

**Development, Administration, and Validity Evidence of a Novel Preparatory Test for
Nephrology Licensure**

BY

JOHN FRANCIS JOHNSON

B.S., York University, 1992

M.D., University of Ottawa, 1996

M.S., Western University, 2007

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Master of Health Professions Education
in the Graduate College of the
University of Illinois at Chicago, 2017
Chicago, Illinois

Defense Committee:

Yoon Soo Park, Chair and Advisor

Alan Schwartz

Matthew Lineberry, University of Kansas

Faisal Rehman, Western University

This thesis is dedicated to my wife, Kim, and to my three sons, Jack, Cooper, and Brady.

Their unconditional love and encouragement enabled me to complete this vision.

ACKNOWLEDGMENTS

I would like to sincerely thank my thesis chair (Yoon Soo Park) and committee members (Alan Schwartz, Matthew Lineberry, and Faisal Rehman) for their tireless guidance and support. I would also like to thank the following individuals for fulfilling valuable project roles; Hariharan Iyer (independent grader), Laila Sadagah (pilot tester), Alaa Alem (pilot tester), and Jennifer Rombouts (administrative assistant).

JFJ

TABLE OF CONTENTS

<u>CHAPTER</u>		<u>PAGE</u>
1.	INTRODUCTION.....	1
2.	PURPOSE.....	2
3.	CONCEPTUAL FRAMEWORKS.....	4
4.	METHODS.....	6
	4.1 Grader Training.....	6
	4.2 Content.....	6
	4.3 Response Process.....	7
	4.4 Internal Structure.....	8
	4.5 Relations to other Variables.....	8
	4.6 Consequences.....	8
5.	RESULTS.....	10
	5.1 Content.....	10
	5.2 Response Process.....	10
	5.3 Internal Structure.....	14
	5.4 Relations to other Variables.....	17
	5.5 Consequences.....	19
6.	DISCUSSION.....	21
7.	LIMITATIONS.....	24
8.	CONCLUSION.....	25

TABLE OF CONTENTS (continued)

<u>CHAPTER</u>	<u>PAGE</u>
CITED LITERATURE.....	26
APPENDIX.....	28
VITA.....	32

LIST OF TABLES

<u>TABLE</u>	<u>PAGE</u>
I. TEST BLUEPRINT FOR THE CONSTRUCTED RESPONSE PREPARATORY TEST: PERCENT WEIGHT.....	7
II. ITEM CLARITY AND LEVEL OF DIFFICULTY.....	11
III. CONSTRUCTED RESPONSE PREPARATORY TEST NEPHROLOGY TRAINEES SCORES (TOTAL POTENTIAL POINTS 261).....	12
IV. CONSTRUCTED REPSONSE PREPARATORY TEST NEPHROLOGY TRAINEES SCORES BREAKDOWN BY CONTENT DOMAIN.....	12
V. CONSTRUCTED RESPONSE PREPARATORY TEST NEPHROLOGY TRAINEES SCORES (TOTAL POTENTIAL POINTS 55).....	13
VI. ITEM DIFFICULTY AND ITEM-TOTAL CORRELATION.....	15
VII. VARIANCE COMPONENT ANALYSIS: G-STUDY (TOTAL POTENTIAL POINTS 261).....	16
VIII. VARIANCE COMPONENT ANALYSIS: G-STUDY (TOTAL POTENTIAL POINTS 55).....	16
IX. CONSTRUCTED RESPONSE PREPARATORY TEST POSTGRADUATE YEAR 4 NEPHROLOGY TRAINEE SCORE BREAKDOWN BY CONTENT DOMAIN.....	17
X. CONSTRUCTED RESPONSE PREPARATORY TEST POSTGRADUATE YEAR 5 NEPHROLOGY TRAINEE SCORE BREAKDOWN BY CONTENT DOMAIN.....	18
XI. CONSTRUCTED RESPONSE PREPARATORY TEST NEPHROLOGY CONSULTANT SCORES (TOTAL POTENTIAL POINTS 261).....	18

LIST OF TABLES (continued)

<u>TABLE</u>	<u>PAGE</u>
XII. CONSTRUCTED RESPONSE PREPARATORY TEST NEPHROLOGY CONSULTANT SCORES BREAKDOWN BY CONTENT DOMAIN.....	19

LIST OF ABBREVIATIONS

PGY	Postgraduate year
CR	Constructed-response
RCPSC	Royal College of Physicians and Surgeons of Canada
WU	Western University
SR	Selected-response
CRPT	Constructed-response preparatory test
KT	Kidney transplantation
RPh	Renal physiology
AKI	Acute kidney injury
RPa	Renal pathology
CKD	Chronic kidney disease
PD	Peritoneal dialysis
HD	Hemodialysis
ICC	Intraclass correlation
SD	Standard deviation

SUMMARY

Western University nephrology program trainees lack validated assessment scores from a preparatory test for their licensing examination. A 55-item constructed-response preparatory test was developed and administered to meet the needs of nephrology trainees and the program at Western University. This study examines validity evidence of assessment scores from the preparatory test.

Messick's unified validity framework was used to gather validity evidence for content, response process, internal structure, relations to other variables, and consequences. Tests were scored independently by three qualified graders using an a priori developed/modified rubric. Questionnaires completed before and after the preparatory test were used to gather data on the trainees' perception of examination preparedness, item clarity, and curriculum adequacy.

There were 10 trainees and 5 faculty volunteers who took this preparatory test. The majority of trainees passed the constructed-response preparatory test. However, many scored poorly on items assessing renal pathology and physiology knowledge. The faculty members who took the preparatory test also had low scores in these two domains. Overall, constructed-response preparatory test scores reflected nephrology exposure (one year versus two years versus many years).

The preparatory test and questionnaire data highlighted shortcomings of the Western University nephrology training program curriculum. Following the preparatory test, trainees expressed an improved sense of preparedness for their upcoming licensing examination.

