

# Validity Evidence for the Medical Student Core Physical Examination

Todd Guth (✉ [todd.guth@CUAnschutz.edu](mailto:todd.guth@CUAnschutz.edu))

University of Colorado Anschutz Medical Campus

Yoon Soo Park

University of Illinois at Chicago

Janice Hanson

Saint Louis University

Rachel Yudkowsky

University of Illinois at Chicago

---

## Research Article

**Keywords:** validity evidence, physical examination, clinical skills assessment, medical education

**Posted Date:** September 28th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-699718/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

## Background

The Core Physical Exam (CPE) has been proposed as a set of key physical exam (PE) items for teaching and assessing PE skills in medical students, and as the basis of a Core + Cluster curriculum. Beyond the initial development of the CPE and proposal of the CPE and the Core + Cluster curriculum, no additional validity evidence has been presented for use of the CPE to teach or assess PE skills of medical students. As a result, a modified version of the CPE was developed by faculty at the University of Colorado School of Medicine (UCSOM) and implemented in the school's clinical skills course in the context of an evolving Core + Cluster curriculum.

## Methods

Validity evidence for the 25-item University of Colorado School of Medicine (UCSOM) CPE was analyzed using longitudinal assessment data from 366 medical students (Classes of 2019 and 2020), obtained from September 2015 through December 2019. Using Messick's unified validity framework, validity evidence specific to content, response process, internal structure, relationship to other variables, and consequences was gathered.

## Results

Content and response process validity evidence included expert content review and rater training. For internal structure, a generalizability study phi coefficient of 0.258 suggests low reliability for a single assessment due to variability in learner performance by occasion and CPE items. Correlations of performance on the UCSOM CPE with other PE assessments were low, ranging from .00-.34. Consequences were explored through determination of a pass-fail cut score. Following a modified Angoff process, clinical skills course directors selected a consensus pass-fail cut score of 80% as a defensible and practical threshold for entry into precepted clinical experiences.

## Conclusions

Validity evidence supports the use of the UCSOM CPE as an instructional strategy for teaching PE skills and as a formative assessment of readiness for precepted clinical experiences. The low generalizability coefficient suggests that inferences about PE skills based on the UCSOM CPE alone should be made with caution, and that the UCSOM CPE in isolation should be used primarily as a formative assessment.

## Introduction

Teaching strategies and methodologies to promote medical students' physical examination (PE) skills include the Head-to-Toe Physical Examination, Hypothesis Driven Physical Examination, and Core + Cluster Physical Examination.<sup>1,2,3</sup> The strengths and limitations of these approaches have been debated in recently published literature with no single best methodology for teaching PE emerging.<sup>4</sup>

The Head-To-Toe physical examination (HTT), a screening PE comprised of about 140 maneuvers performed on Standardized Patients (SPs), has been used to assess the acquisition of foundational PE skills prior to entering clerkships. The HTT has been shown to be useful as a summative assessment of PE skills prior to entering clerkships.<sup>1</sup> The Hypothesis-Driven Physical Examination (HDPE) was developed to promote critical thinking related to the PE in the context of diagnostic challenges within a patient presentation.<sup>2</sup> The HDPE approach provides students with targeted practice in anticipating, eliciting, and interpreting PE maneuvers in the context of patient cases with focused diagnostic challenges. Unlike the HTT method, the HDPE has the potential to promote the development of clinical reasoning through targeted practice and feedback in the process of the selection and interpretation of PE maneuvers using patient based diagnostic challenges.<sup>2</sup>

More recently, the Core Physical Examination (CPE), has been promoted as an instructional and assessment methodology for teaching PE skills as part of a Core + Cluster curriculum.<sup>5</sup> The CPE consists of 37 key PE items based on a survey of internal medicine clerkship directors and clinical skills course directors.<sup>3</sup> Advocates of the CPE intend that the CPE maneuvers should be taught in combination with symptom-driven Clusters of additional PE maneuvers. Beyond the initial development of the CPE and proposal of the CPE and the Core + Cluster curriculum by Gowda et al, no additional validity evidence has been presented for use of the CPE to assess PE skills of medical students.

