

# Validity and Feasibility of the Minicard Direct Observation Tool in 1 Training Program

ANTHONY A. DONATO, MD  
YOON SOO PARK, PhD  
DAVID L. GEORGE, MD  
ALAN SCHWARTZ, PhD  
RACHEL YUDKOWSKY, MD, MHPE

## Abstract

**Background** Availability of reliable, valid, and feasible workplace-based assessment (WBA) tools is important to allow faculty to make important and complex judgments about resident competence. The Minicard is a WBA direct observation tool designed to provide formative feedback while supporting critical competency decisions.

**Objective** The purpose of this study was to collect validity and feasibility evidence for use of the Minicard for formative assessment of internal medicine residents.

**Methods** We conducted a retrospective cohort analysis of Minicard observations from 2005–2011 in 1 institution to obtain validity evidence, including content (settings, observation rates, independent raters); response process (rating distributions across the scale and ratings by month in the program); consequences (qualitative assessment of action plans); and feasibility (time to collect observations).

**Results** Eighty faculty observers recorded 3715 observations of 73 residents in the inpatient ward (43%), clinic (39%), intensive care (15%), and emergency department (3%) settings. Internal medicine residents averaged 28 (SD = 8.4) observations per year from 9 (SD = 4.1) independent observers. Minicards had an average of 5 (SD = 5.1) discrete recorded observations per card. Rating distributions covered the entire rating scale, and increased significantly over the time in training. Half of the observations included action plans with action-oriented feedback, 11% had observational feedback, 9% had minimal feedback, and 30% had no recorded plan. Observations averaged 15.6 (SD = 9.5) minutes.

**Conclusions** Validity evidence for the Minicard direct observation tool demonstrates its ability to facilitate identification of “struggling” residents and provide feedback, supporting its use for the formative assessment of internal medicine residents.

*Editor’s Note: The online version of this article contains the Minicard used in the study.*

## Introduction

In the past decade, outcomes-based educational models for graduate medical education have been implemented in the

United States, Canada, the United Kingdom, and the Netherlands.<sup>1</sup> Learner assessment in this approach generates rich narrative descriptions of the behaviors of a competent professional, which are used to determine progress and achievement toward these goals. The development of workplace-based assessments (WBAs) that are reliable, valid, feasible, educationally useful, and acceptable to faculty raters is critical to making these complex judgments.<sup>2,3</sup> In addition, if the goal is assessment *for* learning,<sup>4</sup> workplace assessment tools should prioritize meaningful narratives of encounters in order to promote formative feedback<sup>5</sup> while providing discrete observations to allow summative evaluators to determine a learner’s progress toward educational goals. Since learner performance may vary with different observers, case content, venue, and case complexity (“context specificity”), WBAs must be flexible enough for use by multiple independent observers<sup>6,7</sup> and must be able to sample the breadth of learner performances.<sup>8,9</sup> The Reading Minicard<sup>10</sup> is a direct observation tool that was developed with the goal of increasing formative feedback while aligning the tool to the expertise and priorities of faculty observers.

**Anthony A. Donato, MD**, is Clinical Associate Professor, Internal Medicine, Jefferson Medical College; **Yoon Soo Park, PhD**, is Assistant Professor, Department of Medical Education, University of Illinois-Chicago; **David L. George, MD**, is Clinical Associate Professor, Internal Medicine, Jefferson Medical College; **Alan Schwartz, PhD**, is Professor, Department of Medical Education, University of Illinois-Chicago; and **Rachel Yudkowsky, MD, MHPE**, is Associate Professor, Department of Medical Education, University of Illinois-Chicago.

**Funding:** The authors report no external funding source for this study.

**Conflict of interest:** The authors declare they have no competing interests.

The authors would like to thank Monica Oliveto for her assistance in collecting and compiling the data for this study.

Corresponding author: Anthony A. Donato, MD, Jefferson Medical College, Department of Medicine, B-2, 6th and Spruce Streets, West Reading, PA 19612, 610.988.8133, donatoa@readinghospital.org

Received September 9, 2014; revision received January 23, 2015; accepted February 11, 2015.