1. INTRODUCTION

In Canada, nephrology is classified as a medical subspecialty for which training entails a 2-year commitment completed during postgraduate years (PGY) four and five. Three months after successful completion of a nephrology training program, trainees complete a 3-hour, 55-item short-answer constructed-response (CR) test administered by the Royal College of Physicians and Surgeons of Canada (RCPSC), the national governing body for medical and surgical specialty training programs. The summative RCPSC CR test is modelled after the medical expert competencies for Canadian nephrology program trainees; it was constructed and updated by the College in 2012. To practice nephrology in Canada, trainees must earn a “Certificate of Special Competence in Nephrology”, which is awarded when one obtains a passing grade on the RCPSC summative examination.

Nephrology trainees at Western University (WU) have been assessed annually with a locally developed 25-item selected-response (SR) test. At present, trainees only receive test scores on their performance upon completion of the formative SR test. Unfortunately, there is no published literature on preparatory tests for the RCPSC licensing examination instituted by nephrology training programs in Canada. As such, the current learner assessment process for nephrology trainees at WU requires significant improvement. Firstly, WU administers an SR-format preparatory test, which does not match the CR format of the RCPSC summative examination. Furthermore, our learners do not receive any constructive test performance feedback or an opportunity to evaluate the examination process.

2. PURPOSE

In this study, we developed, administered, and collected validity evidence for a novel short-answer CR preparatory test (CRPT) for nephrology trainees at WU. The CRPT was administered in June of 2016 to all nephrology trainees at WU (five PGY4 and five PGY5), after they provided written consent for participation. The CRPT was also administered to five practicing nephrology consultants at WU, to examine their response patterns in comparison to the trainees and gain additional data for the answer keys. A scoring rubric was created in advance of the test administration, and three nephrology-medical education-trained individuals graded each test independently.

This project created an assessment instrument that is the same testing format as the RCPSC licensing examination and also provided valuable test performance feedback to the CRPT takers. The study also included an evaluation of the CRPT, in the form of two questionnaires to gauge trainees' perception of the newly developed test and their preparedness for the RCPSC summative examination. The data obtained was used to develop potential improvements of the assessment instrument. Lastly, trainees' CRPT performance was used to identify any curricular deficiencies in the training program.

Test takers received item-specific feedback in the form of test performance reports, identifying area(s) of academic weakness on which the trainee can focus. The CRPT was administered 3 or 15 months prior to the RCPSC summative examination (PGY5 or PGY4, respectively), allowing ample time to modify study habits accordingly.

Examining test taker performance and feedback using the CRPT can guide further revisions and also contribute to assessments that help learners at advanced stages of training to prepare for their licensure examination.

3. CONCEPTUAL FRAMEWORKS

Learner assessment at the “knows” and “knows how” levels of Miller’s Pyramid can be accomplished through written tests.[1] The practice of studying and taking a test has been shown to enhance learning compared to studying alone (e.g. re-reading course material).[2] Furthermore, testing has been shown to slow the rate of forgetting.[2] A non-cued written CR test is an appropriate modality for assessing cognitive knowledge and clinical reasoning.[1] Moreover, the short-answer test format has been shown to strengthen knowledge and improve retention more so than a SR design.[2]

Messick’s unified validity framework provides a systematic approach for seeking construct validity evidence.[3-5] Messick’s framework identifies the sources of validity evidence required to support (or negate) the appropriateness in the use of its test scores. Construct validity represents a summary of the evidence for and consequences of score interpretation and application. Identifying and collecting validity evidence enables one to make more accurate inferences on the usefulness of test scores and whether the test achieves its intended aim. We hope that learners use their CRPT scores to modify their study behaviors in a more effective direction in preparation for the nephrology licensure examination.

Messick’s validity conceptual framework consists of five sources of evidence: content, response process, internal structure, relations to other variables, and consequences. A RCPSC certification examination created in 2014 for General Internal Medicine trainees demonstrated content validity evidence of the performance scores.[6] However, validity evidence for the RCPSC nephrology certification examination has not been reported publicly or studied to date.

An assessment tool without appropriate validity evidence may lead to an unawareness of construct underrepresentation (underrepresented sampling of content domain) and construct-irrelevant variance (non-random error that can bias the measurement of the construct).

This project involved the creation of a validated assessment instrument, provision of test performance feedback in advance of the licensure examination, evaluation of the assessment tool, and the identification of any curricular deficiencies in the nephrology training program at WU.

4. METHODS

4.1 **Grader Training**

The CRPT was scored by three graders who were blinded to the test takers identity/PGY level. Additionally, all graders are certified nephrologists at WU with formal training in the field of medical education. All graders were made aware of the study purpose and had an opportunity to review and provide constructive feedback on the fifty-five test items prior to the examination date. To enhance grader consensus and cross-calibration, all three graders were trained together on how to accurately and consistently apply the scoring rubric.[1]

4.2 **Content**

Content relevance, representativeness, and technical quality are important elements that contribute to content validity.[7] A test blueprint is a detailed description of an examination outline and plan.[1] The RCPSC learning objectives for nephrology training programs was used as the test blueprint of the construct domain being assessed by the CRPT. The RCPSC learning objectives encompass the following seven content domains in nephrology: kidney transplantation (KT), renal physiology (RPh), acute kidney injury (AKI), renal pathology (RPa), chronic kidney disease (CKD), peritoneal dialysis (PD), and hemodialysis (HD). A two-dimensional test blueprint was developed to categorize proportions of items by domain and Bloom's taxonomy of cognitive domains (recall, interpretation, and problem solving).[8] Because learners in Canadian nephrology training programs are at PGY4 or 5, instruments designed to test these learners should assess their interpretation and problem-solving skills (Table I).

