Validity and Feasibility of the Minicard Workplace Direct Observation Tool in a Single Training Program

BY

ANTHONY A. DONATO
B.S., Georgetown University, 1990
M.D., University of Pittsburgh School of Medicine, 1994

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, 2012

Chicago, Illinois

Defense Committee:

Rachel Yudkowsky, Chair and Advisor
Yoon Soo Park
Alan Schwartz, Medical Education
David George, Reading Hospital and Medical Center

This thesis is dedicated to my wife, Karin, whose belief in me makes all things possible.

ACKNOWLEDGEMENTS

I would like to thank my thesis committee for their support of this work. Specifically, I would like to acknowledge Dr. Rachel Yudkowsky, for her specific, directed feedback that helped my hone my ideas; Dr. Yoon Soo Park, whose patience never wavered as he helped me grapple statistics; Dr. Alan Schwartz, who challenged me where I needed it; and David George, for his selfless dedication to help me find my place.

I would also like to acknowledge the assistance of my summer intern, Monica Oliveto, who was essential in helping me collect and organize the data for this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABIM	American Board of Internal Medicine
ACGME	Accreditation Committee on Graduate Medical Education
AERA	American Educational Research Association
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
EFA	Exploratory Factor Analysis
FA	Factor Analysis
ICU	Intensive Care Unit
ITE	In-Training Examination
Mini-CEX	Mini-Clinical Evaluation Exercise
OSCE	Obstructive Structured Cinical Examination
RMSEA	Root Mean Square Error of Approximation
TLI	Tucker-Lewis Index
US	United States
WBA	Workplace-based Assessment

SUMMARY

A retrospective study of the validity of interpretation of the ratings of a direct observational tool, the Minicard, as a measurement of the competence of internal medicine residents at one institution was performed. Validity data was collected from Messick's five sources of validity evidence (content, response process, internal structure, relationships to other variables, and consequences), as well as data on feasibility.

Content validity was demonstrated by showing that the Minicard can be used frequently by multiple independent observers in venues that represent the breadth of resident practice. Response process evidence was established by observations that raters regularly use the majority of the rating range. Internal structure evidence was provided by the identification of three factors on confirmatory factor analysis, interpretable as applied medical knowledge, communication, and professionalism. Relationships to other variables included the improvement seen in ratings over time in training, and positive correlations with OSCE, ITE and ABIM board scores. Consequential validity evidence was demonstrated by the average of 5 behavioral observations per card and written narrative action plans in 50% of all observations. Finally, the Minicard observation system was shown to be time-efficient and feasible for deployment among 80 generalist and subspecialist physicians in one institution.

ABSTRACT

Abstract

Purpose. To collect validity and feasibility evidence for use of the Minicard direct observation tool for assessment of competence of internal medicine residents.

Method. Retrospective cohort analysis of validity evidence and feasibility of the Minicard from 2006-2011 in one institution, including content (settings, observation rates, independent raters), response process (scoring distributions), internal structure (factor analysis), relationships (to time in training, OSCE and medical knowledge exams) and consequences (qualitative analysis of action plans), as well as feasibility (time and financial costs).

Results. 3715 direct observations were analyzed from 80 faculty observers rating 73 residents. Residents averaged 28 (SD 8.4) observations per year from 9 (SD 4.1) independent observers. Scoring distributions used the entire rating scale. Confirmatory factor analysis identified a three-factor fit representing medical knowledge, communication and professionalism ($\chi^2(51)= 107$, $p<0.05$; TLI = 0.97; CFI = 0.98; RMSEA = 0.025). Individual resident scores increased significantly each month. Minicard communication scores correlated weakly with overall OSCE communication Z-scores (r=.11, p<0.001), and Minicard applied medical knowledge average score correlated weakly with in-training exam Z-score of that year (r=.07, p=.02). Action plan reviews identified action-oriented feedback in 50%, observational feedback in 11%, minimal feedback in 9% and no recorded plan in 30%. Observation times averaged 15.6 (SD: 9.5) minutes.

Conclusion. This study demonstrates validity and feasibility evidence for a direct observation system for residents that can produce a broad range of observations, a wide range of scores and substantial formative feedback.

I. INTRODUCTION

In the past decade, educational organizations in the US, Canada, the United Kingdom and the Netherlands have adapted outcomes-based educational models to graduate medical education training[1]. In order to determine achievement of these outcomes, educators must be able to directly measure the highest level of level of Miller's[2] pyramid - what a learner "does" in practice. The development of workplace-based assessment (WBA) tools that are reliable, valid, feasible, educationally useful and acceptable to faculty raters are critical to making these complex judgments[3,4]. However, the development of psychometrically sound, universally applicable tools designed to prioritize formative feedback remains a challenge[5]. Rater bias is a significant source of construct-irrelevant variance, and has been described as 'impervious' to training efforts[6,7]. This bias may be attenuated by employing more independent observers [7,8] as well as cueing observers with prompts[9] and aligning global scales to the priorities and sophisticated cognitive schemes (or 'reality maps'[10]) of its raters. Because learner performance may vary with case content, venue and case complexity ('context specificity'), WBA's must be able to sample the breadth of learner performances[10,11]. Most importantly, workplace assessment tools should promote frequent, written formative feedback[12] as opposed to focusing on summative scores, promoting the concept of 'assessment *for* learning'[13].