**DOI:** <http://dx.doi.org/10.4300/JGME-D-14-00532.1>

In a study in which faculty rated scripted videos of trainees, the Minicard produced more accurate detection of unsatisfactory performances and identified more specific behaviors than the commonly used Mini-CEX direct observation tool, making the Minicard potentially useful as a formative tool.<sup>10</sup> To this point, validity evidence has not been generated to support the interpretation of the Minicard scores. Our project sought to begin to fill this gap.

In this study, we explored the validity and feasibility of the Minicard as implemented in our institution. In particular we sought to determine (1) whether a sufficient number of observations could be collected from a variety of settings by multiple raters, and (2) whether high-quality learner action plans would be generated. We used Mes-sick's<sup>11</sup> validity framework to organize the analysis and results, focusing primarily on collecting evidence for content, response process, relationships to other variables, and consequential validity.

## Methods

We conducted a retrospective cohort analysis of 6 years of direct observations in 1 institution. We sought validity evidence, including content, response process, relationships to other variables, consequences, and feasibility (defined as the financial and opportunity time), as well as costs to implement the Minicard program.

The Minicard is organized into 4 sections that represent commonly observed resident activities: obtaining a history, performing a physical examination, oral presentation of a patient case, and counseling or discussion of findings with the patient (instrument provided as online supplemental material).<sup>10</sup> The development and blueprinting of the Minicard has been described in a prior publication.<sup>10</sup> Each activity provides the clinical context for assessing 3 competency domains: (1) interpersonal communication; (2) medical knowledge in the context of patient care, referred to on the instrument as simply "medical knowledge"; and (3) professionalism. Prompts representing best practices cue observers in each domain and may be used to record the presence or absence of specific behaviors during the observation. Behavioral anchors are given for each of the 4 nominal 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.

## Setting

The study was conducted in an internal medicine residency at Reading Hospital and Medical Center, an independent academic medical center in Pennsylvania. The program comprised 21 internal medicine residents each year during the 6-year study period. Approximately one-third of training took place in the ambulatory setting. Only

### What was known and gap

Faculty need reliable, valid, and feasible tools for workplace-based assessments of residents.

### What is new

A study to assess the reliability, validity, and feasibility of a direct observation tool (the Minicard) in internal medicine residents.

### Limitations

Single site, single specialty study limits generalizability; incentives provided may reduce feasibility.

### Bottom line

The Minicard was able to identify "struggling" residents, making it a useful tool for formative assessment and feedback.

residents with complete 3-year data sets were included in the study. Residents who were in midtraining at the start of the Minicard program and residents who joined the residency after their first year were excluded.

## Minicard Implementation

The Minicard program was piloted in 2005–2007 with full-time general internal medicine faculty, and it was rolled out to physicians in 6 internal medicine subspecialties over the next 4 years. In the final 4 years of the study, financial incentives representing 20% of total compensation include educational metrics that represented 20% of the overall bonus (ie, 4% of total compensation). Completion of 1 direct observation per learner per week was 1 of 3 educational metrics that counted toward the bonus. Attending physicians selected the setting of their direct observation, were encouraged to record observations as they watched and to give verbal feedback (but not ratings), and discussed the action plan immediately after each observation. Minicards were submitted to a central office and recorded on an electronic spreadsheet. Learners received electronic copies of action plans triannually, and reflected on these in their electronic portfolios. Faculty mentors reviewed all Minicards and learner reflections and met with individual residents triannually to discuss learning goals and interventions.

## Minicard Training

Faculty raters were trained in small groups of 1 to 5, in 1-hour sessions led by a single instructor (A.A.D.). The training included a 10-minute introduction to the Minicard followed by 3 video vignettes of resident patient care encounters.<sup>10,12</sup> After this, participants discussed ratings, observations, and action plans. Beginning in 2011, faculty members were given annual report cards indicating their average scores relative to the overall group average, along

with the percentage of “action-oriented” feedback written on their action plans.

Ethical approval for this project was obtained from the Reading Hospital and Medical Center’s Institutional Review Board.

### Data Analysis (Validity Evidence Sought)

**Content** The extent of Minicard use was reported for each of 4 settings (inpatient ward, emergency department, clinic, and intensive care unit) during the last 2 years of the study (2009–2011), by which time the Minicard was fully deployed to generalists and subspecialists. The number of Minicards and independent observers were calculated for each resident and reported by year.

