Effect of Checklist Item on the Reliability of Internal Medicine Residency OSCE Scores

BY

VIJAY DANIELS
B.Sc., University of British Columbia, 2001
B.Sc. (Med), M.D., University of Manitoba, 2005

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, 2013

Chicago, Illinois

Defense Committee:

Rachel Yudkowsky, Chair and Advisor
Georges Bordage
Mark Gierl, Department of Educational Psychology, University of Alberta
DEDICATION

This thesis is dedicated to my wife, Lia. Thank you for your immense support, without which, this work could not have happened. To my two beautiful daughters, Rebecca and Mikayla, I love you to pieces, and should you ever choose to undertake such a huge task when you grow up, I hope you are blessed with the same amount of support that I have been.
ACKNOWLEDGEMENTS

I would like to start by thanking Dr. Ilene Harris, Director of Graduate Studies in the Department of Medical Education. I vividly remember the phone call in April of 2011 in which I presented you three ideas for my thesis project and you pragmatically responded, “Which of these three will you have the most control over?” Thank you for your advice, for now looking at the other two projects, I do not think I would have ever completed them!

I would next like to thank Dr. Rachel Yudkowsky, my thesis supervisor, who has been my guiding light through this project. I am most appreciative of your gentle, kind nature and words of encouragement throughout this endeavor. I am astounded by how quickly you turned around various drafts with such detailed feedback. Thank you again so much for your support.

Thank you to Dr. Georges Bordage, whose visit to Edmonton in 2005 sparked my interest in this program and connected me with a fellow MHPE graduate, Narmin Kassam who has become a mentor to me. I feel very privileged to have had the opportunity to work with you, Georges. Your feedback on my drafts have definitely made me a stronger writer and I also greatly appreciate all your teaching on conceptual frameworks as they now guide everything I do in medical education.

Thank you to Dr. Mark Gierl, my University of Alberta colleague in Educational Psychology. As you said, this is a great time for us as our worlds are coming together, and I feel so lucky to have had the opportunity to work with you on this project. Thank you for your input on this project and your psychometric guidance. It has been invaluable.

I would also like to thank Amin Mousavi and Dr. Todd Rogers from the Centre for Research in Applied Measurement and Evaluation at the University of Alberta for their help with
generalizability analyses. I am also grateful to my colleagues in the Department of Medicine at the University of Alberta: Dr. Cheryl Goldstein for her assistance in the literature review for evidence-based items and to Drs. Bibiana Cujec, Adriana Lazarescu, Fiona Lawson, Fraulein Morales, Uwais Qarni, Irwin Sandu, and Stephanie Smith for their assistance in classifying checklist items. This work was supported in part by the University of Alberta, Department of Medicine’s Medical Education Research Grant.

Finally I would like to thank Janet Settle from the Department of Medical Education at the University of Illinois in Chicago for all of her administrative support throughout the MHPE program. I would have been lost without you!
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>1</td>
</tr>
<tr>
<td>II.</td>
<td>4</td>
</tr>
<tr>
<td>III.</td>
<td>7</td>
</tr>
<tr>
<td>IV.</td>
<td>12</td>
</tr>
<tr>
<td>V.</td>
<td>15</td>
</tr>
<tr>
<td>VI.</td>
<td>16</td>
</tr>
<tr>
<td>VII.</td>
<td>17</td>
</tr>
</tbody>
</table>

**I. INTRODUCTION**

**II. METHODS**

A. Study Setting

B. Design

C. Instruments

D. Data Collection Procedures

E. Analyses

**III. RESULTS**

A. Item Classification

B. Participants

C. Generalizability and Decision Studies

D. Correlation of Checklist Total with Global Rating Scale

**IV. DISCUSSION**

A. Limitations

**V. CONCLUSIONS**

**VI. CITED LITERATURE**

**VII. VITA**
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. ITEM CLASSIFICATION AND MISSING EVIDENCE-BASED ITEMS BY STATION</td>
<td>7</td>
</tr>
<tr>
<td>II. GENERALIZABILITY STUDY RESULTS FOR INTERNAL CONSISTENCY RELIABILITY WITHIN STATIONS</td>
<td>8</td>
</tr>
<tr>
<td>III. GENERALIZABILITY STUDIES – VARIANCE COMPONENTS ACROSS FOUR STATIONS</td>
<td>10</td>
</tr>
<tr>
<td>IV. GENERALIZABILITY AND DECISION STUDY RESULTS FOR RELIABILITY ACROSS FOUR STATIONS</td>
<td>11</td>
</tr>
</tbody>
</table>
SUMMARY