While there are more than 50 available tools for direct observation of medical students and residents[14], the most widely-used instrument for the rating of direct observations of resident physicians is the Mini-CEX[15]. The Mini-CEX's reliability has been demonstrated in studies of inter-item correlations[16–18] and in generalizability studies[7,15,19–22]. Validity evidence includes positive correlations to scripted videos of learners with increasing competence[23], positive correlations to OSCE scores[24], to in-training exam scores[16], and Royal College Exams[25]. Ratings of trainees generally get better with increasing time in training in most[15,18,22] but not all[16]studies of the Mini-CEX. Weaknesses of the Mini-CEX include content underrepresentation as a result of its underuse[9,11,16,17,20,32,33] or resistance to use by faculty[19,22,26], as well as evidence of construct –irrelevant variance (halo effect[7,12,23], range restriction[16,18,27,28], leniency error[7,12,23], and difficulty discriminating between capable and struggling trainees [1,23,25]). In addition, it has been described as a tool designed for summative assessment over formative feedback, given its minimal space for written comments and action plans[27]. The Reading Minicard was developed as alternative to the Mini-CEX with the goals of increasing specific formative feedback while aligning the rating system to the ACGME core competencies and the expertise and priorities of faculty observers. In a study comparing the Minicard and the Mini-CEX in which physicians rated scripted videos of trainees, the Minicard was shown to have higher inter-rater reliability (Fleiss' kappa, 0.52 versus 0.30 for the Mini-CEX)[29].

In this paper we present an argument for the validity of using Minicard scores to assess the competence of internal medicine residents by summarizing both previously published and newly obtained validity evidence from a variety of sources. Using the conceptual framework of validity provided by Messick[30] in the AERA Standards, we present validity evidence based on content, response process, internal structure, relationships to other variables and consequential validity. We will also describe the feasibility and utility of the Minicard as demonstrated in one training program.

II. Methods

Instrument Studied: The Minicard. The Minicard (Appendix 1) is a resident assessment tool organized into four sections that represent commonly observed resident activities: obtaining a history, performing a physical examination, presenting a patient to faculty, and counseling or discussion of findings with the patient[29]. Each activity provides the context for assessing three potentially observable domains drawn from the ACGME competencies: (1) those related to interpersonal communication; (2) those related to applied medical knowledge in the context of patient care, referred to on the instrument as simply "medical knowledge", and (3) professionalism during the encounter. Prompts cue observers in each domain, and may be used to record the presence or absence of specific behaviors during the observation. Behavioral anchors are written for each of the four scoring levels in each domain. Space is provided for free text comments under each domain, and the observer is prompted to produce an action plan at the end.

The Reading Minicard was developed by experienced internal medicine educators using a blueprint of best practices in the areas of interviewing, physical examination and counseling as found in core medicine texts and articles[31–35], and using the domains of the ACGME core competencies, with the stated goal of facilitating formative feedback. It was piloted and refined over 1 year. Faculty raters were trained in 1-hour frame-of reference training sessions. Faculty selected both the patients and the activities to observe, and were financially incentivized to complete one Minicard per learner per week. Action plans were reviewed and uploaded to residents' online structured

portfolio three times a year for review with their institution mentor. Since 2011, observers have received annual feedback on their scoring ranges as compared to peers.

Objective and Design. This study was a retrospective cohort analysis of direct observations in one institution. The objectives of this study were to collect validity evidence for use of the Minicard scores as a measure of a learner's clinical competence, as well as to document feasibility aspects of the observation system. Specific validity evidence sought included (1) evidence of the adequacy of content sampling (the clinical setting of use of the Minicard, number of observations per resident, and number of independent examiners per resident); (2) response process (use of the full range of scores); (3) internal structure (number of independent constructs measured by the Minicard); (4) relationships to other variables (changes in Minicard scores over time for individual residents, and associations of the Minicard to OSCE performance, in-training exam score, and ABIM board scores); (5) consequential validity (the quality of the feedback written on the Minicard); and finally, feasibility estimates, including costs of time and money to train observers and record Minicards.

Participants and Setting. The dataset included direct observations recorded on Minicards over 5 years in one internal medicine residency in an academic independent medical center in the northeastern United States.

Data Analysis. To evaluate the adequacy of content sampling, the percentage of Minicard use was reported for each of four settings (Inpatient ward, Emergency Department, Clinic and ICU) over the last 2 years, after which time the Minicard was

fully deployed in all settings. The number of Minicards and of independent observers were calculated for each resident. Response process was assessed by reporting the distribution of Minicard scores for interns and third year residents. Factor analysis was used to examine the internal structure of the Minicard by randomly dividing the data in half. With one half of the dataset, we performed an exploratory factor analysis (EFA). With the second half, we performed confirmatory factor analysis (CFA) of a hypothesized three-factor structure of the Minicard. The combination of both CFA and EFA provides both hypothesis- and data-driven analyses of Minicard's internal structure[36]. To test relationships to other variables, a mixed-effects linear regression was used to examine individual score changes by month in training. In order to compare scores between the Minicard 4-point ordinal ranking and the continuous scores of the In-traning exam (ITE) and ABIM boards, we used the average score for the medical knowledge portions of the Minicard for the 6-month period preceding the time of the ITE test report or the last six months of residency (For ABIM board scores). The 6-month average score of items representing the communication domain were compared to the Z-score of a 10-station communication OSCE performed at four months into the intern year.

To gather evidence for consequential validity, we examined the quality of the Minicard's action plans. Two investigators (AD, DG) were trained in using a standard coding schema for feedback[27], coding each written plan as "action-oriented" (e.g. "next time, set an agenda first"), "observational" (e.g. he was terse with the patient"), "minimal"(e.g. "good resident"), or "none/blank". Action plans containing two levels of

action were coded using the higher plan. Double-coding was performed for 20% of all observations to test inter-rater reliability, and discrepancies were discussed until agreement was reached. Finally, feasibility in using the Minicard was assessed by calculating observer-reported time spent per observation, time spent in faculty training, and direct and indirect costs (administrative staff recording time, faculty incentive costs, reproduction costs).