TABLE I
TEST BLUEPRINT FOR THE CONSTRUCTED RESPONSE PREPARATORY TEST:
PERCENT WEIGHT

Content category	Recall	Interpretation	Problem-solving	Total
AKI	4	4	5	24%
CKD	3	3	3	16.5%
HD	1	1	3	9%
PD	1	1	1	5.5%
KT	3	3	3	16.5%
RPa	0	1	1	3.5%
RPh	4	5	5	25 %
Total	29%	33%	38%	100%

Test item quality is a fundamental consideration in assessment.[1] Published guidelines for writing SR items were used by CR test developers, specifically recommendations pertaining to item content, style concerns, and stem construction.[9-11] All test items were reviewed for clarity and cultural sensitivity by a medical educator.[12] Additionally, trainees evaluated each test item individually for its clarity and level of difficulty. In parallel with the RCPSC licensing examination, the CRPT consisted of 55 items.

4.3 **Response Process**

A substantive feature of construct validity entails evidence of response consistency and test data integrity. Practices that minimize rater bias include the creation of model answers, the development of a scoring rubric, and the use of well-trained independent graders. [1] Intraclass correlation (ICC) was used to assess the reliability of rater data. A grader scored the consultant CRPT prior to the trainees' tests, to gather additional information on potential answer keys. Also

two PGY6 trainees took the CRPT to pilot test the exam, to ensure the 3-hour time frame was appropriate. Any unclear items revealed prior to the trainees' CRPT were either modified or removed from the test. The response data was analyzed using the points assigned to each test item (e.g. 8/10), followed by a re-analysis where 1 point was assigned to each correctly answered item.

4.4 **Internal Structure**

Two important components of the internal structure of a test are item characteristics and reliability. In addition to descriptive statistics, item characteristics (e.g. item difficulty) were examined. A high reliability indicates consistency in individual test takers' scores.[1] Generalizability theory was used to evaluate the overall reliability of the CRPT and its variance components using a fully-crossed design, taking into account the overall effects of raters and items, candidates (p) x raters (r) x items (i). Both raters and items were assumed random facets.

4.5 **Relations to other Variables**

CRPT scores and response patterns were reviewed for different training levels of test takers, by comparing scores between PGY-4 trainees, PGY-5 trainees, and consultants, hypothesizing learners with more advanced training would have higher performance.

4.6 **Consequences**

The consequential element of construct validity involves appraising both the short- and long-term effects of a test. Scores from a valid assessment instrument have the potential to improve both learning and instruction.[3, 13] An item-based standard setting method using the Extended Angoff procedure was used to set a passing score for the CRPT.[1, 14] Formal written

feedback was provided to all test takers, which included overall score, rank, and a performance breakdown based on content domains.

The short-term impact of the CRPT was revealed by the questionnaire completed by trainees both before and following the formative examination. The questionnaire gauged whether the CRPT impacts the trainees' level of preparedness for the RCPSC licensing examination, study habits, and perception of the training program curriculum. The long-term effects of the CRPT were measured by examining the changes recommended to the nephrology training program curriculum at WU.

5. RESULTS

5.1 **Content**

The test takers evaluated the CRPT for item clarity and level of difficulty. All 10 trainees rated 46 of the 55 items (84%) to be clearly written (Table II). All of the nephrology trainees rated 16 of the 55 items (29%) as “not difficult” or “somewhat difficult”, and all of the test takers found 9 of the 55 items (16%) to be either “difficult”, “more difficult”, or “very difficult” (Table II).

5.2 **Response Process**

With a partial point marking scheme (total potential points 261), the CRPT nephrology trainee (PGY4 and PGY5) score ranges for Graders 1, 2, and 3 were 51-68%, 49-68%, and 54-73% (Table III). The inter-rater reliability of the scores indicated moderate agreement, ICC = 0.55; 95% CI = 0.09, 0.85.[15] Overall, the standard deviation (SD) of the CRPT nephrology trainee scores (all Graders) was 4.33 points.

The CRPT nephrology trainee scores were stratified into the seven content domains outlined in the test blueprint (Table IV). The overall mean scores divided by content domains ranged from 30% to 92%. The highest and lowest test scores were in the domains of peritoneal dialysis and renal pathology, respectively.

TABLE II
ITEM CLARITY AND LEVEL OF DIFFICULTY

Item	Clarity (% clear)	Level of difficulty^a (% reporting)	Item	Clarity (% clear)	Level of difficulty^a (% reporting)
1	100	1 (80); 2 (20)	29	100	1 (20); 2 (20); 3 (30); 4 (30)
2	100	1 (20); 2 (40); 3 (40)	30	100	1 (80); 2 (20)
3	100	1 (40); 2 (40); 3 (20)	31	100	1 (10); 2 (60); 3 (40)
4	100	1 (50); 2 (50)	32	100	1 (10); 2 (10); 3 (30); 4 (30); 5 (20)
5	100	1 (40); 2 (60)	33	100	1 (90); 2 (10)
6	100	2 (20); 3 (60); 4 (20)	34	100	2 (100)
7	100	1 (20); 2 (80)	35	60	2 (30); 3 (40); 4 (30)
8	90	3 (60); 4 (30); 5 (10)	36	100	1 (30); 2 (50); 3 (20)
9	80	2 (40); 3 (50); 4 (10)	37	100	1 (50); 2 (40); 4 (10)
10	100	1 (20); 2 (80)	38	100	2 (50); 4 (50)
11	100	3 (20); 4 (70); 5 (10)	39	70	1 (30); 2 (40); 3 (30)
12	80	1 (30); 2 (50); 3 (20)	40	100	1 (90); 2 (10)
13	100	1 (20); 2 (80)	41	100	2 (30); 3 (70)
14	100	1 (20); 2 (50); 3 (30)	42	100	3 (20); 4 (50); 5 (30)
15	70	2 (50); 3 (50)	43	100	1 (60); 2 (40)
16	100	1 (40); 2 (60)	44	100	1 (70); 2 (30)
17	100	1 (80); 2 (20)	45	100	3 (40); 4 (40); 5 (20)
18	100	1 (10); 2 (50); 3 (40)	46	80	1 (70); 2 (30)
19	100	2 (30); 3 (70)	47	100	1 (70); 2 (20); 3 (10)
20	100	3 (40); 4 (60)	48	60	1 (20); 2 (20); 3 (60)
21	100	1 (20); 2 (30); 3 (40); 4 (10)	49	50	2 (20); 3 (60); 4 (20)
22	100	1 (10); 3 (30); 4 (60)	50	100	1 (80); 2 (20)
23	100	2 (30); 3 (40); 4 (30)	51	100	2 (5); 4 (50)
24	100	3 (10); 4 (50); 5 (40)	52	100	2 (10); 3 (50); 4 (40)
25	100	3 (20); 4 (60); 5 (20)	53	100	2 (60); 3 (20); 4 (20)
26	100	2 (30); 3 (70)	54	100	1 (10); 2 (50); 3 (40)
27	100	2 (30); 3 (50); 4 (20)	55	100	1 (20); 2 (80)
28	100	3 (50); 4 (50)			