A modified version of the CPE was developed by faculty at the University of Colorado School of Medicine (UCSOM) and implemented in the school's clinical skills course in the context of an evolving Core + Cluster curriculum. The purpose of this study was to provide initial validity evidence for the use of the UCSOM CPE, as an exemplar of the CPE approach, in the assessment of PE skills in medical students. We gather validity evidence using Messick's unified validity framework in the methods and use Kane's validity framework to make arguments on score inference and interpretation in the discussion.

## Methods

# Instructional Methods and Assessments

The UCSOM PE curriculum is a body system-based curriculum. The CPE was introduced as the first step towards a Core + Cluster curriculum. The UCSOM CPE consists of a 25-item subset of the originally published 37-item CPE;<sup>3</sup> the remaining 12 items are taught at UCSOM in the context of body-system maneuvers. Most of the remaining 12 items are neurologic items and these items were not included in the first iteration of the CPE as they were not taught until the second year of training when the basic science and anatomy of the neurologic system were introduced. The UCSOM PE curriculum is implemented by PE

teaching assistants (who serve as SPs and raters for the clinical performance center), senior medical students, and clinical faculty in the pre-clinical years in the context of a clinical skills course. See Appendix A for a detailed listing of the UCSOM CPE items and scoring criteria.

In the first year of medical school, UCSOM students learn six complete body systems: head and neck, pulmonary, cardiovascular, abdominal, upper musculoskeletal and lower musculoskeletal body systems. In the second year, the neurologic body system is taught. Each set of body system PE maneuvers is made up of a subset of items contained within the UCSOM CPE plus additional PE maneuvers. The total number of items contained within the six body systems is similar in scope to most versions of the HTT with a total of 104 PE items taught at the UCSOM. Later in the first year, the UCSOM CPE is taught as a 25-item cohesive subset of PE maneuvers that borrows from all six body systems. Appendix A includes the CPE items with instructions that students are provided. UCSOM students are assessed on their clinical skills in each semester during their clinical skills course. In the first two years of training, clinical skills PE assessments emphasize either body systems or the UCSOM CPE. In the third year, a single 10-station clinical skills assessment emphasizes the selection and performance of PE maneuvers in the context of a series of clinical cases.

Table 1 summarizes descriptive information about clinical skills assessments at the UCSOM over the first three years of the curriculum. Students were assessed on three out of the six complete body systems in a PE-only objective structured clinical exam (OSCE) in the fall of their first year (M1-Fall: Systems) as a resource-saving approach to PE skill retention.<sup>6</sup> Students were assessed on the full UCSOM CPE in the spring of first year (M1-Spring: CPE) in the context of a comprehensive medical encounter. In the fall of second year (M2-Fall: Neuro), students were assessed on the neurologic body system examination in the context of a focused medical encounter. In the spring of second year (M2-Spring: CPE+), students were instructed to perform a full UCSOM CPE plus 7 additional items from the abdominal body system as part of a comprehensive medical encounter of a patient presenting with an abdominal complaint. In the spring of the third year of medical school near the end of their core clinical rotations, students participated in a comprehensive performance assessment, an Objective Structured Clinical Exam (OSCE) comprised of 10 focused medical encounters (M3-Spring: OSCE) designed to prepare students for the United States Medical Examination Licensing Examination Step 2 Clinical Skills assessment. Each encounter checklist of the M3-Spring: OSCE assessment includes a selected number of UCSOM CPE items and additional PE items based on the presenting symptoms of the patient. PE items were scored as performed, performed incorrectly (half credit), or not performed. One of the cases was a telephone only case and did not include any PE items.

Table 1  
Summary of UCSOM Clinical Skills Assessments Detailing Number of Core and Non-Core Physical Examination Items