III. Results

A total of 3715 Minicards were collected and analyzed over a period of 5 years (2006- 2011) from an internal medicine residency program with 27 residents (six one-year transitional residents, plus seven categorical medicine residents in each of three years). Eighty physician faculty raters (30 generalists, 50 sub-specialists from six subspecialties) performed observations on 73 different residents during this period.

Content validity- sampling adequacy: The most commonly reported setting for observation was the inpatient ward (43%), followed by the clinic (39%), the intensive care unit (15%) and the emergency department (3%), with less than 1% of observations unidentified. Raters recorded observations in the history section of the Minicard in 30% of the encounters, in the physical exam section in 23%, in the oral presentation section in 52% and in the counseling section in 27%.

Residents were observed an average of 28 (SD 8.4) times per year (range: 15-45).

Response process: Observers most often rated interns as 'good' (56% of ratings), and used the 'marginal' rating for 8% of intern observations, while they most often rated third year residents as 'excellent' (67% of ratings), and used 'marginal' ratings only 2% of the time (Figures 1 and 2). Residents had an average of 15 (SD 9.0) independent observers during residency (range, 3-38), with on average 9 (SD 4.1) independent observers per resident per year (range: 3-20).

Figure 1. Histogram of Intern Scores

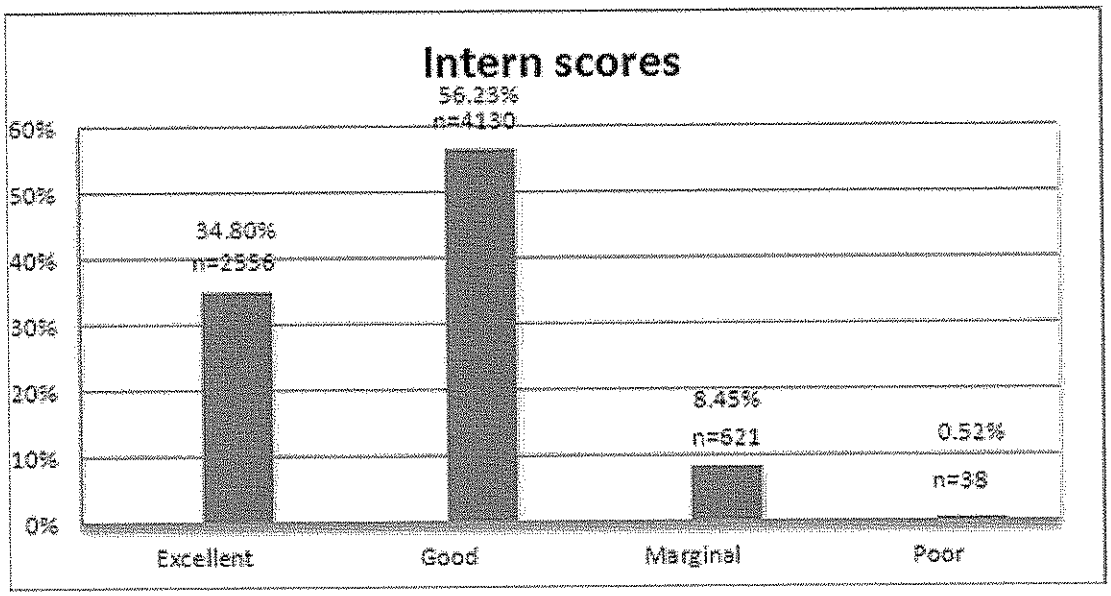
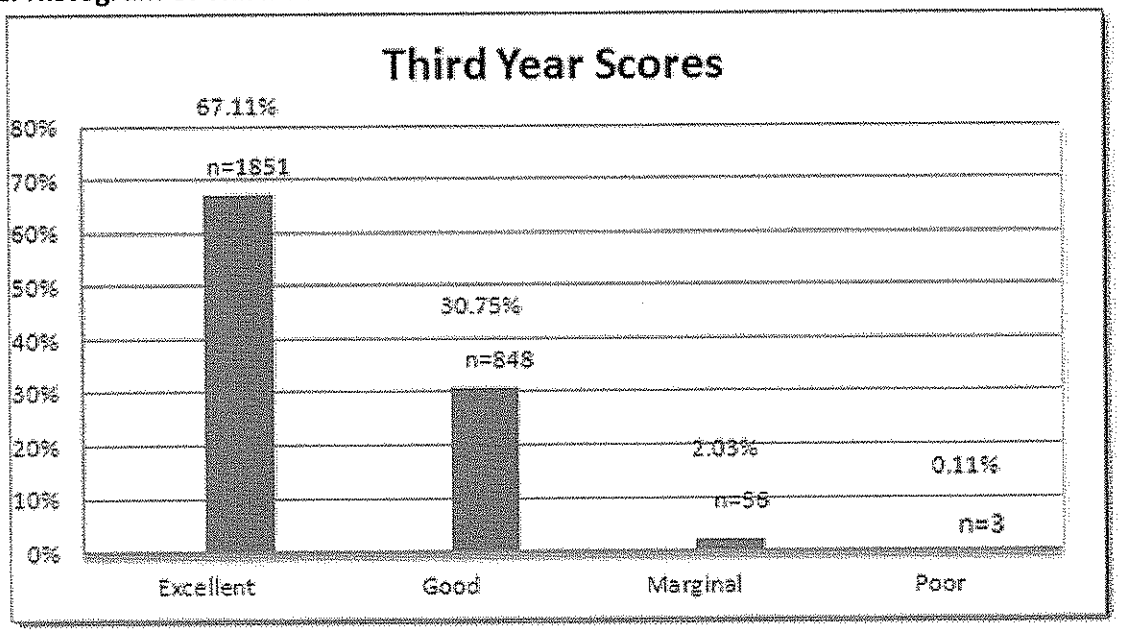


