

# From evidence to argument

## Citation for published version (APA):

Kinnear, B. (2025). *From evidence to argument: bringing validity to life in health professions education assessment*. [Doctoral Thesis, Maastricht University]. Maastricht University.  
<https://doi.org/10.26481/dis.20250619bk>

## Document status and date:

Published: 19/06/2025

## DOI:

[10.26481/dis.20250619bk](https://doi.org/10.26481/dis.20250619bk)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

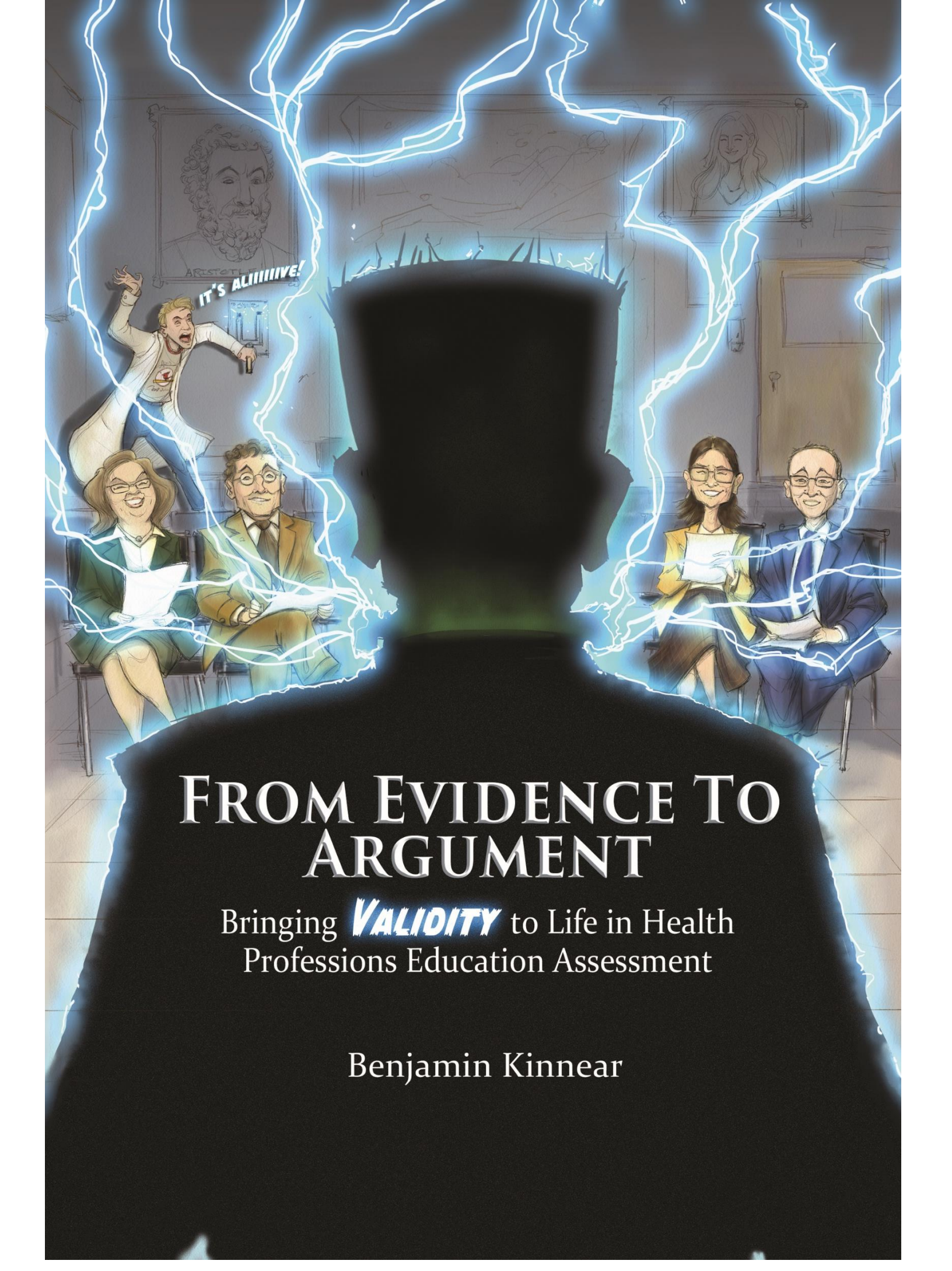
[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

The book cover features a central illustration of a classroom. A large, dark silhouette of a person's head and shoulders is in the foreground, looking towards a blackboard. On the blackboard, there is a drawing of a classical man's head with the word 'ARISTOTLE' written below it. To the left of the blackboard, a man in a white lab coat is jumping in the air, shouting 'IT'S ALIIIIIVE!'. In the foreground, four people are seated at a table: a woman in a green suit on the left, a man in a brown suit next to her, a woman in a yellow jacket on the right, and a man in a blue suit next to her. All four are holding papers and looking towards the blackboard. The background is a classroom with a door on the right and a window on the left. The entire scene is framed by bright, jagged blue lightning bolts.

# FROM EVIDENCE TO ARGUMENT

Bringing **VALIDITY** to Life in Health  
Professions Education Assessment

Benjamin Kinnear







# **Bringing Validity to Life in Health Professions Education Assessment**

Benjamin Kinnear

The research reported here was carried out at



In the School of Health Professions Education



And at Cincinnati Children's Hospital Medical Center



Parts of this work were supported by funding from the American Board of Medical Specialties Visiting Scholars Program.

**ISBN**

978-94-6380-727-2

**Lay-out/print**

Gildeprint, Enschede

**Cover design**

Joseph E Merideth

© Benjamin Kinnear, 2025

All rights are reserved. No part of this book may be reproduced, distributed, stored in a retrieval system, or transmitted in any form or by any means, without prior written permission of the author.

# **From Evidence to Argument**

## **Bringing Validity to Life in Health Professions Education Assessment**

DISSERTATION

to obtain the degree of Doctor at the Maastricht University,  
on the authority of the Rector Magnificus,  
Prof. Dr. Pamela Habibović  
in accordance with the decision of the Board of Deans,  
to be defended in public on Thursday the 19<sup>th</sup> of June, 2025, at 15.45 hours  
by

Benjamin Kinnear

**Supervisors:**

Prof. Dr. Erik Driessen (Maastricht University, The Netherlands)

Prof. Dr. Lara Varpio (Children's Hospital of Philadelphia, USA)

Prof. Dr. Daniel J. Schumacher (Cincinnati Children's Hospital Medical Center, USA)

**Assessment Committee:**

Chair:

Cees van der Vleuten (Maastricht University, The Netherlands)

Members :

Prof Dr. Debbie Jaarsma (Utrecht University, The Netherlands)

Prof Dr. Desirée Joosten-ten Brinke (Maastricht University, The Netherlands)

Prof. Dr. Fedde Scheele (Amsterdam UMC, The Netherlands)

Prof. Dr. Pim Teunissen (Maastricht University, The Netherlands)

## Table of Contents

Chapter 1	<b>Introduction</b>	7
Chapter 2	<b>Constructing a Validity Map for a Workplace-Based Assessment System: Cross-Walking Messick and Kane</b> Published in: <i>Academic Medicine</i> 2021;96(7S):S64-S69	19
Chapter 3	<b>How argumentation theory can inform assessment validity: A critical review</b> Published in: <i>Medical Education</i> 2022; 56(11):1064-1075	31
Chapter 4	<b>How do validity experts conceptualise argumentation? It's a rhetorical question</b> Published in: <i>Medical Education</i> 2024; 58(8):989-997	53
Chapter 5	<b>Legitimation Without Argumentation: An Empirical Discourse Analysis of 'Validity as an Argument' in Assessment</b> Published in: <i>Perspectives on Medical Education</i> 2024; 13(1):469-480	71
Chapter 6	<b>Validity in the Next Era of Assessment: Consequences, Social Impact, and Equity</b> Published in: <i>Perspectives on Medical Education</i> 2024; 13(1):452-459	91
Chapter 7	<b>What the hell is water? Changing medical education's Ideology through validity</b> Published in: <i>Medical Education</i> 2024; 58(3):274-276	105
Chapter 8	<b>MedEd Model: Validity</b> Published online at <a href="https://www.mededmodels.com/validity">https://www.mededmodels.com/validity</a>	111
Chapter 9	<b>Discussion</b>	115
Appendices	Summary	130
	Samenvatting	137
	Scientific Impact	144
	Acknowledgements	150
	Curriculum Vitae	151
	SHE Dissertation Series	153



# **CHAPTER 1**

## **Introduction**

## Validity: the Lurking, Chameleon-like God-Term

Validity holds a powerful place in our lives. Its presence looms in any field in which measurement occurs: Physics, Astronomy, Sociology, Psychology, Politics, Economics, Engineering, and even professional sports. Health professions education (HPE) is no different, particularly with regard to learner assessment. Validity has been called the “sine qua non”<sup>1</sup> and “cardinal virtue”<sup>2</sup> of assessment, as “without evidence of validity, assessments in medical education have little or no intrinsic meaning.”<sup>1</sup> Still others have called validity “the most important term in the educational and psychological measurement lexicon.”<sup>3</sup> Validity is always looming in assessment to hold us accountable in our pursuit of truth, knowledge, and understanding. In some ways, validity has become a god term in HPE, serving as an “educational idol” alongside other deified concepts such as *competence* and *patient safety*.<sup>4</sup> However, I want to signal at the outset of this thesis that my intention in studying validity was not to worship at the altar of this god term nor to elevate my standing within HPE by hitching my academic wagon to an unassailable concept. As you will see, my interest in validity stems from real-world questions and challenges that I have faced as a clinician educator. My hope is that by diving deeply into the assumptions about validity in HPE, I will play a part in nudging validity away from being a god term to which we are beholden and instead toward being a term over which we have agency; a concept that serves us in our ultimate goal of providing high-quality, equitable care to patients.

Validity is not just a god term; it is also a chameleon, changing its appearance based on context and use. Indeed validity has a long history of shapeshifting through time and across disciplines; a story that has been well documented in the HPE literature and does not need repeating to open this thesis. However, even when only considering the *here-and-now* of HPE, validity is not a concept with a concretized, fully agreed-upon meaning.<sup>5</sup> St Onge and Young describe how multiple discourses on the nature of validity coexist within HPE.<sup>6</sup> The first is that of validity as a test characteristic, or an intrinsic property of an assessment instrument. In this view, the label of “valid” is bestowed upon an instrument after empirical (often psychometric) testing, and the instrument is considered valid in all future uses and contexts. This discourse on validity persists in HPE today, but has largely given way to validity as an evidentiary argument. This discourse is derived from other fields such as Psychology<sup>7,8</sup> and Language Testing,<sup>9,10</sup> and views validity as the degree to which an argument supports the interpretation and uses of assessment data. Validity, therefore, is not the property of an instrument but of the decisions being made.<sup>11,12</sup> The third discourse is still relatively novel in HPE, and views validity as social imperative.<sup>13,14</sup> This discourse still embraces validity as an argument, but foregrounds the consequences of assessment for individuals, programs, and society. All three views of validity remain in use in HPE (and other related academic fields) today.

## Choosing Argument Out of the Many Faces of Validity

For my research, I chose to focus on the conceptualization of validity as an argument for multiple reasons. First, the argument-based conceptualization has been adopted by the Standards for Educational and Psychological Testing<sup>15</sup> and has become the predominant view of many HPE scholars<sup>1,11,12,16-18</sup>. Second, argument-based approaches to validity provide flexibility to integrate different assessment methods and data types into validation work. Programmatic assessment is increasingly used in HPE,<sup>19-22</sup> and validity needs to account for the integration of heterogeneous assessment strategies that inform educational decisions. Finally, the argument-based approach was useful in applying validity theory to real-world assessment conversations that were taking place in my day-to-day life as a program director.

It may be helpful to dive more deeply into what the argument-based conceptualization entails. In this view, validation is the process of making a claim about a specific interpretation and use of assessment data, and providing evidence to support such a claim. This argument-based approach to validity has been described in multiple ways. Kane described two types of arguments: 1) An interpretation-use argument (IUA) that lays out the framework of inferences and assumptions that constitute an argument, and 2) A validity argument in which those inferences and assumptions are evaluated for evidence of clarity, coherence, and plausibility.<sup>23</sup> While these two types of arguments are conceptually different, they are “intertwined”<sup>20</sup> and in essence different parts of a larger argument. Bachman similarly lays out two component arguments, though they are different than Kane’s approach. He describes an *assessment validity argument*, which links a moment of performance (presumably by a learner) to an interpretation, followed by an *assessment utilization argument*, which links interpretations to resulting decisions.<sup>10,24</sup> Mislevy’s *evidence-centered design* approach integrates assessment arguments connecting observations to inferences into the design of assessment systems. Kane, Bachman, and Mislevy’s approaches are different ways of organizing and operationalizing pieces (claims, inferences, data, interpretations) of an overarching argument that supports the interpretations and uses of data. The previous argument nomenclatures can quickly become confusing, and from here forward, I will use the term *validity argument* to mean the arguments (claims, inferences, data, interpretations) that constitute validity, regardless of how one slices or dices them.

Kane, Bachman, and Mislevy all draw from the same idea that validity is, fundamentally, an argument that supports or refutes the uses of assessment data. To be clear, this is different than making an argument *about* validity. For example, one might consider validity to be a property of an assessment instrument that is primarily evaluated via psychometric methods. An argument can then be made about the instrument’s validity, but the argument and the validity are two separate ideas. In the argument-based approach to validity (e.g. Kane, Bachman, Messick), the argument *is* validity. They are one and the same. Psychometric evaluation of an instrument might be one piece of evidence to support a validity argument, but validity and argument are one and the same. Therefore, without fully understanding argument and argumentation, we cannot understand validity.

## Argumentation Theory

Arguments are, by their very nature, social.<sup>25</sup> They exist to make an impact in the world such as to influence another's point of view or to change a policy. Argumentation is the process of developing and evaluating an argument. Argumentation Theory is a field of study (not a single theory per se) concerned with the production, analysis, and evaluation of argumentation discourse (spoken or written).<sup>25,26</sup> It draws from multiple fields, including Logic, Speech, Linguistics, Philosophy, Psychology, and Law. There are several different orientations within Argumentation Theory that each have varying argument structures, goals, and evaluation standards. For example, the dialectical orientation emphasizes the *processes* and *rules* of an argumentative dialogue.<sup>26</sup> In this orientation, arguments are evaluated by whether interlocutors follow rules of engagement, whether proper argument forms are used, and if they are acceptable to all parties. In contrast, rhetorical orientations hold that the goal of argumentation is persuasion of a given audience to a standpoint put forth by an arguer. In this orientation, there is less focus on adherence to a predefined argumentation process or evaluative standard, and more focus on meeting the specific needs of the intended audience. Argumentation Theory provides a way to understand which orientation is being used for any given exchange: how are arguments constructed and organized, who evaluates an argument, and what standards of evaluation are used. Argumentation Theory, then, is akin to the rules in a game like chess. Players can sit together at a chessboard with all of the pieces (i.e. validity evidence), but chess does not exist without a shared understanding of the rules of the game (i.e. underlying Argumentation Theory). Without a shared conception of processes and standards, argumentation is futile.<sup>25</sup>

## **Bringing Argumentation to Validity Arguments in HPE: The Work This Thesis Begins**

Argumentation Theory has vanishingly appeared in the validity argument literature. Kane, Bachman, and Mislevy all briefly reference argument structures put forth by Toulmin,<sup>2,10,23,27-30</sup> but they do not more deeply explore Argumentation Theory. These validity scholars cracked the door to Argumentation Theory, and with this PhD work I hope to kick the door open. Specifically, I hope to deepen the discourse on validity in HPE by bringing Argumentation Theory into our field's discourse. While HPE has adopted the ideas of Kane and Mislevy (I've yet to see Bachman make an appearance, though I'm holding out hope), their references to Toulmin or any other aspect of argumentation is conspicuously absent from our field. By more deeply examining the nature of argument and argumentation, helped in part by Argumentation Theory, HPE can more meaningfully operationalize validity arguments. Therefore, while my PhD thesis clearly has a strong philosophical and theoretical bend, my hope is that this work is all in service of helping educators, programs, and learners by providing clearer ideas of what validity is (or could be) in HPE.

## **Reflexivity Statement**

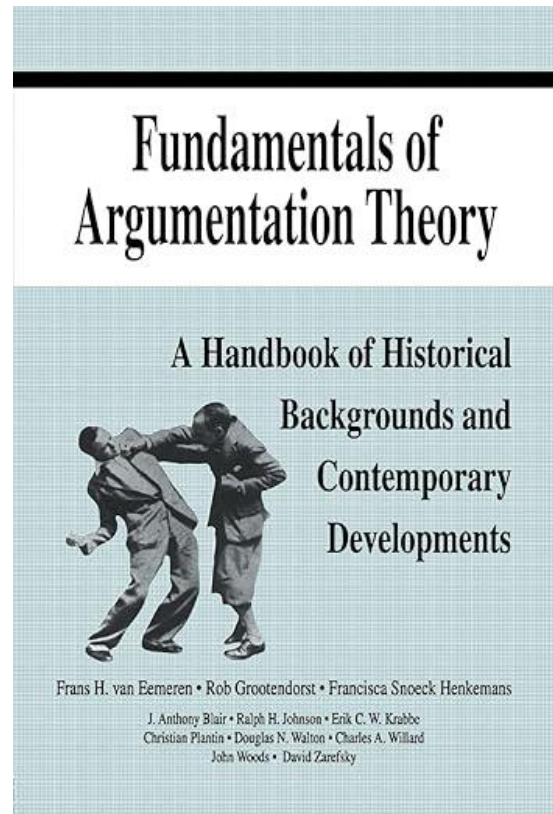
The story of how I became ensnared in the whirlwind of a PhD sheds light on the way I think about validity arguments. When compared with other validity scholars who are often trained in Measurement or

Psychometrics, I come from a clinical background. I began my PhD journey primarily as a clinician who was more equipped to perform a lumbar puncture than develop a research agenda or think deeply about a theoretical concept like validity. My positionality allowed me to *make strange* what others may have taken for granted. I lacked advanced training or understanding of validity theory, which allowed me to ask questions that others might overlook. My decision to follow such questions to Maastricht came after years of learning, a moment of practical uncertainty, and an inspirational meeting.

Over nearly a decade, my home program's education team developed a multifaceted program of assessment<sup>20</sup> consisting of workplace-based entrustment data,<sup>31,32</sup> narrative comments,<sup>33</sup> learning analytics,<sup>34,35</sup> an internal testing program,<sup>36</sup> 360-degree ratings from a year-long longitudinal ambulatory long block,<sup>37</sup> resident self-assessment, outpatient clinical care measures,<sup>38,39</sup> and an evidence-informed clinical competency committee process.<sup>40</sup> By 2019, we felt that our summative decisions about resident readiness for practice were more defensible when grounded in our competency-based assessment system rather than relying on time in training (as usual US-based medical training entails). Therefore, in 2020, we began a competency-based, time-variable training pilot called TIMELESS (Transitioning in Internal Medicine Education Leveraging Entrustment Scores Synthesis).<sup>41</sup> While planning for the pilot, we engaged with multiple stakeholders of our program's summative decisions including our clinical department leadership, our residency program leadership, our institution-wide PGME leadership, certifying boards, and accrediting bodies. A similar question appeared in nearly each conversation: *How do you know that your time-variable summative decisions are defensible?* I could cite nearly a dozen peer-reviewed studies that chronicled our work, but what these stakeholders wanted was something more than a pile of manuscripts. They wanted an *argument* for why our summative decisions were defensible.

I had so many questions. What does an argument look like? How do I organize it, and how will it be evaluated? Do I present the same argument to every stakeholder, or do I make each argument fit for the audience? Is there a framework or theory of argument that could guide me? None of the educators on our team were proficient with validity theory at this point – we were all clinician educators learning as we went. Building from seminal writings of David Cook,<sup>12,16,17</sup> Samuel Messick,<sup>8</sup> and Michael Kane,<sup>9</sup> we invested fully in the conceptualization of validity as an argument that supports the interpretation and uses of data.<sup>12</sup> This was a breakthrough, enabling us to use validity theory to build an argument supporting assessment decisions needed for competency-based, time-variable training. Our first attempt involved collating and organizing all validity evidence we had to support our assessment decisions into a *validity map* (Chapter 2).<sup>42</sup> This map would organize evidence into prominently-used validity frameworks – Messick and Kane – for purposes of showing to stakeholders and to help guide future research activities. While we successfully constructed such a map, and found value in the process, we realized that something was missing. We had created a map of evidence, but we still lacked a cohesive argument.

We lacked a framework and vocabulary to describe what we felt was needed. I began searching for ways to learn more about arguments and argumentation. I found the field of Argumentation Theory, and decided I wanted to read more from pre-eminent argumentation scholars. I chose to begin with the book *Fundamentals of Argumentation Theory: A Handbook of Historical Backgrounds and Contemporary Developments*,<sup>25</sup> both because it had been cited over 2000 times and because I found the cover (depicting a fistfight) to be charming. This book became my gateway into Argumentation Theory, including all the fields therein – Logic, Informal Logic, Dialectics, PragmaDialectics, New Rhetoric, and more. As I read the book and swam through a sea of new ideas and concepts, I struggled to understand if and how Argumentation Theory could help with my very practical and real world validity argument related to time-variable assessment decisions. I was learning new things, but struggled with what to do next.



The final breadcrumb leading me to a PhD came via a podcast. While listening to episode 256 of the original KeyLime podcast, I heard Dr. Lara Varpio mention the name Stephen Toulmin and his framework for argument structure. I had just finished reading a section on Toulmin and his impact on Argumentation Theory in the *Fundamentals* book, and I felt that Dr. Varpio might be able to help funnel my curiosity toward something productive. I reached out via Twitter direct message at the outset of the COVID-19 pandemic (June 4<sup>th</sup>, 2020), and she graciously replied the next day with an offer to meet via Zoom. That one-hour meeting changed my life. I learned that her PhD was in argumentation and rhetoric, and that my message had harkened back to her training. By the end of the meeting, I was committed to pursuing a PhD to explore argument and argumentation in HPE assessment validity.

I am a clinician educator at a large, urban, academic medical center in the United States. I serve as program director for a medium-sized residency program that spans across two specialties – Internal Medicine and Pediatrics (Med-Peds). Med-Peds training involves being embedded in both the world of Internal Medicine and the world of Pediatrics, while at the same time being a separate specialty. Med-Peds therefore requires one to navigate the very different environments and cultures that exist in each specialty, including being attuned to the needs of differing groups of stakeholders. My role as a program director and a member of multiple clinical competency committees means that I frequently am involved in summative decision-making about learner readiness to progress. The defensibility of promotion or graduation of a resident is not a theoretical concern, but a real-world, consequential issue that I face on a regular basis. Therefore, I bring a very pragmatic, utilitarian view to the concept of validity arguments.

I have also spent a considerable amount of time researching competency-based medical education (CBME) and incorporating competency-based approaches into local education programs. I am therefore strongly influenced by multiple concepts related to CBME. First, CBME grew, in part, out of the need for better accountability from healthcare to the public.<sup>43</sup> CBME asserts that medical education has a social contract with the public to graduate physicians who are ready to provide high quality care. This social contract resonates with me and influences how I think about the concept of defensibility and the purpose of validity in assessment. Second, CBME is an outcomes-based approach which emphasizes the need for defensible assessment strategies to determine when desired training outcomes have been met.<sup>44</sup> The concept of validity as an argument relating to defensibility of interpretations and uses of assessment data resonates with CBME principles. Finally, competency-based assessment has been framed as a service that is co-created with stakeholders rather than a product that is developed by educators and delivered to learners and society.<sup>45</sup> The concept of co-creation in CBME influences how I value the social nature of argumentation and the central role that stakeholders (e.g. learners, patients) play in that process.

## Thesis Outline

This thesis explores the overarching research question: *What is the nature of argument and argumentation with regard to assessment validity in health professions education?* I begin by sharing a validity map that outlines evidence supporting summative decisions being made using my home program's competency-based assessment system (Chapter 2). I then tap into the field of Argumentation Theory using narrative review to draw out orientations that may be generative to HPE assessment validity (Chapter 3). Specifically, I suggest that Informal Logic and New Rhetoric may provide useful argumentation orientations for assessment in HPE. To understand how argumentation manifests (or does not) in HPE validity theory and practical validation, I explore both the views of validity experts (Chapter 4) and the contents of peer reviewed publications (Chapter 5). The findings of these studies suggest that HPE validity should be more social (i.e. involving information exchange between people or groups) and participatory (i.e. engaging stakeholders of assessment decisions), and that true argumentation is generally lacking. I then explore how HPE assessment validity might evolve if such changes should occur and if society (as a stakeholder) had its needs foregrounded in validation work (Chapter 6). I conclude with a synthesis of the findings of this research program, implications for stakeholders, and future research directions. Chapter 7 is an invited commentary describing how members of the HPE community have agency to choose how broadly or narrowly validity is defined. I argue that a broader definition of validity allows for equity to become central to validity arguments. Given the high cognitive load of validity and Argumentation Theory, I collaborated with MedEd Models to create high-yield infographics (Chapter 8) to serve as an introduction to validity arguments for non-experts. Finally, I close (Chapter 9) with a discussion on the implications of the research in this thesis, including the potential for validity to be holistic, participatory, community-based, agentic, and equitable. I use the analogy of Frankenstein's monster to call for more life to be brought to HPE validity through argumentation.

## Worldviews Underpinning the Research

I initially came to this PhD journey with a social constructionist worldview<sup>46</sup> which acknowledges that reality is created by and between people. This aligns with my belief that there is no absolute or objective truth or reality about validity. The lack of objective reality with regard to validity is not from lack of measurement ability. Even with an imagined, perfect instrument, there is no *thing* to measure, quantify, or define. Validity (as well as its conceptualization as an argument) is a human-created concept that only has meaning if there is some degree of shared reality between people. This social constructionist worldview influenced all of the papers contained within this thesis.

However, my philosophical worldview has shifted during my PhD journey. I have been influenced by critical researchers like Dr. Abigail Konopasky, Dr. Justin Bullock, Dr. Tasha Wyatt, Hannah Anderson, and many others. Their work, as well as my own personal journey learning about the inequities and power structures that constitute our social world, has brought me to a much more critical view. Like a ship being continually rebuilt as it sails, my worldview has evolved as I've sailed into new ideological waters.<sup>47</sup> This critical worldview shines through in Chapters 5, 6, and 7.

## References

1. Downing SM. Validity: on the meaningful interpretation of assessment data. *Medical education*. 2003;37(9):830-837.
2. Mislevy RJ, Steinberg LS, Almond RG. Focus article: On the structure of educational assessments. *Measurement: Interdisciplinary research and perspectives*. 2003;1(1):3-62.
3. Newton PE, Baird J-A. *The great validity debate*: Taylor & Francis; 2016.
4. Lingard L. What we see and don't see when we look at 'competence': notes on a god term. *Advances in health sciences education*. 2009;14(5):625-628.
5. St-Onge C, Young M. Evolving conceptualisations of validity: impact on the process and outcome of assessment. *Medical education*. 2015;49(6):548-550.
6. St-Onge C, Young M, Eva KW, Hodges B. Validity: one word with a plurality of meanings. *Adv Health Sci Educ Theory Pract*. 2017;22(4):853-867.
7. Cronbach LJ. Five perspectives on validity argument. *Test validity*. 1988:3-17.
8. Messick S. Validity In. R. Linn (Ed.) *Educational measurement* (13-103): New York: Macmillan Publishing; 1989.
9. Kane M. The argument-based approach to validation. *School Psychology Review*. 2013;42(4):448-457.
10. Bachman LF. Constructing an assessment use argument and supporting claims about test taker-assessment task interactions in evidence-centered assessment design. MEASUREMENT-LAWRENCE ERLBAUM ASSOCIATES-. 2003;1:63-65.
11. von Bergmann H, Childs RA. When I say... validity argument. *Medical Education*. 2018;52(10):1003-1004.
12. Cook DA. When I say... validity. *Medical education*. 2014;48(10):948-949.
13. Marceau M, Gallagher F, Young M, St-Onge C. Validity as a social imperative for assessment in health professions education: a concept analysis. *Med Educ*. 2018;52(6):641-653.
14. Marceau M, St-Onge C, Gallagher F, Young M. Validity as a social imperative: users' and leaders' perceptions. *Canadian Medical Education Journal*. 2019.
15. *Standards for educational and psychological testing*: Amer Educational Research Assn; 2014.
16. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical education*. 2015;49(6):560-575.
17. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Adv Simul (Lond)*. 2016;1:31.
18. Govaerts M, van der Vleuten CP. Validity in work-based assessment: expanding our horizons. *Medical education*. 2013;47(12):1164-1174.
19. Pearce J, Tavares W. A philosophical history of programmatic assessment: tracing shifting configurations. *Advances in Health Sciences Education*. 2021:1-20.
20. Schuwirth LW, Van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. *Medical teacher*. 2011;33(6):478-485.
21. van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ*. 2005;39(3):309-317.
22. Uijtdehaage S, Schuwirth LW. Assuring the quality of programmatic assessment: Moving beyond psychometrics. *Perspectives on medical education*. 2018;7(6):350-351.
23. Kane MT. Validating the interpretations and uses of test scores. *Journal of Educational Measurement*. 2013;50(1):1-73.
24. Bachman LF. Building and supporting a case for test use. *Language Assessment Quarterly: An International Journal*. 2005;2(1):1-34.

25. Van Eemeren FH, Grootendorst R, Johnson RH, Plantin C, Willard CA. Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments: Routledge; 2013.
26. Tindale CW. Acts of arguing: A rhetorical model of argument: SUNY Press; 1999.
27. Kane MT. An argument-based approach to validity. *Psychological bulletin*. 1992;112(3):527.
28. Mislevy RJ. Validity by design. *Educational researcher*. 2007;36(8):463-469.
29. Kane M. Validity and fairness. *Language testing*. 2010;27(2):177-182.
30. Mislevy RJ. The case for informal argument. *Measurement: Interdisciplinary Research & Perspective*. 2012;10(1-2):93-96.
31. Warm EJ, Mathis BR, Held JD, et al. Entrustment and mapping of observable practice activities for resident assessment. *Journal of general internal medicine*. 2014;29(8):1177-1182.
32. Warm EJ, Held J, Hellman M, et al. Entrusting Observable Practice Activities and Milestones Over the 36 Months of an Internal Medicine Residency. *Academic Medicine*. 2016;91(10):1398-1405.
33. Kelleher M, Kinnear B, Sall DR, et al. Warnings in early narrative assessment that might predict performance in residency: signal from an internal medicine residency program. *Perspectives on Medical Education*. 2021:1-7.
34. Schauer DP, Kinnear B, Kelleher M, Sall D, Schumacher DJ, Warm EJ. Developing the Expected Entrustment Score: Accounting for Variation in Resident Assessment. *Journal of general internal medicine*. 2022.
35. Kinnear B, Schauer DP, Warm EJ. Development of an entrustment ratings display fit for ordinal data. *Medical Education*. 2023;57(11):1118-1119.
36. Mathis BR, Warm EJ, Schauer DP, Holmboe E, Rouan GW. A multiple choice testing program coupled with a year-long elective experience is associated with improved performance on the internal medicine in-training examination. *Journal of general internal medicine*. 2011;26(11):1253-1257.
37. Warm EJ, Schauer DP, Diers T, et al. The ambulatory long-block: an accreditation council for graduate medical education (ACGME) educational innovations project (EIP). *Journal of general internal medicine*. 2008;23(7):921-926.
38. Warm EJ, Kinnear B, Knopp MI, et al. Ambulatory Long Block: A Model of Precision Education and Assessment for Internal Medicine Residents. *Academic Medicine*. 2023;10.1097.
39. Zafar MA, Diers T, Schauer DP, Warm EJ. Connecting resident education to patient outcomes: the evolution of a quality improvement curriculum in an internal medicine residency. *Academic Medicine*. 2014;89(10):1341-1347.
40. Kinnear B, Warm EJ, Hauer KE. Twelve tips to maximize the value of a clinical competency committee in postgraduate medical education. *Medical teacher*. 2018:1-6.
41. Kinnear B, Santen SA, Kelleher M, et al. How does TIMELESS training impact resident motivation for learning, assessment, and feedback? Evaluating a competency-based time-variable training pilot. *Academic Medicine*. 2023;98(7):828-835.
42. Kinnear B, Kelleher M, May B, et al. Constructing a validity map for a workplace-based assessment system: Cross-walking Messick and Kane. *Academic Medicine*. 2021;96(7S):S64-S69.
43. Hodge S. The Origins of Competency-Based Training. *Australian journal of adult learning*. 2007;47(2):179-209.
44. Gruppen LD, Ten Cate O, Lingard LA, Teunissen PW, Kogan JR. Enhanced Requirements for Assessment in a Competency-Based, Time-Variable Medical Education System. *Academic medicine : journal of the Association of American Medical Colleges*. 2018;93(3S Competency-Based, Time-Variable Education in the Health Professions):S17-S21.
45. Holmboe ES. Work-based Assessment and Co-production in Postgraduate Medical Training. *GMS journal for medical education*. 2017;34(5).

46. Rees CE, Crampton PE, Monrouxe LV. Re-visioning academic medicine through a constructionist lens. *Academic Medicine*. 2020;95(6):846-850.
47. Kinnear B, Weber DE, Schumacher DJ, Edje L, Warm EJ, Anderson HL. Reconstructing Neurath's Ship: A Case Study in Reevaluating Equity in a Program of Assessment. *Academic Medicine: Journal of the Association of American Medical Colleges*. 2023.



# CHAPTER 2

## **Constructing a Validity Map for a Workplace-Based Assessment System: Cross-Walking Messick and Kane**

Published as:

Kinnear B, Kelleher M, May B, Sall D, Schauer DP, Schumacher DJ, Warm EJ.  
Constructing a validity map for a workplace-based assessment system: cross-walking  
Messick and Kane. *Academic Medicine*. 2021 Jul 1;96(7S):S64-9.

# Abstract

## Problem

Health professions education has shifted to a competency-based paradigm in which many programs rely heavily on workplace based assessment (WBA) to produce data for summative decisions about learners. However, WBAs are complex and require validity evidence beyond psychometric analysis. Here, the authors describe their use of a rhetorical argumentation process to develop a map of validity evidence for summative decisions in an entrustment based WBA system.

## Approach

To organize evidence, the authors cross-walked 2 contemporary validity frameworks, one that emphasizes *sources* of evidence (Messick) and another that stresses *inferences* in an argument (Kane). They constructed a validity map using 4 steps: (1) Asking critical questions about the stated interpretation and use, (2) Seeking validity evidence as a response, (3) Categorizing evidence using both Messick's and Kane's frameworks, and (4) Building a visual representation of the collected and organized evidence. The authors used an iterative approach, adding new critical questions and evidence over time.

## Outcomes

The first map draft produced 25 boxes of evidence that included all 5 sources of evidence detailed by Messick and spread across all 4 inferences described by Kane. The rhetorical question–response process allowed for structured critical appraisal of the WBA system, leading to the identification of evidentiary gaps.

## Next Steps

Future map iterations will integrate evidence quality indicators and allow for deeper dives into the evidence. The authors intend to share their map with graduate medical education stakeholders (e.g., accreditors, institutional leaders, learners, patients) to understand if it adds value for evaluating their WBA programs' validity arguments.

Health professions education has shifted to a competency-based paradigm, 1–4 an approach in which the processes used to train health care providers are driven by the needs of the populations they serve. 5–7 The need for competency measurement has made assessment a particularly crucial, yet challenging, aspect of competency based medical education (CBME) implementation. 8–11 Workplace-based assessment (WBA) has become integral to programs of assessment, 8 carrying the advantage of assessing learners performing real-world clinical tasks. However, WBAs are often criticized for lacking validity evidence for summative purposes due to relatively poor psychometric performance. 12 We believe validity should include more than psychometrics alone.

Validity is the *sine qua non* of good assessment, 13 and therefore is central to CBME programs. While multiple descriptions of validity exist, 14 modern frameworks conceptualize it as an interpretive argument that supports a predefined interpretation or use of data. 15 Two contemporary frameworks are commonly used to organize validity evidence, though they differ in emphasis (Table 1). Messick’s framework stresses *sources* of validity evidence, 16 and Kane’s focuses on *inferences* in an evidentiary chain. 17 While these frameworks are not mutually exclusive, often one or the other is used to organize evidence. However, we believe they are complementary. Different sources of evidence (Messick) can support argument inferences (Kane) and provide a more complete picture of a complex validity argument. In this paper, we provide a working example of how both frameworks can be used in combination with a rhetorical argumentation process to develop a validity map for an entrustment-based WBA system. We describe how our mapping process helped to organize validity evidence, identify evidentiary gaps, and guide a research and program improvement agenda.

Table 1

**A Summary of Messick's<sup>16</sup> and Kane's<sup>17</sup> Validity Frameworks**

Source or inference	Description of evidence	Example evidence for WBAs
<b>Messick</b>		
Content	Evidence that the domain intended to be assessed is truly represented	<ul style="list-style-type: none"> <li>- Input from clinical and education experts on what is assessed</li> <li>- Representativeness of skills observed with domain of interest</li> <li>- Deliberate revisions of what is observed based on program evaluation</li> </ul>
Response process	Evidence that an assessor's or respondent's thinking aligns with the intended process	<ul style="list-style-type: none"> <li>- Assessor training</li> <li>- Think-aloud studies with assessors</li> <li>- Use of construct-aligned scales</li> <li>- Direct observation</li> </ul>
Internal structure	Evidence that the structure of response data aligns with intended assessment design	<ul style="list-style-type: none"> <li>- Reliability analysis (e.g., generalizability study)</li> <li>- Factor analysis</li> </ul>
Relationship to other variables	Evidence that resulting data align with external measures of the same construct or related constructs	<ul style="list-style-type: none"> <li>- Association with clinical care measures</li> <li>- Association with other education data (e.g., test scores)</li> </ul>
Consequences	Evidence supporting benefits or harms that result from assessment data	<ul style="list-style-type: none"> <li>- Standard setting for promotion/remediation</li> <li>- Impact on graduation rate</li> <li>- Acceptability to learners</li> </ul>
<b>Kane</b>		
Scoring	Evidence to support how an assessor gives a single score	<ul style="list-style-type: none"> <li>- Rater training</li> <li>- Think-aloud studies with assessors</li> <li>- Use of construct-aligned scales</li> <li>- Direct observation</li> </ul>
Generalization	Evidence that an assessment is representative of performance on all possible items in a theoretical test universe	<ul style="list-style-type: none"> <li>- Reliability analysis (e.g., generalizability study)</li> <li>- Factor analysis</li> <li>- Sampling strategy</li> </ul>
Extrapolation	Evidence that performance in the theoretical test universe reflects performance in the real-world on related constructs	<ul style="list-style-type: none"> <li>- Authenticity of assessment methods</li> <li>- Association with clinical care measures</li> </ul>
Implication	Evidence to support how the assessment data are used	<ul style="list-style-type: none"> <li>- Standard setting for promotion/remediation</li> <li>- Impact on graduation rate</li> <li>- Acceptability to learners</li> </ul>

**Approach**

The University of Cincinnati (UC) Internal Medicine (IM) residency program, based in an urban, tertiary referral medical center, consists of approximately 89 categorical residents who rotate at UC Medical Center, Veterans Affairs Medical Center, and multiple ambulatory clinics. In 2012, we implemented an entrustment-based WBA system that integrated the Accreditation Council for Graduate Medical Education subcompetencies, and we have been collecting validity evidence for using these data for determination of resident competence since that time. 18,19

Our initial validity mapping team consisted of 1 IM program director, 4 associate program directors, 1 chief resident, and 1 medical education researcher with expertise in CBME. For purposes of our validity argument, our intended use of the data was for the determination of competence to inform promotion and graduation decisions within our IM residency program. Thus, we aimed to collect and organize validity evidence for this predetermined interpretation and use of the data. We constructed our validity map using a multistep process that mirrored a rhetorical argument including the following 4 steps: (1) Asking critical questions about the stated interpretation and use, (2) Seeking validity evidence as a response, (3) Categorizing evidence, and (4) Visualizing evidence.