Objective Structured Clinical Examinations (OSCEs) are used worldwide for summative examinations but often lack acceptable reliability. Research has shown that reliability increases if OSCE checklists for medical students include only clinically relevant items. Also, checklists are often missing evidence-based items that high-achieving learners are more likely to use. The purpose of this study was to determine if limiting checklist items to clinically discriminating items and/or adding missing evidence-based items improved score reliability in an Internal Medicine residency OSCE.

Six internists reviewed the traditional checklists of four OSCE stations classifying items as clinically discriminating or non-discriminating. Two independent reviewers augmented checklists with missing evidence-based items. We used generalizability theory to calculate overall reliability of faculty observer checklist scores from 45 first and second-year residents and predict how many 20-item stations would be required to reach a Phi coefficient of 0.8.

Removing clinically non-discriminating items from the traditional checklist did not affect the number of stations (11) required to reach a Phi of 0.8 with 20 items. Focusing the checklist on only evidence-based, clinically discriminating items increased test score reliability, needing eight stations instead of 11 to reach 0.8; adding missing evidence-based clinically discriminating items to the traditional checklist modestly improved reliability (needing 10 instead of 11 stations).

Checklists composed of evidence-based, clinically discriminating items improved the reliability of checklist scores and reduced the number of stations needed for acceptable reliability. Educators should give preference to evidence-based items over non-evidence-based items when developing OSCE checklists.
I. INTRODUCTION

First introduced in the mid 1970s, the Objective Structured Clinical Examination (OSCE) is used worldwide and is now the format of choice for assessing clinical skills, including the Medical Council of Canada’s Qualifying Examination and the United States Medical Licensing Examination. Most OSCEs are scored using dichotomous checklist items (typically 15 to 20 items per station) that are created by content experts. The checklist items depend on the purpose of the station, such as making a diagnosis based on a history or physical exam or managing a condition. For diagnosis stations, the checklist items can be grouped into two categories: history and physical exam items that clinically discriminate between the diagnoses under consideration and those that are routinely collected for reasons of “thoroughness” but do not clinically discriminate between competing diagnoses, hereon referred to as clinically non-discriminating items. Clinically discriminating items can be further subdivided into those with scientific evidence supporting their use to discriminate between conditions (evidence-based, clinically discriminating items) and others that lack supporting evidence (non-evidence-based, clinically discriminating items) but are still used clinically mostly for historical reasons.

Previous studies have shown that expert clinicians tend to perform fewer checklist items than novices (Hodges et al., 1999) due to their ability to recognize patterns of diseases (illness scripts) and quickly compare the case in front of them with these illness scripts (Eva, 2004; Schmidt and Rikers, 2007). In the context of an OSCE we would expect the more competent candidates to take a thoughtful approach to data gathering and perform relatively more clinically discriminating items and evidence-based items than the less competent candidates, who might take a rote approach to data gathering and include many routine, non-
discriminating items. Thus a checklist composed of clinically discriminating items should better
differentiate between more competent and less competent candidates than a checklist that
includes clinically non-discriminating items that may simply add noise or even disadvantage the
more expert clinician. Indeed, Yudkowsky et al. (2009) found that the generalizability of a three-

case medical student OSCE increased from 0.35 to 0.50 if only the clinically discriminating items
were used to generate case scores. Their study preferred evidence-based clinically
discriminating items but did not specify what proportion of their clinically discriminating items
was indeed evidence-based. Other research has found that expert-created checklists included
only 44% of available evidence-based, clinically discriminating items (Hettinga et al., 2010) so
assessing the impact of adding missing evidence-based items would be important.