Figure 2. Histogram of Third Year Scores



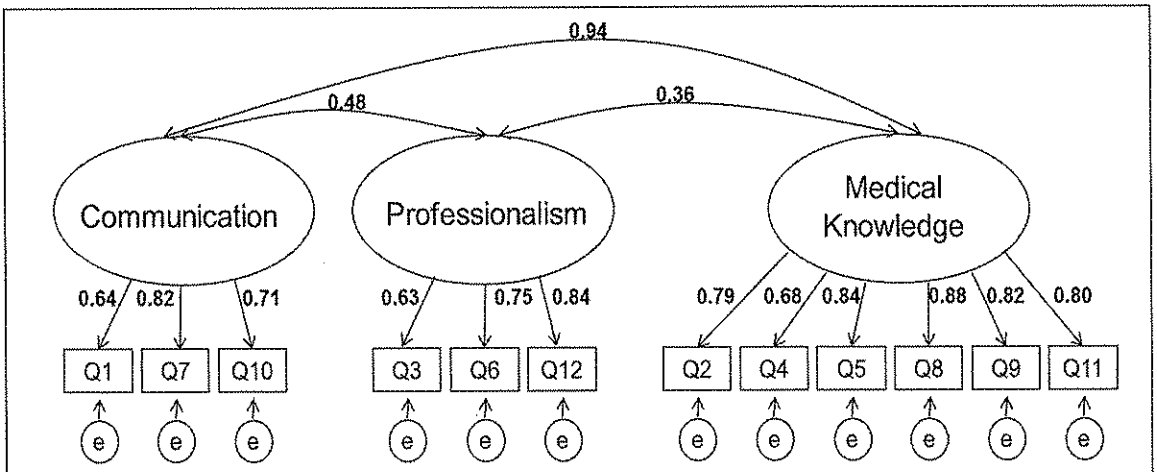
Internal structure: EFA identified a two-factor solution that was interpretable as Factor 1: 'applied medical knowledge and communication' and Factor 2: 'professionalism' (Table 1). CFA was consistent with a two-factor solution but also supported a three-factor model (Figure 3) that showed a better model fit ($\chi^2(51) = 107$,

$p<0.05$; TLI = 0.97; CFI = 0.98; RMSEA = 0.025). The three-factor model showed a strong correlation between applied medical knowledge and communication scores (.94), and a lower association of these two factors with the professionalism factor (.48 and .36 respectively). All factor loadings were over 0.63.

Table 1. Exploratory Factor Analysis, Two-Factor Solution

	Section	2 Factors	
		Factor 1	Factor 2
	Communication questions		
Q1	History: Communication	0.65	0.16
Q7	Presentation: Oral	0.81	0.00
Q10	Counseling: Communication	0.59	0.23
	Medical Knowledge questions		
Q2	History: Medical knowledge	0.78	–0.09
Q8	Presentation: Data synthesis/reasoning	0.88	–0.07
Q11	Counseling: Medical knowledge	0.87	0.00
	Professionalism questions		
Q3	History: Professionalism	0.12	0.47
Q6	Physical exam: Professionalism	0.00	0.83
Q12	Counseling: Professionalism	0.00	0.87
	Physical examination questions		
Q4	Physical exam: Skill	0.68	0.04
Q5	Physical exam: Understanding	0.83	0.01
	Systems-based practice question		
Q9	Presentation: Systems-based practice	0.79	0.03