**1. Asking critical questions about the stated interpretation and use**

Our team first took the approach of acting as an interlocutor who is critical of our own WBA system. We individually wrote questions that might challenge whether decisions made using the WBA data were valid for assessment of competence. Example questions included, “Why is entrustment used as a construct for frontline assessment?”, “What is the reliability of entrustment ratings?”, and “Do entrustment scores correlate with measures of clinical care quality?”. Individually generated questions were collated by B.K., who then removed duplicate questions. Our group then met to review the questions and generate any further questions. These questions then served as a guide for the validity evidence we sought.

## **2. Seeking validity evidence as a response**

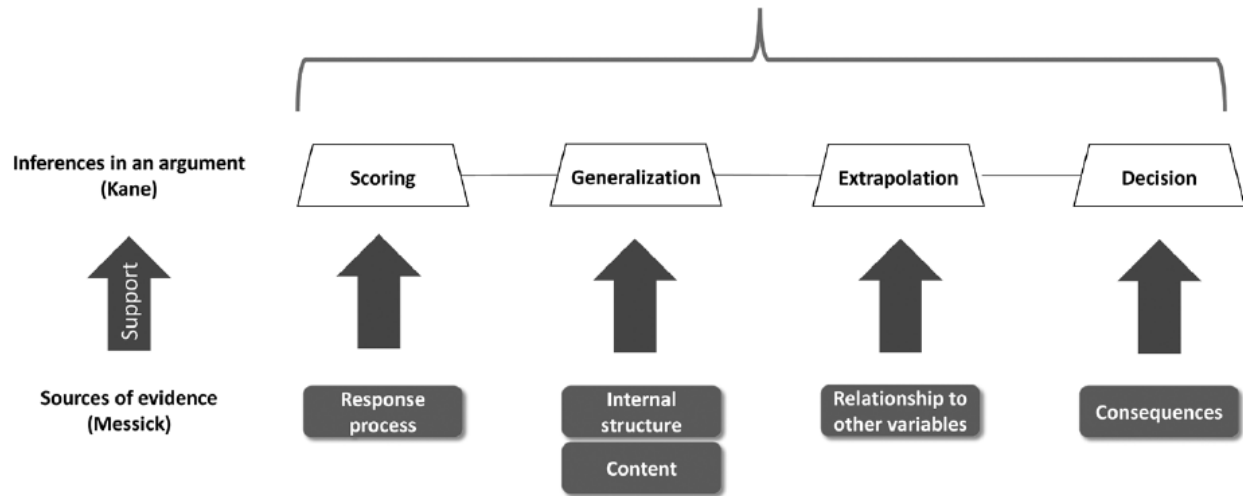
We attempted to respond to each question and provide supportive evidence. We first reviewed existing literature for evidence that would support or refute our response. For example, in response to the first question above, we found literature supporting the importance of construct-alignment using entrustment as a framework for WBAs. 20–22 Other times we used evidence that we had generated through previous work. For example, we had already performed a generalizability study and knew the reliability of our WBA data at different time points in residency training. 23 Many times we were unable to find evidence or provide a reasonable response to questions from step 1, thereby identifying gaps in our validity argument, which were translated into research questions for future validation work. While we intended to address these questions through quantitative or qualitative research, others may choose to address them through programmatic improvement efforts.

This work was organized using a Microsoft Excel (Microsoft Corporation, Redmond, Washington) spreadsheet with the following column headings: *Question, Response, Supporting Evidence, Citations, Gaps.*

## **3. Categorizing evidence**

We organized our evidence by the *source* of evidence (Messick) supporting an *inference* (Kane) in our argument (Figure 1). This process was done using discussion to form group consensus around categorizations. For example, the use of an entrustment framework for frontline WBA instruments was considered *response process* evidence (Messick) to support the *scoring* inference (Kane). The generalizability study was deemed *internal structure* evidence (Messick) to support the *generalization* inference (Kane). The categorization into Messick’s and Kane’s framework was integrated into the Excel spreadsheet. Some validity evidence supported multiple inferences or represented multiple sources. When this occurred, we used group discussion to generate consensus regarding which source/inference was most represented in our system by a given piece of evidence. Our goal was not to find the *only* way to map our evidence, but rather the one that fits best to help guide our future validation work.

# Validity argument



**Figure 1** - An illustration of how we cross-walked the 2 frameworks. We used Messick's 16 sources of evidence to support Kane's 17 inferences in a validity argument. Of note, multiple different sources of evidence could support any given inference. For simplicity, we chose to only connect each source of evidence to one inference in this figure.

## 4. Visualizing evidence: Map building

Excel was a functional means of cataloguing evidence, but we wanted a more visual format for 2 reasons. First, we sought an understandable, high-level review of collected evidence to identify remaining gaps in our argument. This visual could serve as a validity "map" for quick reference, with the Excel document used for deeper dives into the evidence. Second, the map allowed us to organize evidence into that which was already collected, ongoing validation work, and future studies. We used Microsoft Visio (Microsoft Corporation, Redmond, Washington) to create the map and organize the information using spatial orientation and color/shading (Figure 2). Inferences from Kane's framework were placed in 4 rows that crossed 3 columns labeled "Evidence obtained," "Work in progress," and "Gaps/future studies." Each critical question from the Excel sheet was placed in a box that represented the corresponding response/evidence. Boxes were color coded based on which source of evidence they represented from Messick's framework and placed in the appropriate row and column on the map.

2a

Messick's sources of validity evidence				
Content	Response process	Internal structure	ROV	Consequences
Kane	Evidence obtained	Work in progress	Gaps/future work	
Scoring	<div>Why use entrustment as a construct for frontline assessment?</div> <div>Do assessors give quality ratings?</div> <div>Do assessors observe the skills they assess?</div> <div>Do assessors receive initial rater training?</div>	<div>How do assessors give ratings?</div>	<div>Do assessors receive ongoing rater training?</div> <div>Do non-faculty assessors receive training?</div>	
Generalization	<div>How were skills chosen to be included on assessment forms?</div> <div>What narrative data are collected??</div> <div>What is the reliability of ratings?</div>		<div>How are assessment instruments revised over time?</div> <div>Which assessment items are best at learner discrimination?</div> <div>How do ratings at this program compare with other programs with the same system?</div>	
Extrapolation	<div>How were assessment items mapped to the ACGME milestones?</div>	<div>Were the correct mapping decisions made?</div>	<div>Do ratings have an association with post-graduation clinical care quality?</div> <div>Do ratings have an association with in-residency clinical care quality?</div> <div>Do ratings have an association with other educational measures?</div>	
Implications	<div>How were promotions standards determined?</div> <div>Does the system detect critical deficiencies?</div>	<div>How does the CCC account for construct-irrelevant variance?</div>	<div>Are there non-physician members on the CCC?</div> <div>What process does the CCC use for decision-making?</div> <div>How are learners helped?</div> <div>Do CCC members receive training?</div>	

2b

Messick's sources of validity evidence				
Content	Response process	Internal structure	ROV	Consequences
Kane	Evidence obtained	Work in progress	Gaps/future work	
Scoring	<div>Why use entrustment as a construct for frontline assessment?</div> <div>Do assessors give quality ratings?</div> <div>Do assessors observe the skills they assess?</div> <div>Do assessors receive initial rater training?</div> <div>Do assessors have access to just-in-time training materials?</div> <div>How do assessors give ratings?</div>		<div>Do assessors receive ongoing rater training?</div> <div>Do non-faculty assessors receive training?</div>	
Generalization	<div>How were skills chosen to be included on assessment forms?</div> <div>What narrative data are collected??</div> <div>What is the reliability of ratings?</div> <div>How do ratings at this program compare with other programs with the same system?</div> <div>Are there multiple assessors in multiple contexts?</div>	<div>Which assessment items are best at learner discrimination?</div> <div>How are assessment instruments revised over time?</div>		
Extrapolation	<div>How were assessment items mapped to the ACGME milestones?</div> <div>How can clinical care measures be attributed to a resident using the EHR?</div> <div>Are observed skills performed in true clinical context?</div> <div>How do mapping decisions compare with other programs?</div> <div>Which clinical care measures best represent a resident's care?</div>	<div>Do ratings have an association with in-residency clinical care quality?</div> <div>Do ratings have an association with other educational measures?</div>	<div>Do ratings have an association with post-graduation clinical care quality?</div>	
Implications	<div>How were promotions standards determined?</div> <div>Does the system detect critical deficiencies?</div> <div>Do CCC members receive training?</div> <div>What process does the CCC use for decision-making?</div> <div>How are learners improved?</div>	<div>How does the CCC account for construct-irrelevant variance?</div> <div>How are narrative comments used?</div> <div>How is the program improved?</div>	<div>Are there non-physician members on the CCC?</div> <div>How are data displayed for group deliberation?</div>	

## Figure 2

Panel 2a represents an early version of our validity map. Each box contains a critical question from step 1 of the rhetorical process described in the text. Each row represents an inference in Kane's validity framework, 17 and sources of evidence from Messick's framework 16 are represented by different shading within each box. The location of a box within the 3 columns represent whether there is evidence that supports an answer, evidence that is currently under study/development, or evidence that is part of future work. Panel 2b is an updated version of our validity map. As evidence is gathered, more boxes appear in the left-hand column of the map.

## Outcomes

The first draft of our validity map produced 25 boxes of evidence that included all 5 sources of evidence detailed by Messick and spread across all 4 inferences described by Kane (Figure 2a). Our team found value in both the process of map construction and in the map itself. The rhetorical question–response process allowed us to critically appraise our WBA system in a deliberate fashion. Questioning for purposes of validation created a permission structure to ask critical questions without being perceived as “negative” or “dissenting.” Instead, it fostered an atmosphere of constructive skepticism that created a safe space for dialogue and curiosity. This led us to identify several evidentiary gaps that were previously unknown to us and to integrate them into future research endeavors. For example, we realized that we had no *relationship to other variables* evidence (Messick) to support our *extrapolation* inference (Kane). We have since begun work to explore if our entrustment data are associated with other education data. We also found value in taking the time to organize existing validity evidence, both from existing literature and from our own work. While we had been undertaking validation work in pieces, we did not have a holistic view of our argument, and thus our efforts were somewhat disjointed. Collating and organizing relevant evidence helped us to coordinate our validation work moving forward, to integrate new work into our argument, and to identify scholarly projects for interested junior educators.

We found it beneficial to integrate both Messick's and Kane's frameworks into our map. Kane's framework carries the advantage of allowing for the identification of the “weakest link” in the evidentiary chain, helping us to prioritize our work. 15 As previously mentioned, the map showed our argument was lacking in evidence to support the *extrapolation* inference. Thus, we have focused our recent efforts on building evidence by researching if our WBA data are associated with other measures of performance or with clinical care outcomes. Messick's framework has helped us think systematically through sources of evidence that might be sought to support a given inference.

Our validity map is dynamic, and changes as new questions arise and new evidence is collected. Using an electronic format allows us to easily edit, update, and share the map over time. Figure 2b is an updated map, showing how boxes move from right (future work) to left (accrued evidence) as validity studies are completed or evidence is discovered. We anticipate that we will never be able to answer all critical questions, and therefore will always have boxes in the “future studies” column of the map. The benefit of this is that evidentiary gaps are made obvious to allow for better scrutiny of the overall argument.

## Next Steps

A limitation of our current validity map is that it does not allow for an easy deep dive into the evidence undergirding each box's question. As noted above, we used Excel for the initial categorization of evidence, and these spreadsheets contain more in-depth explanations and citations for relevant evidence. Future iterations will involve building our validity map on a platform that is interactive and allows for deep dives. Some current online presentation software programs contain a "zoom-in/ zoom-out" function that could allow for deep dives within a visual map. An online, cloud-based structure would also allow for easy sharing of our map with stakeholders outside of our institution. Another limitation is that our map does not indicate the weight or quality of each piece of evidence to support the overall argument. In future work, we will attempt to include indicators of these variables in our map. Finally, while this process helped us organize our evidence, it does not provide full argumentation structure such as those put forth by Toulmin or Wigmore. 24,25 Mapping out the connections between every element of argumentation (data, warrant, backing, and claim) was beyond our goal of organizing evidence to identify gaps and inform future work.

We found value in collating, categorizing, and organizing validity evidence for our WBA system, but the highest value lies in sharing this work with the stakeholders of our assessment system. We have preliminarily shared our validity map with others in the medical education community to obtain feedback on our process and the acceptability of the evidence provided. 26 Next steps include collecting stakeholders' (e.g., accreditors, institutional leaders, learners, patients) input to understand if our validity map adds value for evaluating our program's validity argument. Stakeholders may prioritize evidence differently or highlight new gaps not initially evident to us. We recognize that stakeholder input is crucial to ensure that we are not "gilding the lily" by only asking questions to which we have a reasonable response. 27 This work will also lay the foundation for future studies exploring the nature of argumentation in health professions validity work, 28 which will be crucial in understanding how comprehensive validity arguments might be evaluated.

## **Conclusion**

We found that constructing a validity map for our WBA system helped identify evidentiary gaps and plan future research and program improvement efforts. Other programs could take a similar approach by defining the intended use of their WBA system, using a rhetorical question–answer approach to generate critical questions and collect relevant evidence, and using the frameworks from Messick and Kane to organize evidence and identify critical gaps. Importantly, education leaders should keep in mind that validation of WBAs is a longitudinal process that incorporates multiple efforts, and they should not be discouraged if initial maps seem relatively empty. Empty space means opportunity for learning and scholarship.

## References

1. Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system— Rationale and Benefits. *N Engl J Med*. 2012;366:1051–1056.
2. Frank JR, Danoff D. The CanMEDS initiative: Implementing an outcomes-based framework of physician competencies. *Med Teach*. 2007;29:642–647.
3. Simpson JG, Furnace J, Crosby J, et al. The Scottish doctor—Learning outcomes for the medical undergraduate in Scotland: A foundation for competent and reflective practitioners. *Med Teach*. 2002;24:136–143.
4. Confederation of Postgraduate Medical Education Councils. Australian Curriculum Framework for Junior Doctors. Melbourne, Australia: CPMEC; 2012.  
[http://www.cpmec.org.au/files/27112017\\_watermark.pdf](http://www.cpmec.org.au/files/27112017_watermark.pdf). Accessed March 22, 2021.
5. Holmboe ES, Sherbino J, Englander R, Snell L, Frank JR; ICBME Collaborators. A call to action: The controversy of and rationale for competency-based medical education. *Med Teach*. 2017;39:574–581.
6. Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: Theory to practice. *Med Teach*. 2010;32:638–645.
7. Carraccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C. Shifting paradigms: From Flexner to competencies. *Acad Med*. 2002;77:361–367.
8. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach*. 2010;32:676–682.
9. Lockyer J, Carraccio C, Chan MK, et al; ICBME Collaborators. Core principles of assessment in competency-based medical education. *Med Teach*. 2017;39:609–616.
10. Lurie SJ. History and practice of competency based assessment. *Med Educ*. 2012;46:49–57.
11. Lurie SJ, Mooney CJ, Lyness JM. Commentary: Pitfalls in assessment of competency-based educational objectives. *Acad Med*. 2011;86:412–414.
12. Govaerts M, van der Vleuten CP. Validity in work-based assessment: Expanding our horizons. *Med Educ*. 2013;47:1164–1174.
13. Downing SM. Validity: On meaningful interpretation of assessment data. *Med Educ*. 2003;37:830–837.
14. St-Onge C, Young M, Eva KW, Hodges B. Validity: One word with a plurality of meanings. *Adv Health Sci Educ Theory Pract*. 2017;22:853–867.
15. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: A practical guide to Kane’s framework. *Med Educ*. 2015;49:560–575.
16. Messick S. Validity. In: Linn RL, ed. *Educational Measurement*. New York, NY: Macmillan Publishing; 1989:13–103.
17. Kane MT. Validating the interpretations and uses of test scores. *J Educ Meas*. 2013;50:1–73.
18. Warm EJ, Mathis BR, Held JD, et al. Entrustment and mapping of observable practice activities for resident assessment. *J Gen Intern Med*. 2014;29:1177–1182.
19. Warm EJ, Held JD, Hellmann M, et al. Entrusting observable practice activities and milestones over the 36 months of an internal medicine residency. *Acad Med*. 2016;91:1398–1405.
20. Rekman J, Gofton W, Dudek N, Gofton T, Hamstra SJ. Entrustability scales: Outlining their usefulness for competency based clinical assessment. *Acad Med*. 2016;91:186–190.
21. Crossley J, Johnson G, Booth J, Wade W. Good questions, good answers: Construct alignment improves the performance of workplace-based assessment scales. *Med Educ*. 2011;45:560–569.
22. Crossley J, Jolly B. Making sense of workbased assessment: Ask the right questions, in the right way, about the right things, of the right people. *Med Educ*. 2012;46:28–37.

23. Kelleher M, Kinnear B, Sall D, et al. A reliability analysis of entrustment-derived workplace-based assessments. *Acad Med*. 2020;95:616–622.
24. Toulmin S. *The Uses of Argument*. Cambridge, UK: Cambridge University Press; 1958.
25. Wigmore JH. *The Science of Judicial Proof, as Given by Logic, Psychology, and General Experience, and Illustrated in Judicial Trials*. Boston, MA: Little, Brown; 1937.
26. Kinnear B, Kelleher M, May B, et al. *Constructing a Validity Map for Workplace Based Assessments*. Paper presented at: Research Group for Pursuit of Validity Colloquium; Montreal, Canada; 2019.
27. Kane MT. Current concerns in validity theory. *J Educ Meas*. 2001;38:319–342.
28. Van Eemeren FH, Grootendorst R, Johnson RH, Plantin C, Willard CA. *Fundamentals of Argumentation Theory: A Handbook of Historical Backgrounds and Contemporary Developments*. New York, NY: Routledge; 2013.



# CHAPTER 3

## **How argumentation theory can inform assessment validity: A critical review**

Published as:

Kinnear B, Schumacher DJ, Driessen EW, Varpio L. How argumentation theory can inform assessment validity: A critical review. *Medical Education*. 2022 Nov;56(11):1064-75.

# **Abstract**

## **Introduction**

Many health professions education (HPE) scholars frame assessment validity as a form of argumentation in which interpretations and uses of assessment scores must be supported by evidence. However, what are purported to be validity arguments are often merely clusters of evidence without a guiding framework to evaluate, prioritise, or debate their merits. Argumentation theory is a field of study dedicated to understanding the production, analysis, and evaluation of arguments (spoken or written). The aim of this study is to describe argumentation theory, articulating the unique insights it can offer to HPE assessment, and presenting how different argumentation orientations can help reconceptualize the nature of validity in generative ways.

## **Methods**

The authors followed a five-step critical review process consisting of iterative cycles of focusing, searching, appraising, sampling, and analysing the argumentation theory literature. The authors generated and synthesised a corpus of manuscripts on argumentation orientations deemed to be most applicable to HPE.

## **Results**

We selected two argumentation orientations that we considered particularly constructive for informing HPE assessment validity: New rhetoric and informal logic. In new rhetoric, the goal of argumentation is to persuade, with a focus on an audience's values and standards. Informal logic centres on identifying, structuring, and evaluating arguments in real-world settings, with a variety of normative standards used to evaluate argument validity.

## **Discussion**

Both new rhetoric and informal logic provide philosophical, theoretical, or practical groundings that can advance HPE validity argumentation. New rhetoric's foregrounding of audience aligns with HPE's social imperative to be accountable to specific stakeholders such as the public and learners. Informal logic provides tools for identifying and structuring validity arguments for analysis and evaluation.

## Case study

*This case represents a real-world example of the importance of argumentation paradigms in HPE assessment validity. The case is referenced throughout the Results and Discussion sections.*

*A PGME programme transitioning to time-variable promotion decisions. A post-graduate medical education programme (PGME) has developed a robust programme of assessment over several years. Programme leaders have collected validity evidence for summative decisions about learner performance via literature review and new validation studies, and are hoping to transition to a time-variable training approach. Rather than being promoted to lesser degrees of supervision on a time-based schedule, summative promotion decisions will solely determine when and how learners progress through the programme. When engaging important stakeholders (e.g. learners, PGME leaders, departmental leaders, accreditation officials, certifying bodies) about this potential change, the programme is universally asked, “How do you know your summative promotion decisions are defensible?” In other words, “What is your argument that you can make the right promotion decision?” Before the programme could begin making their argument, they had several questions that needed to be answered. How should their argument be structured? By what standard would it be evaluated? Which audience(s) would levy judgement?*

*How can new rhetoric inform this case?*

*The programme leaders recognise that their validity arguments needs to resonate with each stakeholder group; therefore, their argument must respond to many different expectations. The programme leaders first identify which stakeholders are most important to serve as audiences, attempt to understand their values, and iteratively construct, evaluate, and refine their validity argument until it is acceptable to each stakeholder group. Accrediting and certifying bodies are interested in the outcomes and consequences of this specific PGME programme. Therefore, the programme's validity argument includes evidence that patients cared for by graduates receive high-quality care. In contrast, learners are concerned about equity and fairness in assessment practices. Therefore, the programme's validity argument includes quantitative and qualitative data on how different learner identity groups (e.g. race, ethnicity, gender identity, disability status) are assessed and whether there is evidence of bias against any groups.*

*How can informal logic inform this case?*

*Programme leaders collect significant amounts of validity evidence, but initially do not have it organised into a cogent argument for stakeholders to evaluate. Using Toulmin's model, the programme leaders signpost claims, data, warrants, rebuttals, and other salient elements, facilitating easier analysis by the audience. Programme leaders and stakeholders agree that the validity arguments will be evaluated in terms of relevance, acceptability, and sufficiency, and that Toulmin's model will be used to identify any aspects of the argument that are deemed inadequate.*

*When is an argument complete?*

*Both stakeholder groups found the programme's initial validity argument to be inadequate, asking for additional evidence and clarification on different argument aspects. Programme leaders worked iteratively with stakeholders to collect and evaluate evidence until both groups found the argument to be acceptable. Programme leaders could then focus resources on engaging other stakeholder groups in similar validity argumentation processes.*

## **1 | Introduction**

Many modern health professions education (HPE) scholars frame assessment validity as a form of argumentation: A claim about a specific interpretation and use of assessment data.<sup>1-4</sup> However, HPE scholars have yet to delve into the philosophical, theoretical, or practical aspects of argumentation that undergird validation. What are purported to be validity arguments are often merely the listing of evidence without a guiding framework to evaluate, prioritise, or debate their merits. This is akin to holding a courtroom trial while only knowing how to organise the evidence, but not knowing how the argument should be structured, who will evaluate the argument, nor how they will evaluate it. We contend that it is vital for HPE to develop a deeper understanding of argumentation theory to support our current understandings and uses of validity.

This is not to say that important insights into validity have not already been generated by HPE scholars. Far from it. Our field's understanding of validity has progressed thanks to the work of many HPE researchers. For instance, Cook, Hatala, and Downing have helped bring argument-based validity frameworks, such as those of Messick and Kane,<sup>5,6</sup> into the HPE assessment sphere.<sup>1-3,7-14</sup> St. Onge and Young have helped map out the current state of our field's understanding of validity.<sup>4,15-17</sup> Govaerts, Schuwirth, van der Vleuten, and others have progressed our fields view of validity in the context of programmatic assessment.<sup>18-26</sup> It is thanks to scholars like these, and others, that HPE has evolved to conceptualise validity as argument. And yet, HPE's use of contemporary validity frameworks lack an explicit description of the audiences, structures, and evaluation standards for validity arguments. This leaves important questions unanswered: How should validity arguments be evaluated and by whom? How should validity arguments be structured? Not knowing the answers to these questions leaves our field ill-equipped to fully embrace the validity-as-argument paradigm.

In this critical review of the literature, we offer a description of argumentation theory—tailored to the HPE audience—articulating the unique insights it can offer to HPE assessment, and presenting how different approaches to argumentation can help us reconceptualize the nature of validity in generative ways. We present two different orientations within argumentation theory and discuss how each creates different ways of understanding, structuring, and evaluating validity arguments. To make these abstract theories more tangible, we apply these approaches to an HPE-relevant case (Box 1). Via this case discussion, we highlight

how argumentation theories can help advance our assessment efforts. To frame our review, we first offer definitions of validity arguments and argumentation theory.

## **1.1 | Validity arguments within and beyond HPE**

Validity is not about assessment scores but rather the interpretations and uses of those scores.<sup>27-30</sup> Validation is the process of developing an argument to support those interpretations and uses.<sup>29</sup> Validity arguments are comprised of claims—i.e. assertions about the phenomena engaged with—that connect to form a defensible chain of reasoning. For instance, we make claims about how competence is defined,<sup>31</sup> how it can be observed,<sup>32</sup> and how it can be assessed.<sup>33</sup> However, these claims are not statements of fact; instead, they are assertions that can be debated, accepted, or refuted. If the claims within a validity argument are viewed as objective, static, and inactive declarations, it is because the argumentation strategies embedded therein have become so pervasively used and accepted that we no longer recognise them as contestable. The validity argument claims upon which HPE assessment decisions rest are merely assertions steeped in argumentation and so are up for debate.

The concept of validity as argument has existed and evolved for decades.<sup>34-36</sup> Although many HPE scholars operate using Messick's or Kane's conceptualization of validity argumentation,<sup>1,3,26,37-43</sup> scholars outside HPE present other approaches.<sup>44</sup> Bachman advocates validation work should entail an assessment use argument, which is comprised of an assessment validity argument linking assessment performance to an interpretation and an assessment utilisation argument linking interpretations to a decision.<sup>45,46</sup> Kane similarly proposes an interpretation/use argument that explicitly states the inferences and assumptions ingrained within use of assessment scores.<sup>6,47</sup> Mislevy describes an evidence-centered design approach to validation in which the arguments underpinning assessment are integrated into the design and implementation of assessment systems.<sup>48,49</sup> These three non- HPE scholars all employ argument structures put forth by Stephen Toulmin,<sup>27,29,45,48-51</sup> but they only superficially address the underlying argumentation orientation utilised for validation. Unfortunately, Toulmin's structures—much less argument orientations—have yet to enter the discourse of HPE validation. This omission generates problems for HPE. Embedded within HPE's validity arguments are assumptions and ways of thinking that we often fail to recognise and critique. This failure risks assuming that the validity arguments are objective truths and not structures for reasoning that bring certain ideologies to life. In other words, the arguments entrenched in our assessment decisions and in our conceptualizations of their validity must be made explicit so that we can deliberately consider when and how each decision is valid. Argumentation theory can help us achieve that clarity.

## **1.2 | Argumentation theory**

Argumentation theory is not a single unified theory; instead, it is a field of study that draws from multiple disciplines (e.g. Logic, Speech, Linguistics, Philosophy, Psychology, and Law) to grapple with the production,

analysis, and evaluation of argumentative discourse (spoken or written).<sup>52,53</sup> Several orientations exist, with scholars from a wide range of academic disciplines adding unique threads to the tapestry of argumentation theory. Van Eemeren et al. provide a definition of argumentation that includes several aspects that are key to argumentation theories:

*Argumentation is a verbal and social activity of reason aimed at increasing (or decreasing) the acceptability of a controversial standpoint for the listener or reader, by putting forward a constellation of propositions intended to justify or refute the standpoint before a rational judge.*<sup>52</sup>

This conceptualization highlights that argumentation is an endeavour (activity) that relies on language (verbal) being exchanged between people (social). This communication is goal oriented (aimed); the goal is to convince an audience (single listener or reader or groups of listeners or readers) to agree (increase acceptability) or disagree (decrease acceptability) with a particular viewpoint or interpretation (standpoint). Importantly, argumentation exists in social situations where the viewpoint being espoused is disputable (controversial) by an audience. Therefore, arguments (constellations of propositions) are created to present a particular interpretation of the viewpoint (justify or refute the standpoint) that will convince the audience (rational judge) to align themselves with the arguer. Without a shared conception of processes and standards, argumentation is futile.<sup>52</sup>

Many different theories of argumentation propose how arguments could be realised. Some foreground the use of deductive reasoning,<sup>54</sup> whereas others accentuate the power of dialogue.<sup>55</sup> Some stress normative standards,<sup>56</sup> whereas others emphasise rhetorical aims.<sup>57</sup> Furthermore, argumentation scholars debate and revise each of these theories, offering different interpretations. Clearly, a review of this entire body of scholarship would not readily offer practical insights for advancing HPE's thinking about validity. However, a more targeted review of argumentation theories that can offer important considerations and concrete applications that could be harnessed by HPE's assessors and assessment researchers. Therefore, we selected theories and scholarly interpretations thereof that could be usefully applied to validity argumentation in HPE.

## 2 | Methods

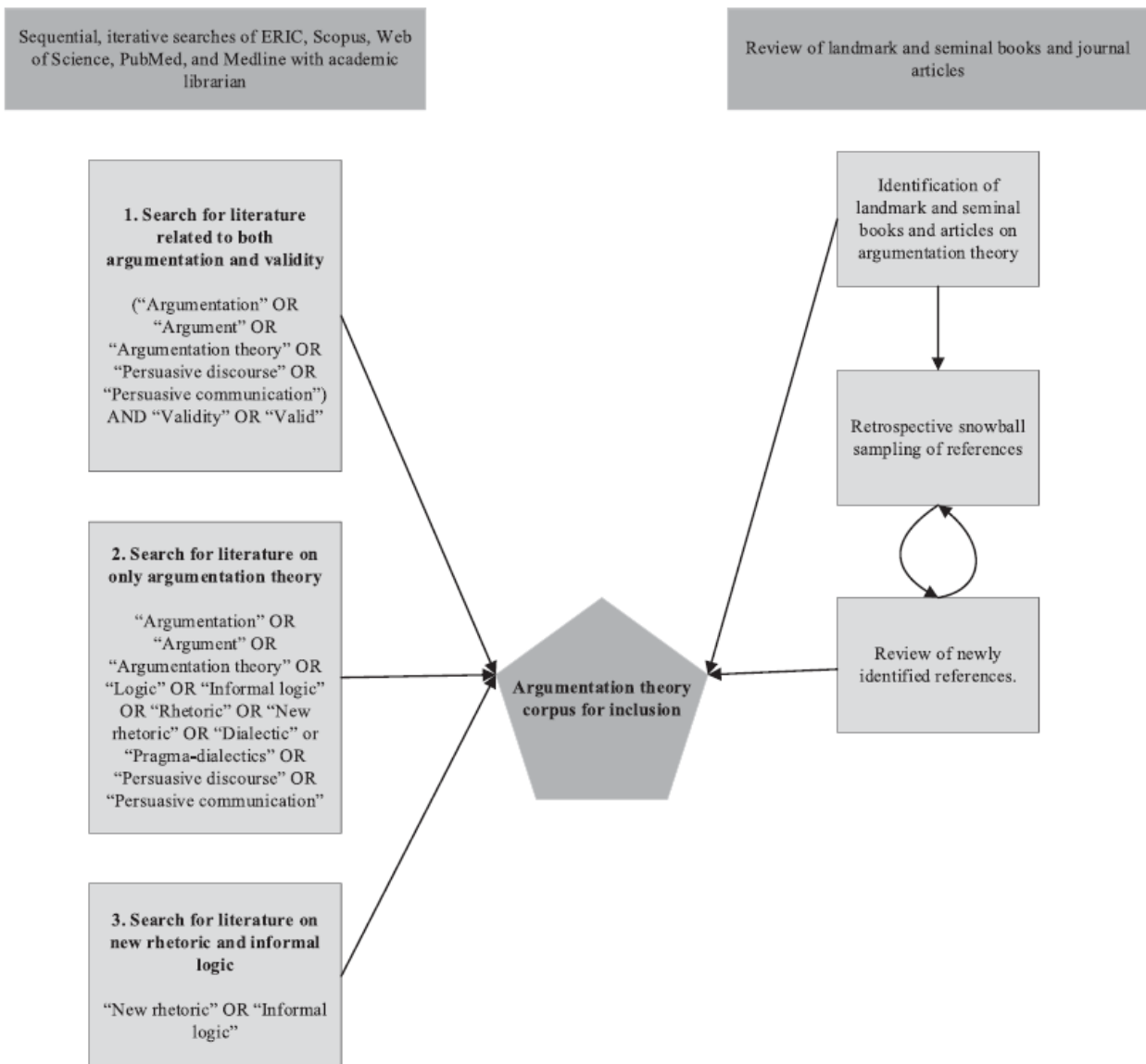
Our investigation asked: How can argumentation theory help us conceptualise the nature of HPE assessment validity in new and generative ways? To answer this question, we conducted a critical review. Critical reviews are not intended to produce generalizable truths<sup>58</sup> like systematic reviews. Rather, critical reviews are rooted in a constructivist ontology and epistemology<sup>59</sup> and “draw on literature and theory from different domains to re-envision current ways of interpreting the problem.”<sup>60</sup> Instead of aiming to synthesise all knowledge relevant to a particular topic, the research team conducts a critical review and subjective interpretation of a body of literature. The research team thus “acts as a research instrument, using their perspective to appraise and interpret the literature uncovered.”<sup>60</sup> This approach aligns with our goal of describing a selection of

argumentation theories that can advance HPE by examining the assumptions and preconceptions around assessment validity. We followed the five-step critical review process detailed by Kahlke et al to generate a corpus of theories and to analyse that corpus.<sup>60</sup> We engaged in iterative cycles of focusing (i.e. constructing/revising the research aims); searching (i.e. targeted exploration for sentinel perspectives in argumentation theory); appraising (i.e. assessing manuscripts for relevance to our aims); sampling (i.e. determining manuscripts potential to offer insights into our research); and analysing (i.e. assessing each theory's applicability to validity argumentation).

Our search strategy (Figure 1) began by locating seminal works in the field of argumentation theory by identifying highly cited peer reviewed journal articles that were written by authors considered leaders in the field of argumentation theory. Given that argumentation theory sits in the humanities, we recognised that, although peer reviewed journal articles would be informative, we would also have to broaden our scope to books and textbooks since these are highly valued forms of dissemination in this domain. We reviewed these manuscripts and mapped the leading strains of thought in argumentation theory that addressed theoretical and/or practical aspects of argumentation from several points in the field's history. We used our expertise in argumentation and HPE assessment validity to “appraise papers for inclusion based on their sense of a source's relevance to the research question and the value added by the information it contains.”<sup>60</sup> In parallel to our review of seminal works, we worked with an academic librarian to search ERIC, Scopus, Web of Science, PubMed, and Medline databases because they index a wide range of literatures—i.e. both journal articles and books—addressing humanities- and education-related publications. We searched for highly cited books and peer-reviewed journal articles that were not included in our original corpus. We used snowball sampling retrospectively (i.e. investigating references listed in seminal works) to identify additional key references.

BK and LV reviewed the corpus to independently create lists of the pre-eminent theories and scholars that have shaped the field of argumentation. These lists were not intended to be comprehensive, but rather to capture the most influential voices and viewpoints within argumentation. We then compared and resolved these into a single list through consensus discussions. We reviewed sources from the corpus that aligned with the consensus list to determine which argumentation theories were most applicable to HPE.

Specifically, we considered which theories would provide beneficial insights to HPE scholars engaged in assessment validation, both in the context of practical validation work and broader scholarly endeavours. Table 1 provides a summary of this work, presenting prominent orientations within argumentation theory which we considered, though some we did not fully discuss in our findings.



**Figure 1** - An iterative process for building an argumentation theory corpus for analysis

## 2.1 | Reflexivity

The critical review research team “acts as a research instrument, using their perspectives to appraise and interpret the literature uncovered.”<sup>60</sup> With this goal in mind, our unique perspectives, areas of expertise, and personal backgrounds informed the synthesis activity. Our team consisted of medical education scholars with diverse interests and experiences. Two team members (BK and DS) are clinician educators who actively engage in HPE research. The study's primary researcher (BK) immersed himself in the argumentation theory literature as part of his PhD in HPE; he also has expertise in competency-based medical education (CBME), learner assessment and validity. The other clinician educator (DS) has expertise in CBME and assessment validity. ED is a PhD-trained researcher with expertise in learner assessment, validity, and CBME. The senior researcher on this paper (LV) is a PhD-trained researcher whose graduate work focused on argumentation

theory, rhetoric, and professional communication. Her insights helped the team to identify significant scholars and theories in argumentation and to reflect on how these scholars and theories aligned with other approaches to validity espoused in HPE. Team research discussions often involved debates about which argumentation theories could be usefully applied to concerns about validity in HPE.

Methodological reflexivity also guided our research. By reviewing the theories from several authors and the common assumptions held within and across schools of argumentation theory, we selected theories we thought held the most generative possibilities for HPE assessment validity. We sought to include a broad representation of argumentation theories, but the guiding principle primarily informing our selections was: Can this theory substantially help us understand validity arguments in the HPE context?

### **3 | Results**

In Table 1, we outline key aspects of five prominent theories of argumentation, along with our assessments of each theory's applicability to HPE's validity arguments. Through our iterative cycles of analysis, we determined that two orientations—new rhetoric and informal logic—offered particularly useful insights into HPE assessment validity. This is not to say that other argumentation orientations are without value for HPE; instead, based on our expertise, we suggest that new rhetoric and informal logic are especially applicable to assessment validity in our field. We first review these orientations, then focus on three aspects of argumentation that we believe are relevant to validity argumentation in HPE: audience, argument structure, and evaluation standards.