^a Difficulty scale: 1, not difficult; 2, somewhat difficult; 3, difficult; 4, more difficult; 5, very difficult.

TABLE III

CONSTRUCTED RESPONSE PREPARATORY TEST NEPHROLOGY TRAINEES SCORES
(TOTAL POTENTIAL POINTS 261)

Test taker	Grader 1	Grader 2	Grader 3
<i>PGY4</i>			
1	51%	49%	54%
2	62%	57%	67%
3	65%	59%	69%
4	66%	62%	70%
5	68%	63%	73%
<i>PGY5</i>			
6	58%	63%	67%
7	62%	68%	66%
8	63%	67%	72%
9	63%	62%	70%
10	64%	68%	70%

TABLE IV

CONSTRUCTED RESPONSE PREPARATORY TEST NEPHROLOGY TRAINEES SCORES
BREAKDOWN BY CONTENT DOMAIN

Test taker	Kidney transplant	Renal physiology	Acute kidney injury	Renal pathology	Chronic kidney disease	Peritoneal dialysis	Hemo-dialysis
<i>PGY4</i>							
1	50%	32%	60%	36%	46%	85%	71%
2	58%	59%	59%	18%	69%	100%	71%
3	75%	63%	57%	18%	63%	100%	75%
4	67%	52%	63%	18%	83%	100%	71%
5	77%	62%	63%	18%	71%	100%	79%
<i>PGY5</i>							
6	73%	46%	54%	36%	54%	77%	71%
7	58%	65%	63%	45%	77%	77%	38%
8	75%	54%	57%	27%	60%	92%	83%
9	67%	56%	62%	36%	63%	100%	67%
10	65%	49%	63%	45%	63%	85%	75%
Mean	67%	54%	60%	30%	65%	92%	70%

With a 1 point per correct item marking scheme (total potential points 55), the CRPT nephrology trainee (PGY4 and PGY5) score ranges for Graders 1, 2, and 3 were 51-78%, 45-76%, and 55-79% (Table V). The inter-rater reliability of the scores indicated moderate agreement, ICC = 0.58; 95% CI = 0.16, 0.86.[15]

TABLE V

CONSTRUCTED RESPONSE PREPARATORY TEST NEPHROLOGY TRAINEES SCORES
(TOTAL POTENTIAL POINTS 55)

Test taker	Grader 1	Grader 2	Grader 3
<i>PGY4</i>			
1	51%	45%	55%
2	75%	62%	75%
3	73%	58%	73%
4	76%	62%	69%
5	78%	69%	76%
<i>PGY5</i>			
6	69%	69%	79%
7	71%	76%	75%
8	67%	67%	73%
9	67%	62%	71%
10	69%	67%	71%

5.3 **Internal Structure**

Item difficulty and item-total correlation was calculated for each of the 55 items on the CRPT (Tables VI). Twenty-nine of the test items were found to have an item-total correlation exceeding 0.20.

Using a partial point marking scheme (total potential points 261), the G-coefficient = 0.76 (normative uses of test scores; e.g., ranking learners), and the ϕ -coefficient = 0.53 (criterion-based uses of test scores; e.g. pass-fail decisions). The largest variance was found in items (29.6%), indicating variability in item difficulty (Table VII). Learner performance varied by item (person by item interaction = 25.5%), which means some learners do well in some items while performing poorly on other items.

Using a 1 point per correct item marking scheme (total potential points 55), the G-coefficient = 0.60, and the ϕ -coefficient = 0.48. The largest variance was again found in items (17%), and learner performance continued to vary by item (33.6%) (Table VIII).

TABLE VI**ITEM DIFFICULTY AND ITEM-TOTAL CORRELATION**

Item	Item Difficulty (0–1.00)	Item-Total Correlation	Item	Item Difficulty (0–1.00)	Item-Total Correlation
1	1.00	0.57	29	0.90	-0.02
2	0.80	0.43	30	0.70	0.20
3	0.30	0.37	31	1.00	0.20
4	0.50	0.58	32	0.50	-0.12
5	0.30	-0.07	33	0.70	0.00
6	0.70	0.46	34	0.90	0.14
7	0.60	0.18	35	0.30	0.18
8	0.20	-0.18	36	0.60	0.59
9	0.70	-0.04	37	0.40	0.42
10	0.90	0.05	38	0.60	0.44
11	0.90	-0.29	39	0.30	0.05
12	1.00	-0.08	40	1.00	0.40
13	0.90	0.31	41	0.90	0.81
14	0.70	0.41	42	0.10	0.61
15	0.10	0.26	43	1.00	-0.34
16	0.30	0.31	44	0.90	0.32
17	0.90	0.30	45	0.70	0.44
18	1.00	0.50	46	1.00	0.19
19	0.90	-0.05	47	1.00	0.29
20	0.40	0.26	48	0.80	0.11
21	1.00	0.30	49	0.50	0.06
22	0.80	0.25	50	0.80	0.61
23	0.40	-0.16	51	1.00	0.67
24	0.60	-0.39	52	0.90	0.07
25	0.80	0.49	53	0.90	0.07
26	0.20	0.37	54	0.70	0.23
27	0.80	0.15	55	0.60	0.60
28	0.90	-0.24			

TABLE VII

VARIANCE COMPONENT ANALYSIS: G-STUDY
(TOTAL POTENTIAL POINTS 261)^a

Effect	df	Variance Component (VC)	% VC
person (p)	9	0.002 (0.001)	2.2%
rater (r)	2	0.001 (0.001)	1.4%
item (i)	54	0.028 (0.006)	29.6%
p x r	18	0.000 (0.000)	0.0%
p x i	486	0.024 (0.002)	25.5%
r x i	108	0.005 (0.001)	5.5%
p x r x i, error	972	0.034 (0.002)	35.8%

^a G-study design: person (p) x rater (r) x item (i) design; values in parentheses are standard errors.