**Response Process** Response process was assessed by (1) reporting the distribution of Minicard scores by year and (2) identifying the number of specific behaviors recorded.

**Relationship to Other Variables** Changes in individual scores by month in training were analyzed using a mixed-effects linear regression.

**Consequential Validity** Educational impact was indirectly explored through qualitative analysis of recorded action plans. Two investigators (A.A.D. and D.L.G.) were trained in using a predetermined coding schema for feedback,<sup>13</sup> coding each written plan as “action-oriented” (eg, “next time, set an agenda first”); “observational” (eg, “he was terse with the patient”); “minimal” (eg, “good resident”); or “none/blank.” Action plans containing 2 levels of action were coded using the higher plan. These investigators also coded the Minicard domain targeted by the action-oriented plans as “communication,” “applied medical knowledge,” or “professionalism.” Double coding was performed for 20% of all observations to assess interrater reliability, and discrepancies were discussed until agreement was reached.

The feasibility of 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, and reproduction costs).

## Results

A total of 3715 Minicards were collected and analyzed over a 6-year period (2005–2011). Eighty physician faculty raters (30 generalists and 50 subspecialists from 6 subspecialties) observed 73 residents during this period.

### Content Sampling

The most commonly reported setting for observation was the inpatient ward (43%, 1597 of 3715), followed by the clinic (39%, 1446 of 3715), the intensive care unit (15%, 555 of 3715), and the emergency department (3%, 110 of 3715), with less than 1% (7 of 3715) of observations

unidentified. Raters recorded observations in the history section of the Minicard in 30% (1115 of 3715) of the encounters, in the physical examination section in 23% (855 of 3715) of the encounters, in the oral presentation section in 52% (1931 of 3715), and in the counseling section in 27% (1003 of 3715) of the encounters.

Residents were observed an average of 28 times per year (SD = 8.4, range, 15–45). There were on average 9 independent observers per resident per year (SD = 4.1, range 3–20), and 15 independent observers (SD = 9.0, range, 3–38) across the 3 years of residency (TABLE).

### Response Process

An average of 5 (SD = 5.1) prompts were checked per Minicard, with each representing a discrete rater comment (range 0–20). About 8% (295 of 3715) of Minicards included ratings only, with no checked prompts or action plan.

### Relationship to Other Variables

Minicard scores increased by 0.021 points per month of training (95% CI 0.019–0.024;  $P < .001$ ; FIGURE 1). Observers most often rated first-year residents as “good” (56% of ratings, 4130 of 7345), and used the “marginal” rating for 8% (621 of 7345) of first-year resident observations, whereas they most often rated third-year residents as “excellent” (67% of ratings, 1851 of 2758), and used “marginal” ratings only 2% (56 of 2758) of the time (FIGURE 2).

### Educational Impact

Eighty action plans were double coded; exact agreement between coders was 93% ( $\kappa = 0.89$ ) for the nature of the action plan, and 91% ( $\kappa = 0.82$ ) for action plan domain. Action-oriented feedback was recorded for 50% (1851 of 3715) of Minicards, observational feedback for 11% (403 of 3715), minimal feedback for 9% (337 of 3715), and no recorded plan for 30% (1124 of 3715). When an action plan was recorded, the plan related to an applied medical knowledge deficiency in 56% (1030 of 1851) of cases, communication deficiencies in 44% (809 of 1851), and professionalism in about 1% (12 of 1851) of cases. Observers documented that they gave verbal feedback in 74% (2749 of 3715) of the encounters.

