมากกว่าค่าเฉลี่ยของกลุ่ม = 0.22 และส่วนเบี่ยงเบนมาตรฐาน มีการกระจายของคะแนนรอบๆ ค่าเฉลี่ยของกลุ่มที่ 4.47 ± 0.13 เรียงตามการกระจายตัวรอบๆ ค่าเฉลี่ยจากมากที่สุดคือคนที่ 1 = $\pm .32$ รองลงมาคนที่ 2 = ± 0.20 และ คนที่ 3 = ± 0.10 ตามลำดับ วิเคราะห์ด้วยวิธีของกิลฟอร์ด ผู้ประเมินคนที่ 1 และคนที่ 2 มีค่าประมาณความคลาดเคลื่อนแบบกตปล่อยคะแนน = -0.14 และ -0.08 คนที่ 3 = 0.22 ตามลำดับ ดังแสดงในตารางที่ 2

ตารางที่ 2 ผลการประมาณค่าความคลาดเคลื่อนแบบกด ปล่อยคะแนน ของอาจารย์ผู้สอนภาคปฏิบัติที่ห้องคลอด

ratee \ rater	1	2	3	28	Max	Min	Mean all rates	X_{kl}	S.D
1	4.00	4.05	4.00	3.76	4.86	3.71	4.33	-0.14	0.32
2	4.62	4.48	4.48	4.00	4.71	3.95	4.39	-0.08	0.20
3	4.76	4.76	4.76	4.67	4.86	4.33	4.69	0.22	0.10
Mean all rater	4.75	4.40	4.57	4.59			4.47		0.13
di	0.26	-0.09	0.09	-0.10					

n = 28, **หมายเหตุ:** ratee แทนนักศึกษาที่ได้รับการประเมินการปฏิบัติงาน จำนวน 28 คน, rater แทนอาจารย์ผู้สอนภาคปฏิบัติ คนที่ 1 ถึง 3, X_{kl} แทน ค่าประมาณความคลาดเคลื่อนแบบกดปล่อยคะแนน

2.2 หน่วยหลังคลอดพบว่า ค่าเฉลี่ยคะแนนของกลุ่ม = 4.24 คนที่ 1 มีค่าเฉลี่ยน้อยกว่าค่าเฉลี่ยของกลุ่ม = -0.29 คน 2 และคนที่ 3 มีค่าเฉลี่ยมากกว่าค่าเฉลี่ยของกลุ่ม = 0.22 และ 0.13 ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานพบว่ามีกระจายของคะแนนรอบๆ ค่าเฉลี่ยของกลุ่มคือ 4.24 ± 0.09 คนที่ 2 คนที่ 1 และคนที่ 3 มีการประเมินคะแนนโดยมีกระจายของคะแนนรอบๆ ค่าเฉลี่ย = ± 0.16 , ± 0.19 และ ± 0.26 ตามลำดับ เมื่อพิจารณาด้วยวิธีของกิลฟอร์ดพบว่า ผู้ประเมินคนที่ 1 มีค่าประมาณความคลาดเคลื่อนแบบกดปล่อยคะแนน -0.30 คนที่ 2 และ 3 = 0.21 และ 0.08 ตามลำดับ ดังตารางที่ 3

ตารางที่ 3 ผลการประมาณค่าความคลาดเคลื่อนแบบกด ปล่อยคะแนน ของอาจารย์ผู้สอนภาคปฏิบัติที่หน่วยหลังคลอด

ratee \ rater	1	2	3	28	Max	Min	Mean all rates	X_{kl}	S.D
1	4.00	4.05	4.00	3.81	4.29	3.43	3.95	0.30	0.19
2	4.57	4.57	4.19	4.43	4.71	4.00	4.46	0.21	0.16
3	4.76	3.81	4.57	4.62	4.76	3.62	4.33	0.08	0.26
Mean all rater	4.44	4.14	4.25	4.29			4.24		0.09
di	0.20	-0.10	0.01	0.04					

n = 28, **หมายเหตุ:** ratee แทนนักศึกษาที่ได้รับการประเมินการปฏิบัติงาน จำนวน 28 คน, rater แทนอาจารย์ผู้สอนภาคปฏิบัติ คนที่ 1 ถึง 3, X_{kl} แทน ค่าประมาณความคลาดเคลื่อนแบบกดปล่อยคะแนน