TABLE VIII

VARIANCE COMPONENT ANALYSIS: G-STUDY
(TOTAL POTENTIAL POINTS 55)^a

Effect	df	Variance Component (VC)	% VC
person (p)	9	0.003 (0.002)	1.3%
rater (r)	2	0.001 (0.001)	0.5%
item (i)	54	0.037 (0.010)	17%
p x r	18	0.000 (0.001)	0.0%
p x i	486	0.073 (0.007)	33.6%
r x i	108	0.015 (0.003)	6.9%
p x r x i, error	972	0.089 (0.004)	40.7%

^a G-study design: person (p) x rater (r) x item (i) design; values in parentheses are standard errors.

5.4 Relations to other Variables

By PGY level, the CRPT scores were as follows: Grader 1, 51-68% for PGY4, and 58-64% for PGY5; Grader 2, 49-63% for PGY4, and 62-68% for PGY5; and Grader 3 54-73% for PGY4 and 66-72% for PGY5. By PGY level, the SDs of the CRPT score means (all Graders) were 5.90 for PGY4 and 5.33 for PGY5.

The CRPT nephrology trainee mean scores by content domains and PGY level were very similar to overall test scores (PGY4 and PGY5) divided by content domains alone. The range of mean CRPT content domains scores was 22-97% for PGY4 trainees and 38-86% for PGY5 trainees (Tables IX, X). Similar to findings for the whole cohort, the highest and lowest test scores were in the domains of peritoneal dialysis and renal pathology, respectively, for both PGY level groups.

TABLE IX

CONSTRUCTED RESPONSE PREPARATORY TEST POSTGRADUATE YEAR 4
NEPHROLOGY TRAINEE SCORE BREAKDOWN BY CONTENT DOMAIN

Test taker	Kidney transplant	Renal physiology	Acute kidney injury	Renal pathology	Chronic kidney disease	Peritoneal dialysis	Hemo-dialysis
1	50%	32%	60%	36%	46%	85%	71%
2	58%	59%	59%	18%	69%	100%	71%
3	75%	63%	57%	18%	63%	100%	75%
4	67%	52%	63%	18%	83%	100%	71%
5	77%	62%	63%	18%	71%	100%	79%
Mean	65%	54%	60%	22%	66%	97%	73%

TABLE X

**CONSTRUCTED RESPONSE PREPARATORY TEST POSTGRADUATE YEAR 5
NEPHROLOGY TRAINEE SCORE BREAKDOWN BY CONTENT DOMAIN**

Test taker	Kidney transplant	Renal physiology	Acute kidney injury	Renal pathology	Chronic kidney disease	Peritoneal dialysis	Hemo-dialysis
1	73%	46%	54%	36%	54%	77%	71%
2	58%	65%	63%	45%	77%	77%	38%
3	75%	54%	57%	27%	60%	92%	83%
4	67%	56%	62%	36%	63%	100%	67%
5	65%	49%	63%	45%	63%	85%	75%
Mean	68%	54%	60%	38%	63%	86%	67%

The CRPT scores for nephrology consultants (N = 5) ranged from 76% to 83% (Table XI). The SD of the nephrology consultants' CRPT mean scores was 2.42.

TABLE XI

**CONSTRUCTED RESPONSE PREPARATORY TEST NEPHROLOGY CONSULTANT
SCORES (TOTAL POTENTIAL POINTS 261)**

Test taker	Grader 1
Consultant A	83%
Consultant B	76%
Consultant C	78%
Consultant D	81%
Consultant E	80%

When the CRPT nephrology consultants' scores were stratified by content domains, their overall mean scores ranged from 69% to 84% (Table XII). Their highest and lowest mean content domain scores were in hemodialysis and renal pathology, respectively.

TABLE XII

**CONSTRUCTED RESPONSE PREPARATORY TEST NEPHROLOGY CONSULTANT
SCORES BREAKDOWN BY CONTENT DOMAIN**

Test taker	Kidney trans-plantation	Renal physiology	Acute kidney injury	Renal pathology	Chronic kidney disease	Peritoneal dialysis	Hemo-dialysis
A	88%	75%	83%	64%	91%	77%	92%
B	73%	61%	85%	55%	74%	92%	83%
C	85%	71%	81%	82%	74%	62%	83%
D	80%	73%	83%	73%	82%	84%	82%
E	79%	74%	80%	70%	85%	82%	82%
Mean	81%	71%	82%	69%	81%	80%	84%

5.5 Consequences

The CRPT raw passing score threshold was determined to be 57% based on application of the Extended Angoff procedure to all three test graders. With that threshold, one nephrology trainee (PGY4) did not pass the CRPT.

Analysis of the results of the questionnaire distributed to learn about self-perceived preparedness for the RCPSC licensing examination prior to taking the CRPT revealed that 8 of the 10 nephrology trainees felt “poorly prepared” or “not well prepared.” This ratio was consistent across PGY levels.

The questionnaire that was distributed after the CRPT revealed that none of the test takers felt “poorly prepared” for the RCPSC examination. All 10 test takers felt that having a preparatory test written in the same format (CR) as the licensing exam was a good idea. For study habits post-CRPT, all 10 test takers felt their studying would increase in duration and be adapted in terms of content focus. When asked if any changes were recommended to the CRPT, seven of the examinees replied “none”, two responded with “more questions”, and one stated “multiple choice questions.”