**TABLE 1** Prominent argumentation theories and their relevance to health professions education

Orientation	Description	Influential scholars	Applicability to validity arguments in health professions education (HPE)
Formal logic <sup>61,62</sup>	Arguments are abstracted from their real-world environments and presented in a standard form of premises and conclusions (called syllogisms) for purposes of deductive reasoning. The goal of logic is to ensure that conclusions necessarily flow from premises.	Aristotle Antoine Arnauld Pierre Nicole John Stuart Mill George Boole	Formal logic statements are disconnected from real-world circumstances, which is antithetical to the highly contextual complex adaptive systems that constitute HPE. Therefore, we deem this theory to be poorly aligned with HPE's context.
Informal logic	Arguments are evaluated with real-world contextual factors included. There is an emphasis on identifying, structuring, and evaluating arguments against normative standards such as relevance, sufficiency, acceptability, and defensibility.	J. Anthony Blair Trudy Govier Ralph Johnson Charles Hamblin Howard Kahane Stephen Toulmin Douglas Walton John Woods	Informal logic provides tools for identifying and structuring arguments for analysis and evaluation without removing context. HPE validity arguments are often complex and not clearly articulated. Therefore, we decided that informal logic could provide useful approaches to HPE validity argumentation.
New rhetoric	Arguments are aimed to persuade and gain an audience's adherence to the arguer's claim. Audience values and norms define the acceptability of an argument. Audiences can be particular or universal and should be clearly identified by arguers.	James Crosswhite Lucie Olbrechts-Tyteca Chaim Perelman	New rhetoric foregrounds the imperative to persuade audiences. Given that HPE is comprised of several different audiences each with unique expectations vis a vis validity, we decided that new rhetoric could offer important insights to HPE.
Formal dialectics <sup>63-65</sup>	Argumentation centres around the resolution of (mostly verbal) disputes between specific interlocutors. A highly-structured normative system of rules guides critical dialogue should unfold. The goal is to defeat the other interlocutor while adhering to the discursive rules.	Roland Barth Charles Hamblin Erik C. W. Krabbe	Formal dialects and pragma-dialectics are a poorer fit given the highly structured rules of engagement and the oppositional nature of arguer and audience (i.e. the goal is to "win" the argument).
Pragma-dialectics <sup>52,66</sup>	Argumentation centres around the resolution of disputes between specific interlocutors. Ten discursive rules govern argumentation, which unfolds in four stages: confrontation, opening, argumentation, and concluding. The goal is to defeat the other interlocutor while adhering to the discursive rules.	Rob Grootendorst Frans van Eemeren	

### 3.1 | New Rhetoric

In the mid-20th century two philosophers, Chaim Perelman and Lucie Olbrechts-Tyteca, developed a theory of argumentation that relied more on value judgements than empirical proof.<sup>67-70</sup> New rhetoric was so named because of similarities to Aristotle's rhetoric, particularly the central importance Aristotle placed on audience and persuasion in argumentation.<sup>61,71</sup> In new rhetoric, the goal of argumentation is not the demonstration of absolute truth but rather to gain an audience's acceptance of a claim being put forth by an arguer.<sup>53,72</sup> Argument validity is dependent upon audience persuasion. Arguers must account for and adapt to an audience's values and beliefs, thereby linking validity to the ability to convince an audience.<sup>57</sup> Validity exists as a matter of degree since an audience might be partially persuaded towards an arguer's claim. In this sense, new rhetoric does not provide an external normative framework for determining argument validity; instead, it relies on an audience to use their own values and norms to determine what is acceptable and plausible. Persuasion does not remove the need for rational argument analysis; it places rationality in the context of an audience's values. Importantly, persuasion is not meant as a way to deceive or manipulate,<sup>73</sup>

but rather to consider and appraise a particular audience's principles and anticipated standards as evaluative criteria.

Given the centrality of an audience's values and standards to validity, arguers must make explicit their intended audience. Perelman and Olbrechts-Tyteca distinguish between arguing to a particular audience and a universal audience. A particular audience consists of a real-world person or group (or representative thereof) to whom an arguer can address their argument.<sup>74</sup> Although the particular audience may only be a subset or sample of a larger group, the argument still occurs using a concrete person's or group's values as the standard. In contrast, the universal audience is conceptual, conceived by the arguer as a group holding the values and ideals of competent and reasonable people. Thus, with a universal audience, the arguer plays a role in determining the values and standards of validity based on their notion of a reasonable audience. There is debate over the specific roles of particular and universal audiences in new rhetoric<sup>74-77</sup>; however, with both particular and universal audience, the arguer's goal is securing audience adherence. New rhetoric compels arguers to explicitly identify their audience, and incentivizes arguers to possess knowledge and understanding of their audience.

Practical implications of New Rhetoric for assessment practices in HPE include the need for programme leaders to tailor their validity arguments to meet the needs of each group they are engaging. When addressing accrediting and certifying bodies, arguments might emphasise validity evidence focused on outcomes and consequences of training. When addressing learners, arguments might emphasise equity and fairness in assessment practices.

### **3.2 | Informal logic**

Informal logic provides procedures and standards for identifying (via structures) and evaluating (via normative standards) arguments that occur in real-world, social environments and public discourse while still valuing logical coherence between premises and conclusions.<sup>56,78</sup> Informal logic is a broad umbrella term covering differing approaches to determining argument validity.<sup>79,80</sup> Before the advent of this theory, formal logic was applied to real-world arguments in a sterile, decontextualized way. Many informal logic scholars rejected this context-free approach because it was fraught with potential distortions that rendered evaluating real-world argumentation impossible. Informal logic was developed to account for the messy, implicit, incomplete nature of real-world argumentation through identification of argument structures without removal from contextual factors. Once an argument's structure is contextually understood, informal logic offers multiple standards for determining the strength of inferences in the argument.<sup>56</sup> One approach relies on the criteria of relevance, acceptability, and sufficiency to evaluate arguments.<sup>81,82</sup> Relevance demands adequate relationship between premises (i.e. claims or propositions), conclusions inferred from these premises, and the overall argument. Acceptability relates to the truth or plausibility of premises as determined by the arguer, audience, and critical community. Sufficiency requires that evidence supporting an argument be appropriate

in type, quantity, and use. Other approaches for determining the strength of an argument in informal logic include excluding fallacies<sup>83,84</sup> (i.e. errors of reasoning), using counterexamples, and formulating argumentation schemes.<sup>85</sup> No one approach is singularly correct for determining an argument's validity; instead, informal logic foregrounds argument strength and cogency as the goal of argumentation.

Philosopher Stephen Toulmin profoundly influenced informal logic with his 1958 book *The Uses of Argument*<sup>86</sup> in which he provided a framework for organising argument structures to facilitate analysis and evaluation.<sup>87,88</sup> In Toulmin's model, arguments are deconstructed into unique component parts of a larger cohesive unit. A claim is a standpoint or opinion on a particular matter taken by an arguer which serves as the point of departure for argumentation. Claims are supported by evidence which Toulmin labels data. However, an arguer must elaborate on how or why particular data support the stated claim. This elaboration is known as a warrant, which provides justification for use of particular data to substantiate a claim. In other words, a warrant provides a bridge between a claim and its supporting data.<sup>52</sup> The data-warrant-claim unit forms the backbone of Toulmin's argumentation structure, though there are other components such as backings (evidence to support a warrant), qualifiers (limitations or restrictions on the universality of a claim), and rebuttals (pre-emptive or reactive responses to counterarguments).

In Toulmin's model, validity hinges on demonstrating that the warrant adequately justifies the leap from data to claim. He believed that similar argument structures could be used regardless of the relevant discipline, or field.<sup>86</sup> Thus, argument structure and the general procedure by which arguments could be analysed are field independent. However, Toulmin believed that many aspects of validity are field dependent, meaning that arguments should be evaluated using the norms, values, and criteria of the relevant field. To that end, Toulmin uses terms such as defensible, acceptable, and sound when describing validity. The audience who levies such determinations can be a specific person or group participating in the argument or an onlooker who is not involved in the argument but can make a judgement nonetheless.

Practical implications of informal logic for assessment practices in HPE include training programme leaders clearly signposting the claims, data, warrants, and backings of their argument, allowing stakeholders to more clearly identify the inferences needing scrutiny to determine the validity of time-variable promotion decisions. Accepted claims then become data or warrants for other arguments related to the programme of assessment informing promotion decisions. Thus, a clearly labelled web of argumentation could be formed to support time-variable decisions. This web would be field dependent, meaning that it would be evaluated using acceptability and soundness criteria that are relevant in HPE, though not necessarily in other disciplines.

## 4 | Discussion

We contend that informal logic and new rhetoric are theories of argumentation that hold promise to deepen our understanding of validity argumentation in HPE. Both approaches acknowledge the importance of

context in argumentation and align with Zumbo's conceptualization of validity as contextualised and pragmatic.<sup>89,90</sup> We are not asserting that HPE's validity practices should be wedded to any particular orientation; instead, we propose that our field could benefit from a useful, reciprocal relationship between HPE validation practices and established argumentation theories.<sup>91</sup> For example, Cook and Hatala provide an excellent practical guide for undertaking validation work in HPE using Kane's framework.<sup>8</sup> We suggest enhancing this guide by including considerations of to whom an argument should be directed (audience), how to organise and signpost arguments within and between inferences (structure), and/or how an argument should be evaluated (standards). With regard to the latter, Cook and Hatala use language that seems to invoke both new rhetoric and informal logic, noting that validity arguments attempt to “persuade others” while also noting that the “relevance, quality, and breadth” of evidence is important. Just as there are few definitive boundaries in the field of argumentation, with scholars within and across traditions debating the meaning, scope, and utility of various orientations, we suggest that HPE should not silo itself into one argumentation orientation. HPE researchers can draw from various traditions based on which will be most helpful to move our field's understanding of validity forward. Therefore, we believe HPE would benefit from using aspects of both new rhetoric and informal logic to deepen our understanding of audience, argument structure, and evaluation standards in validity argumentation.

## **4.1 | New rhetoric to define HPE validity argumentation's audience**

A new and growing discourse in HPE is that of “validity as social imperative,”<sup>4,17</sup> which emphasises accountability to society and learners. In other words, HPE validity arguments centre on the needs of our stakeholder audiences. Informal logic recognises the importance of audience in argumentation, but not to the same degree as new rhetoric.<sup>92</sup> New rhetoric places audience values at the centre of argumentation, and may help us better satisfy the social imperative that HPE is being called to fulfil by better engaging both stakeholders and the broader HPE community.

As mentioned previously, new rhetoric arguments can be directed towards a particular audience or a universal audience. For HPE to adopt a new rhetoric approach to validity argumentation, we would need to better define our audiences. Currently, many formal HPE validity arguments are made in the form of peer-reviewed publications, with journal reviewers, editors, and readers serving as audiences. Journal readers may be the most important audience in the publication world. However readers represent a form of noninteractive audience<sup>93</sup> (with the exception of the very few who write letters or commentaries), which limits the degree to which their values can be known or their judgements shared. Reviewers and editors could serve as a particular audience representing other HPE scholars, or they could be considered a representation of a universal audience because they are most likely to represent competent and reasonable people with sufficient knowledge to levy informed judgements. They could, in fact, serve both roles. However, currently reviewers and editors are not necessarily making explicit judgements on whether a validity argument is acceptable or plausible, and certainly not within the context of a given institution or programme. Rather, they

are making a determination of whether a manuscript, in toto, moves the broader scholarly discussion forward in the HPE community. Reviewers and editors could, in fact, find a validity argument to be inadequate but still recommend publication because the manuscript provides some benefit to the journal's readers, perhaps in terms of innovative thinking, methodological developments, or novel conclusions. Thus, new rhetoric would compel reviewers and editors—were they to be identified as a particular audience—to explicitly judge the plausibility of validity arguments submitted for publication, perhaps with a way to signal their judgement to readers upon publication.

The limited audience interaction that occurs with publication would oblige HPE to expand validity argumentation beyond the pages of journals. HPE has several stakeholders in assessment decisions including learners, programme leaders, accreditors, certifiers, institutions, and payors. Many of these stakeholders may never read the articles in which validity evidence relevant to their context is published. Assessment designers would need to ensure that arguments are plausible to these particular audiences, likely via mechanisms that do not involve peer-reviewed publication. Whereas informal logic presents argument as a product, new rhetoric entreats argumentation to be an activity, or service, that an arguer develops with an audience.<sup>53</sup> This activity-based orientation aligns with Cook and Hatala's description of validation as a process,<sup>8</sup> and with the concept of coproduction, which is gaining traction HPE.<sup>94,95</sup> In coproduction, the consumers' knowledge, experience, and opinions influence the creation of a service rather than passive consumption of a product.<sup>96,97</sup> Adopting a lens of new rhetoric could the coproduction of arguments over time with validity arguers and audiences (i.e. stakeholders) working to best meet the needs of those who HPE aims to serve.

## **4.2 | Informal logic to clarify HPE validity argumentation structure**

HPE scholars regularly organise validity evidence using frameworks from Messick and Kane,<sup>2,3</sup> but do not make full argument structure explicit. Indeed, Kane invokes Toulmin's argumentation structure (warrants, data, backings, rebuttals),<sup>27,29,47,50</sup> though this model has mostly appeared in other fields such as language testing.<sup>98</sup> Toulmin's model exists as a cryptid in HPE, rarely (if ever) surfacing within our corpus. Kane's framework requires arguers to clearly state an intended interpretation and use of assessment data,<sup>6,29</sup> which serves as the argument claim. However, rather than simply using Kane's chain of inferences, HPE scholars could make explicit the full structure of their arguments to augment analysis, particularly for high-stakes decisions that require significant scrutiny. In fact, there are likely to be several arguments embedded within Kane's chain of inferences,<sup>28</sup> which may go overlooked without explicit and structured reporting. Complex validity arguments that are likely to be found in programmatic assessment could be mapped out, showing how an established claim becomes data or a warrant for a subsequent argument in a longer chain of argumentation. Though there is debate amongst validity scholars as to how much structure the validation process should have,<sup>44</sup> currently HPE has no clear or agreed-upon approach for organising, identifying, or signposting validity arguments. Often what are labelled “arguments” are collections of data that lack an

explicit argument structure. Toulmin's model could help to present, interpret, prioritise, and argue the evidence to stakeholders.

### **4.3 | Informal logic or new rhetoric to elucidate standards for evaluating validity arguments**

Neither informal logic nor new rhetoric claim to provide a universal truth, but they could inform when seemingly unending validation efforts have reached defensible stopping points (at least until new arguments are put forth or new audiences are considered). New rhetoric seeks persuasion and audience adherence. Informal logic seeks argument cogency, often in terms of relevance, sufficiency, and acceptability as defined by the arguer's discipline or field. Both lenses would work well within HPE; therefore, choosing which to use may depend on one's philosophical worldview. Informal logic may align well with post-positivist worldviews,<sup>99,100</sup> whereas new rhetoric likely aligns better with constructionist or critical realist views.<sup>101,102</sup> However, informal logic and new rhetoric are not restricted to any specific worldview; each will manifest in different ways when employed with different philosophical beliefs. Given the heterogeneous philosophical worldviews in HPE, is it reasonable to expect one argumentation orientation to fit all arguers and audiences? Likely not. But we should explicitly acknowledge our argumentation paradigms when undertaking validity work, and specifically consider our audience's as well. Doing so can help avoid misunderstanding of worldviews and argumentation paradigms that could cause conflict in HPE validation work.<sup>103</sup> Tavares et al described a compatibility principle in assessment as “the obligation to recognize that different philosophical positions can exist between and within assessment plans and that these positions commit assessment designers to particular ideas and assumptions.”<sup>104</sup> We believe the compatibility principle also applies to the philosophies of argumentation woven into HPE validity.

### **4.4 | Remaining challenges of using argumentation in HPE assessment validation**

Though we believe informal logic and new rhetoric are useful orientations for HPE validity argumentation, several challenges must still be addressed. First, little is known about which argumentation orientations resonate with HPE stakeholders such as learners, programme leaders, accreditors, and certifiers. Future research could begin to unearth the latent argumentation assumptions that these groups may carry. Second, catering to the values of multiple audiences could make validity argumentation a daunting task in a new rhetoric paradigm. Each stakeholder group likely has different values and understandings of assessment and validity. It is unclear if educators can develop a validity argument that is unique to each group, or a single argument that is persuasive and acceptable to all. Third, we must consider and study the impact of infusing informal logic or new rhetoric into validation on hegemony and equity within HPE, particularly with regard to audience values and evaluative standards. Addey et al call for “democratic spaces” in which “legitimately

diverse arguments and intentions can be recognized, considered, assembled, and displayed.”<sup>105</sup> Creating such democratic spaces can ensure that all stakeholder voices are heard. Finally, although we believe Toulmin's argumentation model could allow for improved analysis and evaluation of HPE validity arguments, empirical studies are needed to test this hypothesis. Until the HPE community explores these limitations and challenges, assessment validity argumentation will remain a black box for programmes such as the one in our case study.

## **4.5 | Limitations**

Though our author group selected informal logic and new rhetoric as most useful for HPE validity argumentation, there are several other argumentation orientations that we did not select (a small selection is presented in Table 1). We acknowledge that other scholars might find these other orientations more useful to HPE, particularly if they value discursive structure (formal dialectics and pragma-dialectics) or have positivist worldviews (formal logic). We also acknowledge that given the long and rich history of argumentation theory, it was impossible for us to fully review every orientation and approach therein.

## **4.6 | Conclusion**

HPE scholars have published robust validity evidence<sup>13,38-40,106</sup> for various assessment decisions organised within the commonly used frameworks of Messick<sup>5</sup> or Kane.<sup>27</sup> Although these provide excellent examples of how to organise validity evidence, they do not explicitly describe who should evaluate the evidence (audience), what structure the argument (not just the evidence) should take, and what criteria should be used for evaluation. This omission propagates the notion that evidence equals argument and that a validity judgement has been rendered by simply laying out the evidence. In this manuscript, we have detailed how informal logic and new rhetoric can help advance HPE's ongoing work with validity in assessment. Each theory offers affordances for transforming ambiguous, inert HPE assessment validity evidence into clearer, animate validity arguments.

## References

1. Cook DA. When I say... validity. *Med Educ*. 2014;48(10):948-949. doi:10.1111/medu.12401
2. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006; 119(2):e167-e116. doi:10.1016/j.amjmed.2005.10.036
3. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ*. 2015;49(6):560-575. doi:10.1111/medu.12678
4. St-Onge C, Young M, Eva KW, Hodges B. Validity: one word with a plurality of meanings. *Adv Health Sci Educ Theory Pract*. 2017;22(4):853-867. doi:10.1007/s10459-016-9716-3
5. Messick S. Validity. In: Linn R, ed. *Educational Measurement*. New York: Macmillan Publishing; 1989:13-103.
6. Kane M. The argument-based approach to validation. *School Psychol Rev*. 2013;42(4):448-457. doi:10.1080/02796015.2013.12087465
7. Cook DA, Lineberry M. Consequences validity evidence: evaluating the impact of educational assessments. *Acad Med*. 2016;91(6):785-795. doi:10.1097/ACM.0000000000001114
8. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Adv Simul (Lond)*. 2016;1(1):31. doi:10.1186/s41077-016-0033-y
9. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830-837. doi:10.1046/j.1365-2923.2003.01594.x
10. Downing SM, Haladyna TM. Validity threats: overcoming interference with proposed interpretations of assessment data. *Med Educ*. 2004;38(3):327-333. doi:10.1046/j.1365-2923.2004.01777.x
11. Downing SM, Yudkowsky R. *Assessment in Health Professions Education*. New York, NY: Routledge; 2009. doi:10.4324/9780203880135
12. Cook DA, Kuper A, Hatala R, Ginsburg S. When assessment data are words: validity evidence for qualitative educational assessments. *Acad Med*. 2016;91(10):1359-1369. doi:10.1097/ACM.0000000000001175
13. Hatala R, Cook DA, Brydges R, Hawkins R. Constructing a validity argument for the Objective Structured Assessment of Technical Skills (OSATS): a systematic review of validity evidence. *Adv Health Sci Educ*. 2015;20(5):1149-1175. doi:10.1007/s10459-015-9593-1
14. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Adv Health Sci Educ*. 2014;19(2):233-250. doi:10.1007/s10459-013-9458-4
15. St-Onge C, Young M. Evolving conceptualisations of validity: impact on the process and outcome of assessment. *Med Educ*. 2015;49(6):548-550. doi:10.1111/medu.12734
16. Young M, St-Onge C, Xiao J, Lachiver EV, Torabi N. Characterizing the literature on validity and assessment in medical education: a bibliometric study. *Persp Med Educ*. 2018;7(3):182-191. doi:10.1007/s40037-018-0433-x
17. Marceau M, Gallagher F, Young M, St-Onge C. Validity as a social imperative for assessment in health professions education: a concept analysis. *Med Educ*. 2018;52(6):641-653. doi:10.1111/medu.13574
18. Wilby KJ, Govaerts MJ, Dolmans DH, Austin Z, van der Vleuten C. Reliability of narrative assessment data on communication skills in a summative OSCE. *Patient Educ Couns*. 2019;102(6):1164-1169. doi:10.1016/j.pec.2019.01.018
19. de Jonge LP, Timmerman AA, Govaerts MJ, et al. Stakeholder perspectives on workplace-based performance assessment: towards a better understanding of assessor behaviour. *Adv Health Sci Educ*. 2017;22(5):1213-1243. doi:10.1007/s10459-017-9760-7
20. van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ*. 2005;39(3):309-317. doi:10.1111/j.1365-2929.2005.02094.x

21. van der Vleuten CP, Schuwirth LW, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Med Teach*. 2012;34(3):205-214. doi:10.3109/0142159X.2012.652239
22. Govaerts M, van der Vleuten CP. Validity in work-based assessment: expanding our horizons. *Med Educ*. 2013;47(12):1164-1174. doi:10.1111/medu.12289
23. Driessen EW, Overeem K, Van Tartwijk J, Van Der Vleuten CP, Muijtjens AM. Validity of portfolio assessment: which qualities determine ratings? *Med Educ*. 2006;40(9):862-866. doi:10.1111/j.1365-2929.2006.02550.x
24. Schuwirth LW, Van der Vleuten CP. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach*. 2011;33(6):478-485. doi:10.3109/0142159X.2011.565828
25. Bok HG, de Jong LH, O'Neill T, Maxey C, Hecker KG. Validity evidence for programmatic assessment in competency-based education. *Persp Med Educ*. 2018;7(6):362-372. doi:10.1007/s40037-018-0481-2
26. McGill DA, van der Vleuten CP, Clarke MJ. Construct validation of judgement-based assessments of medical trainees' competency in the workplace using a "Kanesian" approach to validation. *BMC Med Educ*. 2015;15(1):237. doi:10.1186/s12909-015-0520-1
27. Kane MT. An argument-based approach to validity. *Psychol Bull*. 1992;112(3):527-535. doi:10.1037/0033-2909.112.3.527
28. Wools S, Sanders P, Eggen T. Evaluation of validity and validation by means of the argument-based approach. *Eval Valid Validation Means Argument-Based Approach*. 2010;(1):1000-1020. doi:10.3280/CAD2010-001007
29. Kane MT. Validating the interpretations and uses of test scores. *J Educ Meas*. 2013;50(1):1-73. doi:10.1111/jedm.12000
30. LeBaron Wallace T. An argument-based approach to validity in evaluation. *Evaluation*. 2011;17(3):233-246. doi:10.1177/1356389011410522
31. Frank JR, Mungroo R, Ahmad Y, Wang M, De Rossi S, Horsley T. Toward a definition of competency-based education in medicine: a systematic review of published definitions. *Med Teach*. 2010;32(8):631-637. doi:10.3109/0142159X.2010.500898
32. Kogan JR, Hatala R, Hauer KE, Holmboe E. Guidelines: the do's, don'ts and don't knows of direct observation of clinical skills in medical education. *Perspect Med Educ*. 2017;6(5):286-305. doi:10.1007/s40037-017-0376-7
33. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR, Collaborators IC. The role of assessment in competency-based medical education. *Med Teach*. 2010;32(8):676-682. doi:10.3109/0142159X.2010.500704
34. Cronbach LJ. Five perspectives on validity argument. *Test Validity*. 1988;3-17.
35. House ER. *The Logic of Evaluative Argument*. Center for the Study of Evaluation, UCLA Graduate School of Education; 1977.
36. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull*. 1955;52(4):281-302. doi:10.1037/h0040957
37. Hodwitz K, Tays W, Reardon R. Redeveloping a workplace-based assessment program for physicians using Kane's validity framework. *Can Med Educ J*. 2018;9(3):e14-e24. doi:10.36834/cmej.42286
38. Hess BJ, Kvern B. Using Kane's framework to build a validity argument supporting (or not) virtual OSCEs. *Med Teach*. 2021;1-6(9):999-1004. doi:10.1080/0142159X.2021.1910641
39. Tavares W, Brydges R, Myre P, et al. Applying Kane's validity framework to a simulation based assessment of clinical competence. *Adv Health Sci Educ*. 2018;23(2):323-338. doi:10.1007/s10459-017-9800-3
40. Kinnear B, Kelleher M, May B, et al. Constructing a validity map for a workplace-based assessment system: cross-walking Messick and Kane. *Acad Med*. 2021;96(7S):S64-S69. doi:10.1097/ACM.0000000000004112

41. Tavares W, Gofton W, Bhanji F, Dudek N. Reframing the OSCORE as a retrospective supervision scale using validity theory. *J Grad Med Educ.* 2022;14(1):22-24. doi:10.4300/JGME-D-21-00592.1
42. Hamstra SJ, Cuddy MM, Jurich D, et al. Exploring the association between USMLE scores and ACGME milestone ratings: a validity study using national data from emergency medicine. *Acad Med.* 2021;96(9):1324-1331. doi:10.1097/ACM.0000000000004207
43. Hamstra SJ, Yamazaki K. A validity framework for effective analysis and interpretation of milestones data. *J Grad Med Educ.* 2021;13(2s):75-80. doi:10.4300/JGME-D-20-01039.1
44. Lavery MR, Bostic JD, Kruse L, Krupa EE, Carney MB. Argumentation surrounding argument-based validation: a systematic review of validation methodology in peer-reviewed articles. *Educ Meas Issues Pract.* 2020;39(4):116-130. doi:10.1111/emip.12378
45. Bachman LF. Building and supporting a case for test use. *Lang Assess Q.* 2005;2(1):1-34. doi:10.1207/s15434311laq0201\_1
46. Bachman LF. Constructing an assessment use argument and supporting claims about test taker-assessment task interactions in evidence-centered assessment design. *Measurement-Lawrence Erlbaum Associates.* 2003;1:63-65.
47. Kane M. Validating score interpretations and uses: Messick lecture, language testing research colloquium, Cambridge, April 2010. *Lang Testing.* 2012;29(1):3-17. doi:10.1177/0265532211417210
48. Mislevy RJ, Steinberg LS, Almond RG. Focus article: on the structure of educational assessments. *Meas Interdiscip Res Persp.* 2003;1(1):3-62.
49. Mislevy RJ. Validity by design. *Educational Researcher.* 2007;36(8):463-469. doi:10.3102/0013189X07311660
50. Kane M. Validity and fairness. *Language Testing.* 2010;27(2):177-182. doi:10.1177/0265532209349467
51. Mislevy RJ. The case for informal argument. *Meas Interdiscip Res Persp.* 2012;10(1-2):93-96. doi:10.1080/15366367.2012.682525
52. Van Eemeren FH, Grootendorst R, Johnson RH, Plantin C, Willard CA. *Fundamentals of Argumentation Theory: A Handbook of Historical Backgrounds and Contemporary Developments.* Routledge; 2013. doi:10.4324/9780203811306
53. Tindale CW. *Acts of Arguing: A Rhetorical Model of Argument.* SUNY Press; 1999.
54. Corcoran J. Aristotle's demonstrative logic. *Hist Philos Logic.* 2009; 30(1):1-20. doi:10.1080/01445340802228362
55. Krabbe EC, Walton DN. Formal dialectical systems and their uses in the study of argumentation. In: *Keeping in Touch with Pragma-Dialectics*; 2011:245-263. doi:10.1075/z.163.17kra
56. Blair JA. What is informal logic? In: *Reflections on Theoretical Issues in Argumentation Theory.* Springer; 2015:27-42. doi:10.1007/978-3-319-21103-9\_2
57. Perelman C, Olbrechts-Tyteca L. The new rhetoric: A treatise on argumentation, trans. In: John Wilkinson and Purcell Weaver (Notre Dame). Vol. 19. IN: University of Notre Dame Press; 1969.
58. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J.* 2009;26(2):91-108. doi:10.1111/j.1471-1842.2009.00848.x
59. Colliver JA. Constructivism: the view of knowledge that ended philosophy or a theory of learning and instruction? *Teach Learn Med.* 2002;14(1):49-51. doi:10.1207/S15328015TLM1401\_11
60. Kahlke R, Lee M, Eva KW. Critical reviews in health professions education research. *J Grad Med Educ.* PMID: In press.
61. Aristotle JB. *The Complete Works of Aristotle.* Vol. 2. NJ: Princeton University Press Princeton; 1984.
62. Boole G. *The Laws of Thought.* New York (original edition 1854): Dover; 1957.
63. Barth EM, Krabbe EC. Formal dialectics: instruments for the resolution of conflicts about expressed opinions. *Spektator.* 1978;7(307):341.

64. Barth EM, Krabbe EC. *From Axiom to Dialogue: A Philosophical Study of Logics and Argumentation*. Walter de Gruyter; 2010.
65. Krabbe EC. Formal systems of dialogue rules. *Synthese*. 1985;63(3):295-328. doi:10.1007/BF00485598
66. Van Eemeren FH, Grootendorst R. *Argumentation, Communication, and Fallacies: A Pragmatic-Dialectical Perspective*. Routledge; 2016. doi:10.4324/9781315538662
67. Perelman C, Olbrechts-Tyteca L. Logique et rhétorique. *Rev Philos France Let*. 1950;140:1-35.
68. Perelman C, Olbrechts-Tyteca L. Rhétorique et philosophie; pour une théorie de l'argumentation en philosophie. *Revue de Métaphysique et de Morale*. 1953;58(3).
69. Frank DA, Bolduc M. Lucie Olbrechts-Tyteca's new rhetoric. *Q J Speech*. 2010;96(2):141-163. doi:10.1080/00335631003796685
70. Crosswhite J. The new rhetoric project. *Philosophy & Rhetoric*. 2010; 43(4):301-307. doi:10.5325/philrhet.43.4.0301
71. Van Eemeren FH. Argumentation theory after the new rhetoric. *L'analisi Linguistica e Letteraria*. 2009;17(1):119-148.
72. Frank DA. Argumentation studies in the wake of the new rhetoric. *Argumentation Advocacy*. 2004;40(4):267-283. doi:10.1080/00028533.2004.11821612
73. Paso M. Rhetoric meets rational argumentation theory. *Ratio Juris*. 2014;27(2):236-251. doi:10.1111/raju.12043
74. Sigler J. The new rhetoric's concept of universal audience, misconceived. *Argumentation*. 2015;29(3):325-349. doi:10.1007/s10503-015-9349-3
75. Aikin SF. Perelmanian universal audience and the epistemic aspirations of argument. *Philos Rhetoric*. 2008;41(3):238-259. doi:10.5325/philrhet.41.3.0238
76. Gross AG. Misunderstanding the universal audience. *Adv History Rhetoric*. 2019;22(3):290-302. doi:10.1080/15362426.2019.1671704
77. Perelman C. The new rhetoric and the rhetoricians: remembrances and comments. *Q J Speech*. 1984;70(2):188-196. doi:10.1080/00335638409383688
78. Blair JA. The "logic" of informal logic. In: *Groundwork in the Theory of Argumentation*. Springer; 2012:101-117. doi:10.1007/978-94-007-2363-4\_9
79. Johnson RH. Making sense of "informal logic". *Informal Logic*. 2006; 26(3):231-258. doi:10.22329/il.v26i3.453
80. Walton D, Brinton A. *Historical Foundations of Informal Logic*. Routledge; 2016. doi:10.4324/9781315253329
81. Johnson RH, Blair JA. Logical self-defense. *Idea*. 2006.
82. Walton D. Formalizing informal logic. Douglas Walton and Thomas F Gordon, *Formalizing Informal Logic*, *Informal Logic*. 2015;35(4):508. doi:10.22329/il.v35i4.4335
83. Hamblin CL. *Fallacies*. *Tijdschrift Voor Filosofie*. 1970;33(1).
84. Walton D. *Informal Logic: A Pragmatic Approach*. Cambridge University Press; 2008.
85. Hitchcock D. Informal logic and the concept of argument. In: *Philosophy of Logic*. Elsevier; 2007:101-129. doi:10.1016/B978-044451541-4/50007-5
86. Toulmin S. *The Uses of Argument*. Cambridge: Cambridge University Press; 1958.
87. Hitchcock D, Verheij B. The Toulmin model today: introduction to the special issue on contemporary work using Stephen Edelston Toulmin's layout of arguments. *Argumentation*. 2005;19(3):255-258. doi:10.1007/s10503-005-4414-y
88. Simosi M. Using Toulmin's framework for the analysis of everyday argumentation: some methodological considerations. *Argumentation*. 2003;17(2):185-202. doi:10.1023/A:1024059024337
89. Stone J, Zumbo B. Validity as a pragmatist project: a global concern with local application. In: *Trends in Language Assessment Research and Practice*; 2016:555-573.

90. Zumbo BD. Validity as contextualized and pragmatic explanation, and its implications for validation practice. Paper presented at: The Concept of Validity: Revisions, New Directions and Applications, Oct, 2009.
91. Kvernbeek T. Argumentation in theory and practice: gap or equilibrium? *Informal Logic*. 2012;32(3):288-305. doi:10.22329/il.v32i3.3534
92. Johnson RH. The role of audience in argumentation from the perspective of informal logic. *Philosophy & Rhetoric*. 2013;46(4):533-549. doi:10.5325/philrhet.46.4.0533
93. Govier T, Hoaglund J. *The Philosophy of Argument*. Vol. 3. Newport News, VA: Vale Press; 1999.
94. Dollinger M, Lodge J, Coates H. Co-creation in higher education: towards a conceptual model. *J Market Higher Educ*. 2018;28(2):210-231. doi:10.1080/08841241.2018.1466756
95. Holmboe ES. Work-based assessment and co-production in postgraduate medical training. *GMS J Med Educ*. 2017;34(5):Doc58.
96. Elwyn G, Nelson E, Hager A, Price A. Coproduction: when users define quality. *BMJ Qual Saf*. 2020;29(9):711-716. doi:10.1136/bmjqs-2019-009830
97. Turakhia P, Combs B. Using principles of co-production to improve patient care and enhance value. *AMA J Ethics*. 2017;19(11):1125-1131. doi:10.1001/journalofethics.2017.19.11.pfor1-1711
98. Knoch U, Chapelle CA. Validation of rating processes within an argument-based framework. *Language Testing*. 2018;35(4):477-499. doi:10.1177/0265532217710049
99. Park YS, Konge L, Artino AR. The positivism paradigm of research. *Acad Med*. 2020;95(5):690-694. doi:10.1097/ACM.0000000000003093
100. Young ME, Ryan A. Postpositivism in health professions education scholarship. *Acad Med*. 2020;95(5):695-699. doi:10.1097/ACM.0000000000003089
101. Rees CE, Crampton PE, Monrouxe LV. Re-visioning academic medicine through a constructionist lens. *Acad Med*. 2020;95(6):846-850. doi:10.1097/ACM.0000000000003109
102. Ellaway RH, Kehoe A, Illing J. Critical realism and realist inquiry in medical education. *Acad Med*. 2020;95(7):984-988. doi:10.1097/ACM.0000000000003232
103. MacLeod A, Ellaway RH, Paradis E, Park YS, Young M, Varpio L. Being edgy in health professions education: concluding the philosophy of science series. *Acad Med*. 2020;95(7):995-998. doi:10.1097/ACM.0000000000003250
104. Tavares W, Kuper A, Kulasegaram K, Whitehead C. The compatibility principle: on philosophies in the assessment of clinical competence. *Adv Health Sci Educ*. 2020;25(4):1003-1018. doi:10.1007/s10459-019-09939-9
105. Addey C, Maddox B, Zumbo BD. Assembled validity: rethinking Kane's argument-based approach in the context of International Large-Scale Assessments (ILSAs). *Assess Educ Principles Policy Practice*. 2020;27(6):588-606. doi:10.1080/0969594X.2020.1843136
106. 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. doi:10.1097/ACM.0b013e3181eac3e6



# CHAPTER 4

**How do validity experts conceptualise argumentation? It's a rhetorical question**

Published as:

Kinnear B, Martini A, Varpio L, Driessen EW, Schumacher DJ. How do validity experts conceptualise argumentation? It's a rhetorical question. *Medical Education*. 2024 Aug;58(8):989-997.

# **Abstract**

## **Introduction**

Health professions education (HPE) has adopted the conceptualization of validity as an argument. However, the theoretical and practical aspects of how validity arguments should be developed, used and evaluated in HPE have not been deeply explored. Articulating the argumentation theory undergirding validity and validation can help HPE better operationalise validity as an argument. To better understand this, the authors explored how HPE validity scholars conceptualise assessment validity arguments and argumentation, seeking to understand potential consequences of these views on validation practices.

## **Methods**

The authors used critical case sampling to identify HPE assessment validity experts in three ways: (1) participation in a prominent validity research group, (2) appearing in a bibliometric study of HPE validity publications and (3) authorship of recent HPE validity literature. Qualitative semi-structured interviews were conducted with 16 experts in HPE assessment validity from four different countries. The authors used reflexive thematic analysis to develop themes relevant to their research question.

## **Results**

The authors developed three themes grounded in participants' responses: (1) In theory, HPE validity is a social and situated argument. (2) In practice, the absence of audience and evaluation stymies the social nature of HPE validity. (3) Lack of validity argumentation creates and maintains power differentials within HPE. Participants articulated that current HPE validation practices are rooted in post-positivist epistemology when they should be situated (i.e. context-dependent), audience centric and inclusive.

## **Discussion**

When discussing validity argumentation in theory, participants' descriptions reflect an interpretivist lens for evaluation that is misaligned with real-world validity practices. This misalignment likely arises from HPE's adoption of "validity as an argument" as a slogan, without integrating theoretical and practical principles of argumentation theory.