Sequence of Assessments	Exam Content	# UCSOM CPE Items	# Additional PE Items	Mean Percentage (SD)	
				Class of 2019	Class of 2020
<b>M1-Fall: Systems*</b>	Cluster 1			<b>92.5 (5.4)</b>	<b>89.9 (5.2)</b>
	- Head and Neck	6	14		
	- Pulmonary	4	8		
	- Upper Musculoskeletal	1	16		
	<b>Total</b>	<b>11</b>	<b>38</b>		
	Cluster 2				
	- Abdominal	4	10		
	- Cardiovascular	7	6		
	- Lower Musculoskeletal	1	18		
	<b>Total</b>	<b>12</b>	<b>34</b>		
<b>M1-Spring: CPE</b>	Comprehensive Medical Encounter: UCSOM CPE items only	<b>25</b>	<b>0</b>	<b>94.8 (5.6)</b>	<b>95.7 (4.8)</b>
<b>M2-Fall: Neuro</b>	Focused Medical Encounter: Neurologic Body System PE items only	<b>0</b>	<b>15</b>	<b>95.6 (4.6)</b>	<b>91.4 (7.5)</b>
<b>M2-Spring: CPE</b>	Comprehensive Medical Encounter: UCSOM CPE items only; additional abdominal PE items not included in analysis	<b>25</b>	<b>0</b>	<b>90.3 (7.5)</b>	<b>91.2 (7.9)</b>
<b>M3-Spring: OSCE</b>	Ten Focused Medical Encounters: Various UCSOM PE items and Additional Items	<b>13</b>	<b>16</b>	<b>68.1 (9.1)</b>	<b>66.0 (8.8)</b>

\* Students are tested on either Cluster 1 or Cluster 2 (3 out of 6 body systems) for the M1 Fall Systems assessment.

## Study Participants

This study was conducted with longitudinal cohort data from clinical skills assessments of the medical student Classes of 2019 and 2020 during the first three years of training of each cohort. Data collection

occurred from September 2015 through December 2019.

The study was considered exempt by the University of Colorado and University of Illinois Chicago institutional review boards.

## Validity Evidence

Validity evidence was sought based on Messick's unified validity framework.<sup>7</sup>

**Content evidence** was obtained by reviewing the process of developing and selecting the items of the UCSOM CPE.

**Response process evidence** was based on review of the materials provided to students and of the quality assurance processes related to scoring the assessment.

**Internal structure.** Reliability estimates were obtained through a Generalizability (G) study using G-String IV (Hamilton, Canada) across the M1-Spring: CPE assessment and the M2-Spring: CPE assessments for the class of 2020.<sup>8</sup> Persons (p) were the objects of measurement, items (i) were fixed (the set of CPE items assessed), and the occasion (o) for the assessment was considered random (M1-Spring: CPE and M2-Spring: CPE). The design was fully-crossed, person (p) crossed with UCSOM CPE items (i) and occasion (o). Raters were not included as rater data were not available.

**Relationship to other variables.** Spearman correlation coefficients were calculated to measure associations between the five clinical skills assessments across the first three years of the curriculum. Spearman correlations were performed in lieu of Pearson correlations as the results of the assessments were not normally distributed given the high overall means for PE performance.

**Consequences.** The consequences of establishing pass-fail cut scores at the UCSOM using normative standards (1.5 or 2 standard deviations (SD) below the mean), clinical course director determined consensus scores, and an item-level, modified Angoff score were explored. Historically, pass-fail cut scores for assessments had been established as either clinical skills course director determined consensus pass-fail cut scores (80% or 75%) or normative determined pass-fail cut scores. For the UCSOM CPE, an item-level, modified Angoff standard setting exercise was conducted with 8 faculty including 2 clinical preceptors, 2 clinical block directors, and 4 clinical skills course directors. The experts were asked to estimate the percentage of borderline students who would correctly perform each item. The borderline student was defined as a minimally competent student to enter into supervised practice with an individual preceptor in their clinical practice setting. Prior to the start of the standard setting process, judgments were informed by performance data from the initial M1-Spring: CPE assessment. The pass-fail cut score was determined following two iterations of discussion at the item level.<sup>9</sup>

## Results

Performance data were obtained for 366 students comprising the medical student Classes of 2019 (N = 182) and 2020 (N = 184).

**Content Evidence.** The 25-item UCSOM CPE was developed by five content experts, including clinical block directors and clinical skills course directors, through discussion and consensus. The CPE comprises a subset of PE items taught in the UCSOM clinical skills course. The content experts were asked to develop the UCSOM CPE as an assessment of PE skill in medical students in relation to curriculum goals and learning objectives, appropriateness for starting clinical experiences, and relevance to clinical practice. Following a presentation of the UCSOM CPE, the entire group of approximately 20 UCSOM clinical block directors achieved consensus and voted to permit the UCSOM CPE to serve as the essential PE maneuvers of a comprehensive medical encounter. The clinical block directors also approved the item performance instructions in the scoring rubric, and agreed that the UCSOM CPE would be appropriate for clinical rotations as well as preceptorship experiences.