Changes recommended to the nephrology training program curriculum included; “none” (2 PGY4s), “more formative preparatory tests” (2 PGY4s and 2 PGY5s), “more consultant teaching” (1 PGY4), “have a renal physiology course” (1 PGY5), “have a lecture geared towards preparing for the CRPT” (1 PGY5), and “more exposure to glomerulonephritis” (1 PGY5). All nephrology trainees indicated that a written format was the best modality in which to receive CRPT performance feedback.

6. DISCUSSION

The development, administration, and collection of validity evidence for the CRPT provide meaningful information that can provide feedback to learners and the program to help prepare nephrology trainees for the licensure examination. Overall, CRPT scores assigned by the three independent graders showed moderate agreement. Consequently, there does not appear to have been a “dove” or “hawk” effect among the three graders. Acceptable inter-rater reliability may be the result of thorough rater training prior to grading the test, using graders with a similar clinical background, and having graders with formal training in medical education.

The variance between the highest and lowest overall test scores awarded by Grader 1 was 17%. The variances in overall test scores tallied by the other two graders were similar (both 19%). However, when the CRPTs results are grouped by PGY level, the test score ranges are quite different. For example, in the case of Grader 1, the variance between the highest and lowest test score awarded was 17% for PGY4, but is only 6% for PGY5. The smaller test score variance for PGY5s demonstrates better performance consistency within the cohort. Likely explanations for the superior consistency among PGY5 trainees, relative to PGY4 trainees, include their longer duration of formal nephrology training (24 months vs. 12 months), their greater clinical exposure, and more CRPT preparation time due to a less demanding rotation schedule (elective time only available to PGY5s). Predictably, a PGY4 trainee received the lowest overall CRPT score by all graders. Somewhat surprisingly however, a PGY4 trainee received the highest overall test score by two of the three graders.

Creating an assessment tool for learners at two different and advanced levels of medical training is unique and challenging. That being said, the majority of questions on the CRPT have

an appropriate level of difficulty based on item difficulty values. Fifty-three percent of the items on the CRPT are able to discriminate between trainees with high and low test scores. Nine out of the ten test takers passed the CRPT.

Analysis of performance by content domain revealed strengths and weaknesses of both the trainees and training program. Test takers' knowledge base on peritoneal dialysis and hemodialysis appears to be excellent. On the other hand, there appears to be a knowledge deficit in renal pathology and renal physiology. This was partly evident in the variance components analyses using both marking schemes, where person-by-item interaction (context specificity) accounted for over 25% of total variance. Possible reasons to explain the discrepancy in test scores across content domains include trainee rotations (and thus content exposure), varying degrees of item difficulty, and consultant knowledge base and expertise. Unfortunately, the nephrology training program curriculum at WU does not have any dedicated renal pathology or renal physiology rotations. The average degrees of item difficulty for questions pertaining to renal physiology and hemodialysis were 0.62 and 0.80, respectively. Test performance based on content domain may also be explained by the strengths (and weaknesses) of the teaching faculty in the nephrology training program at WU. Similar to the trainees, the nephrology consultants scored very well on items testing peritoneal and hemodialysis knowledge. Like the trainees, the consultants scored lowest on items testing their knowledge of renal pathology and renal physiology.

All test takers rated the vast majority of items as being "clearly written." It appears the thorough test preparation process, which involved creating, reviewing and piloting the CRPT was well worth the effort. The level of difficulty ratings for the majority of test items was highly

variable. The overall trainee perception of item difficulty was found to be consistent with the wide-ranging test performances.

The questionnaire distributed prior to taking the CRPT revealed the vast majority of trainees did not feel well prepared for the upcoming licensing examination. Based on this sentiment alone, the nephrology program curriculum needs to be reviewed and likely realigned to improve program participants' licensure exam preparedness. The questionnaire administered following the CRPT highlights a number of interesting trainee attitudes. Notably, 5 of the 10 test takers no longer felt "poorly prepared" for the RCPSC exam following the CRPT. The opportunity to apply one's knowledge base during a three-hour test seems responsible for instigating the change in attitude towards exam readiness. Not surprisingly, all trainees reported believing that a formative test written in the same format as the upcoming summative exam is a good idea. After completing the CRPT, all of the trainees indicated that their study habits would be modified, specifically in terms of spending more time and altering their content focus. The vast majority of test takers did not recommend that any changes should be made to the CRPT. One trainee had the foresight to request a renal physiology course be implemented into the curriculum. The other curriculum suggestions were quite specific and geared towards individual needs. All trainees preferred a written format for their test performance feedback, which facilitates detailed responses while being less intimidating than a one-on-one oral recap.

7. LIMITATIONS

The limitations of the project include a small sample size, limited number of test items, test takers at two levels of training, unavailable RCPSC nephrology licensure test scores, and low consultant participation. Our sample size was limited by the fact that our WU nephrology program accepts only four to six residents per year. A sample size of 10 limits the generalizability of our findings and increases the variability of the summary statistics. The focus of this work was the development of, administration of, and collection of validity evidence for a novel assessment instrument for nephrology trainees at WU. Thus, the broad generalizability of the findings is not an exigent issue. Without RCPSC summative test scores available, it's impossible to demonstrate whether the CRPT actually predicts nephrology licensure performance. Moreover, the CRPT was modelled after the RCPSC licensing examination with 55 items, a 3-hour time allowance, and assesses a limited aspect of the medical knowledge base required for a successful practice of nephrology.

8. CONCLUSION

Validity data enables inferences to be made regarding the usefulness of test scores and whether a test achieves its intended aim. For our evaluation of the CRPT, we employed Messick's unified validity framework, which provides a systematic approach to construct validity assessment based on five sources of evidence. Validity evidence collected for the CRPT provides useful information regarding the utility of the CRPT as a training tool. Furthermore, we confirmed that we can use CRPT data to collect feedback about the adequacy of our training program curriculum.