# 1 | INTRODUCTION

Within health professions education (HPE), validity is often characterized as an argument about the interpretation and uses of data.<sup>1</sup> While other conceptualizations of validity (e.g. validity as a test characteristic) still exist within and beyond HPE,<sup>2,3</sup> the idea of “validity as an argument” is now widely accepted and used in our field. In practice, this conceptualization usually manifests as health professions educators and scholars seeking evidence to support decisions that are grounded in assessment data. We often use frameworks from prominent validity scholars such as Messick<sup>4</sup> or Kane<sup>5</sup> to categorise types of evidence or inferential links in a chain of reasoning. However, while evidence and inferences are necessary for arguments, they are not sufficient. Arguments are social in nature and, if reduced to simply evidence and inferences, they can be rendered inert.<sup>6</sup> Arguments are meant to have impact on people and therefore require more than static piles of evidence, no matter how neatly categorised. In other words, while we can present an organised compilation of data to an audience, those data need to be explained and contextualised to have influence as a validity argument because the data never speak for themselves. In HPE, we have adopted the “validity as an argument” approach; however, we have not attended to the theoretical and practical aspects of how arguments are developed, used and evaluated. To fully actualize validity as arguments, we must understand the argumentation theory that undergirds this conceptualization.

Notably, multiple types of arguments have been elaborated regarding validity and validation. Kane proposed two argument types that work in tandem to establish validity of decisions.<sup>5,7</sup> The first is an interpretation/use argument (IUA), which lays out the claims, inferences and assumptions that link a moment of assessment to the end interpretations and uses. An accompanying validity argument is the theoretical and empirical evaluation of the IUA. As Kane elaborates, the IUA is descriptive (stating what is being claimed) while the validity argument is evaluative (investigating the plausibility of the claims), but both are used when conducting validation work.<sup>8</sup> While these two types of argument are often presented as distinct, Kane himself notes that separating the two is more due to practical utility than conceptual distinction.<sup>8</sup> Bachman is another prominent validity scholar who proposed the term assessment use argument (AUA).<sup>9,10</sup> An AUA can be broken down into two component arguments: an assessment validity argument, which links assessment performance to an interpretation, and an assessment utilisation argument, which links interpretations drawn from an assessment to resulting decisions. The multiple different types of “arguments” being described can be dizzying. However, all of them agree on the premises that validity is conceptually an argument supporting or refuting the uses and interpretations of an assessment. Thus, this study of argument and argumentation can be applied across any and all argument types described above.

HPE's lack of attention to argumentation is surely at least partially attributable to the very nature of our field. HPE is a relatively nascent scholarly field when compared with its older siblings, such as Medicine, Education and Measurement.<sup>11</sup> While HPE has developed its own ideology over the last century,<sup>12</sup> many of its concepts, methods and beliefs “immigrated” from other fields.<sup>13</sup> Validity is one such concept. HPE's use of “validity as

an argument” draws from scholars in Psychology, Language Testing and Measurement,<sup>4,5,10,14,15</sup> with HPE scholars translating and adapting this conceptualization into their own validity discourse.<sup>1-3,16-27</sup> While some validity scholars in other fields attend explicitly to argumentation,<sup>5,10,28-30</sup> HPE has yet to explore how arguments underpin the validity of assessments and how validity arguments in our field should be structured, evaluated or used.

Argumentation Theory is a field of study (rather than a single theory) that is well suited for exploring HPE validity arguments. Argumentation Theory brings a long-historied, multidisciplinary approach to exploring how arguments are conceptualised, developed and evaluated.<sup>6</sup> We previously reviewed Argumentation Theory, exploring how it could inform HPE's understanding of assessment validity arguments.<sup>31</sup> We found that multiple argumentation orientations, such as Informal Logic or New Rhetoric, were particularly constructive for informing HPE assessment validity. However, that work was theoretical; an evidence base is needed to confirm our premise. The purpose of this study is to explore how HPE validity scholars conceptualise validity arguments and argumentation about assessment.

## **2 | METHODS**

The University of Cincinnati institutional review board approved this study as exempt.

### **2.1 | Research design and participant selection**

Taking a social constructionist worldview,<sup>32</sup> we conducted qualitative semi-structured interviews with subsequent reflexive thematic analysis<sup>33</sup> to answer our research question. We chose reflexive thematic analysis because it offered flexibility to take an inductive analytic approach while being sensitised by theory, as well as the flexibility to capture both semantic (i.e. explicitly expressed) and latent (i.e. implicit or conceptual) meaning in the data.<sup>30</sup> We used critical case sampling to identify participants who would yield the most information about our phenomenon of interest.<sup>34</sup> To that end, we sought participants who lead HPE's scholarly discourse on validity and validation and did so in three ways simultaneously. We emailed members of the Research Group in Pursuit of Validity (RGPV), a community of HPE researchers and knowledge users who investigate issues surrounding validity.<sup>35</sup> We also reviewed a bibliometric study describing the literature on validation of assessments within medical education<sup>25</sup> and emailed a subset of these authors identified by Young et al. as “prolific” for participation. Finally, we reviewed the HPE validity literature in the last 7 years (since the bibliometric study) and contacted first or senior authors who published frequently cited or multiple HPE validity manuscripts. Characteristics of participants are described in Table 1. While these validity scholars may or may not be involved in frontline validation work, they represent a group that leads the discourse on what validity is and how to operationalise it within our field.

**TABLE 1** Participant demographics.

Participant characteristic	N (% of total participants)
<b>Location</b>	
Canada	10 (63)
United States	3 (19)
Netherlands	2 (13)
Australia	1 (6)
<b>Highest level education degree</b>	
PhD	12 (75)
Master	4 (25)
<b>Clinicians</b>	7 (44)

## 2.2 | Data collection

We developed a semi-structured interview guide (see Appendix A) with questions informed by our expertise in assessment, validity and argumentation. We focused our interviews on argument evaluation, audience and structure based on our recent narrative review of argumentation theory.<sup>31</sup> To be transparent with participants during the interview, we began interviews with open-ended questions about their views on validity, specifically the validity of assessment decisions of any type, and then explicitly gave them definitions of validity and argumentation (see Appendix A) and asked them to comment (e.g. Does your view of validity align with this definition?). The guide was pilot tested with five clinician educators with knowledge of HPE validity for clarity and iteratively revised as needed. We used individual semi-structured interviews conducted over Zoom (Banyai, Istvan. Zoom. New York: Viking, 1995) given the geographic distribution of participants. We chose one-on-one interviews to allow for more in-depth exploration of the complex topic of argumentation and to capture each individual's views without social influence from other participants. BK conducted all interviews (n = 16) between January and July 2023, which lasted an average of 44.5 minutes (range 35–63 minutes) per interview. All participants provided verbal consent prior to interviews. Audio files were professionally transcribed verbatim and subsequently de-identified prior to coding (e.g. each participant was assigned a number; P1 = Participant 1).

## 2.3 | Data analysis

Data analysis occurred concomitantly and iteratively with data collection. Three authors (BK, AM, DS) immersed themselves in the transcribed data through active reading, followed by independent generation of initial codes for the first three interviews using the coding software Quirkos (Quirkos 2.5.2 Computer

Software, 2022). We took an inductive approach to open coding without using a predetermined theoretical framework to guide us; concepts and theories from our previous narrative review of Argumentation Theory acted as sensitizing concepts.<sup>36</sup> We then met to discuss and compare open codes, combining them into interrelated concepts. These initial concepts informed subsequent rounds of interviews and analysis. BK conducted three more interviews, which were transcribed and independently coded by the same authors. These authors again met to continue discussing interrelated concepts and begin the development of themes. Having three independent coders interpreting the data was not carried out to achieve triangulation (which would conflict with our social constructivist worldview); instead, this approach allowed for a variety of insights and interpretation to flourish in our analysis. After multiple rounds of coding and discussion, we shifted to having these three authors code a common source text in Quirkos (i.e. meaning that the authors coded independently but could see other authors' codes) and organised data into concepts and themes. Next, LV and ED read a sample of interviews and reviewed codes and themes to support theme development. After several meetings, we combined concepts into themes that were supported by multiple quotes from interviews. After 16 interviews, we were seeing significant amounts of repetition of concepts and felt that we had sufficiently answered our research questions.

## **2.4 | Reflexivity**

Our team members' unique areas of expertise and personal backgrounds informed our data analysis. Two team members (BK and DS) are clinician educators who actively engage in HPE research. BK has expertise in competency-based medical education (CBME), learner assessment and validity. He is also an associate programme director for two residency programmes, and his interpretations were influenced by his experiences as a programme leader and frontline clinician educator who rarely saw validity being addressed in practice. DS is a paediatric emergency medicine physician and PhD-trained research scientist with expertise in CBME and programmatic assessment. He has served in residency administration, led national studies with residency leaders in the area of assessment and works with national regulatory bodies making arguments for the validity of their assessment practices. His extensive experience with the real world range of how assessment is carried out at local and national levels influenced his interpretations, and he regularly considered how participants' thoughts illuminate the differences between theoretical considerations and actual practice of validity of argumentation. ED is an educational scientist, who started his career as an assessment coordinator for a law school. During this period, he experienced the importance of arguing the validity of the assessments with students, teachers and leaders. His PhD research was on portfolios for undergraduate medical education, including the validity of portfolio assessment. In both contexts, the validity of the assessment methods was disputed by different stakeholders, including the Dutch National Accreditation Authority. These experiences made ED sensitive for the power dynamics of assessment validity arguments and argumentation. LV is a PhD-trained researcher whose graduate work focused on argumentation theory, rhetoric and professional communication and who has expertise with qualitative

research's paradigms, methodologies and methods. This background shaped LV's interpretations; she regularly considered how participant comments resonated with different argumentation theories and how participants' descriptions pointed to tensions of negotiating a space between validity's post-positivist history and its more interpretivist use in HPE. AM has a Master of Medical Education and is a clinical research coordinator with experience in qualitative research, learner assessment and entrustment. Her experiences as an HPE learner influenced her interpretations as she regularly considered how participants' responses relate to the experiences of and consequences for current HPE learners and trainees. All authors have expertise in qualitative methodology. Multiple authors immersed themselves in argumentation theory literature prior to beginning this study.<sup>31</sup>

### 3 | RESULTS

Participants noted that validity argumentation was challenging in HPE, both conceptually and from a practical standpoint, with participants labelling it as “messy” (P10) and “mixed up” (P2). We developed themes focusing on how participants conceptualised the arguments and argumentation within HPE assessment validity to make the “complete swamp” (P8) of validity more explicit and understandable. These themes were as follows: (1) In theory, HPE validity is a social and situated argument. (2) In practice, the absence of audience and evaluation stymies the social nature of HPE validity. (3) Lack of validity argumentation creates and maintains power differentials within HPE.

#### 3.1 | Theme 1: In theory, HPE validity is a social and situated argument

Participants agreed with the conceptualization of validity as an argument and that principles of argumentation can inform HPE. Their reflections about HPE validity arguments aligned with the definition we offered from Argumentation Theory—most notably, the idea that validity was an argument and so was a social activity, which involved listeners or readers, and hinged on that audience's reaction to the argument's acceptability.

As this excerpt illustrates, participants noted that validity (when conceptualised as an argument) is inherently social, requiring more than one party to participate:

*But for validity, I think that we're always convincing somebody, whether it's a learner that was fair and they deserve the judgment that came from the assessment, or we're trying to argue with the program director that what they're doing is just not cool and not defensible. We're trying to argue with a licensure body about what should be contained in that activity. I'm wondering if validity is more social than other forms of argumentation. (P12)*

The importance of an active listener or reader (i.e. audience) with agency to critique and evaluate a validity argument was frequently highlighted, though several participants noted that complex contextual and social factors influence the audience. As exemplified by this participant's comment, contextual factors such as the

purposes of assessment, social norms and values, cultural differences, socio-political environments and historical backgrounds make validity a situated argument:

*So, it's a social practice anchored in certain set of values depending on where you're situated and you need to be aware of your audience and their values in order to know where you can go and where you just can't go at this point in time. (P3)*

Intended audiences and their associated values were perceived as extremely important when developing or evaluating a validity argument. Participants noted a wide range of audiences for HPE assessment validity including education administrators (e.g. deans, programme directors and competency committees), institutional leaders, assessors, teachers, regulatory bodies (e.g. accreditors and certifiers), HPE scholars, scholarly journal editors and readers, learners and patients/society. Multiple participants shared the sentiment that “every audience that can either benefit or be harmed by your decision must have a say in that validity argument” (P5). Patients and learners were highlighted as the most important stakeholders of assessment decisions, but they were also recognised as infrequently engaged as audiences for validity arguments. While accrediting and certifying bodies could be seen to represent patients and learners in validity argumentation, some participants questioned if patients should be more directly involved:

*I would argue [involving patients] is not done classically... and I guess that's what regulatory bodies kind of are, but maybe it's the other way around. Patients need to be involved at the front end in the construction of the validity argument so that we make sure that the end product reflects what the patients need and want or maybe need more than want. (P4)*

Audience was emphasised as central to validity argumentation for two reasons. First, participants explained that the acceptability of a validity argument depends heavily on an audience's values and beliefs given the situated nature of argumentation. Second, they noted that each audience has specific communication needs for an argument to be understood and accepted. As this participant mentions:

*The first component's always sussing out your audience. Who are you talking to? What is their approach? What is their thinking, and how can I convey my thoughts or my arguments? How can I communicate them in a way that is acceptable and understandable to them? (P5)*

This theme was developed around how participants believe validity argumentation should exist in HPE, but as our next theme illustrates, a gap exists between current validity practices and true argumentation.

## **3.2 | Theme 2: In practice, the absence of audience and evaluation stymies the social nature of HPE validity**

While participants' views aligned with the conceptualization of validity as an argument, many noted that HPE's lack of audience engagement and ambiguous (or missing) evaluation routines has hindered

argumentation in practice. Participants conveyed that a validity argument represents both a product (the claims, premises and data that constitute an argument) and a process (the exchange of ideas and judgements) and that both are lacking in our field. When considering HPE validity arguments, many participants felt “... they're not even structured as an argument” (P16). As another participant explains:

*I think argumentation is never done [in HPE] ... a systematic review of 400 articles and simulation based assessment published about 10 years ago [showed] that only a small minority of people do a post-interpretation argument ... (P9)*

This lack of argumentation manifests as suboptimal dialogue between those who put forth validity arguments (e.g. programme leaders, assessment developers) and audiences, and it sustains unclear argument evaluation standards or practices. As a result, as one participant noted, “I don't think as a field we really operationalize thinking about validity arguments in a real world sense” (P11).

Despite the importance of audience for validity arguments, an intended audience is rarely explicitly articulated within HPE. Some participants felt that validity argument audiences are explicitly identified for national high stakes exams or for assessments related to minoritized groups, but not for most programme-level assessment decisions. Others felt that, for arguments that are published in scholarly journals, audiences can be inferred based on known journal readership. However, most validity arguments were seen to lack a definitive audience and that audiences “are the things that we just take for granted” (P2).

Without clearly identified audiences, the social and discursive aspects of argumentation were felt to be scarce in HPE validation practices. Many validity arguments are shared via peer-reviewed publication; however, participants noted that the intended audiences (see Theme 1) read these publications “probably very infrequently” (P6). While peer reviewed publication was identified as important for validity scholarship, participants noted the dialectical limitations of this format for purposes of argumentation. Relying on publications limits audience engagement. As this participant laments, journal papers do not solicit evaluation conversations with readers:

*I think [validity arguments are] shared through the research literature as best I can tell ... We all try doing a little bit on tools probably that we use to put it out there, but it never really gets traction or at least I've never had somebody be like, “Let's have a conversation” or, “We over here want to use that tool.” ... It just gets cited randomly in next papers. (P11)*

Multiple participants felt that evaluation of a validity argument's merit is frequently lacking and that “... we don't [evaluate validity arguments]. We don't do it period. No one does it ... We just don't do any appraisal of the argument” (P9). Publication of validity evidence, or citation of previous validity work, was not felt to represent deliberate argument evaluation. Rather than engaging or entreating an evaluation from a particular audience, publication of validity evidence was seen as the generation of an authoritative claim: “It's been published and therefore it's good.” (P9).

Participants struggled to articulate an explicit approach to validity argument evaluation. No clear criteria or rubric was described by participants, with several reiterating the subjective nature of evaluation, and that context and audience influence how each validity argument is evaluated. As one participant stated, “Alas, there's no metric, right? So I would have to say it's a subjective feeling. I would hope that it's a subjective feeling based on some expertise and experience” (P16). Participants frequently used terms intimating an audience-centric approach such as “believing” (P7, P12) and “convincing” (P1, P5, P6, P7, P9, P10, P12, P14, P16). Some participants explicitly acknowledged validation as a rhetorical act, with one noting the importance of “ethos, logos, pathos, and kairos” (P5). One participant articulated that rhetoric is inherent to validation, but raised concerns of academic integrity asking, “How do you not prostitute yourself in terms of trying to sell everything to everyone?” (P3). Several participants used language that focused on characteristics of the validity argument and supporting evidence such as “sufficiency” (P2, P3, P4, P10, P11, P13, P16), “alignment” (P2, P3, P5, P9, P12), “coherence” (P2, P5, P9, P12, P14), “defensibility” (P3, P4, P12, P14) and “comprehensiveness” (P2, P9, P12). Some noted the importance of an “internal logic” (P5, P12) within the argument.

Participants suggested multiple reasons for why argumentation is present in HPE validity in theoretical terms but not in practical ones, including complexity of validity theory and argumentation, lack of standard argumentation approach, lack of protected time to do the work and word count limitations in health professions' scholarly journals. However, participants also acknowledged that these reasons could be causes for or the result of scarce explicit attention to argumentation (or both).

### **3.3 | Theme 3: Lack of validity argumentation creates and maintains power differentials within HPE**

Participants noted that HPE has implicit power dynamics within assessment and validation practices. They explained that differential access to power in HPE validity discourses is created in multiple ways: by treating validity as a universal truth rather than a situated and contestable claim, by limiting who is seen as legitimate participants in any argumentation that does occur and by direct political and hierarchical influences.

When no audience is identified or engaged, rebuttal (or even close scrutiny) of a validity argument is hindered. Participants observed that current HPE validity argumentation practices leave audiences inferred or absent; therefore, validity claims are underscrutinised. Without scrutiny, argument claims gain authority and the veneer of unquestionable truth. This definitive status is bolstered by a post-positivist epistemology<sup>37</sup> that participants perceived as dominating the HPE validity discourse:

*We have had moments or people who have overly romanticized the extent to which concrete reality can be expressed or the extent to which the data alone are sufficient to provide compelling argument. Whether we're talking about evidence-based education or good scientific writing or validity, I think, really, we're talking about storytelling. And there's a derogatory version of that, which I do not intend*

*... all those domains are about good rhetoric and trying to provide a compelling story for people because they need that in order to appreciate and understand, [and] to comprehend what you are trying to convince them of, let alone do something with them. (P6)*

Participants' descriptions of validity argumentation were steeped in interpretivism (i.e. emphasis on social context, audience, dialogue) and pragmatism (i.e. the use of data for specific purposes) that are incongruent with validation practices that espouse the creation of indisputable truth statements. Some saw rhetoric as a way to mitigate the perceived authoritarian nature of assessment validity in HPE:

*... pushing rhetorics forward in the assessment discourse is a way to open the authoritarian black box of assessment. Because we use tools, we use language, we use certain concepts but we never say why. We don't argue, we just make claims. We don't justify our claims and we don't justify the specific audience either ... (P3)*

Power differentials are also created through “gatekeeping” (P12)—i.e. limiting who can participate in validity argumentation. Multiple participants favoured an inclusive approach to HPE validity argumentation (i.e. engagement of several audiences) but cautioned that many audiences do not have the ability to participate. Lack of audience participation was attributed to institutional hierarchy, lack of audience knowledge, lack of audience argumentation skills and lack of opportunity. Such obstructions curtail participation from many audiences including patients, trainees and educators.

*In order to achieve [validity argumentation], you first have to make sure that everybody has sufficient knowledge and background knowledge to be able to contribute to the best of their abilities, which is the big challenge ... it's an uneven play field. (P5)*

Participants explained that power differentials are created and maintained through the influence of existing hierarchy and politics that shape the content (i.e. which claims and evidence are put forth) and purpose (i.e. what is the goal or desired outcome) of HPE validity arguments. One participant noted that “the claims you choose to examine and research to build your validity argument probably have some foundation in making someone in a political chain happy” (P10). Multiple participants felt that validity arguments do not exist separate from power dynamics in HPE, but rather that arguments are constituted in part by existing sociopolitical forces.

*But the work in our space is so political ... You're navigating sociopolitical arguments of the time ... I understand wanting to stay focused on almost the academic construction of the argument, but because you are grounding yourself in argumentation and because argumentation necessitates an audience, that audience sits within a sociopolitical and historical space. (P12)*

Validity argumentation can therefore be used to uphold and maintain specific HPE power structures, even if (perhaps especially if) such structures and values are assumed and implicit.

## 4 | DISCUSSION

Our analysis shows a misalignment between how HPE scholars describe the argumentation aspect of validity and current practices in our field. Study participants described validity arguments theoretically as being situated, audience-centric and inclusive. When discussing validity argumentation in theory, their descriptions reflect an interpretivist lens for evaluation. However, our findings also suggest that current HPE validation practices are rooted in post-positivist epistemology, lack audience engagement or evaluation and create and maintain power differentials amongst stakeholders. This misalignment likely arises from HPE's adoption of “validity as an argument” as a slogan, without integrating theoretical and practical principles of argumentation theory. HPE has adopted “validity evidence” but largely uses the term “validity argument” in an ornamental fashion.

Our data suggest that HPE validity is sheltered under a protective veneer of objective truth; this veneer obscures the contextual, socio-political influences shaping our assessment work. This veneer upholds the notion that validity claims are irrefutable truths rather than contestable assertions. It disempowers important stakeholders through lack of engagement and differential knowledge and is in tension with the inclusive, polyvocal conceptualization of validity noted by participants. We acknowledge that power differentials exist in all social situations, and it is unreasonable to mitigate them only through changes in validation practices. However, unless this veneer is shattered, our validity work exists unquestioned—and thereby enables—the power differentials to persist. By embracing the social nature of validity arguments and re-imagining our field's approach to validation, we can improve transparency of the controvertible nature of validity claims and create more democratic spaces for argumentation.<sup>38</sup>

When describing argument evaluation, participants used language consistent with argumentation orientations such as Informal Logic<sup>39,40</sup> (e.g. sufficient, coherent and internal logic) and New Rhetoric<sup>41,42</sup> (e.g. convincing, persuasion and storytelling). These results align with our previous narrative review of Argumentation Theory<sup>31</sup> that put forth Informal Logic and New Rhetoric as orientations that could inform HPE assessment validity. Both orientations prioritise audience values and beliefs during argumentation,<sup>42,43</sup> which aligns with our results. If our community harnessed these orientations, validity argumentation could be foregrounded as socially constructed judgements open to critique, rather than objective truths that stood above questioning.

Our themes suggest that HPE might benefit from adopting a more interpretivist approach when constructing and evaluating validity arguments and align with Zumbo's conceptualization of validity as contextualised and pragmatic.<sup>30,44</sup> The post-positivist worldview<sup>37</sup> that is currently dominating HPE's validity discourse is in tension with the more social constructionist<sup>32</sup> and pragmatist<sup>45</sup> aspects of arguments that are audience-dependent, situated and oriented towards action (e.g. real-world decisions). This discordance may reflect the shift from classic post-positivist conceptualizations of validity (i.e. validity as a test characteristic)<sup>3</sup> to more

socially constructed views (i.e. arguments). It may also reflect a recent, broader embrace of interpretivism and subjectivity in HPE assessment.<sup>46-48</sup> Tavares et al. have called for compatibility between philosophical positions and assessment choices, noting that “assessment is optimized when its underlying philosophical position produces congruent, aligned and coherent views on constructs, assessment strategies, justification and their interpretations.”<sup>49</sup> This same compatibility principle holds true for validity arguments and is relevant as philosophical worldviews shift in HPE. We believe that one particular worldview (e.g. postpositivism versus social constructionism) is not inherently better or worse than another. However, our study suggests that there may be a lack of compatibility between the philosophical underpinnings of how validity arguments are conceptualised and how validation is operationalized in HPE.

The lack of audience identification and engagement presents practical challenges for HPE assessment validation. Currently, validity evidence is usually shared via peer-reviewed publication, with little chance that important audiences (as identified by our participants) such as learners, patients or local administrators will read them. With publication, audiences include journal reviewers and editors, who may focus more on the overall value of a manuscript to scholarly discourse instead of a validity argument's merit. This audience also include journal readers, who may not be a relevant audience given the situated nature of validity arguments. While peer-reviewed publication is a critical mechanism for the advancement of knowledge,<sup>50</sup> we should consider adjunctive mechanisms to better engage relevant audiences when developing and evaluating validity arguments. Participants identified a dearth of argument evaluation standards and practices as a barrier to HPE validation. While standard approaches to argumentation would not necessarily negate the situated nature of arguments, we did not elucidate how best to manage the tension between standardisation and context-dependence. When developing assessment validity arguments, we should also explicitly identify which stakeholders constitute the audience. This approach would provide more context for why a particular argument was made or why particular evidence was used. Perfect alignment with audience needs and values will be impossible given the heterogeneity within and between stakeholder groups, but identification of intended audiences provides valuable rationale for argument development.

This study had multiple limitations that influenced our results. First, while we used multiple methods to identify leading HPE validity scholars, we may still have missed important viewpoints that could have influenced our analysis. Second, nearly all our participants were from North America, and all were from western cultures. This likely reflects the high representation of North American authors in HPE scholarly discourse and excludes non-western viewpoints. Future research should explore how validity arguments are conceptualised in other cultures and parts of the world and to further investigate which actors may be most active in developing and maintaining dominant views on validity argumentation. Finally, our participants were validity scholars, and their views may not reflect those of frontline educators or programme leaders. However, our research question was aimed at understanding the views of those who lead the broader discourse on validity in HPE. Future studies may seek to understand the views of frontline educators and programme leaders.

## 5 | CONCLUSION

HPE validity experts view validity arguments through an interpretivist lens that values context and audience values. To fully embrace validity as an argument, HPE should untether from the moorings of postpositivist, absolutist views about validity and embrace the social and subjective nature of argumentation.

## References

1. Cook DA. When I say...validity. *Med Educ*. 2014;48(10):948-949. doi:10.1111/medu.12401
2. St-Onge C, Young M. Evolving conceptualisations of validity: impact on the process and outcome of assessment. *Med Educ*. 2015;49(6):548-550. doi:10.1111/medu.12734
3. St-Onge C, Young M, Eva KW, Hodges B. Validity: one word with a plurality of meanings. *Adv Health Sci Educ Theory Pract*. 2017;22(4):853-867. doi:10.1007/s10459-016-9716-3
4. Messick S. Validity. In: Linn R, ed. *Educational Measurement* (13-103). Macmillan Publishing; 1989.
5. Kane M. The argument-based approach to validation. *School Psychol Rev*. 2013;42(4):448-457. doi:10.1080/02796015.2013.12087465
6. van Eemeren FH, Grootendorst R, Johnson RH, Plantin C, Willard CA. *Fundamentals of Argumentation Theory: A Handbook of Historical Backgrounds and Contemporary Developments*. Routledge; 2013.
7. Kane MT. Validating the interpretations and uses of test scores. *J Educ Meas*. 2013;50(1):1-73. doi:10.1111/jedm.12000
8. Kane MT. Validation as a pragmatic, scientific activity. *J Educ Meas*. 2013;50(1):115-122. doi:10.1111/jedm.12007
9. Bachman LF. Building and supporting a case for test use. *Lang Assess Quart Int J*. 2005;2(1):1-34. doi:10.1207/s15434311laq0201\_1
10. Bachman LF. Constructing an assessment use argument and supporting claims about test taker-assessment task interactions in evidence-centered assessment design. *Measure Lawrence Erlbaum Assoc*. 2003;1:63-65.
11. Ten Cate O. Health professions education scholarship: the emergence, current status, and future of a discipline in its own right. *Faseb Bioadv*. 2021;3(7):510-522. doi:10.1096/fba.2021-00011
12. Varpio L. Dismantling medical education's incompatible ideology. *Med Teach*. 2023;45:766-771. doi:10.1080/0142159X.2023.2279902
13. Norman G. Fifty years of medical education research: waves of migration. *Med Educ*. 2011;45(8):785-791. doi:10.1111/j.1365-2923.2010.03921.x
14. Cronbach LJ. Five perspectives on validity argument. *Test Valid*. 1988;3-17.
15. Mislevy RJ. Validity by design. *Ed Res*. 2007;36(8):463-469. doi:10.3102/0013189X07311660
16. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ*. 2015;49(6):560-575. doi:10.1111/medu.12678
17. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J med*. 2006;119(2):166.e7-166.e16.
18. Cook DA, Kuper A, Hatala R, Ginsburg S. When assessment data are words: validity evidence for qualitative educational assessments. *Acad med*. 2016;91(10):1359-1369. doi:10.1097/ACM.0000000000001175
19. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Adv Simul (Lond)*. 2016;1(1):31. doi:10.1186/s41077-016-0033-y
20. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Adv Health Sci Ed*. 2014;19(2):233-250. doi:10.1007/s10459-013-9458-4
21. Hatala R, Cook DA, Brydges R, Hawkins R. Constructing a validity argument for the Objective Structured Assessment of Technical Skills (OSATS): a systematic review of validity evidence. *Adv Health Sci Ed*. 2015;20(5):1149-1175. doi:10.1007/s10459-015-9593-1
22. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830-837. doi:10.1046/j.1365-2923.2003.01594.x

23. Downing SM, Haladyna TM. Validity threats: overcoming interference with proposed interpretations of assessment data. *Med Educ.* 2004;38(3):327-333. doi:10.1046/j.1365-2923.2004.01777.x
24. Yudkowsky R, Park YS, Downing SM. *Assessment in Health Professions Education.* Routledge; 2019. doi:10.4324/9781315166902
25. Young M, St-Onge C, Xiao J, Lachiver EV, Torabi N. Characterizing the literature on validity and assessment in medical education: a bibliometric study. *Persp Med Ed.* 2018;7(3):182-191. doi:10.1007/S40037-018-0433-X
26. Govaerts M, van der Vleuten CP. Validity in work-based assessment: expanding our horizons. *Med Educ.* 2013;47(12):1164-1174. doi:10.1111/medu.12289
27. Driessen EW, Overeem K, van Tartwijk J, van der Vleuten CP, Muijtjens AM. Validity of portfolio assessment: which qualities determine ratings? *Med Educ.* 2006;40(9):862-866. doi:10.1111/j.1365-2929.2006.02550.x
28. Mislevy RJ. The case for informal argument. *Measure Interdisc Res Persp.* 2012;10(1-2):93-96. doi:10.1080/15366367.2012.682525
29. Lavery MR, Bostic JD, Kruse L, Krupa EE, Carney MB. Argumentation surrounding argument-based validation: a systematic review of validation methodology in peer-reviewed articles. *Educ Meas Issues Pract.* 2020;39(4):116-130. doi:10.1111/emip.12378
30. Zumbo BD. Validity as contextualized and pragmatic explanation, and its implications for validation practice. In: *The Concept of Validity: Revisions, New Directions, and Applications.* IAP Information Age Publishing; 2009:65-82.
31. Kinnear B, Schumacher DJ, Driessen EW, Varpio L. How argumentation theory can inform assessment validity: a critical review. *Med Educ.* 2022;56(11):1064-1075. doi:10.1111/medu.14882
32. Rees CE, Crampton PE, Monrouxe LV. Re-visioning academic medicine through a constructionist lens. *Acad med.* 2020;95(6):846-850. doi:10.1097/ACM.0000000000003109
33. Clarke V, Braun V. *Thematic Analysis: A Practical Guide.* 1st ed. Sage Publications Ltd; 2021.
34. Hanson JL, Balmer DF, Giardino AP. Qualitative research methods for medical educators. *Acad Pediatr.* 2011;11(5):375-386. doi:10.1016/j.acap.2011.05.001
35. Research Group in Pursuit of Validity. <http://www.grqv-rgpv.ca/en/about-us/>
36. Bowen GA. *Sensitizing Concepts.* SAGE Publications Limited; 2020.
37. Young ME, Ryan A. Postpositivism in health professions education scholarship. *Acad med.* 2020;95(5):695-699. doi:10.1097/ACM.0000000000003089
38. Addey C, Maddox B, Zumbo BD. Assembled validity: rethinking Kane's argument-based approach in the context of International Large-Scale Assessments (ILSAs). *Assess Ed Principles Policy Pract.* 2020;27(6):588-606. doi:10.1080/0969594X.2020.1843136
39. Blair JA. What is informal logic? In: *Reflections on Theoretical Issues in Argumentation Theory.* Springer; 2015:27-42. doi:10.1007/978-3-319-21103-9\_2
40. Walton D. *Informal Logic: A Pragmatic Approach.* Cambridge University Press; 2008.
41. Perelman C. The new rhetoric and the rhetoricians: remembrances and comments. *Q J Speech.* 1984;70(2):188-196. doi:10.1080/00335638409383688
42. Frank DA, Bolduc M. Lucie Olbrechts-Tyteca's new rhetoric. *Q J Speech.* 2010;96(2):141-163. doi:10.1080/00335631003796685
43. Johnson RH. The role of audience in argumentation from the perspective of informal logic. *Philos Rhetor.* 2013;46(4):533-549. doi:10.5325/philrhet.46.4.0533
44. Stone J, Zumbo B. Validity as a pragmatist project: a global concern with local application. *Trends Lang Assess Res Pract.* 2016;555-573.
45. Goldkuhl G. Pragmatism vs interpretivism in qualitative information systems research. *Eur J Inf Syst.* 2012;21(2):135-146. doi:10.1057/ejis.2011.54

46. Hodges B. Assessment in the post-psychometric era: learning to love the subjective and collective. *Med Teach*. 2013;35(7):564-568. doi:10.3109/0142159X.2013.789134
47. ten Cate O, Regehr G. The power of subjectivity in the assessment of medical trainees. *Acad med*. 2019;94(3):333-337. doi:10.1097/ACM.0000000000002495
48. Virk A, Joshi A, Mahajan R, Singh T. The power of subjectivity in competency-based assessment. *J Postgrad med*. 2020;66(4):200-205.doi:10.4103/jpgm.JPGM\_591\_20
49. Tavares W, Kuper A, Kulasegaram K, Whitehead C. The compatibility principle: on philosophies in the assessment of clinical competence. *Adv Health Sci Ed*. 2020;25(4):1003-1018. doi:10.1007/s10459-019-09939-9
50. Strevens M. *The Knowledge Machine: How Irrationality Created Modern Science*. Liveright Publishing; 2020.



# CHAPTER 5

## **Legitimation Without Argumentation: An Empirical Discourse Analysis of 'Validity as an Argument' in Assessment**

Published as:

Kinnear B, Schumacher DJ, Varpio L, Driessen EW, Konopasky A. Legitimation Without Argumentation: An Empirical Discourse Analysis of 'Validity as an Argument' in Assessment. Perspectives on Medical Education. 2024 Oct 3;13(1):469.

# **ABSTRACT**

## **Introduction**

Validity is frequently conceptualized in health professions education (HPE) assessment as an argument that supports the interpretation and uses of data. However, previous work has shown that many validity scholars believe argument and argumentation are relatively lacking in HPE. To better understand HPE's discourse around argument and argumentation with regard to assessment validity, the authors explored the discourses present in published HPE manuscripts.

## **Methods**

The authors used a bricolage of critical discourse analysis approaches to understand how the language in influential peer reviewed manuscripts has shaped HPE's understanding of validity arguments and argumentation. The authors used multiple search strategies to develop a final corpus of 39 manuscripts that were seen as influential in how validity arguments are conceptualized within HPE. An analytic framework drawing on prior research on Argumentation Theory was used to code manuscripts before developing themes relevant to the research question.

## **Results**

The authors found that the elaboration of argument and argumentation within HPE's validity discourse is scant, with few components of Argumentation Theory (such as intended audience) existing within the discourse. The validity as an argument discourse was legitimized via authorization (reference to authority), rationalization (reference to institutionalized action), and mythopoesis (narrative building). This legitimation has cemented the validity as an argument discourse in HPE despite minimal exploration of what argument and argumentation are.

## **Discussion**

This study corroborates previous work showing the dearth of argument and argumentation present within HPE's validity discourse. An opportunity exists to use Argumentation Theory in HPE to better develop validation practices that support use of argument.

## INTRODUCTION

Validity has long been a revered concept in health professions education (HPE). Demonstration of validity is often seen as imperative since, “without evidence of validity, assessments in medical education have little or no intrinsic meaning [1].” Despite its importance to HPE, assessment validity is conceptualized in multiple ways [2]. Some see validity as a property of an assessment instrument, with a given test or tool being labelled “validated” for use across differing contexts. Over the last two decades, HPE has drawn from other fields such as Psychology and Language Testing to frame validity as an argument that supports the interpretations and uses of assessment data [3, 4]. In this orientation, validity is not an immutable property of an instrument, but a context-specific judgment about interpretations and uses of assessments. This argument-based conceptualization has gained traction in HPE in recent years, for a variety of possible reasons. HPE has growing epistemological diversity (i.e. diversity of theories regarding what is knowledge) [5] with subjectivity and constructionist approaches being increasingly embraced in assessment [6–8]. The importance of context in assessment has similarly gained recognition [9, 10], making the idea that an assessment instrument can be “validated” in a context-independent way less aligned with broader discourses in HPE. Concurrently, programs are increasingly using complex, multifaceted programs of assessment [11] that integrate and synthesize multiple types of assessment data that may not be appropriate for traditional psychometric validation approaches. *Validity as an argument* is a conceptualization that allows for epistemological diversity, context specificity, and multiple assessment approaches.

However, the adoption of *validity as an argument* in HPE has occurred without development of what is meant by *argument*. Arguments exist to have real-world impacts such as advancing an idea, defending a viewpoint, or changing a policy [12]. The process of using an argument in the real world is called *argumentation*. By its very nature, argumentation (and, therefore, an argument) is social (i.e. exists between people or groups), invokes reason (i.e. uses rationality to some degree), and relates to a particular standpoint or opinion [12]. The concepts of *argument* and *argumentation* in the context of validity have been explored through Argumentation Theory, a theoretically and philosophically diverse field of study concerned with the development, execution, and evaluation of various forms of argument [12]. Each argumentation orientation found in Argumentation Theory brings assumptions and expectations. For example, Informal Logic is an orientation that emphasizes identifying and structuring arguments, with evaluation using normative standards such as relevance, sufficiency, and acceptability. In contrast, New Rhetoric is a different orientation that centers on audience values and norms rather than on a particular structure, with the goal of the argument aimed at persuading the audience, not fulfilling normative standards. There are many orientations within Argumentation Theory, each drawing on different aspects of scholarly discourse and each leading to different approaches to argument development, organization, operationalization, analysis, and evaluation.