**Response Process.** Assessment materials created by the clinical skills course directors included the SP case, scoring rubrics, and instructions for students, SPs, and raters. The clinical performance center training process involved both a 4-hour SP and rater training session and 4-hour SP portrayal and rater practice session. A subset of all ratings for each SP and rater were reviewed in real time by another rater as a quality check of rater performance. Expert raters, in a blinded fashion, re-watched and re-scored videos of all the borderline and failing students, corrected any errors in the initial rater scoring, and provided feedback to the raters for any errors identified.

**Internal Structure.** Descriptive statistics (means and standard deviations) for each assessment across the first three years of medical training are shown in Table 1. The results of the G study for the Class of 2020 are shown in Table 2. The largest contributors to score variability were the person-occasion (5.5%) and person-item (5.4%) interactions, indicating variability in learner performance depending on occasion and on specific items performed, respectively. The overall phi coefficient reliability was 0.258 and the G coefficient reliability was 0.308. Decision (D) studies determined that increasing the number of iterations of assessing the UCSOM CPE to six occasions would increase the phi coefficient to 0.486. Increasing the number of items in the UCSOM CPE to 37 items (similar to the published CPE) would increase the phi coefficient to 0.281.

Table 2

Generalizability Study<sup>a</sup> of UCSOM Core Physical Exam Assessment, Class of 2020

Class of 2020			
Effect	Degrees of Freedom	Variance Component	% Variance Component
p <sup>b</sup>	182	0.001	2
o <sup>c</sup>	1	0.001	2.02
i <sup>d</sup>	24	0.001	1.97
po	182	0.002	5.53
pi	4368	0.002	5.41
io	24	0.001	1.58
poi/e	4368	0.036	82.8

The overall phi coefficient was 0.258 and the G coefficient was 0.308.

Notes:

<sup>a</sup> Completely crossed design, p x o x i

<sup>b</sup> Person (p), the object of measurement

<sup>c</sup> Occasion (o) = Spring M1 and Spring M2; a random facet

<sup>d</sup> Items (i) = the 25 items of the CPE, a fixed facet

**Relationship to other variables.** Spearman correlations between the UCSOM clinical skills assessments are detailed in Table 3. The UCSOM CPE assessments showed low correlations to the assessments during years one and two and the M3-Spring: OSCE. Correlations between the M1-Spring: CPE assessment to the M2-Spring: CPE assessment, both of which contain all UCSOM CPE items, were generally higher than correlations to the body system assessments.

Table 3

Relationships to Other Variables: Spearman Correlations Between Assessments for the Physical Examination Assessments by Class

<b>Class of 2019</b>	<b>M1-Fall: Systems</b>	<b>M1-Spring: CPE</b>	<b>M2-Fall: Neuro</b>	<b>M2-Spring: CPE</b>
M1-Fall: Systems N = 182				
M1-Spring: CPE N = 181	0.14 P = .05 N = 179			
M2-Fall: Neuro N = 181	0.16 P = .03 N = 179	0.08 P = .28 N = 179		
M2-Spring: CPE N = 180	0.20 P < .01 N = 175	.13 P = .08 N = 176	0.20 P < .01 N = 177	
M3-Spring: OSCE N = 173	0.20 P = .02 N = 150	0.22 P < .01 N = 150	0.08 P = .31 N = 150	0.08 P = .40 N = 147
<b>Class of 2020</b>	<b>M1-Fall: Systems</b>	<b>M1-Spring: CPE</b>	<b>M2-Fall: Neuro</b>	<b>M2-Spring: CPE</b>
M1-Fall: Systems N = 184				
M1-Spring: CPE N = 183	0.14 P = .05 N = 183			
M2-Fall: Neuro N = 184	0.12 P = .11 N = 183	0.18 P = .06 N = 183		

Class of 2019	M1-Fall: Systems	M1-Spring: CPE	M2-Fall: Neuro	M2-Spring: CPE
M2-Spring: CPE N = 184	0.16 P = .03 N = 183	.23 P < .01 N = 183	0.34 P < .01 N = 184	
M3-Spring: OSCE N = 171	0.07 P = .39 N = 148	0.00 P = .99 N = 148	0.06 P = .49 N = 148	0.19 P = .02 N = 148

Note: Associated p values and numbers of students included in the correlations are included below the correlation. N's are variable because of variable student schedules and progress.