Figure 3. Confirmatory Factor Analysis: Three-factor Assessment Model

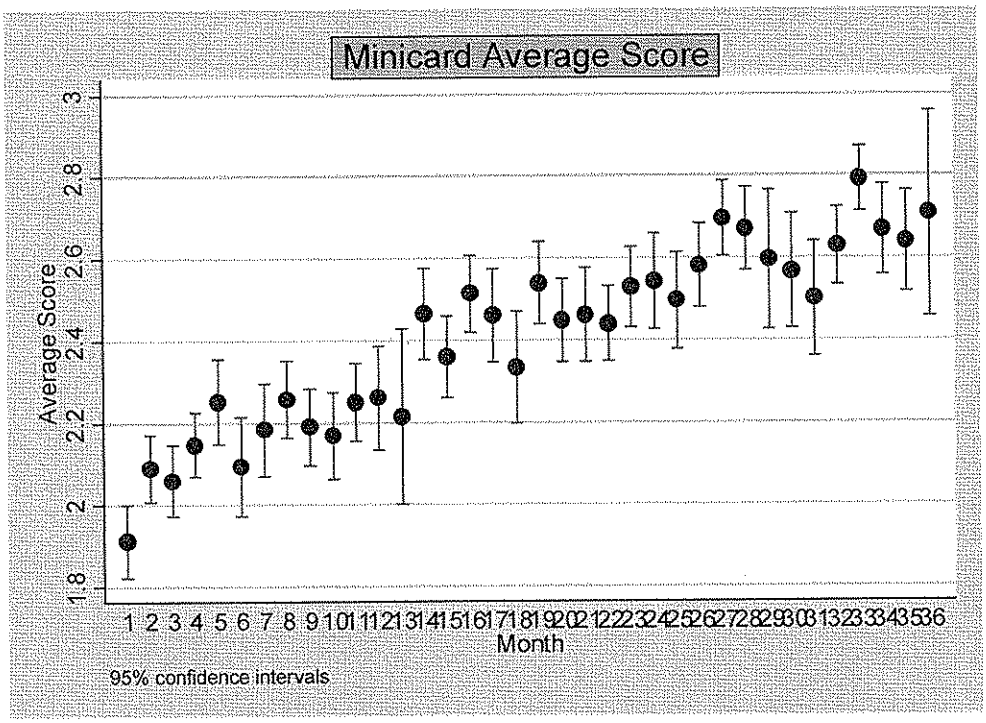


Note: Q1 History: Communication, Q2 History: Medical knowledge, Q3 History: Professionalism, Q4 Physical exam: Skill, Q5 Physical exam: Understanding, Q6 Physical exam: Professionalism, Q7 Presentation: Oral, Q8 Presentation: Data synthesis/reasoning, Q9 Presentation: Systems-based practice, Q10 Counseling: Communication, Q11 Counseling: Medical knowledge, Q12 Counseling: Professionalism

Three-factor measurement model. Tucker-Lewis Index=0.97; Comparative Fit Index=.98; Root Mean Square Error of Approximation=0.025; Chi-Squared=107.43; degrees of freedom=51, $p<0.05$. e=error

Relationships to other variables: (1) To time in training: Minicard scores increased by 0.021 points per month (95% CI: 0.019-0.024, p<0.001) (Figure 4).

Figure 4: Minicard Scores over time



(2) To other measures of communication: Minicard communication scores for the first 6 months of the internship years correlated weakly with overall OSCE communication Z-scores (r=0.11, 95% CI: 0.051-0.175, p<0.001), but not with OSCE counseling scores (r=0.06, 95% CI: -0.004-0.121, p=0.07). (3) To other measures of medical knowledge: The average Minicard applied medical knowledge score in the first six months of the year correlated weakly with in-training exam Z-score of that year (r=0.075, 95% CI: 0.011-0.138, p=0.02) and year-three medical knowledge scores correlated weakly with ABIM board Z-scores (r=0.10, 95% CI: 0.003-0.203, p=0.04) for the subset of 13 participants who took the board examination.

Consequential validity: Exact agreement between coders was 93% (Kappa=0.89). Qualitative analysis of written action plans identified action-oriented feedback in 50% of Minicards, observational feedback in 11%, minimal feedback in 9% and no recorded plan in 30%. When an action plan was recorded, an action related to a medical knowledge deficiency was prescribed in 56%, actions related to communication deficiencies in 44%, and actions related to professionalism in about 1%. There was an average of 5 (SD 5.1) prompts checked per Minicard, with each representing a discrete observation (range: 0-20). About 8% of Minicards included ratings only, with no checked prompts or action plan. Observers documented that they gave verbal feedback in 74% of the encounters.

Feasibility: The average duration of an observed encounter was 15.6 (SD 9.5) minutes (range: 1-120 minutes; interquartile range, 10-20 minutes). Observers spent 1 hour in frame-of-reference training before using the Minicard. Cost of reproduction of the cards was $0.022/card, or $0.61/resident/year. An administrative assistant spent 2 minutes recording each Minicard, or approximately 29 minutes per week to record observations (checked prompts) and action plans on an electronic spreadsheet. Financial incentive costs included 1.2% of annual salary for generalists and 0.6% annual salary for specialists; the total estimated cost of the incentive program for the 80 observers was $140,800/ year.

IV. Discussion

This paper has presented validity evidence from each of Messick's five sources for use of the ratings from the Minicard to assess the competency of Internal Medicine residents. Content validity was demonstrated by showing that the Minicard can be used frequently by multiple independent observers in venues that represent the breadth of resident practice. Response process evidence was established by observations that raters regularly use the majority of the rating range. Internal structure evidence was provided by the identification of three factors on confirmatory factor analysis, interpretable as applied medical knowledge, communication, and professionalism. Relationships to other variables included the improvement seen in ratings over time in training, and positive correlations with OSCE, ITE and ABIM board scores. Consequential validity evidence was demonstrated by the average of 5 behavioral observations per card and written narrative action plans in 50% of all observations. Finally, the Minicard observation system was shown to be time-efficient and feasible for deployment among 80 generalist and subspecialist physicians in one institution.

While the confirmatory factor analysis showed the best fit for three factors, exploratory FA identified only two factors: professionalism, and a combined factor of applied medical knowledge and communication. The high correlation of .94 between the factors of applied medical knowledge and communication in the three-factor model suggests that the case-specific natures of patient care and communication render these factors highly interdependent within a case. Nonetheless, we found the three-factor

interpretation to be educationally useful in identifying opportunities for improvement. Raters distinguished between these two skill domains (factors) in their action plans, which tended to focus on either applied medical knowledge (56%) or communication (44%). Both analyses clearly identified professionalism as a separate competency domain.

The low correlations with OSCE, ITE and ABIM board scores are similar to the low to moderate (r=0.1-0.3) correlations between measures of competence seen in workplace-based assessments in one systematic review[14]. The low correlations suggest that direct observation of patient care provides added value by measuring constructs of applied knowledge and communication skill that may not be measured by OSCE and written tests of medical knowledge.