3. ผลการวิเคราะห์ความสอดคล้องระหว่างผู้ประเมินของอาจารย์ผู้สอนภาคปฏิบัติแบ่งตามหน่วยฝึก

3.1 หน่วยห้องคลอด พิจารณาตามสมรรถนะ พบว่าสมรรถนะด้านความสามารถในการใช้กระบวนการพยาบาล ICC = -0.27 (95% CI = -0.93 - 0.49) ความสามารถในการปฏิบัติกิจกรรมพยาบาลมีความสัมพันธ์ระดับปานกลาง ICC = 0.44 (95.00% CI = -0.04 - 0.72) ความสามารถในการติดต่อสื่อสาร ICC = -0.68 (95% CI = -2.16 - 0.16) จรรยาบรรณวิชาชีพ ICC = -0.26 (95% CI = -1.35 - 0.38) การพัฒนาตนเอง ICC = 0.35 (95.00% CI = -0.22 - 0.68) และการประเมินผลการปฏิบัติการพยาบาล ICC = -0.04 (95.00% CI = -0.95 - 0.48) ดังแสดงในตารางที่ 4

ตารางที่ 4 ผลการวิเคราะห์ความสอดคล้องระหว่างผู้ประเมินของอาจารย์ผู้สอนภาคปฏิบัติหน่วยห้องคลอด

หัวข้อประเมิน	ICC (3,3)	95% CI
1. ความสามารถในการใช้กระบวนการพยาบาล	-0.27	-0.93-0.49
2. ความสามารถในการปฏิบัติกิจกรรมพยาบาล	0.44	-0.04-0.72
3. ความสามารถในการติดต่อสื่อสาร	-0.68	-2.16-0.16
4. จรรยาบรรณวิชาชีพ	-0.26	-1.35-0.38
5. การพัฒนาตนเอง	0.35	-0.22-0.68
6. การประเมินผลการปฏิบัติการพยาบาล	-0.04	-0.95-0.48

3.2 หน่วยหลังคลอดพิจารณาตามสมรรถนะ พบว่าสมรรถนะด้านความสามารถในการใช้กระบวนการพยาบาล ICC = -0.04 (95% CI = -0.96-0.48) ความสามารถในการปฏิบัติกิจกรรมพยาบาล ICC = -0.32 (95.00% CI = -1.47-0.35) ความสามารถในการติดต่อสื่อสาร ICC = -0.15 (95% CI = -1.49-0.47) จรรยาบรรณวิชาชีพ ICC = 0.04 (95% CI = -0.81-0.52) การพัฒนาตนเอง ICC = -0.41 (95.00% CI = -1.65-0.30) และการประเมินผลการปฏิบัติการพยาบาล ICC = -0.18 (95.00% CI = -1.22-0.41) ดังแสดงในตารางที่ 5

ตารางที่ 5 ผลการวิเคราะห์ค่าดัชนีความสอดคล้องระหว่างผู้ประเมินของอาจารย์ผู้สอนภาคปฏิบัติหน่วยหลังคลอด

หัวข้อประเมิน	ICC (3,3)	95% CI
1. ความสามารถในการใช้กระบวนการพยาบาล	-0.04	-0.96-0.48
2. ความสามารถในการปฏิบัติกิจกรรมพยาบาล	-0.32	-1.47-0.35
3. ความสามารถในการติดต่อสื่อสาร	-0.15	-1.49-0.47
4. จรรยาบรรณวิชาชีพ	0.04	-0.81-0.52
5. การพัฒนาตนเอง	-0.41	-1.65-0.30
6. การประเมินผลการปฏิบัติการพยาบาล	-0.18	-1.22-0.41