### Feasibility

The average observation duration of an encounter (not including feedback) was 15.6 minutes (SD = 9.5; range 1–120 minutes; interquartile range 10–20 minutes). Cost of reproduction of the cards was \$0.022 per card, or \$0.61 per resident per year. An administrative assistant spent 2 minutes recording each Minicard, or approximately 29 minutes per week to record comments (checked

| TABLE                                                         | DESCRIPTIVE STATISTICS OF OBSERVATIONS BY ACADEMIC YEAR, 2005–2011 |           |           |           |           |           |
|---------------------------------------------------------------|--------------------------------------------------------------------|-----------|-----------|-----------|-----------|-----------|
| Academic Year                                                 | 2005–2006                                                          | 2006–2007 | 2007–2008 | 2008–2009 | 2009–2010 | 2010–2011 |
| No. of residents observed                                     | 9                                                                  | 15        | 24        | 25        | 27        | 28        |
| No. of faculty members performing observations                | 9                                                                  | 9         | 18        | 41        | 75        | 56        |
| No. of subspecialties performing observations (including GIM) | 1                                                                  | 1         | 2         | 5         | 7         | 7         |
| Total observations                                            | 199                                                                | 421       | 603       | 806       | 868       | 818       |
| Mean (SD) direct observations per resident per year           | 22.3 (7)                                                           | 27.6 (8)  | 25.2 (6)  | 31.0 (12) | 33.3 (10) | 30.8 (8)  |
| Mean (SD) unique observers per resident                       | 5.3 (1)                                                            | 4.8 (2)   | 6.4 (3)   | 9.7 (3)   | 15.1 (5)  | 14.9 (4)  |
| Mean (SD) discrete recorded comments per encounter            | 6.8 (5)                                                            | 5.0 (4)   | 4.9 (4)   | 5.0 (4)   | 3.8 (5)   | 5.1 (7)   |
| Percentage of observations with action plans recorded         | 74.2                                                               | 76.7      | 75.6      | 77.6      | 66.8      | 78.0      |

Abbreviation: GIM, general internal medicine.

prompts) and action plans on an electronic spreadsheet. Total incentive salary at risk for all educational metrics was approximately \$2,000 per attending for each 6-month period. Completion of 1 Minicard per learner per week was 1 of 3 educational metrics used to calculate incentive pay.

## Discussion

This paper presents validity and feasibility evidence to support the use of Minicard ratings for the assessment of the competency of internal medicine residents. The Minicard was used in venues that represent the breadth of resident practice; faculty raters regularly used the majority

of the rating range, and scores increased over the course of training.

The Minicard was particularly effective in identifying “struggling” residents. Minicard ratings included 659 of 7345 (9%) unsatisfactory ratings for first-year residents. In comparison, Mini-CEX unsatisfactory ratings are rarely used: 0 of 1280 in third-year medical students,<sup>14</sup> 1 of 196 resident physicians,<sup>15</sup> 0 of 388 first-year medicine residents,<sup>16</sup> and 0 of 107 first-year medicine residents.<sup>13</sup> This is consistent with studies that have found that narrative descriptors (like the prompts and behavioral anchors in the Minicard) are better in identifying “struggling” learners’ marginal performance than numerically based assessment tools.<sup>17</sup>

The Minicard generated action-oriented feedback in 50% (1851 of 3715) of encounters, compared with the Mini-CEX, where action-oriented feedback was recorded