Validity arguments in HPE rarely (if ever) explicitly attend to the argumentation orientation undergirding validation work [13]. In our interview study of HPE experts regarding their conceptualizations of argument and argumentation related to assessment validity, participants described an absence of explicit reference to argumentation in HPE validation work. Specifically, they noted a shortage of attention to argument audiences, evaluation, and situated context. Participants perceived that peer-reviewed publications (often from North America or Europe) are used to authoritatively justify validity approaches. Trying to create or evaluate arguments without an explicit argumentation orientation is like chess players who have the game pieces but lack the rules. If *validity as an argument* is to continue to be used in real-world HPE assessment work, our field requires a better understanding of the argument and argumentation *discourses* (i.e., “way[s] of signifying a particular domain of social practice from a particular perspective”) [14] that exist within peer reviewed publications. Therefore, in this study, we used discourse analysis to examine how argument and argumentation are manifested in HPE publications around validity. By developing this understanding, we hope to uncover implicit argumentation concepts, and to better understand the discourses driving *validity as an argument* in peer reviewed publications.

## METHODS

### APPROACH

Our work is undergirded by Norman Fairclough’s approach to discourse analysis, which holds that “one cannot properly analyse content without simultaneously analysing form.” [14]. In other words, to understand ways of signifying (i.e., discourses), one must look at the *language*, i.e., at the *signifiers*. Yet, unlike Fairclough, we do not take “discourse” as a social process to be fully deterministic. Instead, we follow Archer in seeing human reflexivity as mediating the forces of structure and agency [15]. We then draw on Fairclough’s approach alongside other tools offered by Gee [16] and Laclau and Mouffe [17] to engage in methodological *bricolage*, using multiple methods of analysis to understand validity in HPE as “part of a historically situated complex system [18].”

We began with the research question, “What aspects of argumentation implicitly or explicitly manifest in influential HPE validity manuscripts published in peer-reviewed journals?”. Our goal was not triangulation with our previous HPE expert interviews [13], but rather to seek a richer understanding of how *argument* and *argumentation* manifest and are shaped by the language used in HPE validity publications. To achieve this goal, we relied on discourse analysis, which allows researchers to understand how language influences and is connected with social practices [19, 20]. Since peer-reviewed publications are a major conduit of discourse within HPE, we used critical discourse analysis to understand how the language in influential manuscripts reflects discourses of arguments and argumentation relating to validity.

### DOCUMENT SAMPLING STRATEGY

The goal of this study was not to conduct a comprehensive discourse analysis on all manuscripts addressing validity in HPE; instead, we sought to compile a corpus of manuscripts that directly addressed discourses on validity arguments within HPE. To that end, we used multiple search strategies to identify candidate manuscripts. First, one author (BK) met with an academic librarian to formulate search strategies to capture as many HPE articles on validity and validation as possible. We chose only to search in PubMed because our research question focused on HPE's discourse, and most HPE journals are indexed within PubMed. Nine different search strings were used (Appendix A), resulting in 800 unique manuscripts. BK reviewed the titles of these manuscripts to identify papers that promised to explore validity or validation. BK read the abstracts of the promising manuscripts to identify papers that explored the philosophical, theoretical, or practical aspects of validity arguments, even if this exploration was only in part of the manuscript (e.g. introduction). Manuscripts focusing on assessment more broadly without rich discussion on validity arguments were excluded. If a manuscript was felt to be useful in answering the research question, it was added to the corpus for full text analysis. To ensure that we had not missed any particularly relevant publications, we contacted the authors of a previous discourse analysis on validity in HPE [2] and obtained their corpus, reviewing it for manuscripts that would contribute to answering our research question. We also mined the references of the sampled publications looking for manuscripts relevant to our research question that had yet to be identified. We limited inclusion to manuscripts that were within the HPE literature.

BK completed an in-depth review of 59 manuscripts after the above process, searching for papers that provided a substantive discussion of the conceptualization or operationalization of validity arguments in HPE. This resulted in 38 articles being selected for analysis. All articles that were excluded by BK were reviewed by a second study author (AK) to ensure that their exclusion would not remove data that would be meaningful in answering our research question. This second author's (AK) screening led to the re-inclusion of one article. Our final corpus was 39 manuscripts spanning in publication date from 2003 to 2022. Table 1 provides a summary of the final corpus, with more details available in Appendix B.

**Table 1:** Summary of the included manuscripts

Publishing journal	Medical Education (n=13) Advances in Health Sciences Education (n=7) Academic Medicine (n=2) Medical Teacher (n=2) Advances in Medical Education and Practice (n=1) Advances in Simulation (n=1) American Journal of Medicine (n=1) American Journal of Pharmaceutical Education (n=1) American Journal of Surgery (n=1) Current Urology Reports (n=1)
--------------------	---

	<p>Currents in Pharmacy Teaching and Learning (n=1)</p> <p>Internal Medicine Journal (n=1)</p> <p>Journal of Graduate Medical Education (n=1)</p> <p>Journal of Clinical Nursing (n=1)</p> <p>Journal of Endourology (n=1)</p> <p>Journal of Pediatrics (n=1)</p> <p>MedEdPortal (n=1)</p> <p>Perspectives on Medical Education (n=1)</p> <p>Surgical Endoscopy (n=1)</p>
Country of first author professional affiliation	<p>United States (n=22)</p> <p>Canada (n=9)</p> <p>Australia (n=2)</p> <p>United Kingdom (n=2)</p> <p>Egypt (n=1)</p> <p>Germany (n=1)</p> <p>Netherlands (n=1)</p> <p>New Zealand (n=1)</p> <p>Sudan (n=1)</p>

*Note that one author listed affiliations in two countries, both of which are listed*

## ANALYSIS

Our bricolage approach to discourse analysis drew on multiple traditions. First, like Laclau and Mouffe [17], we approached discourse as a net of linguistic terms and processes through which meaning is constantly created and negotiated [17]. In particular, based on how we saw terminology being used in our prior work [13], Laclau and Mouffe's concept of nodal points (i.e. words/concepts around which other concepts are ordered) helped us to consider the terms *validity* and *argument* as knots in the net of HPE's discourse; these words, as nodal points, derive their meaning by being situated in particular positions in HPE's ever shifting discourse. By examining how these terms are used in our corpus as nodal points, we were able to understand how these words were prioritized and used to set meaning within HPE's discourse.

Our approach was also informed by Fairclough's critical discourse analysis [21] which frames discourse as a social practice "that both reproduces and changes knowledge" [17]. It emphasizes that individual linguistic choices (i.e., words) "can only be understood in relation to webs of other texts and in relation to the social context" [17]. We used the concept of *legitimation* (i.e. the process by which systems of authority seek to establish a belief as legitimate) to explore how discourses related to *validity as an argument* are established and empowered. This bricolage approach employed multiple tools to understand what the authors of each manuscript in the corpus meant or intended to accomplish with their use of language.

Guided by what we found in our previous work on Argumentation Theory [22], we developed an analytic framework—acting as bricoleurs to bring together Laclau and Mouffe, Fairclough, and Gee —consisting of questions that we asked of the data during coding (Table 2). Each question served to apply various tools from Gee to explore the language used in our discourses of interest. One author (BK) read each of the manuscripts in the final corpus to familiarize himself with the texts. He then reread and followed by a second co-investigator (DS, LV, AK, or ED) independently coding one of the four manuscripts each. We then met to review our coding and to discuss the fit of our analytic framework to our research aim of elucidating how authors used language to set meaning and influence HPE’s understanding of validity arguments. After this meeting, BK coded 16 more manuscripts, assigning a second coder for each (every co-investigator coded 5 of the 20 initial manuscripts). The team met again to discuss codes, language patterns, nodal points, and developing themes. At this meeting, the team further developed a shared mental model for future coding and theme development. BK then read and coded the remaining 18 manuscripts using the established model. We felt that we had reached theoretical sufficiency prior to coding all 39 manuscripts; however, we analyzed the entire corpus to identify rich quotes or examples that might otherwise be missed. We used Quirkos (Quirkos 2.5.2 Computer Software, 2022) to facilitate primary coding, to develop a visual representation of how codes could be grouped and related, and to facilitate reflections and interpretations throughout the process.

**Table 2:** Analytic framework used by the authors for primary coding

Question	Gee’s discourse analysis tools (16)	Rationale for asking
How are audiences of validity arguments addressed?	<b>Fill in tool</b> – asking what is not being explicitly said but is inferable  <b>Making strange tool</b> – acting as an outside in the discourse	Audience was identified as central in validation in study 2
What evaluation criteria is explicitly stated or implied by the authors?	<b>Fill in tool</b> – asking what is not being explicitly said but is inferable  <b>Making strange tool</b> – acting as an outside in the discourse	<i>Logic</i> = truth between premises/conclusions  <i>Informal logic</i> = relevance, sufficiency, acceptability  <i>New Rhetoric</i> = convincing, persuasion  <i>Pragmadiialectics</i> = winning against an interlocutor
Who is in the author group? What authority is claimed or implied for their writing of the topic?	<b>Identities building tool</b> – asking which social identities are assumed or enacted  <b>Significance building tool</b> – asking how language builds or diminishes significance	These papers set the discourse for our field. Writing them is an implicit claim to authority.

What argumentation ideas/concepts/terms (e.g. claim, warrant, rebuttal) are explicitly used by authors? What might this imply?	<b>Fill in tool</b> – asking what is not being explicitly said but is inferable  <b>Making strange tool</b> – acting as an outside in the discourse	This may help understand how much argumentation is emphasized or used in HPE validity
What argumentation concepts are being assumed by authors?	<b>Fill in tool</b> – asking what is not being explicitly said but is inferable  <b>Making strange tool</b> – acting as an outside in the discourse	Much about argumentation is not fully explored in HPE validity writing
What is being problematized about argumentation? Who is addressed as having power to ‘fix’ the outlined problem?	<b>Activities building tool</b> – asking what activity a discourse is seeking to perform  <b>Identities building tool</b> – asking which social identities are assumed or enacted	Study 2 suggest argumentation isn’t really happening. I’d like to know if anything about arguments/argumentation has already been problematized
How do the authors use intertextuality?  How much dialogicity do they use (i.e. description vs naming vs quotes).  What references are made or what sources of knowledge are employed?	<b>Intertextuality tool</b> – Asking how a discourse draws from other texts	Lots of intertextuality and using names (e.g. Messick, Kane, Cook) of validity scholars noted throughout. Why do they use intertextuality that way?
What are the “nodal points” (privileged sign around which other signs are ordered) that help set meaning related to argumentation? How are they used?	<b>Significance building tool</b> – asking how language builds or diminishes significance	Certain terms, concepts, or ideas may be important in describing/defining arguments/argumentation

## REFLEXIVITY

Our team members’ unique areas of expertise and personal backgrounds informed our data analysis and interpretation. AK is a critical qualitative researcher with a background in linguistics. She came to this work with a particular interest and expertise in authors’ discursive techniques. ED, LV, DS, and BK brought social constructionist views to the work, viewing knowledge as being constructed by and between people. DS and BK are clinician educators who actively engage in HPE research. BK is a PhD candidate with expertise in programmatic assessment whose interpretations were influenced by prior research in the area of validity argumentation [13, 22], and by his experience as a frontline clinician educator and residency program leader.

DS is a pediatric emergency medicine physician and PhD-trained research scientist with expertise in competency-based medical education (CBME) and programmatic assessment. ED is an educational scientist whose PhD research included examining the validity of portfolio assessment. His experience as editor-in-chief of a HPE journal has influenced his view of the discourses created through the publication process. LV is a PhD-trained HPE researcher with expertise in argumentation theory, rhetoric, and professional communication, and in qualitative research's paradigms, methodologies, and methods. This expertise shaped her interpretations of the manuscripts.

As a team, our varied educational backgrounds, areas of research expertise, and work experiences in HPE meant that we often pushed each other's interpretations and questioned the various kinds of legitimacy that were being harnessed in any given manuscript. Also, given that this project was undertaken by BK as part of his thesis research and that the other members of the team included supervisors of his thesis research and collaborators invited to the project because of their expertise, the team was sensitive to the power dynamics shaping the collaboration. The team focused on simultaneously meeting two goals: to support BK in developing research skills by having him lead the analysis of the corpus; and to ensure that rigorous, theory informed discourse analysis was carried out.

## RESULTS

Although *validity as an argument* is a prominent discourse within HPE literature (as noted previously by St. Onge et al) [2], we found the elaboration of *argument* and *argumentation* within this discourse to be scant. Specifically, we found that the discourse of validity arguments neglected or minimized audiences for those arguments. The discourse was legitimated via multiple methods that cemented *validity as an argument* in HPE despite minimal exploration of what argument and argumentation are. Below, we elaborate on four aspects of the discourse on HPE validity arguments: lack of argumentation, absence of audience, epistemological tension, and legitimation.

### VALIDITY ARGUMENTS WITHOUT ARGUMENTATION

Relatively few manuscripts explained the concepts of argument or argumentation. In nearly all papers, the *validity as an argument* discourse centered around the concepts of *evidence* and *inference*. These concepts served as nodal points when describing validity as an argument and were often linked to two specific frameworks: evidence was often described within Messick's framework [23] and inference was often used in the context of Kane's framework [24]. Most articles focused significantly on evidence and did not describe other aspects of argument or argumentation. Multiple articles did not describe nor provide detailed explanations on arguments at all. Rarely, articles explicitly noted that validity requires more than evidence collection. For example, one author noted: "Contrary to popular opinion, evidence does not speak for itself. To convince a jury, evidence must be selected, interpreted and presented in a carefully orchestrated

argument [4].” Despite the prevalence of the *validity as argument* discourse, an approach to building an argument was rarely articulated.

No explicit articulation of an argumentation theory, orientation, or philosophy (e.g. Informal Logic, New Rhetoric) appeared anywhere in our corpus. Perhaps as a corollary, papers in our corpus used general and heterogenous language when describing criteria for evaluating validity arguments such as “coherent” [11, 25–31] “credible” [26, 32, 33], “clear” [11, 31], “complete” [11], defensible [28, 30, 32–35], plausible [3, 11, 25, 31], justifiable [4, 34, 36–39], well-grounded [4, 36], trustworthy [33, 36, 40], reasonable [1, 33, 36], sound [32, 41], accurate [39, 40], sufficient [2, 29, 32, 40, 42], acceptable [35, 43], and specific [11]. One author explicitly acknowledged the difficulty in elaborating evaluation criteria for validity arguments, writing: “In theory it is difficult to define the point at which we can claim that a test is valid.” [44] Across the corpus, there was considerable ambiguity about how validity arguments should be developed and evaluated.

### **ARGUMENTS WITHOUT AUDIENCE**

Similarly, the *validity as an argument* discourse did not describe who can or should be the audiences addressed by such arguments. By *audience* we mean the person, group, or institution who receives, interprets, and evaluates validity arguments. The term audience only appeared one time within the entire corpus in reference to validity arguments. While this manuscript did not explicitly state who the audience is or should be, it acknowledged the centrality of audience values in determining validity of arguments: “[an argument] is not ‘valid’ or ‘invalid’, but varies by degrees and will be more or less convincing to different audiences.” [40] Some manuscripts used terms such as users [26, 29, 45] or stakeholders [32, 35, 44], but these terms were often subordinate to the nodal points of evidence and inference. One author made explicit that readers of journal articles are the target audience, writing “it is up to the reader to evaluate whether sufficient evidence is provided to support the adaptation of a tool [46].”

Several manuscripts did not address audience at all; instead, they focused entirely on evidence. In fact, a common manuscript organization strategy was to have subsections focused entirely on evidence types or sources without attending to the broader argument or audience. Some manuscripts intimated the existence of an audience through language that implied an exchange of information such as “argue for and against” [1], “theory, hypotheses, and logic which are presented” [1], “no validation work can possibly address all concerns” [46]. However, these manuscripts gave no clear description to whom the arguments are argued or presented, nor whose concerns should be addressed. The minimization (or absence) of audience reduced the engagement of social exchanges as part of HPE validity arguments.

### **EPISTEMOLOGICAL TENSIONS IN VALIDITY DISCOURSE**

The language used to describe nodal points revealed a diversity of epistemological worldviews within the *validity as an argument* discourse. By *epistemology* we mean implicit beliefs about how knowledge is generated or used [5], in this case specifically regarding validity arguments. For example, the nodal point of

evidence was frequently described as being “obtained” [37], “compiled” [28], “gathered” [39], “sought” [33, 39], or “collected” [4, 26, 27, 33, 47], implying that evidence objectively exists and, like a buried fossil, waits to be discovered. In contrast, arguments were frequently described as being built [11, 34], formulated [29], or constructed [26], implying that they rely on human agency and decisions to exist. These characterizations often co-occurred in the same text or even within the same sentence: “The process of building an argument and collecting evidence supporting that interpretation and its decision-making conclusions, with each different learning assessment use and among different test-takers, is termed validation [28].”

These descriptions imply an objective, more post-positivist view of *evidence* with a more subjective, interpretivist conceptualization of *arguments*. Such implicit views were reflected in how evidence and argument were seen to be operationalized in validity. Authors described “closing gaps” [4, 27] in evidence, again implying an objective and ideal type of evidence that researchers should strive to uncover. One author used terms that are often used in qualitative research but have a postpositivist underpinning [48], such as “triangulate” [32] and “saturation” [32], in describing the collection of validity evidence. Another author highlighted the interpretive nature of arguments, noting: “Kane makes no claims about representing a ‘truth’, and instead argues for a justified belief obtained using whatever means necessary, leaving truth in the background [38].” In sum, the language used within HPE’s *validity as argument* discourse contains heterogeneous epistemological groundings that can be contradictory across manuscripts and even within a single manuscript.

## LEGITIMATING VALIDITY AS AN ARGUMENT

Legitimation is the use of textual explanations and justifications for defending or upholding how things are done [21]. Authors in our corpus used multiple methods of legitimation to cement the *validity as an argument* discourse. Nearly every text used authorization (i.e. reference to authority) via intertextuality (i.e. relating different texts to one another). While referencing previous literature is part of scholarly writing’s tradition, authors in our corpus nearly always referenced high-profile validity scholars by name. While some scholars were named multiple times (e.g. Cook, Cronbach, Downing), Messick and Kane appeared in almost every paper, sometimes appearing with direct quotes from their original work. This demonstrates a high degree of intertextuality for purposes of authorization. High intertextuality includes directly naming or quoting other authors [21]. One author in our corpus even portrayed Kane as speaking directly with readers using the present tense, writing “Kane reminds us...” [34] In addition to citing specific scholars, authors in our corpus also used high intertextuality in referencing the Standards for Educational and Psychological Testing [49] as a method of authorization: “All validity is construct validity in this current framework, described most eloquently by Messick and embodied in the current Standards of Educational and Psychological Measurement” [1]. By referencing Messick and Kane by name and frequently, and by citing authoritative standards, the manuscripts in the corpus create the impression that an accepted consensus exists around

the *validity as an argument* discourse which bolstered the discourse's legitimacy, despite the fact that no such consensus exists.

Another legitimization approach we observed in the corpus was mythopoesis, which conveys legitimacy through narrative [21]. Numerous papers in our corpus included a historical narrative about how the conceptualization of validity has changed over time [1, 4, 25, 26, 28–30, 34, 36, 37, 44, 45, 50, 51]. There were striking similarities in the narratives, often beginning with validity's conceptualization in the 1950s and invoking the same validity scholars' names (e.g., Messick and Kane) as characters in the narrative. Multiple papers described how validity has “evolved” over time [25, 26, 30, 33, 37, 52, 53]. The term *evolve* implies that validity changed into an improved, more advanced state, thereby lending legitimacy to the current argument-based conceptualization. This narrative often framed argument-based approaches as “contemporary” [1, 32, 33, 35–37, 41, 43, 45] or “modern” [2, 32, 34, 35, 45, 50, 52], while non-argument based approaches were labelled as “outdated” [28, 35, 37, 45], “misconceptions and malpractices” [37], and “antiquated” [2, 54]. The evolutionary narrative was often presented with language that implied how the world ought to be such as “should” [27, 30, 32, 34–36, 39, 40, 53, 55, 56] or “must” [1, 11, 32, 41, 47, 57]; implied a request or command such as “Provide information such as who created the instrument” [57]; or gave explicit recommendations about how validity ought to be conceptualized or operationalized, such as “Recommendation 1: Use current validity, reliability, and validation language that reflects today's understanding of these foundational concepts” [34]. The narratives in HPE's validity discourse (e.g. modern vs outdated, evolution, should/must, commands, recommendations) further reduced the possibility of questioning the *validity as an argument* discourse.

Rationalization, which invokes societal or institutional knowledge [21], is a final legitimization approach we found in the corpus. The language used to describe aspects of validity arguments often echoed the scientific method—a stalwart of knowledge and progress. For example, multiple authors used terms like “hypothesis” [1, 25, 29, 30, 33, 36, 37, 45] and “theory” [3, 11, 25, 33–35, 58] in describing validity and validation, grounding the discourse in traditional, accepted scientific language. One author wrote: “Validity refers to the impartial, scientific collection of data, from multiple sources, to provide more or less support for the validity hypothesis and relates to logical arguments, based on theory and data, which are formed to assign meaningful interpretations to assessment data [1].” This quote borrows language found in scientific, empirical papers, and so reinforces the legitimacy of the *validity as an argument* discourse. This use of terminology legitimizes validity by entrenching it in the language of science.

## DISCUSSION

This study builds on our previous work in which experts explained that argument and argumentation are largely absent in HPE validation work [13]. Our analysis highlights how the *validity as an argument* discourse was dominated by *evidence* without significant attention paid to the argumentation principles that might tie

such evidence together. This study also corroborates previous findings [13] in which validity experts stated that audience, a key component of argument, is often minimized or ignored entirely in HPE validity. Most argumentation orientations explicitly debate the identities, roles, and centrality (or lack of centrality) of audiences for arguments [12]. Some argumentation orientations minimize audience importance (e.g. Logic) [59, 60] while others entirely center on audience values and needs (e.g. New Rhetoric) [61, 62]. In HPE, this debate about the role of audience to influence validity's argument has yet to occur in a meaningful way. Some manuscripts in this study's corpus briefly referenced or indirectly implied the existence of an audience, but none deeply explored how audiences should interact with or evaluate validity arguments, nor how much a specific audience's views and values should guide validation work. Our findings do not suggest the presence of conflicting views on audience in HPE validation, but rather an absence of views. Audience seems to have been left behind in the immigration of argument into HPE validity. Ignoring or minimizing audience neuters the power of argument, transforming it simply into a validity claim. As argumentation theorist Frans van Eemeren wrote, "By itself, holding an opinion is not enough to initiate argumentation. Arguing makes sense only if there is a listener or reader who entertains doubt about an opinion or has a diverging opinion" [12]. In order to move from validity opinions to validity arguments, HPE must better define the role and identity of its audiences.

Our corpus was comprised of published manuscripts, and one might argue that reviewers, editors, or readers are serving as audiences. However, given the contextual and situated nature of validity [13], those audiences are unlikely to be meaningful stakeholders for most published validation work. Journal reviewers and editors are meaningful and necessary audiences for knowledge advancement, but the argumentation mechanisms used in HPE validation should allow for the identification of and engagement with context-relevant and -specific audiences. For that to happen, audiences need to be identified and engaged in co-creating validity arguments. Co-creation might include having validity committees that meet to review and discuss validity arguments, asynchronous sharing and evaluation of validity arguments between argument developers and stakeholders, or other approaches that best engage intended audiences. Given the context-dependent nature of validity arguments [2, 29, 37], as well as the dearth of diversity in the HPE journal editorial world [63], actual stakeholders and end users impacted by assessment decisions (e.g. learners, local program leaders, patients) should be highly sought after audiences. Additionally, diverse audiences that represent multiple backgrounds and experiences should be prioritized to ensure equitable arguments are being made. By deliberately attending to the function of audience, perhaps HPE validity can more tangibly meet its social imperative to serve learners and patients [32].

Our analysis found that published HPE manuscripts use multiple approaches to legitimate the discourse of *validity as an argument*. While we cannot say for certain why legitimization is so frequently used, we have several hypotheses. Legitimation may simply be part of the natural process of translating new ways of thinking across disciplines. The *validity as an argument* discourse did not originate in HPE, but, like many concepts, it immigrated [64] to HPE from fields such as Measurement and Language Testing. Authors may be using legitimization as a rhetorical strategy to promote uptake and integration of new ideas into HPE.

Legitimation may also be an indirect way to offset the lack of philosophical and theoretical grounding of argument and argumentation in HPE's validity discourse. Such legitimation could entrench *validity as argument* as the authoritative discourse in HPE without having to fully reconcile aspects of argument or argumentation that may or may not fit with current practices. Our previous work suggested that HPE has largely adopted *validity as an argument* without fully integrating theories or practices used in other forms of argumentation [13, 22]. Legitimation may be a form of compensation—in other words, legitimation may uphold the conceptualization of *validity as argument* despite the fact that we are unsure of what that actually means.

Rather than a weakness, we see the lack of argument and argumentation in HPE as an opportunity. HPE can pull from the rich history of Argumentation Theory to answer questions such as: By which qualities should validity arguments be evaluated and using what threshold? What argument structures might best facilitate evaluation? How should audience values impact argument content? Should specific or general/universal audiences be prioritized? With regard to argument evaluation, the language used in HPE's discourse implies multiple different argumentation orientations such as Informal Logic (e.g. sufficient, acceptable, defensible) or New Rhetoric (e.g. plausible, convincing). Clarifying which argumentation orientation is being used helps set the ground rules for what criteria are used to evaluate validity. In other words, more explicit argumentation expectations would provide the 'rules of chess' to those who hold the game pieces. Exploring such questions may also help ease epistemological tensions. We are not advocating for any particular worldview, and indeed believe that the multiple epistemologies used in HPE enrich our field [65]. Rather, we hope that better articulated argumentation orientations could improve compatibility [38] between philosophical worldviews and argumentation approaches being used.

Our study has multiple limitations. First, we began with the presumption that *validity as an argument* is the predominant view of validity in HPE. We recognize that other views on the nature of validity exist concomitantly in our field [2]. However, we believe that *validity as an argument* is currently the most widely used conceptualization. Second, we only analyzed a sample of the HPE validity literature; we do not offer an interpretation that represents every published manuscript. Our focused approach is reflective of our critical and social constructionist worldviews in which our goal was not to seek an objective truth, but to develop understanding informed by our experiences and expertise. Third, we did not include specific validation studies (e.g. those of a specific assessment instrument), instead focusing on manuscripts that we felt were intended to influence our field's views of validity and validation. Different discourses may exist in those studies that could lead to different results.

## CONCLUSION

Argumentation is missing from HPE's discourse on validity arguments, fostering ambiguity in terms of who the audiences are (or should be) and how arguments are evaluated (or should be). More explicit use of

Argumentation Theory in HPE, including the exploration of which orientations (e.g. Informal Logic, New Rhetoric) best support validity argumentation, may deepen our field's understanding of how *validity as an argument* can best serve the interests of our learners, programs, and patients. Despite a lack of argumentation and minimization of argument (in favor of evidence), *validity as an argument* is legitimated via multiple mechanisms which serve to entrench it as a definitive conceptualization throughout HPE publications. Future research should include the use of critical theory to explore who this legitimation favors, and who it marginalizes or excludes.

## REFERENCES

1. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003; 37(9): 830–7. DOI: <https://doi.org/10.1046/j.1365-2923.2003.01594.x>
2. St-Onge C, Young M, Eva KW, Hodges B. Validity: one word with a plurality of meanings. *Adv Health Sci Educ Theory Pract*. 2017; 22(4): 853–67. DOI: <https://doi.org/10.1007/s10459-016-9716-3>
3. von Bergmann H, Childs RA. When I say... validity argument. *Med Educ*. 2018; 52(10): 1003–4. DOI: <https://doi.org/10.1111/medu.13592>
4. Cook DA. When I say... validity. *Med Educ*. 2014; 48(10): 948–9. DOI: <https://doi.org/10.1111/medu.12401>
5. MacLeod A, Ellaway RH, Paradis E, Park YS, Young M, Varpio L. Being edgy in health professions education: Concluding the philosophy of science series. *Acad Med*. 2020; 95(7): 995–8. DOI: <https://doi.org/10.1097/ACM.0000000000003250>
6. ten Cate O, Regehr G. The power of subjectivity in the assessment of medical trainees. *Acad Med*. 2019; 94(3): 333–7. DOI: <https://doi.org/10.1097/ACM.0000000000002495>
7. Virk A, Joshi A, Mahajan R, Singh T. The power of subjectivity in competency-based assessment. *J Postgrad Med*. 2020; 66(4): 200. DOI: [https://doi.org/10.4103/jpgm.JPGM\\_591\\_20](https://doi.org/10.4103/jpgm.JPGM_591_20)
8. Hodges B. Assessment in the post-psychometric era: learning to love the subjective and collective. *Med Teach*. 2013; 35(7): 564–8. DOI: <https://doi.org/10.3109/0142159X.2013.789134>
9. Govaerts M, van der Vleuten CP. Validity in work-based assessment: expanding our horizons. *Med Educ*. 2013; 47(12): 1164–74. DOI: <https://doi.org/10.1111/medu.12289>
10. van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ*. 2005; 39(3): 309–17. DOI: <https://doi.org/10.1111/j.1365-2929.2005.02094.x>
11. Schuwirth LW, van der Vleuten CP. Programmatic assessment and Kane's validity perspective. *Med Educ*. 2012; 46(1): 38–48. DOI: <https://doi.org/10.1111/j.1365-2923.2011.04098.x>
12. Van Eemeren FH, Grootendorst R, Johnson RH, Plantin C, Willard CA. *Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments*. Routledge; 2013. DOI: <https://doi.org/10.4324/9780203811306>
13. Kinnear B, Martini A, Varpio L, Driessen EW, Schumacher DJ. How do validity experts conceptualise argumentation? It's a rhetorical question. *Med Educ*. 2024. DOI: <https://doi.org/10.1111/medu.15311>
14. Fairclough N. *Critical discourse analysis: The critical study of language*. Routledge; 2013. DOI: <https://doi.org/10.4324/9781315834368>
15. Archer MS. *Structure, agency and the internal conversation*. Cambridge University Press; 2003. DOI: <https://doi.org/10.1017/CBO9781139087315>
16. Gee JP. *How to Do Discourse Analysis: A Toolkit*. New York and London: Routledge; 2011. DOI: <https://doi.org/10.4324/9780203850992>
17. Jorgensen MW, Phillips LJ. *Discourse analysis as theory and method*. Sage; 2002. DOI: <https://doi.org/10.4135/9781849208871>

18. Wyatt TR, Zaidi Z. Bricolage: A tool for race-related, historically situated complex research. *Med Educ*. 2022; 56(2): 170–5. DOI: <https://doi.org/10.1111/medu.14629>
19. Kuper A, Whitehead C, Hodges BD. Looking back to move forward: using history, discourse and text in medical education research: AMEE guide no. 73. *Med Teach*. 2013; 35(1): e849–e60. DOI: <https://doi.org/10.3109/0142159X.2012.748887>
20. Hodges BD, Kuper A, Reeves S. Discourse analysis. *BMJ*. 2008; 337: a879. DOI: <https://doi.org/10.1136/bmj.a879>
21. Fairclough N. *Analysing discourse: Textual analysis for social research*. Psychology Press; 2003. DOI: <https://doi.org/10.4324/9780203697078>
22. Kinnear B, Schumacher DJ, Driessen EW, Varpio L. How argumentation theory can inform assessment validity: A critical review. *Med Educ*. 2022; 56(11): 1064–75. DOI: <https://doi.org/10.1111/medu.14882>
23. Messick S. Validity. In: R Linn (Ed.), *Educational measurement (13–103)*. New York: Macmillan Publishing; 1989.
24. Kane MT. An argument-based approach to validity. *Psychological bulletin*. 1992; 112(3): 527. DOI: <https://doi.org/10.1037/0033-2909.112.3.527>
25. Colliver JA, Conlee MJ, Verhulst SJ. From test validity to construct validity... and back? *Med Educ*. 2012; 46(4): 366– 71. DOI: <https://doi.org/10.1111/j.1365-2923.2011.04194.x>
26. Clauser BE, Margolis MJ, Holtman MC, Katsufakis PJ, Hawkins RE. Validity considerations in the assessment of professionalism. *Adv Health Sci Educ*. 2012; 17: 165–81. DOI: <https://doi.org/10.1007/s10459-010-9219-6>
27. Cook DA, Lineberry M. Consequences Validity Evidence: Evaluating the Impact of Educational Assessments. *Acad Med*. 2016; 91(6): 785–95. DOI: <https://doi.org/10.1097/ACM.0000000000001114>
28. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Adv Health Sci Educ*. 2014; 19(2): 233–50. DOI: <https://doi.org/10.1007/s10459-013-9458-4>
29. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Adv Simul (Lond)*. 2016; 1: 31. DOI: <https://doi.org/10.1186/s41077-016-0033-y>
30. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane’s framework. *Med Educ*. 2015; 49(6): 560–75. DOI: <https://doi.org/10.1111/medu.12678>
31. Uijtdehaage S, Schuwirth LW. Assuring the quality of programmatic assessment: Moving beyond psychometrics. *Perspect Med Educ*. 2018; 7(6): 350–1. DOI: <https://doi.org/10.1007/S40037-018-0485-Y>
32. Marceau M, Gallagher F, Young M, St-Onge C. Validity as a social imperative for assessment in health professions education: a concept analysis. *Med Educ*. 2018; 52(6): 641–53.
33. Cook DA, Kuper A, Hatala R, Ginsburg S. When assessment data are words: validity evidence for qualitative educational assessments. *Acad Med*. 2016; 91(10): 1359–69. DOI: <https://doi.org/10.1097/ACM.0000000000001175>
34. Peeters MJ, Martin BA. Validation of learning assessments: a primer. *Currents in Pharmacy Teach Learn Med*. 2017; 9(5): 925–33. DOI: <https://doi.org/10.1016/j.cptl.2017.06.001>

35. Goldenberg M, Lee JY. Surgical education, simulation, and simulators—updating the concept of validity. *Curr Urol Rep*. 2018; 19: 1–5. DOI: <https://doi.org/10.1007/s11934-018-0799-7>
36. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006; 119(2): 166 e7–16. DOI: <https://doi.org/10.1016/j.amjmed.2005.10.036>
37. Gasmalla HEE, Tahir ME. The validity argument: Addressing the misconceptions. *Med Teach*. 2021; 43(12): 1453–5. DOI: <https://doi.org/10.1080/0142159X.2020.1856802>
38. Tavares W, Kuper A, Kulasegaram K, Whitehead C. The compatibility principle: on philosophies in the assessment of clinical competence. *Adv Health Sci Educ*. 2020; 25(4): 1003–18. DOI: <https://doi.org/10.1007/s10459-019-09939-9>
39. Peeters MJ, Beltyukova SA, Martin BA. Educational testing and validity of conclusions in the scholarship of teaching and learning. *Am Journal Pharm Educ*. 2013; 77(9): 186. DOI: <https://doi.org/10.5688/ajpe779186>
40. Andreatta PB, Gruppen LD. Conceptualising and classifying validity evidence for simulation. *Med Educ*. 2009; 43(11): 1028–35. DOI: <https://doi.org/10.1111/j.1365-2923.2009.03454.x>
41. Downing SM, Haladyna TM. Validity threats: overcoming interference with proposed interpretations of assessment data. *Med Educ*. 2004; 38(3): 327–33. DOI: <https://doi.org/10.1046/j.1365-2923.2004.01777.x>
42. Cook DA. Much ado about differences: why expert-novice comparisons add little to the validity argument. *Adv Health Sci Educ*. 2015; 20(3): 829–34. DOI: <https://doi.org/10.1007/s10459-014-9551-3>
43. Downing SM. Face validity of assessments: faith-based interpretations or evidence-based science? *Med Educ*. 2006. DOI: <https://doi.org/10.1111/j.1365-2929.2005.02361.x>
44. Knorr M, Klusmann D. The trouble with validity: what is part of it and what is not? *Med Educ*. 2015; 49(6): 550–2. DOI: <https://doi.org/10.1111/medu.12738>
45. Noureldin YA, Lee JY, McDougall EM, Sweet RM. Competency-based training and simulation: making a “valid” argument. *J Endourol*. 2018; 32(2): 84–93. DOI: <https://doi.org/10.1089/end.2017.0650>
46. Adler M, Kshetrapal A, Calhoun A. Convince Me! How a Well-Constructed Validity Argument Supports the Decisions We Make About Learners. *J Peds*. 2022; 245: 11–2. DOI: <https://doi.org/10.1016/j.jpeds.2022.03.029>
47. Tavares W, Brydges R, Myre P, Prpic J, Turner L, Yelle R, et al. Applying Kane’s validity framework to a simulation based assessment of clinical competence. *Adv Health Sci Educ*. 2018; 23(2): 323–38. DOI: <https://doi.org/10.1007/s10459-017-9800-3>
48. Varpio L, Ajjawi R, Monrouxe LV, O’Brien BC, Rees CE. Shedding the cobra effect: problematising thematic emergence, triangulation, saturation and member checking. *Med Educ*. 2017; 51(1): 40–50. DOI: <https://doi.org/10.1111/medu.13124>
49. Standards for educational and psychological testing: Amer Educational Research Assn; 2014.
50. Royal KD. Four tenets of modern validity theory for medical education assessment and evaluation. *Adv Med Educ Pract*. 2017: 567–70. DOI: <https://doi.org/10.2147/AMEP.S139492>
51. Schuwirth LW, van der Vleuten CP. A history of assessment in medical education. *Adv Health Sci Educ*. 2020; 25(5): 1045–56. DOI: <https://doi.org/10.1007/s10459-020-10003-0>

52. St-Onge C, Young M. Evolving conceptualisations of validity: impact on the process and outcome of assessment. *Med Educ*. 2015; 49(6): 548–50. DOI: <https://doi.org/10.1111/medu.12734>
53. Cor MK. *Trust me, it is valid: Research validity in pharmacy education research*. *Currents in Pharm Teach Learn*. 2016; 8(3): 391–400. DOI: <https://doi.org/10.1016/j.cptl.2016.02.014>
54. Royal K. “Face validity” is not a legitimate type of validity evidence!. *Am J Surg*. 2016; 212(5):1026-1027. DOI: <https://doi.org/10.1016/j.amjsurg.2016.02.018>
55. Bannigan K, Watson R. Reliability and validity in a nutshell. *J Clin Nurs*. 2009; 18(23): 3237–43. DOI: <https://doi.org/10.1111/j.1365-2702.2009.02939.x>
56. Gallagher A, Ritter E, Satava R. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surgical endoscopy and other interventional techniques*. 2003; 17: 1525–9. DOI: <https://doi.org/10.1007/s00464-003-0035-4>
57. Sullivan GM. A primer on the validity of assessment instruments. *J Grad Med Educ*; 2011. pp. 119–20. DOI: <https://doi.org/10.4300/JGME-D-11-00075.1>
58. Tavakol M, Dennick R. The foundations of measurement and assessment in medical education. *Med Teach*. 2017; 39(10): 1010–5. DOI: <https://doi.org/10.1080/0142159X.2017.1359521>
59. Boole G. *The Laws of Thought*. Dover, New York (original edition 1854); 1957.
60. Aristotle JB. *The complete works of Aristotle*. Princeton, NJ: Princeton University Press; 1984.
61. Perelman C, Olbrechts-Tyteca L. *The new rhetoric: a treatise on argumentation, trans*. John Wilkinson and Purcell Weaver (Notre Dame, IN: University of Notre Dame Press, 1969). 1969; 19.
62. Frank DA, Bolduc M. Lucie Olbrechts-Tyteca's new rhetoric. *Quarterly Journal of Speech*. 2010; 96(2): 141–63. DOI: <https://doi.org/10.1080/00335631003796685>
63. Kusurkar RA. The leaky pipeline of publications and knowledge generation in medical education. *Perspect Medical Educ*. 2022; 11(2): 70–2. DOI: <https://doi.org/10.1007/S40037-022-00700-4>
64. Norman G. Fifty years of medical education research: waves of migration. *Med Educ*. 2011; 45(8): 785–91. DOI: <https://doi.org/10.1111/j.1365-2923.2010.03921.x>
65. Varpio L, MacLeod A. Philosophy of science series: Harnessing the multidisciplinary edge effect by exploring paradigms, ontologies, epistemologies, axiologies, and methodologies. *Acad Med*. 2020; 95(5): 686–9. DOI: <https://doi.org/10.1097/ACM.0000000000003142>



# CHAPTER 6

## **Validity in the Next Era of Assessment: Consequences, Social Impact, and Equity**

Published as:

Kinnear B, St-Onge C, Schumacher DJ, Marceau M, Naidu T. Validity in the Next Era of Assessment: Consequences, Social Impact, and Equity. Perspectives on Medical Education. 2024 Sep 11;13(1).