Validity evidence for the Minicard scores compared positively to the published evidence for the Mini-CEX. While the Minicard produced 28 observations per resident annually, published reviews of the Mini-CEX produce annual rates of observations per trainee of only 0.4-7.9[1,15–17,19,24,25,27,37], well below the minimum number per trainee (12-14) needed in one year to reach an acceptable reliability[38]. Mini-CEX scores are noted to have range restriction, with the majority of authors reporting average scores of 6 or 7 on a 9-point with standard deviations all less than 1.1[16,18,27,28]. This study demonstrated Minicard observation scores that used a broader range of the scale, including 659/7345 (9%) unsatisfactory ratings for intern year residents. Mini-CEX unsatisfactory ratings are rarely used: 0/1280 in 3rd year medical students[18], 1/196 resident physicians[19], 0/388 medicine interns[17] , and 0/107 medicine interns[27]. While

the Minicard has three interpretable subscales, studies of the Mini-CEX found that a single construct may explain all of the scoring variance[7,12,21]. Observation times for Mini-CEX's in the literature are longer than the Minicard's (19-31.5 minutes on average)[16,19,21,24,25,37,39], which may limit its feasibility for busy faculty observers. Most importantly, Minicard users produced a written action plan in 50% of cards, significantly higher than action plan rates of 8% in one study using the Mini-CEX[27].

This study is limited by the fact that it was performed at a single institution with a strong commitment to education and direct observation. Whether this observation system can be successful in other departments of medicine or other specialties is unclear. Whether faculty acceptance was a function of the physician-designed instrument or the assessment system (including training, support structure and incentives) cannot be determined, but should always be considered in adapting a novel assessment tool to a new environment[4]. A Generalizability study is in progress to determine the number of Minicards required to obtain a reliable estimate of resident competence. Whether similarly-designed tools can be implemented in other facilities and residencies, and the degree of contribution of faculty champions, feedback to raters, and financial incentives to implementation success are important areas for future study.

Workplace-based assessment tools can be unpopular among faculty, evoking reactions from 'widespread cynicism'[10] to 'immediate resistance' and faculty revolt[19], and leading some to argue that feasibility is the major barrier to implementation[38]. This study shows a direct observation system that can produce a broad range of observations, a wide range of scores and substantial formative feedback. Finding ways

to use this formative feedback to direct learners' educational efforts and enhance future performance ('assessment for learning'[13]) will be critical to medical educators and an important next step in assessment research.

APPENDICES

Appendix A: Minicard

Reading Hospital Mini-CEX Rating Instrument **Date**____/___/____
Student:_____ **Observer:**_____
Case description:

Directions: **circle features done CORRECTLY, place "X" over ERRORS noted**

History

Interpersonal/Communication skills

1 Greeting Set agenda, "*anything else*?" Uses open-ended , non-leading questions
Gives /responds to patient's non-verbal cues Uses summarizing/clarifying/reflective statements
Demonstrates empathy "*that must have been upsetting*" Avoids medical jargon Attentive

Excellent	**Good**	**Marginal**	**Poor**
Demonstrated all of above, outstanding interaction	missed 1-2 items without egregious mistake	missed >2 or borderline egregious mistake; marginal connection	offended patient, obviously negative interaction

Comments:

Data Collection: Medical Knowledge

2 Elicits focused chief complaint General-to-specific questioning Got relevant PMH/SH
Asked discriminatory questions that prioritized differential

Excellent	**Good**	**Marginal**	**Poor**
understands historical nuances; no irrelevant data collected;	collected enough to correctly rank ddx, rarely tangential	Missed 1 or more vital data points; failed to discriminate Ddx or prioritize complaints	tangential data collector; missed major topics; "lost" in data

Senior resident/staff **Resident/Intern** **Intern/Med student**
Comments:

Professional Conduct

3 Non-judgmental Does not make pt. "prove" illness Respectful to person/privacy/spirituality

Excellent/Good	**Marginal /Poor**
Patient pleased with the interaction	any above feature

Comments:

Physical Exam

Medical knowledge: physical diagnosis skills

Technically proficient at exam maneuvers — Avoided irrelevant exam portions

4 Did not omit necessary elements of exam — Used tools/positioning appropriately

Excellent	**Good**	**Marginal**	**Poor**
No omissions	1-2 less important omissions or 1 irrelevant exam feature	missed or botched major item or non-focused exam	appeared not to understand relevant exam

Comments:

Medical reasoning/exam interpretation

Understood extenuating circumstances that limit exam's usefulness (e.g. steroids/peritonitis)

5 Understood general sensitivity and specificity of findings

Excellent	**Good**	**Marginal**	**Poor**
Can use findings to effectively rank Ddx; aware of limitations Of exam findings	understands relation between disease suspected and test performed	did general physical of that organ system; omitted/did not comprehend discriminators	not able to use exam to refine historical inquiry
Senior res. / staff	**Res./Intern**	**Med student**	

Comments:

Professional Conduct

Asked permission/ explained exam — Respects comfort/modesty — Washes hands

6

Excellent/Good	**Marginal /Poor**
No or minor omissions	any major infraction

Assessment of findings

Oral case presentation

Could logically organize all relevant data Omitted irrelevant data

7 Incorporated pertinent pos/neg data Data given aids listener in assembling/ranking ddx

Excellent	**Good**	**Marginal**	**Poor**
Flowing, relevant presentation; top and next ddx items obvious from data given	minor ddx item or finding neglected; major ddx captured poss. out of order	rambling presentation, all data captured; major ddx item missed but organ system correct	student lost or unfamiliar with relevant features; dangerous misses
Sr. res/staff	**Res./Intern**	**Med student**	