**Consequences.** The outcome of the modified Angoff pass-fail score determination was 90%, which would have resulted in a failure rate of 10–13% in the M1-Spring CPE and 36–39% for the M2-Spring CPE. Failure rates for the 1.5 SD below the mean pass-fail cut score were in the range of 6–8% in the M1-Spring CPE and 5–10% for the M2-Spring CPE. Failure rates for the 80% consensus pass-fail cut score were in the range of 1–2% in the M1-Spring CPE and 8–10% for the M2-Spring CPE. Table 4 shows the numbers of students who would have failed each year based upon the various pass-fail cut scores.

Table 4  
Impact of Standard Setting for UCSOM CPE in M1 Spring CPE and M2 Spring CPE Assessments

M1 Spring CPE			M2 Spring CPE								
Class of 2019			Class of 2020			Class of 2019			Class of 2020		
N = 181			N = 183			N = 181			N = 185		
Mean(SD) = 94.8(5.6)			Mean(SD) = 95.7(4.8)			Mean(SD) = 90.3(7.6)			Mean(SD) = 91.2(7.8)		
Standard Setting Method	Cut Score	Number of Failures (%)	Cut Score	Number of Failures (%)	Cut Score	Number of Failures (%)	Cut Score	Number of Failures (%)	Cut Score	Number of Failures (%)	
Modified Angoff	90%	24 (13%)	90%	18 (10%)	90%	66 (36%)	90%	69 (38%)	90%	69 (38%)	
1.5 SD Below the Mean	86%	11 (6%)	88%	14 (8%)	79%	9 (5%)	80%	18 (10%)	80%	18 (10%)	
2 SD Below the Mean	84%	9 (5%)	86%	10 (5%)	75%	4 (2%)	76%	6 (3%)	76%	6 (3%)	
Consensus cut score: 80%	80%	5 (2%)	80%	2 (1%)	80%	14 (8%)	80%	18 (10%)	80%	18 (10%)	

## Discussion

This paper presents initial validity evidence for the use of the UCSOM CPE as an assessment of PE competence in medical students. In the results we detailed the validity evidence obtained from each of Messick’s five sources. Here we interpret validity evidence gathered using Messick’s sources using Kane’s validity framework to summarize the inference of CPE assessment scores and the overall validity argument for the use of the CPE as an assessment of PE competence in medical students and identify evidence gaps. The argument follows a stepwise approach through each of the four inferences in Kane’s validity framework—scoring, generalization, extrapolation, and implications.<sup>10</sup>

The *scoring* inference (translating an observation into a score) was supported by expert review of UCSOM CPE items and ongoing quality assurance processes in the clinical performance center. Reliance on these quality assurance processes as an assessment of response process requires that quality assurance processes occur annually to ensure acceptable rater performance and training. Formal evaluation of inter-rater reliability of both the real-time quality checks and the video reviews of the borderline and failing students would further strengthen this inference. High overall means of PE performance across the assessments in the first two years of medical training suggest that students are able to perform recently learned PE skills in a clinical performance center assessment setting.

The *generalization* inference involves the extent to which a score on a given assessment is representative of performance in a testing setting. The overall phi coefficient for the G study of 0.258 suggests low reliability for a single assessment, and that sources of error not considered as facets in this G study contribute significantly to the variance in scores. The largest contributors to score variability were the person-occasion (5.5%) and person-item (5.4%) interactions, indicating that individual student performance varied by occasion (from Spring M1 to Spring M2) and across items. Learner-specific factors contributing to score differences between occasions may include learning or unlearning (decay) of PE skills from Spring of M1 to Spring of M2 and/or changes in the motivation of individual learners to prepare for the assessment. Rater-specific factors related to the scoring of specific items are another likely source of error contributing to variance. The low generalizability coefficient suggests that inferences about PE skills based on the UCSOM CPE alone should be made with caution, and that the UCSOM CPE in isolation should be used primarily as a formative assessment.