## **ABSTRACT**

Validity has long held a venerated place in education, leading some authors to refer to it as the “sine qua non” or “cardinal virtue” of assessment. And yet, validity has not held a fixed meaning; rather it has shifted in its definition and scope over time. In this Eye Opener, the authors explore if and how current conceptualizations of validity fit a next era of assessment that prioritizes patient care and learner equity. They posit that health profession education’s conceptualization of validity will change in three related but distinct ways. First, consequences of assessment decisions will play a central role in validity arguments. Second, validity evidence regarding impacts of assessment on patients and society will be prioritized. Third, equity will be seen as part of validity rather than an unrelated concept. The authors argue that health professions education has the agency to change its ideology around validity, and to align with values that predominate the next era of assessment such as high-quality care and equity for learners and patients.

Validity has long held a venerated place in education, leading some authors to refer to it as the “sine qua non” or “cardinal virtue” of assessment [1, 2]. And yet, validity has not held a fixed meaning; rather it has shifted in its definition and scope over time. How will validity change as health professions education (HPE) assessment evolves? In this Eye Opener, we explore if and how current conceptualizations of validity fit the values in a next era of assessment that focuses on ensuring high-quality care for patients. Specifically, we explore what might be required for validity to support a world in which assessment is more socially accountable and equity focused.

## **A BRIEF OVERVIEW OF VALIDITY AND SOME CONTEMPORARY CONCEPTUALIZATIONS**

Validity conceptualizations in HPE have evolved over time. In 2017, St-Onge et al. [3] made explicit three different, co-existing conceptualizations of validity in the HPE literature: validity as a test characteristic, validity as an argument-based evidentiary chain, and validity as a social imperative (a conceptualization still nascent in HPE).

The first conceptualization, validity as a test characteristic, is strongly tied to measurement models, namely Classical Test Theory, Generalizability Theory and Item-Response Theory [4, 5]. These theories and models aim to quantify measurement error and infer individuals’ “true” scores [4, 6, 7]. Reliability and validity are significantly intertwined, with the pursuit of a true score (Classical Test Theory), or a generalizability coefficient (reliability of score given a specified universe of generalization) [4]. Additional pursuit of score precision can be seen in Item Response Theory, which focuses on individual item-level difficulty [5, 6]. In this view of validity, quantitative evidence to support an assessment score’s reliability, generalizability, or precision is highly valued. Validity is a characteristic attributed to a test, indicating that “it measured what it intended to measure” [8, 9]. This conceptualization of validity still exists in HPE, most often with regard to sellable assessment products.

The second conceptualization, validity as an argument-based evidentiary chain, focuses on documenting the appropriateness of the interpretations and decisions made based on assessment data [3, 10]. Two argument-based approaches have been predominantly imported into HPE, Messick’s unified theory of validity [11] and Kane’s approach to validation [12, 13]. Authors that imported these approaches into HPE translated abstract validity conceptualizations into more concrete validation practices (e.g., Cook and Hatala [14], Kinnear et al. [15]). With the multi-faceted and complex programs of assessment that are increasingly found in competency-based education (CBE) [16], argument-based approaches allow for multiple, different types of evidence to be developed and integrated into fit-for-purpose arguments about the validity of assessment decisions. Validity as a social imperative (the third conceptualization) has grown out of argument-based approaches and, as we detail below, aligns well with the next era of assessment.

## **THE NEXT ERA OF VALIDITY IN HPE ASSESSMENT**

The next era of validity will be shaped by broader forces and trends in HPE assessment. As a result, we believe validity will change in three related but distinct ways. First, HPE has already integrated the consequences of assessment decisions into validity conceptualizations, though consequences remain mostly unaddressed in real-world validation work. In the next era, consequences of assessment decisions will play a central role in validity. Second, the proliferation of CBE has foregrounded assessment's role in social accountability. In the future, validity arguments will be more directly connected to impacts on patients and society. Third, equity considerations have become central in many HPE spaces. Similarly, equity will become central to future validity arguments. While consequences of assessment, social accountability, and equity are not novel concepts in HPE assessment, they are not central to most work on validity. Below we expand on how each will play a central role in the next era of validity.

### **FOCUS ON CONSEQUENCES OF ASSESSMENT DECISIONS**

Messick's unified theory of validity brought with it the concept of different sources of evidence that can be sought to support the validity of assessment decisions [11]. One such source of evidence was the social consequence of test uses, which Messick called "consequential validity" evidence [17], now sometimes called "consequential evidence" [18]. Cook and Lineberry explored consequential evidence in HPE, describing it as "the impact, beneficial or harmful and intended or unintended, of assessment." [18]. Consequences include impacts on learners, educators, programs, patients, and other systems and people. The concept of consequential validity evidence has gained acceptance in HPE, regularly appearing in HPE manuscripts describing the concept of validity and the process of validation [1, 19–24]. Even the oft-cited Standards for Educational and Psychological Testing includes consequential evidence as important for validity arguments [10].

Cook et al. argue that "evidence of consequences is ultimately the most important source of validity evidence" [18]. The authors take a teleological stance, drawing an analogy with clinical diagnostic tests. Regardless of a diagnostic test's sensitivity or specificity, its ultimate value will depend on consequences to patients, hospitals, and society. Similarly, while all sources of validity evidence have value, consequential evidence should be central to any validity argument. Despite this, consequential evidence is one of the least reported types of validity evidence. Across three systematic reviews on HPE validation work [25–27], consequential evidence was reported in only 5–20% of studies [18]. While one cannot say for certain why consequential validity evidence is relatively rare in HPE, contributing reasons likely include challenging study designs, limited resources for validation work, and vestigial preferences for more psychometric data such as reliability (i.e. internal structure) or criterion (i.e. relationship to other variables) evidence.

By making consequential evidence part of (or central to) validity arguments, we expand validity's reach. Validation becomes more than 'demonstrating that you are measuring what you think you are measuring' to also include downstream effects of assessment. The next era of validity should include more widespread

integration of consequential evidence into validity arguments. In the following sections, we explore how studying social impact and equity in assessment can provide meaningful consequential evidence.

## **EVIDENCE OF THE SOCIAL IMPACT OF ASSESSMENTS**

As noted above, consequential validity evidence includes impacts of assessment decisions on patients and society, and hence represents a form of social accountability. CBE, the predominant training philosophy in HPE in many countries, is rooted in social accountability [28, 29]. Marceau et al. recently made explicit the concept of validity as a “social imperative” [30, 31], in which validation is a mechanism to ensure that assessment decisions are linked to societal impacts. This view brings a deontological lens to consequential validity evidence by emphasizing HPE’s social contract with the patients it ultimately serves. To that end, the next era of validity will require evidence that assessment ensures trainees and graduates of HPE programs are providing high-quality care.

While the connection between HPE and patient outcomes is complex and non-linear [32], promising approaches are emerging to develop such consequential validity evidence. Clinical care measures that are seen as sufficiently attributable to individual trainees are being developed in multiple medical specialties [33–36]. Improving technology, such as haptics and artificial intelligence, could provide real-time assessment of procedural, communication, and teamwork skills [37]. Better understanding of interdependence of competence could unlock new ways of assessing team-based care outcomes [38, 39]. All of these approaches can be integrated into current programs of assessment to bolster social accountability by connecting education to patient care.

## **CENTERING EQUITY IN ASSESSMENT VALIDITY**

By recognizing consequential evidence as essential to validity arguments, we also make equity central to validity. We define equity in assessment as the opportunity to demonstrate and develop one’s knowledge, skills and abilities without negative influence by “structural or interpersonal bias related to personal or social characteristics of learners or assessors.” [40]. Equity goes beyond impartiality and includes efforts to ensure that each learner is afforded the resources and opportunities that they need to succeed, acknowledging that individuals need different types and levels of support and face different societal and system biases [41–44].

Equity is certainly not new to assessment, with scholars and advocates noting the many biases and injustices that have plagued HPE assessment for years [45–49]. Taking a sociocultural view, assessment has played a key role in creating and maintaining hegemony via control of patronage and access to educational and professional opportunities [50]. Performance on any particular assessment favors the dominant social order which influenced an instrument’s creation, while establishing what is accepted as truth and knowledge [50]. Viewed this way, assessment reinforces power structures while normalizing judgment. Thus, attending to equity is critical to promote fairness and justice for everyone impacted by assessment, particularly those who have been marginalized in a society.

Contemporary advocacy efforts such as #MeToo; Black Lives Matter; and advancing LGBTQ+, Feminist and Indigenous rights have brought equity to the fore of many discourses in HPE. Inequitable assessment is increasingly recognized as a driver of significant and tangible negative effects on learners that amplify and compound over time [45]. Equitable assessment should include choices of assessments that are inclusive of learners who require accommodations [41, 51]. Current assessment accommodations often require learners to come forward, self-identify, and justify their requests [52]. However, education systems are rarely designed to help learners feel comfortable enough to do this [41, 44]. Inequitable assessment also stands to harm patients by reducing the diversity of healthcare professionals that are available to serve diverse patient populations [53–55]. The next era of assessment brings a growing urgency to foreground equity in assessment [41, 44, 56, 57], and validation practices should align with such goals.

Evidence of equity in assessment can be sought by examining the design of assessment tools (i.e. intrinsic equity), the learning environment (i.e. contextual equity), or the uses of assessment data to create equitable opportunities (i.e. instrumental equity) [46]. Onumah et al provide an example of how assessment systems can be designed with all three facets of equity in mind [58]. Equity also means programs should seek to understand how colonialism, racism, and Global North Euro-American principles have shaped HPE's ideology and propagated inequities [47]. Including equity in validity arguments means that if assessment decisions are shown to worsen inequity for learners or patients, then we deem those decisions not valid.

Notably, we are not implying that equity is secondary or subordinate to validity, nor that equity is only important if examined through the lens of validity. We also do not believe that all of the richness, complexity, and nuance of equity initiatives can be captured within a validity argument. However, equity has long been treated as an afterthought in HPE assessment. By making equity a central part of validity arguments, it too becomes a 'cardinal virtue' of assessment. Therefore, in the next era of assessment, equity can function much like the concept of reliability – standing as a distinct concept while also being an integral part of validity arguments.

## **BROAD VS NARROW CONCEPTUALIZATIONS OF VALIDITY: CHOOSING OUR IDEOLOGY**

HPE is not a monolithic group, and we do not expect that everyone will agree with our call to foreground consequences, social impact, and equity in assessment validity. Some validity scholars disagree with the assertion that consequential evidence should be part of validity at all, instead advocating that validity should focus only on construct representation, not the downstream impacts of assessment [59–64]. We are not implying that those scholars do not care about the consequences of assessment such as societal impact or equity. Rather, they see consequences as being different from validity, to be considered separately under categories such as “utility” or “acceptability”. The varying conceptualizations of validity reflect the many disciplines that comprise HPE (e.g. Psychology, Sociology, Measurement, Education) [65], and this diversity of viewpoints represents a strength for our field. Which view will predominate in the next era of assessment?

The wonderful news is that HPE has agency in such a choice. As Varpio points out, “Fortunately, ideology is maintained by our decisions and actions; therefore, we can change our decisions and thereby modify the ideology to work for us, not against us.” [66]. As the next era unfolds, we can align our conceptualizations of validity with the values underpinning our assessment work. As noted at the outset of the article, validity holds a long and tenured position as being the “sine qua non” of assessment. We believe consequences, social impact and equity deserve the same status in the next era or assessment.

A significant challenge will involve anticipating and identifying all relevant consequences of given assessments. When considering the equity consequences of assessment, where do we begin and end? Equity initiatives often range from efforts to improve equitable access to medical education, to ensuring diverse individuals have proper resources to succeed in training programs, to monitoring for negative downstream consequences of assessment decisions. Real-world decisions will be needed to determine where to focus energy and resources in developing validity arguments with a seemingly infinite amount of evidence to be collected. To navigate these discussions, a socio-constructivist approach to validity could be embraced to co-construct the argument to support the defensibility of these assessment choices. Or perhaps critical theory approaches can help ensure that we are scrutinizing who gets to decide what is equitable, whose voice is being valued, and why.

The next era of validity will require adopting a more inclusive perspective. Most, if not all, current authors and leaders in validity are from Europe and North America [67, 68]. Thus, we must be mindful before transposing these validity conceptualizations and practices in other contexts and regions. Anecdotal evidence suggests that Global North customs and practices may not always transpose well in the Global South context. This should be further investigated. We can learn from the current and ongoing challenges of applying Global North conceptualizations in the Global South to challenge our assumptions about validity and validation and inform future development. We should also be open to the idea of validity expanding even further to include not just equity, but ideas such as social responsiveness and awareness. We have agency to determine where the boundaries of validity lie, how they can and should change over time, and which approaches best serve our learners and patients.

## **CONCLUSION**

The values of consequences, social accountability and equity will significantly influence assessment and validity in their next era. These values will undoubtedly challenge the current approach to validation and may require some to reconsider what falls under the purview of validity. A broader conceptualization of validity and validation that incorporates equity concepts in the purpose, design, and use of assessment data could contribute to assessments that are not just technically and psychometrically sound, but also socially accountable and equitable for learners and patients. Like so many required changes in assessment, changing our conceptualization of validity requires a shift in ideology. The next era in assessment has the potential to

catalyze novel ways to develop, share, and evaluate validity arguments with impacts on our patients and learners at the center of what we do.

## References

1. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003; 37(9): 830–7. DOI: <https://doi.org/10.1046/j.1365-2923.2003.01594.x>
2. Mislevy RJ, Steinberg LS, Almond RG. Focus article: On the structure of educational assessments. *Measurement: Interdisciplinary research and perspectives*. 2003; 1(1): 3–62. DOI: [https://doi.org/10.1207/S15366359MEA0101\\_02](https://doi.org/10.1207/S15366359MEA0101_02)
3. St-Onge C, Young M, Eva KW, Hodges B. Validity: one word with a plurality of meanings. *Adv Health Sci Educ Theory Pract*. 2017; 22(4): 853–67. DOI: <https://doi.org/10.1007/s10459-016-9716-3>
4. Cronbach LJ, Rajaratnam N, Gleser GC. Theory of generalizability: A liberalization of reliability theory. *Br J Stat Psychol*. 1963; 16(2): 137–63. DOI: <https://doi.org/10.1111/j.2044-8317.1963.tb00206.x>
5. Schuwirth LW, van der Vleuten CP. General overview of the theories used in assessment: AMEE Guide No. 57. *Med Teach*. 2011; 33(10): 783–97. DOI: <https://doi.org/10.3109/0142159X.2011.611022>
6. Hambleton RK, Swaminathan H. Item response theory: Principles and applications. Springer Science & Business Media; 2013.
7. Embretson SE, Reise SP. Item response theory. Psychology Press; 2013. DOI: <https://doi.org/10.4324/9781410605269>
8. Smith HL, Wright WW. Tests and measurements. Silver, Burdett; 1928.
9. Garrett HE. Statistics in psychology and education. Longmans, Green and Company; 1926.
10. Standards for educational and psychological testing. Amer Educational Research Assn; 2014.
11. Messick S. Validity. In R. Linn (Ed.) *Educational measurement* (13–103). New York: Macmillan Publishing; 1989.
12. Kane MT. An argument-based approach to validity. *Psychol Bull*. 1992; 112(3): 527. DOI: <https://doi.org/10.1037/0033-2909.112.3.527>
13. Kane MT. Validating the interpretations and uses of test scores. *J Educ Meas*. 2013; 50(1): 1–73. DOI: <https://doi.org/10.1111/jedm.12000>
14. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Adv Simul (Lond)*. 2016; 1: 31. DOI: <https://doi.org/10.1186/s41077-016-0033-y>
15. Kinnear B, Kelleher M, May B, Sall D, Schauer DP, Schumacher DJ, et al. Constructing a validity map for a workplace-based assessment system: Cross-walking Messick and Kane. *Acad Med*. 2021; 96(7S): S64–S9. DOI: <https://doi.org/10.1097/ACM.0000000000004112>
16. Schuwirth LW, van der Vleuten CP. Programmatic assessment and Kane's validity perspective. *Med Educ*. 2012; 46(1): 38–48. DOI: <https://doi.org/10.1111/j.1365-2923.2011.04098.x>
17. Messick S. Validity of test interpretation and use. *Research Report for the Educational Testing Service*; 1990. DOI: <https://doi.org/10.1002/j.2333-8504.1990.tb01343.x>

18. Cook DA, Lineberry M. Consequences Validity Evidence: Evaluating the Impact of Educational Assessments. *Acad Med*. 2016; 91(6): 785–95. DOI: <https://doi.org/10.1097/ACM.0000000000001114>
19. Sullivan GM. A primer on the validity of assessment instruments. Sullivan GM. A primer on the validity of assessment instruments. *J Grad Med Educ*. 2011 Jun 1; 3(2): 119–20. DOI: <https://doi.org/10.4300/JGME-D-11-00075.1>
20. Peeters MJ, Martin BA. Validation of learning assessments: a primer. *Curr Pharm Teach Learn*. 2017; 9(5): 925–33. DOI: <https://doi.org/10.1016/j.cptl.2017.06.001>
21. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006; 119(2): 166 e7–16. DOI: <https://doi.org/10.1016/j.amjmed.2005.10.036>
22. Andreatta PB, Gruppen LD. Conceptualising and classifying validity evidence for simulation. *Med Educ*. 2009; 43(11): 1028–35. DOI: <https://doi.org/10.1111/j.1365-2923.2009.03454.x>
23. Gasmalla HEE, Tahir ME. The validity argument: Addressing the misconceptions. *Med Teach*. 2021; 43(12): 1453–5. DOI: <https://doi.org/10.1080/0142159X.2020.1856802>
24. Noureldin YA, Lee JY, McDougall EM, Sweet RM. Competency-based training and simulation: making a “valid” argument. *J Endour*. 2018; 32(2): 84–93. DOI: <https://doi.org/10.1089/end.2017.0650>
25. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Adv Health Sci Educ Theory Pract*. 2014; 19(2): 233–50. DOI: <https://doi.org/10.1007/s10459-013-9458-4>
26. Beckman TJ, Cook DA, Mandrekar JN. What is the validity evidence for assessments of clinical teaching? *J Gen Intern Med*. 2005; 20: 1159–64. DOI: <https://doi.org/10.1111/j.1525-1497.2005.0258.x>
27. 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): 1316–26. DOI: <https://doi.org/10.1001/jama.2009.1365>
28. Morcke AM, Dornan T, Eika B. Outcome (competency) based education: an exploration of its origins, theoretical basis, and empirical evidence. *Adv Health Sci Educ Theory Pract*. 2013; 18(4): 851–63. DOI: <https://doi.org/10.1007/s10459-012-9405-9>
29. Hodge S. The Origins of Competency-Based Training. *Aust J Adult Learn*. 2007; 47(2): 179–209.
30. Marceau M, Gallagher F, Young M, St-Onge C. Validity as a social imperative for assessment in health professions education: a concept analysis. *Med Educ*. 2018; 52(6): 641–53. DOI: <https://doi.org/10.1111/medu.13574>
31. Marceau M, St-Onge C, Gallagher F, Young M. Validity as a social imperative: users’ and leaders’ perceptions. *Can Med Educ J*. 2022; 13(3): 22–36. DOI: <https://doi.org/10.36834/cmej.73518>
32. Varpio L, Sherbino J. Demonstrating causality, bestowing honours, and contributing to the arms race: Threats to the sustainability of HPE research. *Med Educ*. 2024; 58(1): 157–63. DOI: <https://doi.org/10.1111/medu.15148>
33. Schumacher DJ, Martini A, Holmboe E, Carraccio C, van der Vleuten C, Sobolewski B, et al. Initial implementation of resident-sensitive quality measures in the pediatric emergency department: A wide range of performance. *Acad Med*. 2020; 95(8): 1248–55. DOI: <https://doi.org/10.1097/ACM.0000000000003147>

34. Schumacher DJ, Holmboe ES, van der Vleuten C, Busari JO, Carraccio C. Developing resident-sensitive quality measures: A model from pediatric emergency medicine. *Acad Med*. 2018; 93(7): 1071–8. DOI: <https://doi.org/10.1097/ACM.0000000000002093>
35. Burk-Rafel J, Sebok-Syer SS, Santen SA, Jiang J, Caretta- Weyer HA, Iturrate E, et al. TRainee Attributable & Automatable Care Evaluations in Real-time (TRACERs): A Scalable Approach for Linking Education to Patient Care. *Perspect Med Educ*. 2023; 12(1): 149. DOI: <https://doi.org/10.5334/pme.1013>
36. Kalet AL, Gillespie CC, Schwartz MD, Holmboe ES, Ark TK, Jay M, et al. New measures to establish the evidence base for medical education: identifying educationally sensitive patient outcomes. *Acad Med*. 2010; 85(5): 844–51. DOI: <https://doi.org/10.1097/ACM.0b013e3181d734a5>
37. Kinnear B, Caretta-Weyer H, Lam AC, Tang B, Ginsburg S, Wong BM, et al. Finding Medicine's Moneyball: How Lessons from Major League Baseball Can Advance Assessment in Precision Education. *Acad Med*. 2023; 10.1097.
38. Sebok-Syer SS, Chahine S, Watling CJ, Goldszmidt M, Cristancho S, Lingard L. Considering the interdependence of clinical performance: implications for assessment and entrustment. *Med Educ*. 2018; 52(9): 970–80. DOI: <https://doi.org/10.1111/medu.13588>
39. Sebok-Syer SS, Lingard L, Panza M, Van Hooren TA, Rassbach CE. Supportive and collaborative interdependence: Distinguishing residents' contributions within health care teams. *Med Educ*. 2023 Oct; 57(10): 921–31. DOI: <https://doi.org/10.1111/medu.15064>
40. Rashid MA, Ali SM, Dharanipragada K. Decolonising medical education regulation: a global view. *BMJ Glob Heal*. 2023; 8(6): e011622. DOI: <https://doi.org/10.1136/bmjgh-2022-011622>
41. Ajjawi R, Tai J, Boud D, Jorre de St Jorre T. Assessment for inclusion in higher education: Promoting equity and social justice in assessment. Taylor & Francis; 2023. DOI: <https://doi.org/10.4324/9781003293101>
42. Aquario D. Through the lens of justice. A systematic review on equity and fairness in learning assessment. *Educ Sci Soc*. 2021(2021/2). DOI: <https://doi.org/10.3280/ess2-2021oa12405>
43. Hanesworth P, Bracken S, Elkington S. A typology for a social justice approach to assessment: learning from universal design and culturally sustaining pedagogy. *Teach High Educ*. 2019; 24(1): 98–114. DOI: <https://doi.org/10.1080/13562517.2018.1465405>
44. Tai J, Ajjawi R, Boud D, de St Jorre TJ. Promoting Equity and Social Justice through Assessment for Inclusion. *Assessment for Inclusion in Higher Education*. Routledge; 2022. p. 9–18. DOI: <https://doi.org/10.4324/9781003293101-3>
45. Teherani A, Hauer KE, Fernandez A, King Jr TE, Lucey C. How small differences in assessed clinical performance amplify to large differences in grades and awards: a cascade with serious consequences for students underrepresented in medicine. *Acad Med*. 2018; 93(9): 1286–92. DOI: <https://doi.org/10.1097/ACM.0000000000002323>
46. Lucey CR, Hauer KE, Boatright D, Fernandez A. Medical education's wicked problem: achieving equity in assessment for medical learners. *Acad Med*. 2020; 95(12S): S98–S108. DOI: <https://doi.org/10.1097/ACM.0000000000003717>

47. Anderson N, Nguyen M, Marcotte K, Ramos M, Gruppen LD, Boatright D. The long shadow: a historical perspective on racism in medical education. *Acad Med*. 2023; 98(8S): S28–S36. DOI: <https://doi.org/10.1097/ACM.0000000000005253>
48. Anderson H, Abdulla L, Balmer D, Govaerts M, Busari J. Inequity is woven into the fabric: a discourse analysis of assessment in pediatric residency training. *Adv Health Sci Educ Theory Pract*. 2024 Mar; 29(1): 199–216. DOI: <https://doi.org/10.1007/s10459-023-10260-9>
49. Gipps C. What do we mean by equity in relation to assessment? *Assess Educ*. 1995; 2(3): 271–81. DOI: <https://doi.org/10.1080/0969595950020303>
50. Gipps C. Chapter 10: Socio-cultural aspects of assessment. *Rev Res Educ*. 1999; 24(1): 355–92. DOI: <https://doi.org/10.3102/0091732X024001355>
51. McArthur J. Assessment for social justice: The role of assessment in achieving social justice. *Assess Eval High Educ*. 2016; 41(7): 967–81. DOI: <https://doi.org/10.1080/02602938.2015.1053429>
52. Kinnear B, Weber DE, Schumacher DJ, Edje L, Warm EJ, Anderson HL. Reconstructing Neurath’s Ship: A Case Study in Reevaluating Equity in a Program of Assessment. *Acad Med*. 2023 Aug 1; 98(8S): S50–6. DOI: <https://doi.org/10.1097/ACM.0000000000005249>
53. Snyder JE, Upton RD, Hassett TC, Lee H, Nouri Z, Dill M. Black representation in the primary care physician workforce and its association with population life expectancy and mortality rates in the US. *JAMA Net Open*. 2023; 6(4): e236687–e. DOI: <https://doi.org/10.1001/jamanetworkopen.2023.6687>
54. Cooper LA, Roter DL, Johnson RL, Ford DE, Steinwachs DM, Powe NR. Patient-centered communication, ratings of care, and concordance of patient and physician race. *Ann Intern Med*. 2003; 139(11): 907–15. DOI: <https://doi.org/10.7326/0003-4819-139-11-200312020-00009>
55. Saha S, Komaromy M, Koepsell TD, Bindman AB. Patient-physician racial concordance and the perceived quality and use of health care. *Arch Intern Med*. 1999; 159(9): 997–1004. DOI: <https://doi.org/10.1001/archinte.159.9.997>
56. Holmboe ES, Osman NY, Murphy CM, Kogan JR. The urgency of now: Rethinking and improving assessment practices in medical education programs. *Acad Med*. 2023; 98(8S): S37–S49. DOI: <https://doi.org/10.1097/ACM.0000000000005251>
57. Larson T. Josiah Macy Jr. Foundation Conference on Ensuring Fairness in Medical Education Assessment: Conference Recommendations Report. *Acad Med*. 2023; 98(8S): S3–S15. DOI: <https://doi.org/10.1097/ACM.0000000000005243>
58. Onumah CM, Pincavage AT, Lai CJ, Levine DL, Ismail NJ, Alexandraki I, et al. Strategies for advancing equity in frontline clinical assessment. *Acad Med*. 2023; 98(8S): S57–S63. DOI: <https://doi.org/10.1097/ACM.0000000000005246>
59. Clauser BE, Margolis MJ, Holtman MC, Katsufakis PJ, Hawkins RE. Validity considerations in the assessment of professionalism. *Adv Health Sci Educ Theory Pract*. 2012; 17: 165–81. DOI: <https://doi.org/10.1007/s10459-010-9219-6>
60. Borsboom D. Whose consensus is it anyway? Scientific versus legalistic conceptions of validity. *Meas Interdiscip Res Perspect*. 2012; 10(1–2): 38–41. DOI: <https://doi.org/10.1080/15366367.2012.681971>

61. Borsboom D, Wijsen LD. Frankenstein's validity monster: The value of keeping politics and science separated. *Assess Educ Princ Policy Pract*. 2016; 23(2): 281–3. DOI: <https://doi.org/10.1080/0969594X.2016.1141750>
62. Newton PE, Shaw SD. Disagreement over the best way to use the word 'validity' and options for reaching consensus. *Assess Educ Princ Policy Pract*. 2016; 23(2): 178–97. DOI: <https://doi.org/10.1080/0969594X.2015.1037241>
63. Newton PE, Baird J-A. The great validity debate. Taylor & Francis; 2016. DOI: <https://doi.org/10.1080/0969594X.2016.1172871>
64. Newton PE. Questioning the consensus definition of validity. *Meas Interdiscip Res Perspect*. 2012; 10(1–2): 110–22. DOI: <https://doi.org/10.1080/15366367.2012.688456>
65. Norman G. Fifty years of medical education research: waves of migration. *Med Educ*. 2011; 45(8): 785–91. DOI: <https://doi.org/10.1111/j.1365-2923.2010.03921.x>
66. Varpio L. Dismantling medical education's incompatible ideology. *Med Teach*. 2023: 1–6.
67. Young M, St-Onge C, Xiao J, Lachiver EV, Torabi N. Characterizing the literature on validity and assessment in medical education: a bibliometric study. *Perspect Med Educ*. 2018; 7(3): 182–91. DOI: <https://doi.org/10.1007/S40037-018-0433-X>
68. Kinnear B, Martini A, Varpio L, Driessen EW, Schumacher DJ. How do validity experts conceptualise argumentation? It's a rhetorical question. *Med Educ*; 2024. DOI: <https://doi.org/10.1111/medu.15311>



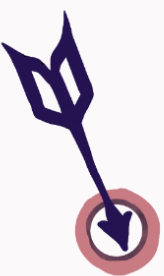


# **CHAPTER 7**

## **Validity Arguments: A MedEd Model**

Published online at <https://www.mededmodels.com/validity>

MedEd Models is a team dedicated to developing “eye-catching #tweetorials on key #meded concepts and theories for busy medical educators”. In a collaboration with MedEd Models’ found, Dr. Timothy Dyster, we developed a visual primer on validity arguments that integrated components of Argumentation Theory. This chapter includes the text and visual content from <https://www.mededmodels.com/validity>

VALIDITY

OBJECTIVES

By the end of this @MedEdModels #Tweetorial, learners will be able to:

1. Summarize three discourses on validity from the education literature.
2. Define the three core components of an argument.
3. Describe four inferences of Kane's validity argument framework.

This #MedEd Model is on validity. Validity describes how justifiable the decisions are that result from a test, whether we are referring to an exam, a rating scale, or a clinical diagnostic.

We'll focus on describing current validity discourses, then list the components of a validity argument, and wrap up by applying validity to #MedEd.

VALIDITY DISCOURSES

1 A TEST CHARACTERISTIC	2 AN EVIDENTIARY ARGUMENT	3 A SOCIAL IMPERATIVE
		
<p>validity is an <b>intrinsic property</b> of a test, instrument, or tool</p> <ul style="list-style-type: none"> <li>• a 'validated' tool retains that quality indefinitely, independent of context</li> <li>• validity is a <b>static property</b> (e.g., "this multiple choice test <i>has been</i> validated; this is a <i>validated</i> measure of X")</li> </ul>	<p>validity is the <b>evidentiary argument</b> that supports a proposed use/interpretation of assessment data</p> <ul style="list-style-type: none"> <li>• iterative &amp; process-based, involves ongoing collection of validity evidence</li> <li>• validity is <b>fluid</b> (e.g., "this use of these scores is defensible <i>right now</i> based on this current evidence")</li> </ul>	<p>validity is a social imperative that deliberately considers the <b>consequences of assessment</b></p> <ul style="list-style-type: none"> <li>• emphasizes a high-level view of individual &amp; societal impacts of assessment</li> <li>• validity is <b>contingent on deliberate, holistic use</b> (e.g., "these persons will be impacted in <i>this way</i> by <i>this use</i>")</li> </ul>

Let's review 3 validity discourses in #MedEd:

1. **Validity is a property of a given test**
2. **Validity is an argument supporting a proposed interpretation & use of assessment data**
3. **Validity is a social imperative that considers individual & societal consequences of assessment**

St Onge & Young 2016. Validity: one word with a plurality of meanings. *Advances in Health Sciences Education*. 22(4), 853-867.

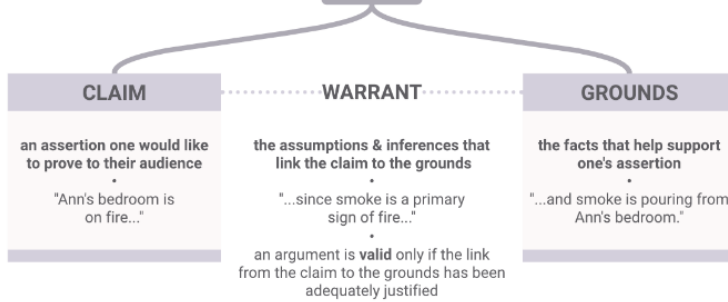
@MedEdModels

# ARGUMENTATION

## ar·gu·ment / n./

a reason, or set of reasons, articulated to persuade others that an idea, action, or interpretation is right or wrong

### CORE COMPONENTS



If a **CLAIM** is accepted, it can subsequently serve as the **GROUND**s or **WARRANT** for other related arguments. This creates sets of nested arguments that form a complex **WEB OF ARGUMENTATION**.

Karbach 1987. "Using Toulmin's model of argumentation." *Journal of Teaching Writing*. 6(1), 61-92.  
Kinnear et al 2022. How argumentation theory can inform assessment validity: a critical review. *Medical Education*. 56, 1064-1075.

@MedEdModels

Let's dive deeper into the second discourse: validity as an argument. First, we need to understand what we mean by *argument*.

The core components of an argument have been described in the argumentation literature by Toulmin in a 3-part framework: 1 – Claim, 2 – Warrant, 3 – Grounds/Data

Now, let's consider what it would mean for validity to be presented as an argument

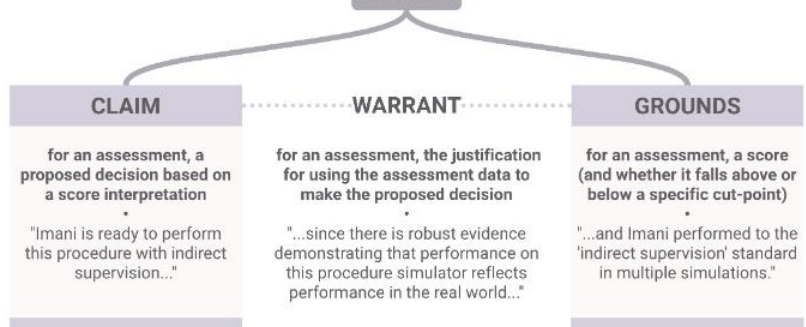
Check out the image to see how the claim, warrant, and grounds connect to validity in [#MedEd](#).

# VALIDITY AS ARGUMENT

## va·lid·i·ty ar·gu·ment / n./

an organized presentation that interprets collected validity evidence & advocates for a proposed use of scores

### CORE COMPONENTS



A robust **VALIDITY ARGUMENT** justifies the link between an assessment/test score (the **GROUND**s) and a proposed interpretation/decision based on the score (the **CLAIM**). Like any other argument, accepted claims can become the warrant or grounds for related arguments, forming a complex web of argumentation.

Cook 2014. When I say... validity. *Medical Education*. 48(10), 948-949.  
Karbach 1987. Using Toulmin's model of argumentation. *Journal of Teaching Writing*. 6(1), 61-91.  
Kinnear et al 2022. How argumentation theory can inform assessment validity: a critical review. *Medical Education*. 56, 1064-1075.

@MedEdModels

In other words, whenever we use an assessment to make a decision in #MedEd, we need to 1) articulate the inferences that justify using that assessment for the proposed decision and 2) generate evidence supporting those inferences.

This matters, because if we use an assessment to make a decision \*without\* ensuring we have evidence supporting our inferences, we could end up making misguided (or even harmful) decisions in #MedEd.

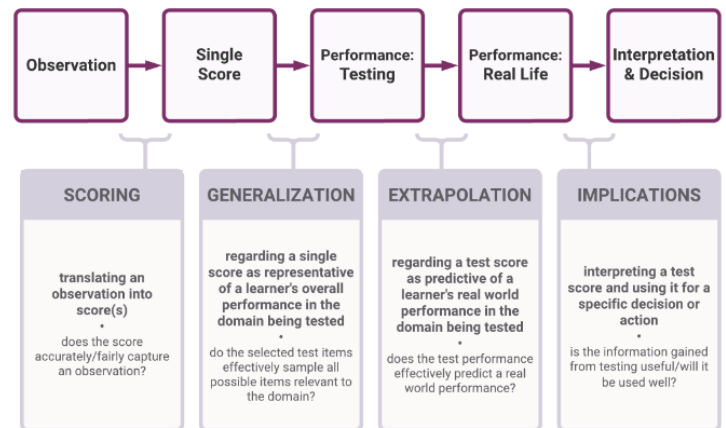
Making a plan to gather evidence for a validity argument can be daunting. Luckily, Kane's Framework can help us!