Comments:

Data synthesis/reasoning (medical knowledge components)

Logic, prioritization of differential is consistent, accurate Values datapoints appropriately

8 Analysis of prevalence of disease, test sensitivity/specificity obvious in discussion

Not reliant on single data point No omission of relevant data points that may refute diagnosis

Recognizes knowledge gaps, formulates appropriate clinical questions Avoids early closure

Excellent	**Good**	**Marginal**	**Poor**
No omissions, Clear, accurate logic for ddx, formulates approp. clin ?'s	correct ddx, possibly miss or omit data, did not use/understand prev/sens/spec	Got major ddx item and correct organ system but 1 or greater major error; or can't see error	unable to synthesize data or faulty reliance on bad data point
Sr. res/staff	**Res./Intern**	**Intern/Med student**	

Comments:

Plan: systems-based practice

9 Able to incorporate comorbid conditions into test/ treat. choices

Cost-conscious, ethical approach to testing Correctly identifies level of urgency of evaluation

Understands what to do with (pos or neg) test results Uses ancillary staff/resources appropriately

Understands limitations of tests chosen (sens/spec/ risks of false pos results)

Excellent	**Good**	**Marginal**	**Poor**
Mature, forward-thinking decisions consideration of patient's unique circumstances	Orders correct tests relevant to disorder without considering comorbidities, cost	"shotguns" tests, not aware of dz history, fails to use anc. staff, fail to consider pt issues	Makes 2 or more major mistakes
Sr. res/staff	**Res./Intern**	**Intern/Med student**	

Comments:

Presentation of plan to patient/Counseling/Behavioral Change

Interpersonal/Communication Skills

Defines issue Shared decisionmaking "*Let's do this together*" Good pace

10 Common ground/patient education/understanding evaluated "*what do you understand about..*"

Avoids medical jargon Explores variables that would affect pt's choice

Pauses for/invites questions Respects pt. opinions and preferences Summarizes

Gives and responds to patient's non-verbal cues

Excellent	**Good**	**Marginal**	**Poor**
Found common ground, shared decision/uncertainty comfortably	missed some minor (defining/shaping discussion) issues overall positive	missed 1 major (defining "where pt is", stud. not aware they are not understood)	dictatorial; patient with negative experience

Comments:

Medical Knowledge components

Addresses uncertainties with choice (limitations of testing/therapy/varied patient response to tx)

11 Discussion of pros/cons of options (incl. nothing) Conveys risk in testing/treating

Demonstrates understanding of limitations in test/tx

Excellent	**Good**	**Marginal**	**Poor**
Thorough understanding of all diagnostic and therapeutic options; comfort with uncertainty	knows major options may miss minor nuances of tx/ less important side fx	can name 1-2 options and basic dz course; unaware of major alternatives	makes >2 major errors
Sr. Res/Staff	**Res./Intern**	**Med Student**	

Comments:

Professionalism

12 Demonstrated bias Condescending Ignored pt's preferences Disrespectful

Excellent/Good	**Marginal /Poor**
No or minor omissions	any major infraction

Comments:

Total time observed____ **Feedback given? Y / N**

ACTION PLAN_____

CITED LITERATURE

1. Mitchell C, Bhat S, Herbert A, Baker P. Workplace-based assessments of junior doctors: do scores predict training difficulties? *Med Educ.* 2011;45(12):1190–1198.

2. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.* 1990;65(9 Suppl):S63–67.

3. Weinberger SE, Pereira AG, Iobst WF, Mechaber AJ, Bronze MS. Competency-based education and training in internal medicine. *Ann. Intern. Med.* 2010;153(11):751–756.

4. Schuwirth LWT, Van Der Vleuten CPM. Changing education, changing assessment, changing research? *Medical Education.* 2004;38(8):805–812.

5. Norcini J, Anderson B, Bollela V, et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach.* 2011;33(3):206–214.

6. Cook DA, Dupras DM, Beckman TJ, Thomas KG, Pankratz VS. Effect of rater training on reliability and accuracy of mini-CEX scores: a randomized, controlled trial. *J Gen Intern Med.* 2009;24(1):74–79.

7. Margolis MJ, Clauser BE, Cuddy MM, et al. Use of the mini-clinical evaluation exercise to rate examinee performance on a multiple-station clinical skills examination: a validity study. *Acad Med.* 2006;81(10 Suppl):S56–60.

8. Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance ratings. *Teach Learn Med.* 2003;15(4):270–292.

9. Nathan BR, Lord RG. Cognitive categorization and dimensional schemata: A process approach to the study of halo in performance ratings. *Journal of Applied Psychology. Vol 68(1).* 1983;68:102–114.

10. Crossley J, Jolly B. Making sense of work-based assessment: ask the right questions, in the right way, about the right things, of the right people. *Medical Education.* 2012;46(1):28–37.

11. Kane MT. The assessment of professional competence. *Eval Health Prof.* 1992;15(2):163–182.

12. Hawkins RE, Margolis MJ, Durning SJ, Norcini JJ. Constructing a validity argument for the mini-Clinical Evaluation Exercise: a review of the research. *Acad Med.* 2010;85(9):1453–1461.

13. Shepard LA. The Role of Assessment in a Learning Culture. *EDUCATIONAL RESEARCHER.* 2000;29(7):4–14.

14. Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *JAMA*. 2009;302(12):1316–1326.

15. Norcini JJ, Blank LL, Arnold GK, Kimball HR. The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Ann. Intern. Med.* 1995;123(10):795–799.