In this study, our first objective was to replicate the findings from Yudkowsky et al. (2009) in
a sample of Internal Medicine residents to assess if limiting the OSCE checklist to clinically
discriminating items improved the reliability of test scores. Our second objective was to
evaluate the impact of adding missing evidence-based items both to the traditional checklist
and to a focused checklist of clinically discriminating items. Our third objective was to explore
how limiting a clinically discriminating checklist to only evidence-based items (i.e. eliminating
non-evidence-based, clinically discriminating items) impacted test score reliability. If these
changes improve test score reliability this would guide future checklist development and could
result in OSCEs that require fewer stations, and thus fewer resources, to achieve acceptable
reliability.
Our final objective was to explore whether our examiners valued a focused approach or a thoroughness approach to data gathering, that is, whether their global ratings would correlate more strongly with a checklist composed exclusively of clinically discriminating items than with the traditional checklist.
II. METHODS

Study setting

This study was conducted at the University of Alberta as part of the annual Internal Medicine residency OSCE for first and second-year residents. This ten-station OSCE (10 minutes per station) is a moderate-stakes examination of history-taking and physical examination (PE) with a formative focus, scored with both a dichotomous checklist and a six-point global rating scale (1=Needs a lot of improvement, 2=Needs improvement, 3=Borderline Fail, 4=Borderline Pass, 5=Good, 6=Excellent).

Design

The study focused on test score reliabilities as a function of the type of checklist items used to calculate the test scores. We used generalizability theory (Brennan, 2001) as the data analysis framework. Our study design was \((r \times i) : s \times p\) where \(r=\text{rater}, i=\text{item types}, s=\text{station}, p=\text{person}\). For each of the five stations (out of 10) that were part of the study, two raters rated each candidate on all the checklist items to allow us to determine variance due to raters and avoid a confounded design (Shavelson and Webb, 1991). Checklist items were station-specific. The ethics internal review boards from the University of Alberta and the University of Illinois at Chicago approved this study.

Instruments

Five cases were selected from our preexisting database of 30 physical examination cases. Three criteria were used to select the cases: each prompted the candidate to use the physical examination to diagnose a suspected condition (as opposed to a simple demonstration of technique), each was missing evidence-based items, and as a group the cases tested different
content areas. The content areas were: congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), cirrhosis, acute painful knee, and Parkinsonism.

Six internists from the Department of Medicine (from the specialties of cardiology, gastroenterology, geriatrics, hematology, infectious diseases and nephrology) reviewed the original station checklists for any missing items based on their content expertise, resulting in one item being added to two of the stations (both evidence-based). These reviewers then independently classified the items as clinically discriminating (i.e. the presence or absence of that finding would impact the probability of the condition) or non-discriminating. If the group was split on classifying an item, a seventh internist acted as a tiebreaker.

The principle investigator and another physician from the Department of Medicine searched for missing evidence-based clinically discriminating items (defined as items having predictive likelihood ratios greater than 3.0 or less than 0.3 – McGee, 2007, p. 24) in the JAMA series of articles on the Rational Clinical Examination (JAMA Evidence, 2012) and in McGee’s Evidence Based Physical Diagnosis (McGee, 2007; McGee 2012). They also performed a PubMed search but did not find items beyond those from JAMA or McGee. They found on average six missing evidence-based items per station (range: 2-10, SD=3).

Data Collection Procedures

Examiner training included reviewing the station objectives, the meaning of each checklist item and the operational definition of success, and how to use the global rating scale. The two examiners in each station were asked to score each candidate independently. The rating sheets were scanned and the data was imported into a spreadsheet and de-identified for analysis. Only consenting residents had their data included for analysis.
Analyses

To assess inter-rater agreement during the classification of item types, we used intraclass correlations coefficients (SPSS v.20, IBM Corp., 2011). We also used SPSS to analyze the Spearman correlation coefficient between the checklist ratings and the global rating for each station and used Simple Interactive Statistical Analysis (SISA) to assess for statistically significance between correlation coefficients.

We used G-String IV (Bloch and Norman, 2011) to conduct generalizability and decision studies to estimate the number of 20-item cases needed to reach a Phi coefficient of 0.8 appropriate for moderate-stakes, local, summative examinations (Downing, 2004) with the various checklist options across stations using the \((r x i : s) x p\) design, and to assess the internal consistency of each station (similar to Coefficient Alpha) using a \(p x r x i\) design. We controlled for number of items per case to allow us to assess the quality of items without being confounded by the quantity of items.
III. RESULTS

*Item Classification*

Ninety-one items were used across the five stations: the raters classified 48 as clinically discriminating and 43 as non-discriminating. The faculty members incorrectly classified 7 evidence-based items as non-discriminating (see Table 1). The inter-rater reliability of the six raters (intraclass correlation) was 0.70. These checklists included 52% of available evidence-based items. Twenty-nine missing evidence-based items then were added to create the final the checklists.