Kane's framework identifies four inferences:

- Scoring
- Generalization
- Extrapolation
- Implications

## KANE'S FRAMEWORK

KANE offers a structured approach to constructing a **VALIDITY ARGUMENT**, based on **FOUR POINTS OF INFERENCE** common to all proposed uses of an assessment: **SCORING, GENERALIZATION, EXTRAPOLATION, & IMPLICATIONS.**

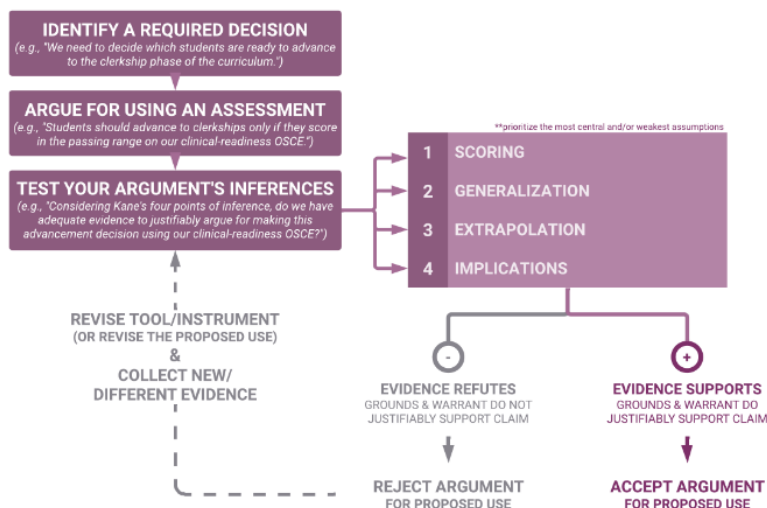


Cook et al 2015. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education*. 49(6), 560-575.

@MedEdModels

## APPLYING KANE'S

**KANE'S FRAMEWORK** can help plan a **VALIDITY ARGUMENT**, which involves identifying a required decision, arguing to use a specific assessment to make the decision, and collecting evidence to test that argument's **INFERENCES**.



Cook et al 2015. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education*. 49(6), 560-575.

@MedEdModels

By applying Kane's Framework, we can uncover the weakest assumptions in a validity argument, and then gather evidence to explicitly/robustly address those assumptions.

If the evidence is refuting, we revise our plan.

# APPLYING KANE'S

What constitutes evidence for each of the four inferences?

We're glad you asked! Check out the table for some ideas/examples of quant & qual validity evidence!

Sources of evidence to support or refute a validity argument's **INFERENCES** can be **QUANTITATIVE** or **QUALITATIVE**, which makes **KANE'S FRAMEWORK** powerful & transferable across many contexts in Med Ed.

EVIDENCE FOR VALIDITY ARGUMENTS	
Scoring	<ul style="list-style-type: none"> <li>Item analysis (e.g., item difficulty, reliability)</li> <li>Rater training</li> <li>Rater/rating match (e.g., does the rater have adequate experience to render the assessment?)</li> <li>Adequate rater exposure to learners</li> <li>Quality of prompts for narrative assessments</li> </ul>
Generalization	<ul style="list-style-type: none"> <li>Test blueprint (i.e., document describing domains to be assessed that enumerates how many items will test each domain)</li> <li>Reliability/generalizability analyses</li> <li>Sampling strategy (e.g., items, tasks, occasions, no. of observers, complementary data sources)</li> <li>Transparency of interpretation (e.g., transparency/reflexivity of interpretative process)</li> </ul>
Extrapolation	<ul style="list-style-type: none"> <li>Authenticity of testing (e.g., context of assessment is authentic, items reflect actual knowledge required/real-life decision making, scenarios tested are rated as authentic by experts, etc.)</li> <li>Scope of testing (e.g., clearly defined constructs/domains are assessed)</li> <li>Discrimination evidence (i.e., testing reliably discriminates between known groups)</li> <li>Responsiveness evidence (i.e., testing scores change in response to interventions)</li> <li>Member checking (i.e., asking raters if final interpretation reflects their perceptions)</li> </ul>
Implications	<ul style="list-style-type: none"> <li>Robust process for setting pass/fail standard</li> <li>Clear remediation processes</li> <li>Expert agreement about final judgment/decisions</li> <li>Intended (and unintended) consequences of using test for decision-making</li> </ul>

Cook et al 2015 A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education*. 49(6), 560-575.

@MedEdModels

# APPLYING VALIDITY



Remember that while "validity as argument" and "validity as social imperative" are the most prominent discourses in medical education currently, different disciplines and fields may conceptualize validity differently. Some fields (e.g. engineering) may have a prominent discourse of "validity as test characteristic", which while somewhat passé in medical education, is a completely acceptable view.

The goal is to understand how the term "validity" is being used to allow for meaningful discussion about how assessment, evaluation, and measurement is used.

Now let's practice applying the concept of validity to #MedEd.

Check out the prompt below. We can't wait to hear your thoughts!

## PROMPT

Reflect on a recent experience that involved designing an assessment plan. Which discourses did you draw on most heavily when considering validity in your assessment plan?

@MedEdModels

# VALIDITY



COMMENTS FROM

**Benjamin Kinnear, MD, MEd**  
[@Midwest\\_MedPeds](#)

***"Validity is the sine qua non of assessment in medical education. Without validity, assessment decisions have no grounds on which to stand.***

*"Consider which evidence is most important to support decisions that result from assessment data, using Kane's chain of inferences to target evidence for the weakest link in the chain.*

*"Which evidence is most important may depend on to whom your validity argument is being made, so consider your audience."*

@MedEdModels

# CHAPTER 8

**What the hell is water? Changing medical education's ideology through validity**

Published as:

Kinnear, Benjamin, and Daniel J. Schumacher. "What the hell is water? Changing medical education's ideology through validity." *Medical Education* 58.3 (2024): 274-276.

In 2005, American novelist David Foster Wallace gave a commencement speech to the graduating class at Kenyon College in the United States. He opened the speech with a parable about two fish:

*There are these two young fish swimming along and they happen to meet an older fish swimming the other way, who nods at them and says ‘Morning, boys. How’s the water?’ And the two young fish swim on for a bit, and then eventually one of them looks over at the other and goes ‘What the hell is water?’<sup>1</sup>*

Of course, the wisdom in this story is that the ideologies that constitute our realities are often transparent to us, but may not be to others. Assumptions that are baked into our worldviews can keep us from questioning what may be noticeable or unusual to outsiders or novices.

In this issue of Medical Education, Coyle et al challenge us to examine the waters in which we swim.<sup>2</sup> They note that efforts to improve widening participation and access to medicine for people from under-privileged or minoritised backgrounds are at tension with medical education’s preoccupation with academic excellence as a key metric for applicant selection. The authors poignantly write, ‘We suggest that it is time for medical schools to acknowledge that some of the drivers for ever higher academic thresholds for entry to medicine are artifacts of managing the number of applicants rather than anything more noble’. Such a bold call should rouse medical education to scrutinise the entrenched use of academic excellence as a selection standard by re-examining the rationale for doing so.

Academic performance has ruled medical education selection for decades, embedding itself as an ideological norm. We no longer question why it is used in applicant selection. It has become part of medical education’s ideology, often passing ‘unseen as normal or as factual’<sup>3</sup> like water to our parabolic fish. However, the suboptimal diversity, equity and inclusion of medical education’s assessment and selection practices are being increasingly recognised as a wicked problem,<sup>4</sup> leading to more frequent scrutiny of sacred (or unseen) ideologies. In response, we believe medical education should consider removing academic excellence as the gatekeeping metric to our profession.

Ostensibly, the most important stakeholders of applicant selection are learners and patients. For learners, selection represents the culmination of years of study, service and research. Selection presents a high stakes branchpoint that dictates much of learners’ future career. For patients, selection represents an accountability mechanism to ensure that future physicians are prepared for the rigours of medical training and capable of providing high-quality care. We should, then, question if relying on academic excellence serves these groups. Coyle et al’s work suggests that academic excellence presents a roadblock for learners from under-privileged or minoritized backgrounds who are unfairly disadvantaged due to systematic bias.

This unfairness certainly harms such learners, indicating that academic excellence is not beneficial to one of our key stakeholder groups. Patients are also harmed. Multiple studies have shown that perceived care

quality and improved clinical outcomes are associated with patient-physician racial concordance for minoritised populations.<sup>5-7</sup> Therefore, admitting diverse training cohorts best positions the health professions to serve diverse patient populations. While the reduction of physician workforce diversity due to biased selection metrics (such as academic excellence) is just one factor contributing to widespread racial and ethnic healthcare disparities,<sup>8</sup> it is a factor that is within medical education's sphere of control. As Varpio wrote, 'Fortunately, ideology is maintained by our decisions and actions; therefore, we can change our decisions and thereby modify the ideology to work for us, not against us'. In other words, we choose the waters in which we swim, and we have agency to change.

One way to modify our ideology is to consider diversity and equity as part of validity arguments for selection decisions. Medical education has largely adopted that validity is not a property of a specific instrument or tool, but rather an argument with supporting evidence that a given decision, interpretation or use of data is justifiable or defensible.<sup>9,10</sup> One type of evidence that is often overlooked<sup>11</sup> but critical to validity arguments relates to the consequences of decisions. Consequences evidence 'looks at the impact, beneficial or harmful and intended or unintended, of assessment'.<sup>12</sup> Coyle et al show how academic excellence works against diversity and equity policies and initiatives. If the inequitable treatment of under-privileged and minoritised learners or the negative downstream consequences to patients are unacceptable (as they should be), then selection decisions that centre on academic excellence are not valid. Hauer et al have previously argued that diversity and equity considerations should be included in medical education assessment validity arguments.<sup>13</sup> We agree and believe that the same considerations should be extended selection decisions.

To be clear, we are not arguing that learners from underprivileged or minoritised backgrounds are not excellent. Rather, we believe that the societal and systemic inequities and biased assessment strategies make academic performance an indefensible metric to use as the crux of selection for physician training. Selection decisions relying primarily on academic excellence are not valid if diversity and equity become key aspects of our validity arguments. We are also not implying that toppling academic excellence is an easy task. Changing medical education's ideology can seem daunting, but we must remember that it is within our control. We are actors in this network, with agency to swim to new ideological waters that embrace selection metrics which are more beneficial for both learners and patients. But first we must ask the fraught question, 'What the hell is water?'

## References

1. Wallace DF. *This Is Water: Some Thoughts, Delivered on a Significant Occasion, About Living a Compassionate Life*. Hachette UK; 2009.
2. Coyle M, Bullen J, Poobalan A, Sandover S, Cleland J. Follow the policy: an actor network theory study of widening participation to medicine in two countries. *Med Educ*. 2024;58(3):288–298. doi:10.1111/medu.15178
3. Varpio L. Dismantling medical education's incompatible ideology. *Med Teach*. 2023;1-6.
4. Lucey CR, Hauer KE, Boatright D, Fernandez A. Medical education's wicked problem: achieving equity in assessment for medical learners. *Acad Med*. 2020;95(12S):S98-S108. doi:10.1097/ACM.0000000000003717
5. Snyder JE, Upton RD, Hassett TC, Lee H, Nouri Z, Dill M. Black representation in the primary care physician workforce and its association with population life expectancy and mortality rates in the US. *JAMA Netw Open*. 2023;6(4):e236687. doi:10.1001/jamanetworkopen.2023.6687
6. Cooper LA, Roter DL, Johnson RL, Ford DE, Steinwachs DM, Powe NR. Patient-centered communication, ratings of care, and concordance of patient and physician race. *Ann Intern Med*. 2003; 139(11):907-915. doi:10.7326/0003-4819-139-11-200312020-00009
7. Saha S, Komaromy M, Koepsell TD, Bindman AB. Patient-physician racial concordance and the perceived quality and use of health care. *Arch Intern Med*. 1999;159(9):997-1004. doi:10.1001/archinte.159.9.997
8. 2022 national healthcare quality and disparities report [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2022. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK587182/>
9. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830-837. doi:10.1046/j.1365-2923.2003.01594.x
10. Cook DA. When I say...validity. *Med Educ*. 2014;48(10):948-949. Doi:10.1111/medu.12401
11. Beckman TJ, Cook DA, Mandrekar JN. What is the validity evidence for assessments of clinical teaching? *J Gen Intern Med*. 2005;20(12): 1159-1164. doi:10.1111/j.1525-1497.2005.0258.x
12. Cook DA, Lineberry M. Consequences validity evidence: evaluating the impact of educational assessments. *Acad Med*. 2016;91(6):785-795. doi:10.1097/ACM.0000000000001114
13. Hauer KE, Park YS, Bullock JL, Tekian A. "My assessments are biased!" Measurement and sociocultural approaches to achieve fairness in assessment in medical education. *J Assoc Am Med Coll*. 2023;98(8S):S16-S27. doi:10.1097/ACM.000000



# **CHAPTER 9**

## **Discussion**

## **Overview**

This PhD journey has challenged me not only to learn new research skills and ways of thinking, but also to keep a highly theoretical line of research grounded in practical utility. I am not suggesting that highly theoretical research is without value, but I aimed to tether this work to practical implications for health professions educators. At times, that tether has been under high tension, like a mooring line stretched to its limits while holding a ship at port. Validity theory and Argumentation Theory do not always lend themselves to easy translation to the world or frontline educators. However, three mechanisms have strengthened my tether between theoretical and practical. First, I have repeatedly considered a question during this PhD work: So what? In other words, I recurrently paused to consider how my research is relevant to education program leaders? Second, I have had the privilege of presenting my PhD research at multiple conferences and invited talks where very few audience members are experts in validity or argumentation. I quickly realized at my first presentation that dissemination of my PhD work would be limited to a very small number of validity scholars unless I connected results to meaningful, real-world education issues. Finally, my role as a residency program director grounds all of my thinking in daily work with residents. Validity arguments are not inconsequential for me; instead, they underpin highly consequential (and sometimes extremely difficult) conversations that I have with residents about whom I care deeply. The validity of assessment decisions is personally important to me. Therefore, in this discussion chapter, I will describe the theoretical implications of this thesis for HPE assessment validity, and connect those implications to practical impacts on frontline education leaders, learners, and programs.

## **Summary of Findings**

The flow of chapters in this thesis reflects my personal journey of exploration of validity arguments. In constructing a validity map for my home program's assessment system (Chapter 2), I attempted to organize evidence supporting summative assessment decisions in a complex program of assessment. This proved challenging, particularly when multiple different facets of the system required evidence (e.g., observation, learning analytics, sampling, committee decision-making). A combination of rational explanation, theory, and empirical studies provided evidence in support of summative decision-making, all organized within commonly used validity frameworks. However, I found that the mechanism that created a sense of meaning and accountability was not the organization of validity evidence into types (Messick) or inferences (Kane), but rather in the dialectical approach we took. Rather than simply putting forth evidence of our choosing, we started with critiques of our system in the form of questions, and then attempted to answer those critiques to the best of our ability. This was the beginning of us bringing argument to our validity work. It laid the foundation for further questions about who the audience/interlocutor should be for such an argumentation process (rather than serving as our own audience), and what rules or principles should guide such argumentation. Chapter 2 provided impetus for the empirical work in the thesis.

Chapters 3, 4, and 5 explore the various conceptualizations of argument and argumentation, both outside of HPE (Chapter 3) and within HPE (Chapters 4 and 5). The critical review in Chapter 3 gave form to many of the nascent argumentation concepts that had been elicited by the work in Chapter 2. Argumentation Theory provided scaffolding for thinking about how arguments are developed, organized, analyzed, and evaluated. Among the myriad orientations within Argumentation Theory, I found that Informal Logic and New Rhetoric would be most useful for shaping validity arguments in HPE. Both orientations elevate the importance of context and audience in argumentation and provide insight into how arguments might be evaluated. While Informal Logic provides context-dependent normative standards (relevance, sufficiency, acceptability), New Rhetoric centers on persuasion of a given audience, hence focusing on audience values and norms. I do not go as far as to be proscriptive in which orientation is preferable for HPE, nor do I declare how they *should* be integrated in to validity and validation. Rather, I propose that both bring useful ideas that might make validity arguments more impactful.

I used Chapter 3 to translate useful Argumentation Theory concepts into HPE's validity discourse. Though I did not see argument or argumentation being meaningfully used in HPE validity, this was purely based on my experiences and opinion. I wanted to empirically explore how argument and argumentation were currently understood and manifested in HPE assessment validity. To do this, I turned to HPE validity experts (Chapter 4) and HPE peer reviewed publications (Chapter 5).

The thematic analysis of HPE validity scholars' views on validity arguments (Chapter 4) revealed that my experiences with building the validity map were not unique. Participants described how validity in HPE assessment was not social (i.e. there was minimal interaction between argument developers and their audiences), despite the fact that argument is a social concept. Participants noted that audiences were absent from HPE validation work, meaning that the most important stakeholders of a given assessment decision rarely (if ever) had opportunities to review or evaluate validity arguments. It seemed that HPE has adopted validity *argument* in name only; it has reduced arguments to piles of evidence. The validity map exercise in Chapter 2 chipped away at this problem by infusing actual dialogue into the process of building a validity argument, but we still lacked a true audience and we were left with organized pieces of evidence rather than a coherent argument. Participants noted that the social aspect of argumentation is an important part of building and evaluating a defensible validity argument, but, in practice, the social aspect is absent. Our discourse analysis of HPE publications in Chapter 5 supported our findings from Chapter 4. Argument and argumentation were not prominent aspects of HPE's published validity discourse, nor were audiences. Instead, organizing evidence and legitimating the concept of *validity as an argument* were more prominent. Thus, Chapter 2 unearthed questions and challenges, Chapter 3 provided a conceptual framework for ongoing inquiry, and Chapters 4 and 5 brought empirical evidence of what is lacking in HPE's validity arguments.

I do not simply want to problematize validity with this thesis. I want to generate ideas of how HPE might more fully understand and operationalize validity arguments to improve education and patient care. Chapters 6 and 7 provide a look forward, exploring how validity might evolve should there be meaningful engagement with argument audiences, leading to the foregrounding of issues like learner consequences and equity. Chapter 6 describes how validity could better align with the values of HPE in the near future. Equity and accountability to society are values that are increasingly foregrounded in HPE assessment conversations.<sup>1-8</sup> By truly making validity arguments social and audience-centered, educators could engage with their most meaningful audiences such as learners and patients to co-produce validity arguments that align with such values. Argumentation Theory could provide the scaffolding on which such a process is built by providing guidance for how arguments can be developed, organized, and evaluated. Chapter 7 continues this line of thinking, using a specific example from an article by Coyle et al to show how validity arguments could be used to promote equity in HPE applicant selection. Coyle et al suggest that HPE's focus on academic achievement when selecting applicants for medical training works against many diversity efforts given the systematic bias that is woven into the fabric of education and society. I frame this tension through the lens of validity of selection decisions, and note that we have agency to choose whether validity arguments should reflect our values and the values of our learners (e.g. equity).

Chapter 8 is a series of infographics created in collaboration with MedEd Models (<https://www.mededmodels.com/>) to disseminate the concepts of validity arguments and Argumentation Theory in an easily accessible way.

### **Argument and Argumentation: HPE Validity Cryptids**

One theme that echoes throughout this thesis is the conspicuous absence of actual arguments or argumentation from HPE validity. The *validity as an argument* conceptualization truly has taken hold in HPE, and is even used as a shibboleth to enter into contemporary validity discussions. And yet, very little ink has been spilt to truly explore what is meant by *argument* despite having centuries of Argumentation Theory to draw from. If validity truly is the *sine qua non*<sup>9</sup> or *cardinal virtue*<sup>10</sup> of assessment, as has been advocated, then perhaps we ought to think deeply about what it means to adopt the view of validity being an argument. As the work in this thesis has demonstrated, there are some potentially seismic changes that may come with fully embracing the concept of validity being an argument. I believe HPE should deeply explore the consequences of embracing argument and argumentation in assessment validity, and determine if and when *validity as an argument* is a conceptualization that should be used.

Why is argument a word that frequently appears in HPE assessment validity literature, but lacks depth of meaning? My research for this thesis did not explain causality, but I have multiple hypotheses. First, argument and argumentation are not deeply explored in the source material from which HPE draws for validity theory.

Some scholars such as Kane,<sup>11,12</sup> Bachman,<sup>13</sup> and Mislevy<sup>10,14,15</sup> address aspects of argument, but they mostly reference argument structure à la the work of philosopher Stephen Toulmin work.<sup>16</sup> Toulmin's work provides useful ways of diagramming arguments and identifying component parts such as data, warrants, claims, and rebuttals, but these structures are only one part of argumentation. A deeper exploration of Argumentation Theory, the multiple orientations therein, and which aspects of argumentation might be considered with validity as an argument is lacking. To be clear, I am not casting aspersions at these validity scholars. Their work was foundational to modern validity theory. They drew (implicitly and explicitly) from the tradition of Informal Logic in conceptualizing arguments, which is an argumentation orientation that my thesis work also found would be generative for HPE to consider in assessment validity. However, I contend that one potential reason for the lack of deeper exploration of Argumentation Theory in HPE assessment validity may be because the source materials on which our field draws also does not describe such ideas in depth.

Second, validity, as a concept, serves as a *boundary object* between HPE and other specialties such as Language Assessment, Psychology, and Measurement. A boundary object is an artifact or idea that serves a bridging function between different sites or groups.<sup>17</sup> As we noted in the opening chapter, validity is present in myriad disciplines given its centrality to measurement. Different disciplines abut one another ideologically, with ideas and concepts migrating between them. For example, many ideas from the field of implementation science<sup>18,19</sup> and complexity science<sup>20,21</sup> have migrated into HPE. Related ideas such as *fidelity of implementation* and *complex adaptive systems* serve as boundary objects connecting HPE with these fields, respectively. Similarly, validity is a boundary object, and as such it must be “both plastic enough to adapt to local needs... yet robust enough to maintain a common identity across sites”.<sup>22</sup> This plasticity allows for changes to occur as ideas like *validity as an argument* are shared between other fields and HPE. Argument and argumentation are concepts that were diminished in the migration of ideas to HPE via the boundary object that is validity.

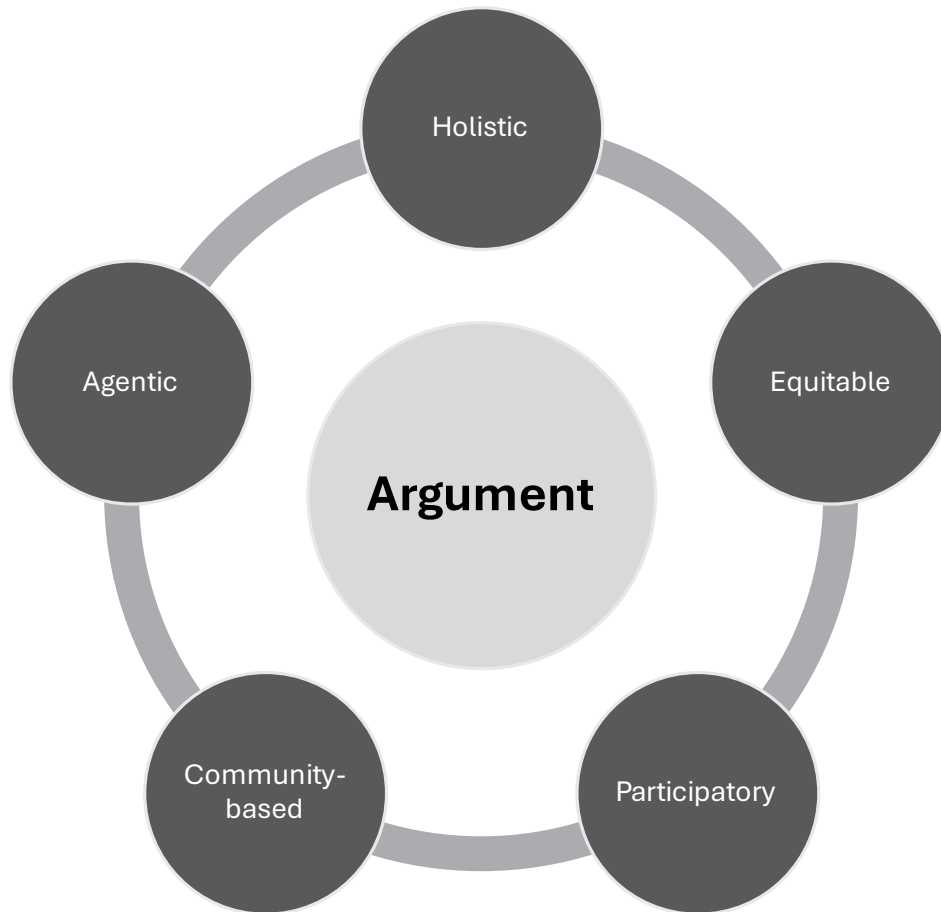
Third, HPE's current processes and structures may inherently minimize argument and argumentation with regard to assessment. As the Canadian communication theorist Marshall McLuhan stated “The medium is the message”.<sup>23</sup> Much of the discourse relating to validity arguments happens via peer-reviewed publication. Perhaps the process of peer and editorial review does not foreground argumentation in ways that other approaches (e.g. committee presentations, stakeholder discussions) might. HPE currently does not have the mechanisms in place for a social exchange of ideas with many of the audiences that would be considered stakeholders of assessment decisions.

In sum, I do not believe that argument and argumentation have been intentionally disregarded or flouted in HPE validity work – I believe they were inadvertently minimized by a combination of the mechanisms listed above (and likely several others).

## Argument and Argumentation Can Transform Validity

The work in this thesis suggests that if HPE enriches assessment validity by integrating aspects of Argumentation Theory, the fundamental way in which we conceptualize validity would change. By fully embracing validity as an argument, validity takes on other characteristics that are worth exploring (Figure).

# Validity as...



### *Validity as community-based*

As I emphasized throughout this thesis, argumentation necessitates meaningful social exchange, ideally with the audiences (i.e. stakeholders) who are most impacted by the decisions (e.g., interpretations and uses) being validated. This empowers any community group who is an audience of interest to take part in the process of determining what is valid and what is not. In other words, validity arguments could engage

stakeholders that have previously been left out of validity discussions such as learners and patients, echoing approaches seen in community-based research.<sup>24</sup> This would embrace the views and opinions of groups that are outside typical validity discussion circles.

#### *Validity as participatory*

Validity arguments could also be more participatory, calling the aforementioned communities (as well as others) to take an active role in co-creating what are valid or defensible assessment decisions. This co-creation<sup>25</sup> could be a cyclical, iterative process such as what is seen in community-based participatory research.<sup>24</sup> Stakeholder/audience groups could be engaged from the very beginning of developing validity arguments, helping to shape the direction of argument and inquiry from the very beginning. A participatory version of validity arguments would call us to move beyond peer reviewed publication and toward more interactive, dialogical approaches to validation work such as verbal exchange via focus groups, interviews, or committees. Peer reviewed publication would still serve a key role in sharing the work within the academic community, but participatory engagement with meaningful audiences/stakeholders could be part of what is reported in published validation work. Mislevy's *validity by design* already calls for assessment processes to be designed with validity considered at every stage of development rather than validity being a post-hoc consideration. This proactive approach could be infused with a participatory nature that invites meaningful stakeholders to be part of co-creating validity arguments.

#### *Validity as holistic*

Validity has historically been heavily influenced by psychometric considerations and focused on reliability.<sup>26</sup> No doubt, reliability and psychometric evaluation still play an important role in many validity arguments. However, argument and argumentation invite a more holistic approach to determining the validity of a given decision. Particularly as programmatic assessment approaches are becoming more prominent in HPE,<sup>27-31</sup> varied assessment strategies (and resulting data) often inform the same downstream interpretations and uses. This requires a holistic approach to building and evaluating validity arguments. For example, if summative decisions made by a competency committee are informed by narrative data, workplace-based entrustment ratings, simulation scores, and test scores, then a validity argument supporting the final decision should integrate evidence and inferences related to the interpretation and use of each type of assessment data. This use of *holistic* is similar to calls for holistic review of medical trainee applications when applying for medical school or residency training.<sup>32</sup> In medical trainee selection, holistic review avoids overreliance on isolated metrics (e.g., test scores) when choosing candidates, instead taking into account multiple facets of an applicant's experiences, attributes, and academic metrics.<sup>33</sup> Similarly, argument-based validity allows for audiences to take a holistic approach to weighing different lines of argument and supporting evidence rather than relying solely on psychometric performance.

### *Validity as equitable*

I have argued in this thesis that validity arguments should foreground the values and judgments of a diverse range of audiences. Consequently, the equity of assessment decisions should be part and parcel to validity arguments. Historically, if an assessment was shown to measure the intended construct of interest, it was considered valid regardless of equity considerations. The argument-based conceptualization of validity invites us to consider equity part of consequential validity<sup>34</sup> evidence, and hence part of a validity argument. Therefore, even if a construct of interest is measured accurately, the assessment decision could be considered invalid if inequity was identified and rejected by a given audience. This brings both a deontological and teleological lens to validity by inviting in the moral and practical goods of equity in validity theory.

### *Validity as agentic*

Argumentation invites participants to have agency in the assessment through debate, analysis, and evaluation of validity arguments. We can empower argument builders and audiences not just to participate in validity arguments, but also to shape the boundaries of where validity begins and ends. I have suggested that validity should include equity, essentially turning validity into an expansive concept that goes well beyond the question of whether an instrument measures the intended construct. Many people within and outside of HPE would disagree with this expansive view. Some consider the argument-based approach too subjective and lacking a scientific perspective.<sup>35</sup> Others state that conflating objective knowledge with moral judgments turns validity into “Frankenstein’s monster”.<sup>36</sup> One’s perspective on the boundaries of validity likely depends heavily on their field of study and educational background. For example, Newton and Baird suggest that psychometric scientists tend to favor a “narrower, more scientific” view of validity while educational practitioners tend to favor a “broader, more ethical” view.<sup>37</sup> These authors call it a “worrying observation” that there is lack of consensus around validity’s definition.

While consensus certainly has advantages in academic pursuits, it can also be limiting. I would reframe the uncertain boundaries of validity arguments as a strength. HPE has a rich history of pulling from many fields such as Psychology, Measurement, Sociology, Education, and more.<sup>38</sup> The field of HPE is disciplinarily diverse, with scholars therein upholding an eclectic array of epistemologies, ontologies, and axiologies.<sup>39</sup> We (everyone in HPE) get to choose whether validity arguments are expansive (including consequences and equity) or narrow (only construct relevance). HPE is a constructed environment – contingent upon our decisions. As Varpio writes, “We made medical education”.<sup>40</sup> We get to decide how to conceptualize validity (expansive or narrow) in HPE; it does not need to be uniform across our field. Validity arguments invite us to take agency in choosing which ideology we ascribe to in our assessment work. Varpio again writes, “Since we made the ideology by creating the structures, practices and policies of medical education, we can change our minds and change our choices.” I contend that validity arguments should foreground audiences’ values

such as impacts on learners and equity. I am not claiming that my conceptualization is *truth* or *right*; instead, I offer this position as an example of how we can agentially use validity arguments to make HPE a better place. I hope others in HPE will read my research and be inspired to do the same.

### **Confronting Frankenstein's Monster**

Does a broad, argumentation-focused view of validity transform it to “Frankenstein’s monster”? Perhaps in some ways. Maybe by broadening validity’s reach it truly does become a monster, lurking around every corner, subsuming all concepts relevant to assessment such as *acceptability* or *educational impact*. Or maybe, there is something more complex with this monster than simply a lumbering personification of malice. In the film adaptation of *Frankenstein*, Victor Frankenstein’s first words upon animating the monster are “It’s alive!”. This reflects an important subtext of this thesis: HPE assessment validity arguments need life. Like the lightning in *Frankenstein*, the social and dynamic nature of argumentation can awaken HPE validity to be participatory and agentic, engaging with real-world audiences and inviting them to help co-create validity arguments. Frankenstein’s monster also reflects many aspects of its creator, much like how we conceptualize validity reflects aspects of HPE’s ideologies. As author Sheila English writes “Shelley’s Creature is not just a physical being brought to life by science, but a symbolic representation of deeper questions about what it means to be human.”<sup>41</sup> Validity argumentation can similarly provoke deeper questions about HPE’s educational values. Validity can be agentic, allowing us to integrate values such as equity and societal accountability into our validity arguments. I therefore do not ignore nor rebut the claim that animating validity arguments with argumentation could create a Frankenstein’s monster. I prefer a monster with life, engagement, and values to a lifeless, inconsequential husk.

### **Implications and future research**

Multiple practical implications result from this research. First, HPE assessment validation would be enriched by integrating Argumentation Theory. Specifically, validation processes could include the explicit articulation of the argumentation orientation being used by argument developers. For example, an argument developer might explicitly state that they are using Informal Logic, and therefore they will organize their validity argument using Toulmin’s model and evaluate the argument using standards of relevance, sufficiency, and acceptability. Or they might choose New Rhetoric and focus on identifying a particular audience for their argument, describing that audience’s values, and evaluating their argument based on whether the intended audience is persuaded. My research suggests that explicitly stating one’s argumentation orientation could serve a similar purpose to stating one’s worldview (ontology and epistemology) in research.<sup>42</sup> Compatibility between worldview and methodology is critical to ensuring that research is sound and trustworthy.<sup>43</sup>

Second, validation should be a more social process. Audiences should be identified early in the process and engaged in co-creation of validity arguments. The most important audiences, learners and patients, should

also be engaged. This obviously presents a logistical challenge: most learners and patients have little-to-no expertise in assessment or validity and would struggle to truly participate in validity argumentation. One model to address this challenge might be to create validation committees that include all stakeholders of a program's assessment decisions. This committee could receive education and training on assessment and validity theory, provide a forum for sharing of arguments and gathering audience feedback, and create longitudinal touchpoints for ongoing validation work. Peer reviewed publication would still play a role in sharing validation work, but perhaps publication could include the processes used by validation committees and the input gained by engaging with meaningful audiences.

Third, validation of assessment decisions should occur at the level of training programs. Even when broader assessment frameworks are used (e.g. entrustable professional activities, subcompetencies, etc.), validity arguments should be developed by local programs and address relevant audiences. The work in this thesis shows that validity is context-dependent, and programs should not rely only on context-independent evidence from national studies to support the validity of their assessment decisions. This does not mean that validity evidence in one context cannot support decisions in another context. It certainly could. But a validity argument should include a rationale for why evidence is relevant to a given program's validity argument and intended audience. Many training programs likely do not have the expertise or resources to conduct such work at present. Additional resources in terms of training, protected time, and research funds may be needed to develop validity arguments at programs that currently lack infrastructure to do so. However, any program can begin by using existing theory as evidence for a nascent validity argument. Empirical evidence can be added as resources become available.

Fourth, equity should be part of HPE assessment validity arguments. This is not to say that until perfect equity is achieved, no assessment decisions are valid. That level of rigor would put HPE into a state of paralysis in which nothing is ever good enough. Humans are too flawed and biased to achieve perfect equity. However, validity arguments could include evidence of monitoring for equitable learning outcomes, processes that support equitable decisions, and ongoing efforts of continuous improvement regarding inequity and bias. The degree to which these (and other) types of evidence need to be demonstrated depends on the audience of interest.

This thesis did not include a proscription of how to implement these ideas. Rather than a weakness, I think this is a strength. I cannot (nor do I want to) provide the *how* of this work, as much of it will be context and resource dependent. However, three considerations may be useful for future work. First, engagement with implementation scientists would be useful for translating these theoretical ideas into practice.

Implementation science uses rigorous methods to identify barriers to the uptake of educational evidence and design interventions to reduce research-practice gaps.<sup>18,44</sup> Engagement with implementation scientists could promote rigorously designed translation practices. Second, a broader array of research methods could be

useful in understanding the impacts of more social and audience-centric validation practices. For example, realist inquiry is a context-sensitive approach for understanding the mechanisms that drive change in a given context.<sup>45</sup> It asks: What works for whom, in what circumstances, and why? Since different contexts will have variable affordances and barriers to validity argumentation, realist inquiry could be a useful tool in navigating that context and understanding the underlying mechanisms to create transferable knowledge. Community-based participatory research (CBPR) is another potentially useful approach that forms a collaborative partnership with members of a community to build knowledge based on their strengths and actions.<sup>46</sup> CBPR empowers stakeholders, is iterative, and involves frequent engagement and communication. This research approach aligns well with understanding social, iterative, stakeholder-centric forms of validity argumentation. Third, future research should explore what role regulatory bodies such as accreditors and certifiers should play in helping programs develop validity arguments.

### **Reflexivity Epilogue**

As my thesis work has unfolded, I have grappled with my personal role within HPE's validity discourse. As I mentioned in the opening chapter, my worldview has shifted through time, taking on a more critical stance. I am more attuned to questioning power structures and hierarchy, as well as who the loudest voices are in academic discourse. An ironic tension has weighed on me with regard to my role in helping to shape our field's understanding of validity with my PhD work. In Chapter 3, the vast majority of participants were white academicians from the Global North and Western cultures. Similarly, most of the manuscripts included in our discourse analysis in Chapter 4 were written by Global North and Western authors. In fact, one quarter of the manuscripts in the corpus analyzed in Chapter 4 were written by two prolific authors, both of whom are white, Western men. By writing about validity, I am adding to the predominantly white, Western, Global Northern viewpoint that is woven into HPE's dominant validity discourse.

In Chapter 6, I explicitly call for better engagement with validity scholars in parts of the world and from backgrounds which are underrepresented in HPE. However, even by writing that manuscript, I in some ways contributed to the representation problem in HPE academia. I still brought my white, male, cisgender, Western, Global Northern views to every paper I published in this thesis. I cannot help feeling like a hypocrite - simultaneously calling for more diverse views to shape HPE's validity discourse while piling on articles written from my overrepresented background and identity. Honestly, I have not found a way to resolve this tension; I carry it as a continual internal narrative. My hope is that others who have unique and important viewpoints with underrepresented identities and backgrounds will take the ideas put forth in this thesis and make them even better. I have benefited enormously from the mentorship and sponsorship of my PhD supervisors and other academic sponsors. I hope to pay that support forward: to sponsor those whose voices would contribute uniquely and meaningfully to HPE's validity discourse. Rather than becoming a gatekeeper of validity arguments in HPE, I want to be a person who unlocks the gate and leaves it open for others.