16. Durning SJ, Cation LJ, Markert RJ, Pangaro LN. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. *Acad Med.* 2002;77(9):900–904.

17. Norcini JJ, Blank LL, Duffy FD, Fortna GS. The mini-CEX: a method for assessing clinical skills. *Ann. Intern. Med.* 2003;138(6):476–481.

18. Kogan JR, Bellini LM, Shea JA. Implementation of the mini-CEX to evaluate medical students' clinical skills. *Acad Med*. 2002;77(11):1156–1157.

19. Jackson D, Wall D. An evaluation of the use of the mini-CEX in the foundation programme. *Br J Hosp Med (Lond)*. 2010;71(10):584–588.

20. Nair BR, Alexander HG, McGrath BP, et al. The mini clinical evaluation exercise (mini-CEX) for assessing clinical performance of international medical graduates. *Med. J. Aust.* 2008;189(3):159–161.

21. Hill F, Kendall K, Galbraith K, Crossley J. Implementing the undergraduate mini-CEX: a tailored approach at Southampton University. *Med Educ*. 2009;43(4):326–334.

22. Alves de Lima A, Barrero C, Baratta S, et al. Validity, reliability, feasibility and satisfaction of the Mini-Clinical Evaluation Exercise (Mini-CEX) for cardiology residency training. *Med Teach*. 2007;29(8):785–790.

23. Holmboe ES, Huot S, Chung J, Norcini J, Hawkins RE. Construct validity of the miniclinical evaluation exercise (miniCEX). *Acad Med.* 2003;78(8):826–830.

24. Boulet JR, McKinley DW, Norcini JJ, Whelan GP. Assessing the comparability of standardized patient and physician evaluations of clinical skills. *Adv Health Sci Educ Theory Pract*. 2002;7(2):85–97.

25. Hatala R, Norman GR. In-training evaluation during an internal medicine clerkship. *Acad Med*. 1999;74(10 Suppl):S118–120.

26. Morris A. Practical experience of using directly observed procedures, mini clinical evaluation examinations, and peer observation in pre-registration house officer (FY1) trainees. *Postgraduate Medical Journal*. 2006;82(966):285–288.

27. Holmboe ES, Yepes M, Williams F, Huot SJ. Feedback and the mini clinical evaluation exercise. *J Gen Intern Med*. 2004;19(5 Pt 2):558–561.

28. Torre DM, Simpson DE, Elnicki DM, Sebastian JL, Holmboe ES. Feasibility, reliability and user satisfaction with a PDA-based mini-CEX to evaluate the clinical skills of third-year medical students. *Teach Learn Med*. 2007;19(3):271–277.

29. Donato AA, Pangaro L, Smith C, et al. Evaluation of a novel assessment form for observing medical residents: a randomised, controlled trial. *Med Educ*. 2008;42(12):1234–1242.

30. Messick S. *Educational Measurement*. 3rd ed. New York: American Council on Education and Research; 1989.

31. (MD.) LG, Ausiello DA. *Cecil textbook of medicine*. Saunders; 2004.

32. Fiebach NH. *Principles of ambulatory medicine*. Lippincott Williams & Wilkins; 2007.

33. MD LSB. *Bates' Guide to Physical Examination & History Taking*. Eighth. Lippincott Williams & Wilkins; 2002.

34. Braddock CH 3rd, Edwards KA, Hasenberg NM, Laidley TL, Levinson W. Informed decision making in outpatient practice: time to get back to basics. *JAMA*. 1999;282(24):2313–2320.

35. Snyder L, Leffler C, for the Ethics and Human Rights Committee AC of P. Ethics Manual. *Annals of Internal Medicine*. 2005;142(7):560 –582.

36. Fabrigar LR1, Wegener DT, MacCallum RC, Strahan EJ1 QU. Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*. 1999;4(3):272–299.

37. Overeem K, Lombarts MJMH, Arah OA, et al. Three methods of multi-source feedback compared: a plea for narrative comments and coworkers' perspectives. *Med Teach*. 2010;32(2):141–147.

38. Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med Teach*. 2007;29(9):855–871.

39. Hawkins RE, Margolis MJ, Durning SJ, Norcini JJ. Constructing a validity argument for the mini-Clinical Evaluation Exercise: a review of the research. *Acad Med*. 2010;85(9):1453–1461.

VITA

NAME:	Anthony Andrew Donato
EDUCATION:	BS, Biology, Georgetown University, Washington, DC, 1990
	MD, Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, 1994
TEACHING:	Clinical Assistant Professor, Uniformed Services University, 1999-2001
	Clinical Associate Professor, Jefferson Medical College, 2007-present
HONORS:	Blockley-Osler Award for Distinguished Teaching 2009 Awarded by Jefferson Medical Students for Clerkship teaching
PROFESSIONAL MEMBERSHIP:	Jefferson Academy of Distinguished Educators
PUBLICATIONS:	Donato, AA, Pangaro, L, Holmboe E, et al. Evaluation of a Novel Assessment Tool for Observing Residents, A Randomized, Controlled Trial. *Medical Education* 2008 42(12): 1234-1242.
	Donato, AA and George, D. "A Blueprint for Implementation of a Structured Portfolio in an Internal Medicine Residency" *Acad Med*, 87(2): 185-91 , Feb 2012
	Motz L, Lloyd B, Donato A, Chaudhary A, Kaliyadan A, Stavarski D, et al. Interdisciplinary Curriculum and Simulation Cases for Teaching Leadership and Communication to Medical Rapid Response Teams. MedEdPORTAL; 2012. Available from: www.mededportal.org/publication/9145