Prior to taking the CRPT, trainees reported feeling poorly prepared for the RCPSC licensing examination. Following the CRPT, the same trainees expressed feeling more prepared for the RCPSC and reported that they were adjusting their study habits in response to taking the CRPT. With respect to providing information about our training program, the CRPT results indicated program weaknesses in the instructional areas of renal physiology and renal pathology, which can serve to guide curricular revisions.

CITED LITERATURE

1. Downing, S.M., and Yudkowsky, R.: Written tests: constructed-response and selected-response formats. In: *Assessment in Health Professions Education*. Downing, S.M. (ed.), pp.149-184, New York, Routledge, 2009.
2. Rohrer, D., and Pashler, H.: Recent research on human learning challenges conventional instructional strategies. *Educational Researcher*, 39:406-412, 2010.
3. Messick, S.: Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50:741-749, 1995.
4. Downing, S.M.: Validity: on the meaningful interpretation of assessment data. *Medical Education*, 37:830-837, 2003
5. American Educational Research Association, American Psychological Association, National Council on Measurement and Education.: In: *Standards for Educational and Psychological Testing*. Washington, American Educational Research Association, 2014.
6. Pugh, D., Govig, B., Gibson, N., and Cole, G.: The general internal medicine certification examination: does it measure what it should? *Canadian Journal of General Internal Medicine*, 11:30-33, 2016.
7. Cronbach, L.J.: Test validation. In: *Educational Measurement*. Thorndike, R. (ed.), pp. 443-507, Washington, American Council on Education, 1971.
8. Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., and Krathwohl, D.R.: Taxonomy of educational objectives: the classification of educational goals. In: *Handbook I: cognitive domain*. New York, Longmans Green, 1956.
9. Haladyna, T.M., Downing, S.M., and Rodriguez, M.C.: A review of multiple-choice item writing guidelines for classroom assessment. *Applied Measurement in Assessment*, 15:309-334, 2002.
10. Case, S.M., and Swanson, D.B.: Issues regarding format and structure of test questions. In: *Constructing written test questions for the basic and clinical sciences*. Case, S.M. (ed.), pp. 7-27, Philadelphia, National Board of Medical Examiners, 2002.
11. Downing, S.M., and Haladyna, T.M.: Item and prompt development in performance testing. In: *Handbook of test development*. Welch, C. (ed.), pp. 354-373, Mahwah, Lawrence Erlbaum Associates, 2006.
12. Downing, S.M., and Haladyna, T.M.: Fairness review in assessment. In: *Handbook of test development*. Zieky, M. (ed.), pp. 419-435, Mahwah, Lawrence Erlbaum Associates, 2006.

13. Cook, D.A., Lineberry, M.: Consequences validity evidence: evaluating the impact of educational assessments. *Academic Medicine*, 91:785-795, 2016.
14. Angoff, W.H.: Scales, norms, and equivalent scores. In: *Educational Measurement*. Thorndike, R. (ed.), pp. 508-600, Washington, American Council on Education, 1971.
15. Shrout, P.E., and Fleiss, J.L.: Intraclass correlation: uses in assessing rater reliability. *Psychol Bull.*, 86:420-428, 1979.

APPENDIX

UNIVERSITY OF ILLINOIS AT CHICAGO

Office for the Protection of Research Subjects (OPRS)
Office of the Vice Chancellor for Research (MC 672)
203 Administrative Office Building
1737 West Polk Street
Chicago, Illinois 60612-7227

Exemption Granted

May 31, 2016

John Johnson, MD, MSc
UIC Medical Education Student
Western University
927 Wellington Street
London, Ontario, Canada N6A-3T1
Phone: (519) 685-8500 / Fax: (915) 663-3449

RE: Research Protocol # 2016-0542
“Lending a formative hand: Creation and implementation of a novel test for nephrology trainees.”

Sponsors: None

Dear Dr. Johnson:

Your Claim of Exemption was reviewed on May 30, 2016 and it was determined that your research protocol meets the criteria for exemption as defined in the U. S. Department of Health and Human Services Regulations for the Protection of Human Subjects [(45 CFR 46.101(b))]. You may now begin your research.

UIC Exemption Period: May 30, 2016 – May 30, 2019
Lead Performance Site: Western University (London, Ontario, Canada)
Other Performance Site(s): UIC
Subject Population: Adult subjects only
Number of Subjects: UIC: 0; Western University: 10; Total: 10

The specific exemption category under 45 CFR 46.101(b) is:

(1) Research conducted in established or commonly accepted educational settings, involving normal educational practices such as (i) research on regular and special education instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.

You are reminded that investigators whose research involving human subjects is determined to be exempt from the federal regulations for the protection of human subjects still have responsibilities for the ethical conduct of the research under state law and UIC policy. Please be aware of the following UIC policies and responsibilities for investigators:

Phone: 312-996-1711

<http://www.uic.edu/depts/ovcr/oprs/>

Fax: 312-413-2929

2016-0542

Page 2 of 2

May 31, 2016

1. Amendments You are responsible for reporting any amendments to your research protocol that may affect the determination of the exemption and may result in your research no longer being eligible for the exemption that has been granted.
2. Record Keeping You are responsible for maintaining a copy all research related records in a secure location in the event future verification is necessary, at a minimum these documents include: the research protocol, the claim of exemption application, all questionnaires, survey instruments, interview questions and/or data collection instruments associated with this research protocol, recruiting or advertising materials, any consent forms or information sheets given to subjects, or any other pertinent documents.
3. Final Report When you have completed work on your research protocol, you should submit a final report to the Office for Protection of Research Subjects (OPRS).
4. Information for Human Subjects UIC Policy requires investigators to provide information about the research protocol to subjects and to obtain their permission prior to their participating in the research. The information about the research should be presented to subjects as detailed in the Claim of Exemption application and research protocol utilizing the approved recruitment and consent process and documents only.