**TABLE I**

ITEM CLASSIFICATION AND MISSING EVIDENCE-BASED ITEMS BY STATION

<table>
<thead>
<tr>
<th>Item Type</th>
<th>CHF</th>
<th>COPD</th>
<th>Cirrhosis</th>
<th>Knee</th>
<th>Parkinsonism</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-discriminating</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-evidence-based</td>
<td>11 (55%)</td>
<td>9 (36%)</td>
<td>5 (20%)</td>
<td>7 (29%)</td>
<td>4 (15%)</td>
<td>36 (30%)</td>
</tr>
<tr>
<td>Discriminating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-evidence-based</td>
<td>2 (10%)</td>
<td>2 (8%)</td>
<td>1 (4%)</td>
<td>7 (29%)</td>
<td>11 (42%)</td>
<td>23 (19%)</td>
</tr>
<tr>
<td>Evidence-based</td>
<td>5c (25%)</td>
<td>7c (28%)</td>
<td>13d (52%)</td>
<td>6 (25%)</td>
<td>1 (4%)</td>
<td>32 (27%)</td>
</tr>
<tr>
<td>Missing</td>
<td>2 (10%)</td>
<td>7 (28%)</td>
<td>6 (24%)</td>
<td>4 (17%)</td>
<td>10 (38%)</td>
<td>29 (24%)</td>
</tr>
<tr>
<td>Missing Evidence-based</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>20</td>
<td>25</td>
<td>25</td>
<td>24</td>
<td>26</td>
<td>120</td>
</tr>
</tbody>
</table>

*Includes one item misclassified by raters as non-discriminating that was an evidence-based, clinically discriminating item.*

*Includes five items misclassified by raters as non-discriminating that were evidence-based, clinically discriminating items.*
Participants

Forty-five residents took part in the OSCE and all 45 residents consented to participate.

Generalizability and Decision Studies

Analysis for internal consistency reliability at a station level revealed that one of the five stations (Cirrhosis) was performing poorly regardless of checklist composition (coefficients range: 0.19-0.31). Consequently this station was removed from the subsequent generalizability and decision studies. All other stations’ traditional checklist had internal consistency reliability coefficients greater than 0.5 (range .506 - .763). Modifying the composition of the checklist had a variable and unsystematic effect on internal consistency. (Table 2.)

TABLE II

GENERALIZABILITY STUDY RESULTS FOR INTERNAL CONSISTENCY RELIABILITY WITHIN STATIONS

<table>
<thead>
<tr>
<th>Checklist Type</th>
<th>CHF</th>
<th>COPD</th>
<th>Cirrhosis</th>
<th>Knee</th>
<th>Parkinson’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional Checklist</td>
<td>Phi</td>
<td>Phi (20)</td>
<td>Phi</td>
<td>Phi (20)</td>
<td>Phi</td>
</tr>
<tr>
<td></td>
<td>0.506</td>
<td>0.526</td>
<td>0.743</td>
<td>0.763</td>
<td>0.302</td>
</tr>
<tr>
<td>Clinically Discriminating Items</td>
<td>0.329</td>
<td>0.529</td>
<td>0.642</td>
<td>0.783</td>
<td>0.143</td>
</tr>
<tr>
<td>Traditional Checklist with Missing Evidence-based Items</td>
<td>0.564</td>
<td>0.564</td>
<td>0.781</td>
<td>0.743</td>
<td>0.327</td>
</tr>
<tr>
<td>Clinically Discriminating Items with Missing Evidence-based Items</td>
<td>0.468</td>
<td>0.643</td>
<td>0.750</td>
<td>0.784</td>
<td>0.238</td>
</tr>
<tr>
<td>All Evidence-based Items</td>
<td>0.490</td>
<td>0.708</td>
<td>0.725</td>
<td>0.782</td>
<td>0.232</td>
</tr>
</tbody>
</table>

*Phi (20) = Phi coefficient with 20 items per checklist
The main source of variance (39%-42%) was due to the interaction of persons with items nested in stations. Items nested in stations accounted for 18%-28% of the variance. Station variance was seen only in checklists composed solely of evidence-based items (1.1%). Raters nested in stations contributed little variance (.11%-47%). (Table 3.)
## TABLE III