สรุปและอภิปรายผลการวิจัย

สรุปผลการวิจัย

ผลการศึกษาความคลาดเคลื่อนแบบกด ปล่อยคะแนน ของอาจารย์ผู้สอนภาคปฏิบัติ ณ หน่วยห้องคลอด อาจารย์คนที่ 1 มีการกดคะแนนมากกว่าคนที่ 2 ในขณะที่คนที่ 3 มีการปล่อยคะแนน การกระจายตัวของคะแนนรอบๆ ค่าเฉลี่ยของกลุ่มพบว่าคนที่ 1 มีการกระจายตัวของคะแนนมากที่สุด รองลงมาคือ คนที่ 2 และคนที่ 3 น้อยที่สุด เมื่อพิจารณาด้วยวิธีของกิลฟอร์ด อาจารย์คนที่ 1 และคนที่ 2 มีค่าประมาณความคลาดเคลื่อนแบบกด ปล่อยคะแนนต่ำกว่าความเป็นจริง (negative leniency error) หรือกดคะแนน คนที่ 3 มีค่าประมาณความคลาดเคลื่อนแบบกด ปล่อยคะแนนสูงกว่าความเป็นจริง (positive leniency error) หรือปล่อยคะแนน ซึ่งให้ผลสอดคล้องเป็นไปในทิศทางเดียวกัน ณ หน่วยหลังคลอด จากค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานพบว่า คนที่ 1 มีการกดคะแนน ส่วนคนที่ 2 และคนที่ 3 มีการปล่อยคะแนนโดยคนที่ 2 มีการปล่อยคะแนนมากกว่าคนที่ 3 การกระจายตัวของคะแนนรอบๆ ค่าเฉลี่ยของกลุ่ม อาจารย์คนที่ 3 มีการกระจายตัวของคะแนนรอบๆ

ค่าเฉลี่ยมากที่สุด รองลงมาคือ คนที่ 1 และคนที่ 2 ตามลำดับ ผู้ประเมินคนที่ 1 มีค่าประมาณความคลาดเคลื่อนแบบกดปล่อยคะแนนต่ำกว่าความเป็นจริง (negative leniency error) หรือกดคะแนน ผู้ประเมินคนที่ 2 และ 3 มีค่าประมาณความคลาดเคลื่อนแบบกดปล่อยคะแนนสูงกว่าความเป็นจริง (positive leniency error) หรือปล่อยคะแนน

ผลการศึกษาความสอดคล้องระหว่างผู้ประเมินของอาจารย์ผู้สอนภาคปฏิบัติ ณ หน่วยห้องคลอด และหน่วยหลังคลอด มีค่า ICC ตีติลลเป็นส่วนใหญ่แสดงว่ามีความไม่สอดคล้องของการประเมินอย่างเป็นระบบ (systematic disagreement) ระหว่างผู้ประเมิน (Hallgren, 2012) แสดงถึงผลการประเมินที่ไม่ไปในทิศทางเดียวกันของอาจารย์ผู้ประเมิน

อภิปรายผล

ข้อมูลเบื้องต้นที่พบว่าอาจารย์ผู้สอนภาคปฏิบัติในหน่วยห้องคลอด และหลังคลอดมีประสบการณ์การสอนที่แตกต่างอย่างมากตั้งแต่ไม่มีประสบการณ์การสอนภาคทฤษฎีไปจนถึงมีประสบการณ์มาก ปัจจัยนี้มีความสำคัญต่อผลคุณภาพการประเมินการปฏิบัติงาน จากการทบทวนวรรณกรรมเกี่ยวกับการประเมินภาคปฏิบัติในนักศึกษาพยาบาลทั้งในและต่างประเทศพบการศึกษาด้านคุณภาพของผู้ประเมินน้อยมาก ส่วนใหญ่เป็นการพัฒนาเครื่องมือที่ใช้ในการประเมินมากกว่าที่จะศึกษาผู้ประเมิน (ปิยะรัตน์ หยกสุริยันต์, 2549; จิราภา คงสัมฤทธิ์, 2544; อนัญญา คูอาริยะกุล, 2554) จึงยกตัวอย่างการศึกษาในสาขาทางการศึกษา เช่น การศึกษาของควัฟแมน, เบเออร์, โคลและ เซ็กซ์ตัน (Kaufman, Baer, Cole, & Sexton, 2008) ทำการศึกษาความสอดคล้องของผู้ประเมินในการประเมินการเขียน ใน 2 กลุ่มคือกลุ่มที่มีความเชี่ยวชาญ และกลุ่มผู้ที่ไม่มีความเชี่ยวชาญในการประเมิน ผลการประเมินไม่พบความสอดคล้องของการให้คะแนนระหว่างผู้ประเมินที่เชี่ยวชาญกับผู้ประเมินที่ไม่มีความเชี่ยวชาญ และพบว่าค่าความสอดคล้องกันระหว่างผู้ประเมินที่ไม่มีความเชี่ยวชาญมีค่าต่ำ (low interrater reliability) และไม่สอดคล้องกับคะแนนของผู้เชี่ยวชาญ หรือการศึกษาของ บาร์เกา ยูอิ และการศึกษาของ ลิน (Kaufman, 2010; Lim, 2011) พบว่าผู้ประเมินที่มีความเชี่ยวชาญจะมีการกดคะแนน ส่วนผู้ประเมินที่ไม่มีความเชี่ยวชาญจะมีการปล่อยคะแนน เนื่องจากผู้ประเมินที่มีความเชี่ยวชาญย่อมมีความเข้าใจในเนื้อหาสาระของการประเมิน เข้าใจในเกณฑ์การประเมินและเกิดความมั่นใจในการใช้เกณฑ์การประเมิน และสามารถแยกคนเก่งคนอ่อนออกจากการได้สังเกตได้จากการกระจายตัวของคะแนนทั่วถึงตลอดมาตการประเมิน ในขณะที่ผู้ประเมินมือใหม่จะไม่มี ความเข้าใจที่ชัดเจน ไม่มี ความมั่นใจในการให้คะแนน และกลัวการเกิดความรู้สึกผิดที่ประเมินให้ผู้ถูกประเมินต่ำจึงมักให้คะแนนกระจุกตัวที่ระดับคะแนนกลางหรือสูง ทำให้ไม่สามารถแยกคนเก่ง คนอ่อนออกจากกันได้ (Swedler-Brown, 1985 cited in Barkaoui, 2010)