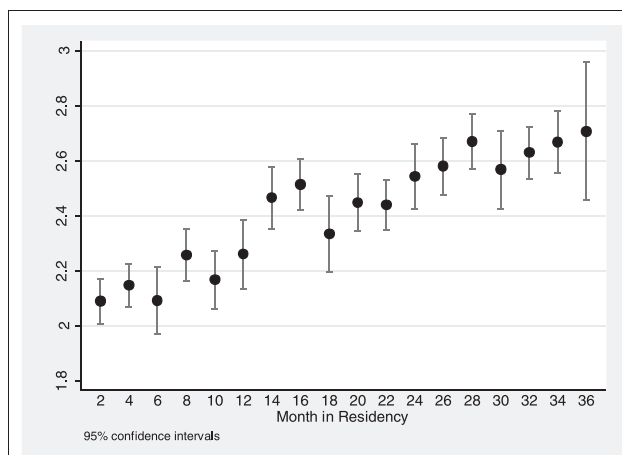


FIGURE 1 AVERAGE MINICARD SCORE CHANGES BY MONTH

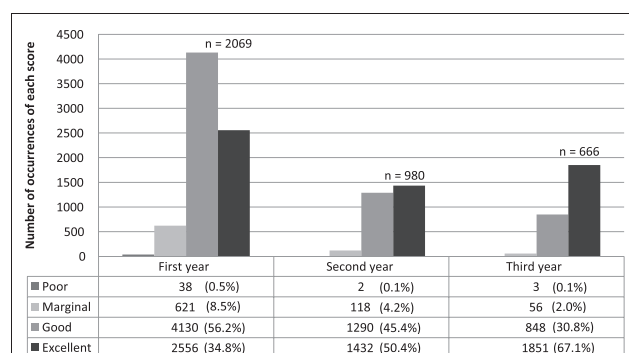


FIGURE 2 MINICARD SCORE DISTRIBUTIONS BY YEAR

only 8% of the time.<sup>13</sup> However, despite incentives and feedback to observers, 30% (1124 of 3715) of Minicards were completed without an action plan.

By using the embedded prompts, Minicard raters also recorded an average of 5 specific comments per encounter, each denoting the presence or absence of a microskill. These comments may be critical for program directors tasked with matching performance to the narratives of the Accreditation Council for Graduate Medical Education milestone goals, although the accuracy of each recorded finding was not able to be confirmed.

This study had several limitations. It was performed in a single internal medicine program, limiting the generalizability of the results. The validity evidence presented represents only a part of the potential evidence that could be collected to challenge the interpretation of scores received on the Minicard. The addition of an incentive program during the last 4 years of the study may confound feasibility assumptions in residency programs attempting this intervention without incentives. However, performance-based incentive plans have become more common in academic centers.<sup>18</sup> Action plans included almost no comments on professionalism, and professionalism scores were nearly universally in the acceptable range. It is possible that direct observation by supervising staff may have suppressed residents' unprofessional behavior. Severity error was identified in a prior study of the Minicard.<sup>10</sup> Although the Minicard is used primarily for formative feedback, the unintended consequences of potentially more negative ratings on learners are unknown. While a significant amount of recorded feedback was coded as action oriented, it is not known how this feedback was received and acted on by the learners, and understanding the effectiveness of this tool on changing learners' behavior is an important avenue for future study.

## Conclusion

The Minicard direct observation tool may efficiently generate useful feedback and specific descriptions of learners' behaviors in real-world clinical settings. This information can support the identification and remediation

of struggling residents. Residency program directors tasked with generating action-oriented feedback as well as capturing specific details of resident performances should consider incorporating the Minicard as 1 element of their assessment system.

## References

- 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.
- Weinberger SE, Pereira AG, Iobst WF, Mechaber AJ, Bronze MS. Alliance for academic internal medicine education redesign task force, II: competency-based education and training in internal medicine. *Ann Intern Med*. 2010;153(11):751–756.
- Schuwirth LWT, Van Der Vleuten CP. Changing education, changing assessment, changing research? *Med Educ*. 2004;38(8):805–812.
- Shepard LA. The role of assessment in a learning culture. *Educ Res*. 2000;29(7):4–14.
- van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, et al. A model for programmatic assessment fit for purpose. *Med Teach*. 2012;34(3):205–214.
- 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.
- Margolis MJ, Clauser BE, Cuddy MM, Ciccone A, Mee J, Harik P, 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(suppl 10):56–60.
- 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. *Med Educ*. 2012;46(1):28–37.
- Kane MT. The assessment of professional competence. *Eval Health Prof*. 1992;15(2):163–182.
- Donato AA, Pangaro L, Smith C, Rencic J, Diaz Y, Mensinger J, et al. Evaluation of a novel assessment form for observing medical residents: a randomised, controlled trial. *Med Educ*. 2008;42(12):1234–1242.
- Messick S. *Educational Measurement*. 3rd ed. New York, NY: American Council on Education and Research; 1989.
- 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.
- 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.
- Kogan JR, Bellini LM, Shea JA. Implementation of the mini-CEX to evaluate medical students' clinical skills. *Acad Med*. 2002;77(11):1156–1157.
- 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.
- 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.
- Regehr G, Regehr C, Bogo M, Power R. Can we build a better mousetrap? Improving the measures of practice performance in the field practicum. *J Soc Work Educ*. 2007;43(2):327–343.
- Reece EA, Nugent O, Wheeler RP, Smith CW, Hough AJ, Winter C. Adapting industry-style business model to academia in a system of performance-based incentive compensation. *Acad Med*. 2008;83(1):76–84.