**GENERALIZABILITY STUDIES – VARIANCE COMPONENTS ACROSS FOUR STATIONS**

<table>
<thead>
<tr>
<th></th>
<th>Traditional Checklist</th>
<th>Clinically Discriminating Items</th>
<th>Traditional Checklist with Missing Evidence-based Items</th>
<th>Clinically Discriminating Items with Missing Evidence-based Items</th>
<th>All Evidence-based Items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Var Comp</td>
<td>% Var</td>
<td>Var Comp</td>
<td>% Var</td>
<td>Var Comp</td>
</tr>
<tr>
<td>Person (p)</td>
<td>0.008</td>
<td>3.53%</td>
<td>0.009</td>
<td>3.86%</td>
<td>0.008</td>
</tr>
<tr>
<td>Stations (s)</td>
<td>0.000</td>
<td>0%</td>
<td>0.000</td>
<td>0%</td>
<td>0.000</td>
</tr>
<tr>
<td>Raters: Stations (r:s)</td>
<td>0.001</td>
<td>0.47%</td>
<td>0.001</td>
<td>0.42%</td>
<td>0.001</td>
</tr>
<tr>
<td>Items: Stations (i:s)</td>
<td>0.059</td>
<td>24.77%</td>
<td>0.041</td>
<td>18.29%</td>
<td>0.069</td>
</tr>
<tr>
<td>ps</td>
<td>0.012</td>
<td>4.82%</td>
<td>0.014</td>
<td>6.13%</td>
<td>0.010</td>
</tr>
<tr>
<td>pr:s</td>
<td>0.001</td>
<td>0.44%</td>
<td>0.002</td>
<td>0.84%</td>
<td>0.001</td>
</tr>
<tr>
<td>pi:s</td>
<td>0.095</td>
<td>39.71%</td>
<td>0.095</td>
<td>41.92%</td>
<td>0.101</td>
</tr>
<tr>
<td>ri:s</td>
<td>0.001</td>
<td>0.57%</td>
<td>0.001</td>
<td>0.61%</td>
<td>0.001</td>
</tr>
<tr>
<td>pri:s</td>
<td>0.063</td>
<td>26.50%</td>
<td>0.064</td>
<td>28.33%</td>
<td>0.059</td>
</tr>
</tbody>
</table>
Traditional and clinically discriminating checklists both required 11 stations to reach a Phi coefficient of 0.8, holding the number of items per checklist constant at 20 items. Checklists composed of traditional plus missing evidence-based items required 10 stations to reach a Phi of 0.8. A focused checklist consisting only of evidence-based, clinically discriminating items provided the highest reliability, requiring only eight stations to reach a Phi of 0.8. (Table 4.)

**TABLE IV**

<table>
<thead>
<tr>
<th>Checklist Types</th>
<th>Average items per station</th>
<th>Phi coefficient</th>
<th>Phi coefficient with 20 items per station</th>
<th># of stations to reach a Phi coefficient of 0.8 with 20 items per station</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional Checklist</td>
<td>18</td>
<td>0.595</td>
<td>0.606</td>
<td>11</td>
</tr>
<tr>
<td>Clinically Discriminating Items</td>
<td>11</td>
<td>0.524</td>
<td>0.595</td>
<td>11</td>
</tr>
<tr>
<td>Traditional Checklist with Missing Evidence-based Items</td>
<td>24</td>
<td>0.634</td>
<td>0.616</td>
<td>10</td>
</tr>
<tr>
<td>Clinically Discriminating Items with Missing Evidence-based Items</td>
<td>16</td>
<td>0.563</td>
<td>0.588</td>
<td>11</td>
</tr>
<tr>
<td>All Evidence-based Items</td>
<td>11</td>
<td>0.600</td>
<td>0.668</td>
<td>8</td>
</tr>
</tbody>
</table>

*Correlation of Checklist Total with Global Rating Scale*

Correlations between the various checklists and the global ratings ranged from 0.37 to 0.88 (Spearman’s rho). Correlations based on different checklist options did not differ significantly after Bonferroni correction.
IV. DISCUSSION

The purpose of this study was to explore the impact on the reliability of OSCE scores of adding missing-evidence based items to the checklist, and of focusing the checklist on all clinically discriminating items vs. only on evidence-based clinically discriminating items. We found that the addition of missing-evidence based items to the traditional checklist improved score reliability modestly; focusing the checklist on all clinically discriminating items did not improve reliability, but focusing on evidence-based, clinically discriminating items did.