ค่า ICC ตีติลลแสดงถึงความขัดแย้งกันระหว่างผู้ประเมินอย่างเป็นระบบ (systematic) คือมีความขัดแย้งกันอย่างทั่วถึง ไม่ได้เกิดแบบสุ่ม (random) ค่า ICC ที่มีค่าต่ำมาก หรือมีค่าตีติลลสะท้อนว่าการประเมินไม่ได้คุณภาพจากสาเหตุสำคัญ 2 ประการคือ 1) ผู้ประเมินได้รับการเตรียมศักยภาพด้านการประเมินไม่เพียงพอ (not well trained rater) หรือ 2) เครื่องมือหรือแบบประเมินที่ใช้ในการประเมินไม่ได้รับการนิยามคุณลักษณะที่ดีพอ (Rousson, Gasser, & Seifert, 2003) สอดคล้องกับการศึกษาปัญหาของการประเมินภาคปฏิบัติในนักศึกษาพยาบาลของโดเนลสัน และเกร ที่พบว่าความเป็นอัตโนมัติของผู้ประเมินเป็นปัญหาหลักของการประเมินและส่งผลถึงการกำหนดเกรดการศึกษา จึงได้สรุปแนวปฏิบัติในการแก้ปัญหา 3 แนวทางคือ 1) พัฒนาเครื่องมือในการทดสอบทักษะปฏิบัติ และ ควรใช้รูปรีคในการประเมิน 2) ใช้วิธีการประเมินหลายวิธีไม่ควรใช้วิธีการเดียว และ 3) การฝึกอบรมอาจารย์ผู้สอนภาคปฏิบัติให้มีความเข้าใจให้ตรงกันในเกณฑ์การประเมิน (Donaldson & Gray, 2012) ดังนั้นข้อมูลจากผลการศึกษาครั้งนี้สรุปได้ว่าผลของความสอดคล้องกันระหว่างผู้ประเมินที่มีค่าตีติลล

นั้นมีผลสืบเนื่องจากระบบการเตรียมผู้ประเมินยังไม่มีประสิทธิภาพเพียงพอ และ หรือแบบประเมินที่ใช้ในการประเมินครั้งนี้ยังไม่เป็นไปตามมาตรฐานเนื่องจากยังไม่เคยได้รับการตรวจสอบความเที่ยงก่อนการนำไปใช้จริง