The *extrapolation* inference relates to using the score as a predictor of real-world performance.

Absent performance measures in clinical settings, the relationship of UCSOM CPE scores to other assessments of PE skills may provide some indication of the transfer of skills beyond the UCSOM CPE. The correlations between the various UCSOM clinical skills assessments are similar to the correlations between cases of an OSCE, which has been shown to be in the range of 0.1 to 0.3 between stations.<sup>11,12</sup> These correlations between different PE assessments are consistent with case specificity, since each of the system-based assessments included different subsets of CPE and non-CPE items. Correlations between performance on these PE assessments and measures of PE skill during clerkships would strengthen this inference.

The relatively low M3 OSCE PE scores are striking, and consistent with low PE scores in other studies of end-of-M3 OSCEs.<sup>13,14,15</sup> The lower scores in M3 may be due to a decay in PE skills during clinical clerkships; alternatively, M3 students may know how to perform the maneuvers but struggle to select the appropriate PE items to perform in a given encounter, indicating the need for more practice in clinical reasoning. This could be addressed by adding a hypothesis-driven (i.e. HDPE) component to PE instruction, to provide practice in the use of the PE in the service of accurate diagnosis of the patient.<sup>2</sup>

The *implication* inference (applying the score to inform a decision) was probed by exploring the impact of different passing standards. The consensus pass-fail cut scores and the normative cut scores were significantly lower than the cut score established using the modified Angoff procedure, which resulted in an unacceptably high failure rate. Our experience at UCSOM suggests that the Angoff 90% cut score may represent the PE competence of a student who is *well prepared* for entering into supervised clinical practice in the clinical clerkships, rather than that of the targeted *minimally competent* or borderline student who is preparing for entry into a pre-clinical supervised preceptorship. Repeating the Angoff exercise after a more detailed discussion of the target student and the intended inference might correct this disjunct. In the meantime, the clinical skills course directors considered the 80% consensus pass-fail cut score for entry into supervised practice within a clinical preceptorship experience as both defensible,

because of the lack of high correlations of the CPE to other PE competence assessments, and practical, because of the costs involved in remediating large numbers of students.

Next steps in the evolution of the UCSOM PE curriculum towards a Core + Cluster curriculum include a transition away from body systems to specific PE clusters related to specific chief complaints with a continuing emphasis on the UCSOM CPE. Additional assessments of PE performance should be considered in clerkships experiences. Composite reliability of PE assessments across the pre-clinical and clinical years may allow for high stakes decisions to be made about PE performance. Programmatic assessment is an emerging approach to assessment in which multiple assessments over time may be combined to make high stakes decisions about advancement and promotions.<sup>16</sup> Scholarly work related to the development and incorporation of PE clusters as part of the PE curriculum for the Core + Cluster curricula at UCSOM, an expansion of the items included in the UCSOM CPE, and correlations to additional variables, such as performance in clerkships or to the USMLE Step 2 Clinical Skills assessment would be reasonable next steps in the evolving considerations for teaching and assessing PE competence in medical students using a programmatic approach.

## **Limitations**

Our study uses the UCSOM CPE as a representative exemplar of the Core Physical Exam approach to teaching and assessing the PE. As the UCSOM CPE is an institution-specific, 25-item version of the published 37-item CPE, the applicability and generalizability of these results to the full CPE and to other settings is limited. The low generalizability coefficient suggests that inferences about PE skills based on the UCSOM CPE should be made with caution.

## **Conclusion**

This paper presents the initial argument for the use of UCSOM CPE in the assessment of the PE skills of pre-clinical medical students. Initial validity evidence supports the use of the UCSOM CPE as an instructional strategy for teaching medical students physical examination skills and as a formative assessment of physical exam skills in readiness for precepted clinical experiences.