Please be sure to use your research protocol number (listed above) on any documents or correspondence with the IRB concerning your research protocol.

We wish you the best as you conduct your research. If you have any questions or need further help, please contact me at (312) 355-2908 or the OPRS office at (312) 996-1711. Please send any correspondence about this protocol to OPRS at 203 AOB, M/C 672.

Sincerely,

Charles W. Hoehne, B.S., C.I.P.
Assistant Director, IRB #7
Office for the Protection of Research Subjects

cc: Ilene B. Harris, Medical Education, M/C 591
Yoon Soo Park (faculty sponsor), Medical Education, M/C 591



Research Ethics

**Western University Non-Medical Research Ethics Board
NMREB Delegated Initial Approval Notice**

Principal Investigator: Dr. John Johnson
Department & Institution: Schulich School of Medicine and Dentistry\Medicine-Dept of, Western University

NMREB File Number: 107819

Study Title: Lending a formative hand: Creation and implementation of a novel test for nephrology trainees.

NMREB Initial Approval Date: March 04, 2016

NMREB Expiry Date: March 04, 2017

Documents Approved and/or Received for Information:

Document Name	Comments	Version Date
Letter of Information & Consent		2016/02/17
Recruitment Items	Email script	2016/02/17
Western University Protocol	Received February 26, 2016	
Instruments	Assessment Tool	2016/02/17

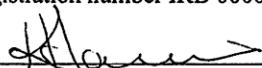
The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the above named study, as of the NMREB Initial Approval Date noted above.

NMREB approval for this study remains valid until the NMREB Expiry Date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario.

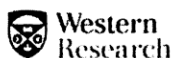
Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB.

The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.


 Ethics Officer, on behalf of Riley Hinson, NMREB Chair or delegated board member

Ethics Officer to Contact for Further Information: Erika Basile ☐ Nicole Kaniki ☐ Grace Kelly ☐ Katelyn Harris ☐ Vikki Tran ☒

This is an official document. Please retain the original in your files



Office of Research Ethics
Support Services Building, Western University
London, Ontario, Canada, N6A 3K7

RECOMMENDATIONS FORM
NON-MEDICAL RESEARCH ETHICS BOARD
Delegated Review | Prof. Riley Hinson, NMREB Chair

Ethics Officer: Grace Kelly

***DO NOT** begin any study related activities until you receive final notification of approval from the Office of Research Ethics (ORE). If this study involves Lawson, you must also ensure you have received Lawson's Institutional Approval (IA).*

SUBMISSION DETAILS

Review Date	March 3, 2016
REB ID*	107819
Study Title	Lending a formative hand: Creation and implementation of a novel test for nephrology trainees.
Principal Investigator	J. Johnson

*This number must be quoted on all modifications, revisions and correspondence.

SUBMISSION STATUS

The proposed study submission has been reviewed and the following decision has been reached:

APPROVED - NO CHANGES REQUIRED

This submission has been reviewed by the REB and no recommendations have been identified at this time. This letter is being sent as acknowledgment that no changes are required. You **MUST NOT** begin any study related activities until you have received final notification of approval from our office.

VITA

NAME: John Francis Johnson

EDUCATION: B.S., Biology, York University, Toronto, Ontario, Canada, 1992

M.D., University of Ottawa, Ottawa, Ontario, Canada, 1996

M.S., Epidemiology and Biostatistics, Western University, London, Ontario, Canada, 2007

M.H.P.E., University of Illinois at Chicago, Chicago, Illinois, 2017

TEACHING: Assistant Professor of Medicine, Department of Medicine, Division of Nephrology, London Health Sciences Centre, London, Ontario, Canada, 2007-2014

Associate Professor of Medicine, Department of Medicine, Division of Nephrology, London Health Sciences Centre, London, Ontario, Canada, 2014-2017

Co-Chair Epidemiology course, Undergraduate Medical Education, Western University, London, Ontario, Canada, 2015-2017

HONORS: University Students' Council Honor Roll Teaching Certificate Undergraduate Medical Education, Western University, London, Ontario, Canada, 2007-2009, 2014-2015.

Department of Medicine Junior Teaching Award of Excellence, Western University, London, Ontario, Canada, 2011.

PROFESSIONAL MEMBERSHIP: American Society of Nephrology
American Society of Transplantation
Canadian Society of Nephrology

PUBLICATIONS: Johnson, J., Jevnikar, A., Mahon, J., Muirhead, N., and House, A.A.: Fate of the mate: the influence of delayed graft function in renal transplantation on the mate recipient. *American Journal of Transplantation*, 9:1796-1801, 2009.

Iordanous, Y., Seymour, N., Young, A., Johnson, J., Iansavichus, A.V., Cuerden, M., Gill, J.S., Poggio, E., and Garg, A.X.: Transplant outcomes for recipients of kidneys from medically complex living donors: a systematic review. *American Journal of Transplantation*, 9:1558-1573, 2009.

Oliver, M.J., Garg, A.X., Blake, P.G., Johnson, J., Verrelli, M., Zacharias, J.M., Pandeya, S., and Quinn, R.R.: Impact of contra-indications, barriers to self-care and support on incident peritoneal dialysis utilization. *Nephrology Dialysis Transplant*, 8:2737-2744, 2010.

Oliver, M.J., Verrelli, M., Zacharias, J.M., Blake PG, Garg A.X., Johnson, J., Pandeya, S., Perl, J., Kiss, A.J., and Quinn, R.R.: Choosing peritoneal dialysis reduces the risk of invasive access interventions. *Nephrology Dialysis Transplant*, 2:810-816, 2012.

Quinn, R.R., Ravani, P., Zhang, X., Garg, A.X., Blake, P.G., Austin, P.C., Zacharias, J.M., Johnson, J., Pandeya, S., Verrelli, M., and Oliver, M.J.: Impact of modality choice on rates of hospitalization in patients eligible for both peritoneal dialysis and hemodialysis. *Peritoneal Dialysis International*, 1:41-48, 2014.