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

นำผลการวิจัยซึ่งสะท้อนคุณภาพการประเมินที่ยังไม่เป็นไปตามมาตรฐาน เป็นข้อมูลพื้นฐานในการวางแผนปรับปรุงและ พัฒนาระบบการเตรียมอาจารย์ผู้สอนภาคปฏิบัติ ให้เกิดคุณภาพมากขึ้น และเป็นรูปธรรมมากขึ้น รวมถึงการพัฒนาเครื่องมือประเมินที่มีความเป็นปรนัยมากขึ้น

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

ควรทำการศึกษายืนยันคุณภาพของเครื่องมือที่ใช้ในการประเมินตามมาตรฐานการประเมินรูบริคที่ใช้อยู่ในปัจจุบัน ซึ่งอาจทำในรายวิชาปฏิบัติอื่นๆ และควรจัดกระทำให้มีความถูกต้องตามหลักการของโมเดลการประมาณค่า Intraclass correlation coefficient โมเดล 3 คือผู้ประเมินตั้งแต่ 2 คนขึ้นไปทำการประเมินผู้ถูกประเมินคนเดียวกันหรือกลุ่มเดียวกัน ในช่วงเวลาเดียวกัน ในงานการประเมินที่มีคุณลักษณะเดียวกัน ซึ่งหากมีข้อจำกัดเรื่องเวลาและทรัพยากรด้านงบประมาณอาจจัดกระทำในช่วงสั้นๆ ช่วงใดช่วงหนึ่งของการฝึก เพื่อความชัดเจนของผลการวิจัย

กิตติกรรมประกาศ

รายงานการวิจัยฉบับนี้ได้รับทุนอุดหนุนการวิจัยปีงบประมาณ 2556 ของมหาวิทยาลัยราชภัฏสวนดุสิต

เอกสารอ้างอิง

- จิราภา คงสมฤทธิ์. (2544). การพัฒนาแบบประเมินการสอนภาคปฏิบัติในคลินิกของอาจารย์พยาบาล. วิทยานิพนธ์ปริญญาศึกษาศาสตรมหาบัณฑิต คณะศึกษาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์.
- ปิยะรัตน์ หยกสุริยันต์, 2549 ปรัชญา ไทยแท้. (2550). การพัฒนาแบบประเมินสมรรถนะของพยาบาลวิชาชีพ โรงพยาบาลราชวิถี. วิทยานิพนธ์พยาบาลศาสตรมหาบัณฑิต คณะพยาบาลศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- อนัญญา คูอาริยะกุล. (2554). การพัฒนารูปแบบการประเมินการฝึกภาคปฏิบัติสำหรับนักศึกษาพยาบาลสังกัดสถาบันพระบรมราชชนก. วิทยานิพนธ์ปริญญาการศึกษาดุษฎีบัณฑิต คณะศึกษาศาสตร์ มหาวิทยาลัยนครสวรรค์.
- Barkaoui, K. (2010). Do ESL essay raters' evaluation criteria change with experience? A mixed-methods, cross-sectional study. *TESOL Quarterly*, 31-57.
- Donaldson, J.H.,and Gray, M. (2012). Systematic review of grading practice: is there evidence of grade inflation?. *Nurse education in practice*, 12: 101-104.
- Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York : McGraw-Hill.
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, 8(1), 23.
- Johnson, R. L., Penny, J. A., & Gordon, B. (2008). *Assessing performance: Designing, scoring, and validating performance tasks*: Guilford Press.

- Kaufman, J. C., Baer, J., Cole, J. C., & Sexton, J. D. (2008). A comparison of expert and nonexpert raters using the consensual assessment technique. *Creativity Research Journal*, 20(2), 171-178.
- Lim, G. S. (2011). The development and maintenance of rating quality in performance writing assessment: A longitudinal study of new and experienced raters. *Language Testing*, 28(4) 543 –560.
- Reynolds, C. R., Livingston, R. B., & Willson, V. L. (2009). *Measurement and assessment in education*. Upper Saddle River: Pearson Merrill Prentice Hall.
- Rousson, V., Gasser, T., & Seifert, B. (2003). Confidence Intervals for Intraclass Correlation in Inter-Rater Reliability. *Scandinavian journal of statistics*, 30(3), 617-624.
- Sweedler-Brown, C. O. (1985). The influence of training and experience on holistic essay evaluation. *English Journal*, 74, 49-55.
- Wolfe, E. W. (2004). Identifying rater effects using latent trait models. *Psychology Science*, 46, 35-51.