## **Abbreviations**

*Core Physical Exam (CPE)*

*Physical Exam (PE)*

*University of Colorado School of Medicine (UCSOM)*

*Head-To-Toe physical examination (HTT)*

*Standardized Patients (SPs)*

*Hypothesis-Driven Physical Examination (HDPE)*

*Generalizability (G)*

*Standard Deviations (SD)*

*Decision (D)*

## Declarations

**Author's Contributions:** TG, YSP, JH, and RY contributed to the study concept and design, analysis and interpretation of the data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, statistical expertise, and final approval of the manuscript. TG, YSP, JH, and RY have agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

**Acknowledgments:** I would like to acknowledge Shimaa Basha MPH, Director of the Center for Advancing Professional Excellence, for her support in providing ready access to the data gathered for our Foundations of Doctoring assessments and Monica McNulty MPH, MSPH for her expertise in providing statistical support for this manuscript.

**Ethical approval:** The study was considered exempt by the University of Colorado and University of Illinois Chicago institutional review boards. This determination was based upon local and federal exemption guidelines for exempt research as research that is conducted in commonly accepted educational settings and normal educational practices from the institutional boards.

**Consent for publication:** The authors of this study have consented for publication of the manuscript as detailed in this submission. No other individual data is included in this manuscript and, as such, the need for consent from the students for use of their de-identified data was waived by the institutional review boards at both of these institutions.

**Availability of Data and Materials:** Data and analyses for this manuscript are available upon request.

**Competing Interests and Disclosure Statement:** To the authors' knowledge, no conflict of interest, financial or other, exist.

**Funding:** None

## References

1. Yudkowsky R, Downing S, Klamen D, Valaski M, Eulenberg B, Popa M. Assessing the head-to-toe physical examination skills of medical students. *Med Teach*. 2004;26(5):415–419.
2. Yudkowsky R, Otaki J, Lowenstein T, Riddle J, Nishigori H, Bordage G. A hypothesis-driven physical examination learning and assessment procedure for medical students: initial validity evidence. *Med Educ*. 2009;43(8):729–740.
3. Gowda D, Blatt B, Fink M, Kosowicz L, Baecher A, Silvestri R. A core physical exam for medical students: results of a national survey. *Acad Med*. 2014;89(3):436–442.
4. Uchida T, Farnan J, Swartz J, Heiman H. Teaching the physical examination: a longitudinal strategy for tomorrow's physicians. *Acad Med*. 2014;89(3):373–375.
5. Gowda D, Blatt B, Kosowicz L. Addressing concerns about a “core + cluster” physical exam. *Acad Med*. 2014;89(6):834.
6. Williams R, Klamen D, Mayer D, Valaski M, Roberts N. A sampling strategy for promoting and assessing medical student retention of physical examination skills. *Acad Med*. 2007;82(10):S22-25.
7. Messick S. Validity. In: Linn R, editor. *Educational measurement*. New York, NY: American Council on Education and Macmillan; 1989.
8. Bloch R, Norman G. Generalizability theory for the perplexed: A practical introduction and guide: AMEE Guide No. 68. *Med Teach*. 2012;34:960–992.
9. Norcini J. Setting standards on educational tests. *Med Educ*. 2003;37:464–469.
10. Cook D, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ*. 2015;49:560–75.
11. Folque de Mendoca Patricio M. A Best Evidence Medical Education (BEME) Systematic Review on the feasibility, reliability and validity of the Objective Structured Clinical Examination (OSCE) in undergraduate medical studies [PhD dissertation]. Lisbon, Portugal: University of Lisbon; 2012.
12. Elstein A, Shulman L, Sprafka S. *Medical Problem Solving: An Analysis of Clinical Reasoning*. Cambridge, MA: Harvard University Press; 1978.
13. Haring C, Cools B, van der Meer J, Postma C. Student performance of the general physical examination in internal medicine: an observational study. *BMC Med Educ*. 2014;14:73–79.
14. Wilkerson L, Ming L. Assessing Physical Examination Skills of Senior Medical Students: Knowing How versus Knowing When. *Acad Med*. 2003;78(10):S30-32.
15. Peitzman S, Cuddy M. Performance in Physical Examination on the USMLE Step 2 Clinical Skills Examination. *Acad Med*. 2015;90(2): 209–13.
16. Van der Vleuten C, Schuwirth L, Driessen E, Dijkstra J, Tigelaar D, Baartman L, van Tartwijk J. A model for programmatic assessment fit for purpose. *Med Teach*. 2012;34(3): 205–214.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [GuthValidityEvidenceforCPEBMCMedicalEducationAppendixAFinalResubmission.docx](#)