Consistent with previous findings from Norman and colleagues (2006), the main source of variance in our OSCE was due to items, not cases; the variance due to person interaction with items nested in cases was eight times the person interaction with cases, indicating that case specificity may be a function of the specific items probed. Only the high-reliability evidence-based checklist was able to demonstrate case specificity.

Previous research showed that expert-created checklists included less than half of known evidence-based items (Hettinga et al., 2010). Our traditional checklists similarly included only 52% of available evidence-based items, highlighting the importance of grounding checklist development on a review of the literature as well as clinical expertise.

When missing evidence-based items were added to the traditional checklist, controlling for the number of items, score reliability improved such that a local summative OSCE could achieve acceptable reliability with 10 stations as compared to 11 stations with the traditional checklist. A focused checklist including only evidence-based, clinically discriminating items improved reliability further such that only 8 stations would be needed to reach acceptable reliability (a 27% reduction in the number of stations); this is consistent with our expectation that more
competent residents will use the higher-yield evidence-based items preferentially over other items. This is also in line with the previous finding from a medical student OSCE (Yudkowsky et al., 2009) that a checklist composed of clinically discriminating items improved reliability over the usual traditional checklist. In that study, the clinically discriminating items were primarily evidence-based.

Our findings suggest a potential reduction of 25-30% in the number of stations required to reach acceptable reliability based on checklist composition. Not only would this make the OSCE more clinically relevant, but would also significantly reduce the time and resources allocated to an OSCE (cost of examiners, standardized patients, utilizing the examination center) as well as the candidates’ time away from other learning and patient care, suggesting that OSCE checklist developers should give strong preference to evidence-based items over non-evidence-based items.

We hypothesized that examiners would value a focused approach to data gathering; we expected the checklist composed of clinically discriminating items to show the strongest correlation with the examiners’ global rating scale. We were unable to determine whether the global ratings were based on a thoroughness model, a clinically discriminating (clinical reasoning model) or a combination of both. Faculty development is essential to ensure examiners understand the desired performance.

Limitations

This study has some limitations in that it was done at one Internal Medicine residency program in Canada with first and second-year internal medicine residents, and the OSCE was a moderate-stakes examination with formative purposes. Different results may be found at other
institutions, with more advanced residents, in summative exams, or with different OSCE stations. Focus groups with residents and faculty could help reveal resident and faculty preferences for having a thorough versus focused approach in examination and real-life situations.
V. CONCLUSION

In conclusion, checklists composed of evidence-based, clinically discriminating items improved the reliability of checklist scores and reduced the number of stations needed for acceptable reliability. Educators should give strong preference to evidence-based, clinically discriminating items over non-evidence-based items when developing OSCE station checklists.
VI. CITED LITERATURE


VII. VITA

NAME: Vijay John Daniels

EDUCATION: B.Sc., Biochemistry, University of British Columbia, Vancouver, British Columbia, Canada, 2001

B.Sc.(Med), M.D., University of Manitoba, Winnipeg, Manitoba, Canada, 2005

Core Internal Medicine Residency and General Internal Medicine Fellowship
University of Alberta, Edmonton, Alberta, Canada, 2009

Master of Health Professions Education,
University of Illinois at Chicago, Chicago, Illinois, 2013

MAJOR
TEACHING /
EDUCATIONAL
LEADERSHIP:

Internal Medicine OSCE Director,
Faculty of Medicine and Dentistry – Postgraduate Medical Education
University of Alberta, Edmonton, Alberta, Canada, 2010-present

Physical Examination Lead for Year 1 Physicianship course
Faculty of Medicine and Dentistry – Undergraduate Medical Education
University of Alberta, Edmonton, Alberta, Canada, May 2013-present

GRANTS:

Daniels V.J.: Effect of clinically discriminating, evidence-based checklist items on the reliability of scores from an Internal Medicine residents’ OSCE.
University of Alberta, Department of Medicine Medical Education Research Grant. $5000.00 awarded April, 2012.


PUBLICATIONS:

PRESENTATIONS: Daniels V.J., Yudkowsky R., Bordage G., Gierl, M.J. Effect of Clinically Discriminating, Evidence-Based Checklist Items on Score Reliability for an Internal Medicine Resident OSCE. Presented at Canadian Conference on Medical Education, Quebec City, Quebec, Canada, April 22, 2013.