I similarly wonder about how my PhD work might impact validity's standing in HPE. In the opening, I claim that it is often used as a god term that can be held up as unassailable. I am hopeful that this thesis does not further deify validity, nor is it my intention to deify Argumentation Theory. Instead, I hope this work opens *validity as an argument* up to new lines of questioning. I hope this work has peeled away some of the veneer of certainty that HPE understands what validity is, could be, or even *should* be. I recognize that I have called for validity to subsume equity in an attempt to call attention to the ongoing bias and injustice in HPE's assessment practices. My call for broadening validity's boundaries is intended to help address problems that are critical to HPE right now. The future may bring more equitable assessment approaches and improved attention to downstream patient impacts of assessment decisions. Perhaps then, validity would better serve HPE if it were narrowed to focus on construct relevance and not include consequences of assessment or equity. Just as validity arguments are situated, so are the ways we choose to conceptualize validity. I recognize that the suggestions I have made in Chapters 6 and 7 will be scrutinized and likely dismantled over time. If this means progress for our field, then I support that dismantling. As Eva wrote in 2016, "Health professional education would be an uninspiring place if every idea put forward was guaranteed to be long lasting."<sup>47</sup>

## References

1. Lucey CR, Hauer KE, Boatright D, Fernandez A. Medical education's wicked problem: achieving equity in assessment for medical learners. *Academic Medicine*. 2020;95(12S):S98-S108.
2. Anderson HL, Abdulla L, Balmer DF, Govaerts M, Busari JO. Inequity is woven into the fabric: a discourse analysis of assessment in pediatric residency training. *Advances in Health Sciences Education*. 2024 Mar;29(1):199-216.
3. Ajjawi R, Tai J, Boud D, Jorre de St Jorre T. *Assessment for inclusion in higher education: Promoting equity and social justice in assessment*: Taylor & Francis; 2023.
4. Onumah CM, Pincavage AT, Lai CJ, et al. Strategies for advancing equity in frontline clinical assessment. *Academic Medicine*. 2023;98(8S):S57-S63.
5. Tai J, Ajjawi R, Boud D, de St Jorre TJ. *Promoting Equity and Social Justice through Assessment for Inclusion*. *Assessment for Inclusion in Higher Education*: Routledge; 2022.
6. Aquario D. Through the lens of justice. A systematic review on equity and fairness in learning assessment. *Education Sciences and Society*. 2021(2021/2).
7. Holmboe ES, Sherbino J, Englander R, Snell L, Frank JR, Collaborators I. A call to action: The controversy of and rationale for competency-based medical education. *Med Teach*. 2017;39(6):574-581.
8. Schumacher DJ, Kinnear B, Carraccio C, Holmboe E, Busari JO, van der Vleuten C, Lingard L. Competency-based medical education: The spark to ignite healthcare's escape fire. *Medical Teacher*. 2024 Jan 2;46(1):140-6.
9. Downing SM. Validity: on the meaningful interpretation of assessment data. *Medical education*. 2003;37(9):830-837.
10. Mislevy RJ, Steinberg LS, Almond RG. Focus article: On the structure of educational assessments. *Measurement: Interdisciplinary research and perspectives*. 2003;1(1):3-62.

11. Kane MT. Validating the interpretations and uses of test scores. *Journal of Educational Measurement*. 2013;50(1):1-73.
12. Kane MT. An argument-based approach to validity. *Psychological bulletin*. 1992;112(3):527.
13. Bachman LF. Constructing an Assessment Use Argument and Supporting Claims About Test Taker–Assessment Task Interactions in Evidence-Centered Assessment Design 1. In *The Writings of Lyle F. Bachman* 2024 Jul 31 (pp. 83-85). Routledge.
14. Mislevy RJ. The case for informal argument. *Measurement: Interdisciplinary Research & Perspective*. 2012;10(1-2):93-96.
15. Mislevy RJ. Validity by design. *Educational researcher*. 2007;36(8):463-469.
16. Toulmin S. *The uses of argument*: Cambridge: Cambridge University Press; 1958.
17. Akkerman SF, Bakker A. Boundary crossing and boundary objects. *Review of educational research*. 2011;81(2):132-169.
18. Thomas A, Bussi eres A. Leveraging knowledge translation and implementation science in the pursuit of evidence informed health professions education. *Advances in Health Sciences Education*. 2021;26(3):1157-1171.
19. Thomas A, Ellaway RH. Rethinking implementation science for health professions education: a manifesto for change. *Perspectives on Medical Education*. 2021;10(6):362-368.
20. Mills T, Lawton R, Sheard L. Advancing complexity science in healthcare research: the logic of logic models. *BMC Medical Research Methodology*. 2019;19(1):1-11.
21. Warm E, Hirsh DA, Kinnear B, Besche HC. The Shadow Economy of Effort: Unintended Consequences of Pass/Fail Grading on Medical Students' Clinical Education and Patient Care Skills. *Academic Medicine*. 2024;10.1097.
22. Star SL, Griesemer JR. Institutional ecology, translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social studies of science*. 1989;19(3):387-420.
23. McLuhan M. *Understanding media: The extensions of man*: MIT press; 1994.
24. Israel BA, Schulz AJ, Parker EA, Becker AB. Review of community-based research: assessing partnership approaches to improve public health. *Annual review of public health*. 1998;19(1):173-202.
25. Dollinger M, Lodge J, Coates H. Co-creation in higher education: towards a conceptual model. *Journal of Marketing for Higher Education*. 2018;28(2):210-231.
26. Tavakol M, Dennick R. The foundations of measurement and assessment in medical education. *Medical Teacher*. 2017;39(10):1010-1015.
27. Misra S, Iobst WF, Hauer KE, Holmboe ES. The importance of competency-based programmatic assessment in graduate medical education. *Journal of Graduate Medical Education*. 2021;13(2s):113-119.
28. Schuwirth LW, Van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. *Med Teach*. 2011;33(6):478-485.
29. van der Vleuten CP, Schuwirth LW, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Med Teach*. 2012;34(3):205-214.
30. van der Vleuten CP, Schuwirth LW, Driessen EW, Govaerts MJ, Heeneman S. 12 tips for programmatic assessment. *Med Teach*. 2014:1-6.
31. Pearce J, Tavares W. A philosophical history of programmatic assessment: tracing shifting configurations. *Advances in Health Sciences Education*. 2021:1-20.
32. Association of American Medical Colleges. *Holistic Review*. <https://www.aamc.org/services/member-capacity-building/holistic-review>. Accessed 12/19/2024.

33. Gottlieb M, Davenport D, Landry A, Bailey J, Westrick J, Daniel M. Holistic Review in Applicant Selection: A Scoping Review. *Academic Medicine*. 2023;10.1097.
34. Cook DA, Lineberry M. Consequences Validity Evidence: Evaluating the Impact of Educational Assessments. *Acad Med*. 2016;91(6):785-795.
35. Borsboom D. Whose consensus is it anyway? Scientific versus legalistic conceptions of validity. *Measurement: Interdisciplinary Research & Perspective*. 2012;10(1-2):38-41.
36. Borsboom D, Wijsen LD. Frankenstein's validity monster: The value of keeping politics and science separated. *Assessment in Education: Principles, Policy & Practice*. 2016;23(2):281-283.
37. Newton PE, Baird JA. The great validity debate. *Assessment in education: principles, policy & practice*. 2016 Apr 2;23(2):173-7.
38. Norman G. Fifty years of medical education research: waves of migration. *Medical education*. 2011;45(8):785-791.
39. MacLeod A, Ellaway RH, Paradis E, Park YS, Young M, Varpio L. Being edgy in health professions education: Concluding the philosophy of science series. *Academic Medicine*. 2020;95(7):995-998.
40. Varpio L. Dismantling medical education's incompatible ideology. *Medical Teacher*. 2023;1-6.
41. English S. The Tragic Figure of Frankenstein's Creature: a Reflection on Humanity. Vol 2024.
42. Kinnear B, Beck J, Schumacher DJ, Zhou C, Balmer D. Building a Solid House of Scholarship: The Importance of Foundational Worldviews. *Hospital Pediatrics*. 2024:e2023007515.
43. Tavares W, Kuper A, Kulasegaram K, Whitehead C. The compatibility principle: on philosophies in the assessment of clinical competence. *Advances in Health Sciences Education*. 2020;25(4):1003-1018.
44. Thomas A, Bussi eres A. Knowledge translation and implementation science in health professions education: time for clarity? *Academic Medicine*. 2016;91(12):e20.
45. Ellaway RH, Kehoe A, Illing J. Critical realism and realist inquiry in medical education. *Academic Medicine*. 2020;95(7):984-988.
46. Wynn TA, Taylor-Jones MM, Johnson RE, Bostick PB, Fouad M. Using community-based participatory approaches to mobilize communities for policy change. *Family & Community Health*. 2011;34:S102-S114.
47. Eva KW. The more things stay the same, the more they change. *Medical Education*. 2016 Jan 1;50(1).

# **APPENDICES**

**Summary**

**Samenvatting**

**Scientific Impact**

**Curriculum Vitae**

**Acknowledgements**

**SHE Dissertation Series**

## Summary of Findings

Ideas hold power because they influence our thinking, actions, and policies. Validity is a concept that holds significant power in HPE. It has been called “the most important term in the educational and psychological measurement lexicon”.<sup>1</sup> Validity is especially crucial in HPE learner assessment, a process in which measurement and observation are used to make consequential decisions about readiness for patient care with decreasing levels of supervision. Patient health, and indeed their lives, are quite literally impacted by assessment decisions. The validity of such decisions is, therefore, of paramount importance for every person who seeks health care. This thesis explores the philosophical, theoretical, and practical aspects of a widely used conceptualization of validity in HPE assessment: validity as an argument.

### Chapter 1

The opening chapter argues for the power that validity carries in academia broadly and HPE specifically. I position validity as a god-term that serves as an “educational idol”<sup>2</sup> in our field. I also position validity as a chameleon that takes on different forms depending on the context, use, and philosophical beliefs.<sup>3</sup> The most common form that this chameleon takes is that of an argument supporting the interpretations and uses of assessment data (i.e. decisions).<sup>4</sup> This conceptualization was adopted and adapted from eminent validity scholars in other fields such as Cronbach,<sup>5</sup> Messick,<sup>6</sup> Kane,<sup>7</sup> and Bachman.<sup>8</sup> In such a view, validity is, fundamentally, an argument that supports or refutes the decisions made informed by assessment data. I note, however, that there is a dearth of understanding in HPE of what *argument* means in this context, and what theoretical and philosophical assumptions need to be considered when operationalizing validity as an argument. To that end, I introduce the field of Argumentation Theory,<sup>9</sup> a field of study concerned with the development and evaluation of spoken or written arguments. Argumentation Theory holds promise to promote deeper exploration of what it means to conceptualize validity as an argument. The chapter includes an in-depth reflexivity statement and background on why this topic is of practical importance, as well as an outline of studies included therein. The overarching research question that this thesis explores is: “*What is the nature of argument and argumentation with regard to assessment validity in HPE?*”

### Chapter 2

This study describes how I used a dialectical (i.e. discussion of ideas) approach to develop and organize validity evidence for decisions made using my residency program’s assessment system. Our team used a question-and-answer approach to build a validity argument by interrogating the assessment system through queries. Questions included, for example: “Do assessors truly observe trainees performing the skills that are being assessed?” and “How were the skills that are being assessed chosen for inclusion on assessment forms?”. We then provided responses and supporting evidence using existing literature, our own research, or simply by providing a rationale for what was done. These responses were then categorized using Messick’s<sup>10</sup>

and Kane's<sup>7</sup> frameworks as a way to organize our thinking, and to be able to present our work to various stakeholders who might prefer to use one framework or the other when discussing validity. The resulting visual representation of evidence organized by Messick's and Kane's frameworks was deemed our *validity map*. This validity map exercise was a watershed moment in my journey with validity. It allowed me to operationalize prominent validity frameworks in a tangible way within my program's assessment system. The work opened my eyes to the importance of actual argumentation (i.e. levying claims and opening them to scrutiny and response) in building a validity argument. Our dialectical approach of asking questions, providing responses, and actively critiquing the evidence showed the value in the social exchange of ideas as part of validation. However, it also showed that we needed more guidance on developing actual arguments (rather than simply evidence), and that engaging with audiences outside of our own research group would be fruitful. This work was the springboard into the subsequent studies within this thesis.

### Chapter 3

The critical review in Chapter 3 explored the field of Argumentation Theory to understand the various theoretical, philosophical, and practical assumptions that constitute different approaches to argument development, critique, and evaluation. This study revealed to me the depth and richness of Argumentation Theory, as well as the long historical narrative of how argumentation has been conceptualized across time. I consulted with an academic librarian to develop multiple search strategies for ERIC, Scopus, PubMed, and Web of Science seeking literature relevant to our critical review. I also used snowball sampling to find landmark papers and books that would be additive to our review. I found multiple prominent argumentation orientations that would not be useful for HPE validation work. For example, Formal Logic's<sup>11</sup> agnostic approach to context and rigid adherence to premises and conclusions makes it impractical to use for the context-dependent nature of HPE assessment. Dialectics<sup>12</sup> and Pragma-Dialectics<sup>13</sup> both focus on rules of engagement and resolution of disputes between two interlocutors. HPE assessment validation does not involve disputes per se, nor should validity depend on rules of interaction or be oppositional in nature. However, I determined that two argumentation orientations would be useful to incorporate into HPE assessment validity: Informal Logic<sup>14</sup> and New Rhetoric.<sup>15</sup> Informal Logic embraces contextual factors and local norms and values in the construction and evaluation of arguments, while providing standards by which arguments can be evaluated such as acceptability (plausibility of premises), relevance (coherence between premises and conclusions), and sufficiency (adequacy of supporting evidence). Informal logic also introduces structures that are useful for organizing arguments such as Toulmin's framework (claims, warrants, data, rebuttals). New Rhetoric foregrounds audience values in the construction and evaluation of arguments. Rather than having normative standards for argument evaluation, the goal of New Rhetoric is persuasion of an audience to agree with a particular claim. I proposed that the structure and normative standards of Informal Logic, combined with the flexibility and audience-centeredness of New Rhetoric,

provides useful framing for validity arguments relating to HPE assessment. However, I hoped to empirically test these ideas by exploring the views of HPE validity experts as well as published HPE validity literature.

## **Chapter 4**

This qualitative study explored how experts in HPE assessment validity conceptualize the argumentation aspect of validity arguments. I used critical case sampling to seek HPE validity experts using multiple criteria and methods, resulting in semi-structured interviews with 16 participants. I used thematic analysis to develop 3 themes relevant to our research question. First, participants believed that validity arguments *should* be social in nature and contextually situated. This theme aligned with the audience-focused approach of New Rhetoric as well as the context-sensitivity of both Informal Logic and New Rhetoric. Second, participants noted that in practice, validity arguments lack audiences and are rarely evaluated. They stated that audiences are rarely identified or considered when validity argument are developed, and therefore arguments are almost never actually evaluated. Third, participants noted that the lack of audience participation with validity arguments created and maintained power differentials within HPE. In particular, the audiences deemed most important (learners and patients) were never engaged in developing or evaluating validity arguments. The contrast between the first and second themes in this study demonstrates a gap between theory (validity arguments should be social and audience-centric) and practice (validity arguments lack audiences or evaluation), and the third theme demonstrated an underrecognized consequence of this gap (hegemony). This study provided a richer and more nuanced understanding of the views of HPE experts on the role (or lack thereof) of argument and argumentation in assessment validity; I next wanted to explore the relevant validity argument discourses existing in one of academia's principal venues of information exchange: peer reviewed publication.

## **Chapter 5**

In this chapter, I used critical discourse analysis of HPE peer reviewed publications to explore how language in influential manuscripts has shaped HPE's understanding of validity arguments and argumentation. After screening over 800 manuscripts from HPE journals, the final corpus for analysis included 39 manuscripts that were identified as influential in shaping our field's understanding of validity arguments. Most of the final papers were review articles, perspectives, and commentary pieces. I drew from multiple discourse analysis traditions including Laclau and Mouffe,<sup>16</sup> Fairclough,<sup>17</sup> and Gee<sup>18</sup> to create a bricolage approach to analysis. The analysis was also sensitized by our previous work on Argumentation Theory, shaping the questions we asked of the data to construct an understanding of the discourses found therein. I found that the discourse on argument and argumentation is scant in HPE peer reviewed publications. Most of the discourse centers on evidence (rather than argument) and which validity frameworks to use (most often Messick or Kane). Few components of Argumentation Theory were found in HPE's published validity discourse, with a notable

absence of argument audiences. The concept of argument generally was underdeveloped and underexplored. Despite this, the *validity as an argument* discourse was legitimized via multiple methods including authorization (often using explicit references to well-known validity scholars), rationalization (often referencing institutional norms around validity), and mythopoesis (including building a narrative of ‘outdated’ vs ‘modern’ approaches to validity). Thus, despite a dearth of exploration regarding the role of argument or argumentation in HPE assessment validity, the concept of *validity as an argument* has been cemented in HPE’s published discourse. This study added to the empirical evidence that perhaps argument is used ornamentally rather than meaningfully in HPE assessment validity.

## Chapter 6

This perspective piece was a collaboration with other validity and equity scholars to imagine what validity might look like in the future. Specifically, as competency-based education, which foregrounds HPE’s accountability to the public,<sup>19,20</sup> continues to spread and as equity becomes more central to our field’s discourse,<sup>21</sup> how might validity evolve? In the manuscript, I argue that HPE validity should focus more on the consequences of assessment decisions<sup>22</sup> when developing and evaluating validity arguments. This includes consequences for learners and patients, effectively engaging them as audiences in the process of determining if assessment decisions are valid and defensible. Consequences also include the equity of assessment decisions, both at the level of the learner (i.e. is assessment done in an equitable way?) and the patient (i.e. is HPE developing a diverse workforce that can serve the needs of our diverse patients?). This manuscript serves as an explicit call to embrace the “broader, more ethical”<sup>1</sup> view of validity for purposes of aligning with HPE’s values of societal accountability and equity. I also acknowledge that some validity scholars, particularly those with psychometric or post-positivist backgrounds, may prefer a narrower view of validity that focuses on construct relevance. However, in this paper, I call for HPE to take a broad, agentic approach to validity--to deliberately choose which values are most important for our highest priority stakeholders/audiences (learners and patients) and integrate those values into how we conceptualize and operationalize validity arguments. Validity should serve HPE rather than vice versa.

## Chapter 7

This invited commentary on a manuscript by Coyle et al<sup>23</sup> builds upon the ideas put forth in Chapter 6. Coyle et al argue that HPE’s emphasis on academic achievement (e.g. grades, class rank, test scores) when selecting applicants for medical training can hinder the diversification of our field. Learners from minoritized or marginalized backgrounds disproportionately face systematic disadvantages such as racism, sexism, bias, economic disparities, and many other headwinds that unfairly hinder their academic performance at a higher rate than their privileged counterparts from majority identity backgrounds. Coyle et al suggest that the entrenched ideology of academic excellence being the criteria par excellence for selecting medical trainees

should be questioned. Coyle's argument directly challenges over a century of ideology and practice in academics that hold academic metrics as the center of performance assessment in education. In this commentary, I suggest that one way to operationalize Coyle's argument would be to include equity considerations in validity arguments. I highlight that the concept of validity holds significant power in HPE. By making equity of selection decisions part of the validity argument, we harness validity's influence and label inequitable selection outcomes as *invalid*. This is a concrete example of how HPE can take agency in choosing how to conceptualize validity arguments to align with our field's values and beliefs. If HPE truly values equity, validity is one tool to move toward more equitable assessment and selection decisions.

## **Chapter 8**

My collaboration with Dr. Tim Dyster of MedEd Models (<https://www.mededmodels.com/>) led to the development of visual infographics that describe the concepts of validity arguments and Argumentation Theory. These infographics were designed for easy sharing (e.g. social media) and for readers who are not experts in validity theory. This project pushed me to distill very complex ideas into simple, digestible aliquots that were understandable for frontline health professions educators.

## **Chapter 9**

Chapter 9 summarizes the research in this thesis, and proposes that embracing argument and argumentation in HPE validation could transform validity by making it more community-based (e.g., by engaging learners and patients), participatory (i.e., the co-creation of validity arguments), equitable (i.e., making equity a core part of validity), holistic (i.e., including many types of evidence and argument), and agentic (i.e., empowering HPE to choose how validity arguments are operationalized to align with its ideals and values). I propose that future research should include the use of implementation science and realist evaluation to understand how a more argumentation-based approach to validity could be operationalized in the real world.

## References

1. Newton PE, Baird J-A. The great validity debate: Taylor & Francis; 2016.
2. Lingard L. What we see and don't see when we look at 'competence': notes on a god term. *Advances in health sciences education*. 2009;14(5):625-628.
3. St-Onge C, Young M, Eva KW, Hodges B. Validity: one word with a plurality of meanings. *Adv Health Sci Educ Theory Pract*. 2017;22(4):853-867.
4. von Bergmann H, Childs RA. When I say... validity argument. *Medical Education*. 2018;52(10):1003-1004.
5. Cronbach LJ. Five perspectives on validity argument. *Test validity*. 1988:3-17.
6. Messick S. Validity of test interpretation and use: Research Report for the Educational Testing Service; 1990.
7. Kane M. The argument-based approach to validation. *School Psychology Review*. 2013;42(4):448-457.
8. Bachman LF. Constructing an assessment use argument and supporting claims about test taker-assessment task interactions in evidence-centered assessment design. MEASUREMENT-LAWRENCE ERLBAUM ASSOCIATES-. 2003;1:63-65.
9. Van Eemeren FH, Grootendorst R, Johnson RH, Plantin C, Willard CA. *Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments*: Routledge; 2013.
10. Messick S. Validity In. R. Linn (Ed.) *Educational measurement* (13-103): New York: Macmillan Publishing; 1989.
11. Boole G. *The Laws of Thought*: Dover, New York (original edition 1854); 1957.
12. Barth EM, Krabbe EC. *Formal Dialectics: Instruments for the Resolution of Conflicts about Expressed Opinions*. *Spektator* 7: 307. 1978;341.
13. Van Eemeren FH, Grootendorst R. *Argumentation, communication, and fallacies: A pragma-dialectical perspective*: Routledge; 2016.
14. Walton D. *Informal logic: A pragmatic approach*: Cambridge University Press; 2008.
15. Perelman C, Olbrechts-Tyteca L. *The new rhetoric: a treatise on argumentation*, trans. John Wilkinson and Purcell Weaver (Notre Dame, IN: University of Notre Dame Press, 1969). 1969;19.
16. Laclau E, Mouffe C. *Hegemony and socialist strategy: Towards a radical democratic politics*. Vol 8: Verso Books; 2014.
17. Fairclough N. *Analysing discourse: Textual analysis for social research*: Psychology Press; 2003.
18. Gee JP. *How to Do Discourse Analysis: A Toolkit*: New York and London: Routledge; 2011.
19. Hodge S. The Origins of Competency-Based Training. *Australian journal of adult learning*. 2007;47(2):179-209.
20. Holmboe ES, Sherbino J, Englander R, Snell L, Frank JR, Collaborators I. A call to action: The controversy of and rationale for competency-based medical education. *Med Teach*. 2017;39(6):574-581.
21. Lucey CR, Hauer KE, Boatright D, Fernandez A. Medical education's wicked problem: achieving equity in assessment for medical learners. *Academic Medicine*. 2020;95(12S):S98-S108.
22. Cook DA, Lineberry M. Consequences Validity Evidence: Evaluating the Impact of Educational Assessments. *Acad Med*. 2016;91(6):785-795.
23. Coyle M, Bullen J, Poobalan A, Sandover S, Cleland J. Follow the policy: An actor network theory study of widening participation to medicine in two countries. *Medical Education*. 2023.



## Samenvatting

Ideeën hebben macht omdat ze ons denken, handelen en beleid beïnvloeden. Validiteit is een belangrijk concept binnen HPE. Het wordt "de belangrijkste term in het lexicon van educatieve en psychologische metingen" genoemd.<sup>1</sup> Validiteit is vooral cruciaal bij de beoordeling van HPE-studenten, waarbij beoordelingen worden gebruikt om te beslissen over de gereedheid van de student voor patiëntenzorg. Patiëntveiligheid wordt beïnvloed door dergelijke beslissingen. De validiteit van dergelijke beoordelingen is daarom van het grootste belang. Dit proefschrift verkent de filosofische, theoretische en praktische aspecten van een veelgebruikte conceptualisering van validiteit in HPE: validiteit als argument.

## Hoofdstuk 1

In het openingshoofdstuk bespreek ik de plaats die validiteit heeft in de academische wereld in het algemeen en HPE in het bijzonder. Ik positioneer validiteit als een "god-term" die functioneert als een "educatieve afgod"<sup>2</sup> voor ons vakgebied. Ik positioneer validiteit ook als een kameleon die verschillende vormen aanneemt, afhankelijk van de context, het gebruik en de filosofische positie.<sup>3</sup> De meest voorkomende vorm die validiteit aanneemt is die van een argument ter ondersteuning van de interpretaties en het gebruik van beoordelingsgegevens (d.w.z. beslissingen).<sup>4</sup> Deze conceptualisering van validiteit werd overgenomen en aangepast door eminente validiteitswetenschappers, zoals Cronbach,<sup>5</sup> Messick,<sup>6</sup> Kane,<sup>7</sup> en Bachman.<sup>8</sup> In een dergelijke opvatting is validiteit in wezen een argument dat de op basis van beoordelingsgegevens genomen beslissingen ondersteunt of weerlegt. HPE heeft zich echter weinig verdiept in wat *argument* in deze context nu eigenlijk betekent, en welke theoretische en filosofische aannames in overweging moeten worden genomen bij het operationaliseren van validiteit als argument. Daartoe introduceer ik het vakgebied van argumentatietheorie,<sup>9</sup> een vakgebied dat zich bezighoudt met de ontwikkeling en evaluatie van gesproken of geschreven argumenten. De overkoepelende onderzoeksvraag van dit proefschrift is: "*Wat is de aard van argumentatie met betrekking tot beoordelingsvaliditeit in HPE?*"

## Hoofdstuk 2

Deze studie beschrijft hoe ik een dialectische (d.w.z. bespreking van ideeën) benadering heb gebruikt om validiteitsbewijs te ontwikkelen en te organiseren voor beslissingen die zijn genomen met behulp van het beoordelingssysteem van mijn opleidingsprogramma voor medisch specialist. Ons team gebruikte een vraag-en-antwoord aanpak om een validiteitsargument op te bouwen door het beoordelingssysteem systematisch te ondervragen. Vragen waren bijvoorbeeld: "Observeren beoordelaars daadwerkelijk dat AIOS de vaardigheden uitvoeren die worden beoordeeld?" en "Hoe zijn de vaardigheden die worden beoordeeld geoperationaliseerd op de beoordelingsformulieren?". Vervolgens hebben we antwoorden en ondersteunend bewijs gezocht met behulp van literatuur, eigen onderzoek of door uitleg te geven. Deze antwoorden werden vervolgens gecategoriseerd met behulp van Messick's<sup>10</sup> en Kane's<sup>7</sup> frameworks als een manier om ons

denken te organiseren en om ons werk te kunnen presenteren aan verschillende belanghebbenden die er misschien de voorkeur aan geven het ene of het andere framework te gebruiken bij het bespreken van validiteit. De resulterende visuele weergave van bewijsmateriaal, georganiseerd volgens Messick's en Kane's frameworks, werd beschouwd als onze *validity map*. Deze oefening met de validity map was een keerpunt in mijn denken over validiteit. Het stelde me in staat om prominente validiteit frameworks te operationaliseren binnen het beoordelingssysteem van mijn opleiding. Het werk opende mijn ogen voor het belang van daadwerkelijke argumentatie (d.w.z. het doen van claims en het kritisch onderzoeken van de onderbouwing van claims) bij het opbouwen van een validiteitsargument. Onze dialectische benadering van het stellen van vragen, het geven van antwoorden en het actief bekritisieren van het bewijs toonde de waarde aan van de sociale uitwisseling van ideeën als onderdeel van het validatie proces. Het toonde echter ook aan dat we meer kennis nodig hadden bij het ontwikkelen van daadwerkelijke argumenten (in plaats van alleen bewijs). Dit werk was de springplank naar de vervolgstudies binnen dit proefschrift.

### Hoofdstuk 3

De kritische review in hoofdstuk 3 verkende het terrein van de argumentatietheorie om de verschillende theoretische, filosofische en praktische aannames te begrijpen die verschillende benaderingen vormen voor argumentatieontwikkeling, kritiek en evaluatie. Deze studie onthulde voor mij de diepte en rijkdom van de argumentatietheorie, evenals de lange historie van hoe argumentatie in de loop van de tijd is geconceptualiseerd. Ik heb een academische bibliothecaris geraadpleegd om meerdere zoekstrategieën te ontwikkelen voor ERIC, Scopus, PubMed en Web of Science, op zoek naar literatuur die relevant is voor onze kritische review. Ik gebruikte ook snowball sampling om papers en boeken te vinden die een aanvulling zouden kunnen zijn op onze search. Ik vond meerdere stromingen in de argumentatietheorie die minder geschikt zijn voor het HPE-veld. De agnostische benadering van de context en de rigide vasthoudendheid aan premissen en conclusies van de formele logica maken het bijvoorbeeld onpraktisch om te gebruiken voor de contextafhankelijke aard van HPE-beoordelingen. Dialectiek<sup>12</sup> en Pragma-Dialectiek<sup>13</sup> richten zich beide op regels voor betrokkenheid en het oplossen van geschillen tussen twee gesprekspartners. Validatie van HPE-beoordelingen betreft niet het oplossen van geschillen, en de geldigheid mag ook niet afhankelijk zijn van interactieregels of oppositioneel van aard zijn. Ik vond echter twee argumentatie stromingen die nuttig zouden kunnen zijn voor het denken over validiteit van HPE-beoordelingen: informele logica<sup>14</sup> en nieuwe retoriek.<sup>15</sup> Informele logica omvat contextuele factoren en lokale normen en waarden bij de constructie en evaluatie van argumenten, terwijl het normen biedt waarmee argumenten kunnen worden beoordeeld, zoals aanvaardbaarheid (plausibiliteit van premissen), relevantie (samenhang tussen premissen en conclusies) en toereikendheid (toereikendheid van ondersteunend bewijs). Informele logica introduceert ook structuren die nuttig zijn voor het organiseren van argumenten, zoals het raamwerk van Toulmin (claims, warrants, data, weerleggingen). Nieuwe retoriek plaatst de waarden van een publiek op de voorgrond bij de constructie en

evaluatie van argumenten. In plaats van normatieve normen te hebben voor de evaluatie van argumenten, is het doel van New Rhetoric het overtuigen van een publiek voor een bepaalde bewering. De structuur en de normen van informele logica, gecombineerd met de flexibiliteit en de focus op een publiek van New Rhetoric, bieden een relevant kader voor validiteitsargumenten met betrekking tot HPE-beoordelingen. Ik wilde deze ideeën empirische testen door de standpunten van HPE-validiteitsexperts en de gepubliceerde HPE-validiteitsliteratuur te onderzoeken.

## **Hoofdstuk 4**

Deze kwalitatieve studie onderzocht hoe experts op het gebied van HPE-beoordelingsvaliditeit validiteitsargumenten conceptualiseren. Ik gebruikte critical case sampling om HPE-validiteitsexperts te zoeken, wat resulteerde in 16 deelnemers aan deze interview studie. Aan de hand van een thematische analyse heb ik 3 thema's ontwikkeld die relevant zijn voor onze onderzoeksvraag. Ten eerste waren de deelnemers van mening dat validiteitsargumenten sociaal van aard en contextueel gesitueerd moesten zijn. Dit thema sloot aan bij de publieksgerichte benadering van Nieuwe Retoriek en de contextgevoeligheid van zowel Informele Logica als Nieuwe Retoriek. Ten tweede merkten de deelnemers op dat validiteitsargumenten in de praktijk geen publiek hebben en zelden worden geëvalueerd. Ze stelden dat het publiek zelden wordt geïdentificeerd of in overweging wordt genomen wanneer validiteitsargumenten worden gebruikt, en daarom worden argumenten bijna nooit daadwerkelijk geëvalueerd. Ten derde merkten de deelnemers op dat het gebrek aan publieksparticipatie machtsverschillen binnen HPE creëerde en in stand hield. In het bijzonder waren de doelgroepen die het belangrijkst werden geacht (studenten en patiënten) nooit betrokken bij het ontwikkelen of evalueren van validiteitsargumenten. Het contrast tussen het eerste en tweede thema in deze studie toont een kloof tussen theorie (validiteitsargumenten moeten sociaal en publieksgericht zijn) en praktijk (validiteitsargumenten missen publiek of evaluatie), en het derde thema toonde een gevolg van deze kloof (hegemonie). Deze studie verschafte een rijker en genuanceerder inzicht in de opvattingen van HPE-experts over de rol (of het gebrek daaraan) van argumenten en argumentatie bij het valideren van beoordelingen. Vervolgens wilde ik de relevante discoursen over validiteitsargumentatie in peer reviewed publicaties onderzoeken.

## **Hoofdstuk 5**

In dit hoofdstuk heb ik kritische discoursanalyse van HPE peer-reviewed publicaties gebruikt om te onderzoeken hoe taal in invloedrijke artikel de conceptualisering van validiteitsargumenten en argumentatie heeft gevormd. Na het screenen van meer dan 800 artikelen uit HPE-tijdschriften, omvatte het uiteindelijke corpus voor analyse 39 manuscripten die werden geïdentificeerd als invloedrijk bij het vormgeven van het denken over validiteitsargumenten in ons vakgebied. De meeste van de uiteindelijke papers waren overzichtsartikelen, perspectieven en commentaarstukken. Ik putte uit meerdere tradities van

discoursanalyse, waaronder Laclau en Mouffe,<sup>16</sup> Fairclough,<sup>17</sup> en Gee<sup>18</sup>. De analyse werd ook beïnvloed door ons eerdere werk over argumentatietheorie. Ik ontdekte dat het discours over validiteitsargumentatie zeer beperkt een plek heeft in HPE peer reviewed publicaties. Het grootste deel van het discours draait om bewijs (in plaats van argumenten) en welke frameworks moeten worden gebruikt (meestal Messick of Kane). Er werden weinig componenten van argumentatietheorie gevonden in de artikelen, met een opmerkelijke afwezigheid van argumentatiepubliek. Het concept van argumentatie was over het algemeen weinig uitgewerkt. Desondanks werd het validiteitsdiscours als argument gelegitimeerd via meerdere methoden, waaronder autorisatie (vaak met expliciete verwijzingen naar bekende validiteitsexperts), rationalisatie (vaak verwijzend naar institutionele normen rond validiteit) en mythopoesis (inclusief het opbouwen van een verhaal van 'verouderde' versus 'moderne' benaderingen van validiteit). Dus, ondanks een gebrek aan onderzoek naar de rol van argument of argumentatie in de validiteit van HPE-beoordelingen, is het concept van *validiteit als argument* belangrijk in het gepubliceerde discours van HPE.

## Hoofdstuk 6

Dit perspectiefstuk was een samenwerking met andere validiteits- en diversiteitswetenschappers om een beeld te vormen hoe validiteit er in de toekomst uit zou kunnen zien. Met name naarmate competentiegericht onderwijs, dat de verantwoordingsplicht van HPE tegenover het publiek op de voorgrond plaatst,<sup>19,20</sup> zich blijft verspreiden en naarmate diversiteit en inclusiviteit meer centraal komt te staan in het discours van ons vakgebied.<sup>21</sup> In het hoofdstuk betoog ik dat HPE-validiteit zich meer zou moeten richten op de gevolgen van beoordelingsbeslissingen<sup>22</sup> bij het ontwikkelen en evalueren van validiteitsargumenten. Gevolgen voor lerenden en patiënten, waardoor zij als publiek effectief worden betrokken bij het proces om te bepalen of beoordelingsbeslissingen valide en verdedigbaar zijn. De gevolgen omvatten ook de billijkheid van beoordelingsbeslissingen, zowel op het niveau van de student (d.w.z. wordt de beoordeling op een billijke manier uitgevoerd?) als op het niveau van de patiënt (d.w.z. ontwikkelt HPE een diverse beroepsgroep dat kan voldoen aan de behoeften van onze diverse patiënten?). Dit manuscript dient als een expliciete oproep om een "breder en meer ethisch"<sup>1</sup> perspectief op validiteit te omarmen om aan te sluiten bij HPE's maatschappelijke verantwoordelijkheid. Ik erken ook dat sommige validiteitswetenschappers, met name degenen met een psychometrische of post-positivistische achtergrond, misschien de voorkeur geven aan een beperktere perspectief op validiteit die zich richt op constructrelevantie. In dit artikel roep ik HPE echter op om een bredere, activistische benadering van validiteit te hanteren - om bewust te kiezen welke waarden het belangrijkst zijn voor onze belanghebbenden/doelgroepen met de hoogste prioriteit (studenten en patiënten) en die waarden te integreren in de manier waarop we validiteitsargumenten conceptualiseren en operationaliseren. Validiteit moet HPE dienen in plaats van andersom.

## Hoofdstuk 7

Dit uitgenodigde commentaar op een manuscript van Coyle et al.<sup>23</sup> bouwt voort op de ideeën die in hoofdstuk 6 naar voren zijn gebracht. Coyle et al. stellen dat HPE's nadruk op academische prestaties (bijv. cijfers en testscores) bij het selecteren van kandidaten voor medische training diversiteit kan belemmeren. AIOS met een minderheids- of gemarginaliseerde achtergrond worden onevenredig geconfronteerd met systematische obstakels zoals racisme, seksisme, vooroordelen, economische ongelijkheden en vele andere belemmeringen die hun academische prestaties op oneerlijke wijze belemmeren. Coyle et al. suggereren dat de diepgewortelde ideologie van academische excellentie als criterium bij uitstek voor het selecteren van AIOS s in twijfel moet worden getrokken. In dit commentaar suggereer ik dat een manier om Coyle's argument te operationaliseren zou zijn om inclusiviteitsoverwegingen op te nemen in validiteitsargumenten. Ik benadruk dat het concept van validiteit invloedrijk is in HPE. Door inclusiviteit en diversiteit van selectiebeslissingen onderdeel te maken van het validiteitsargument, benutten we de invloed van validiteit en bestempelen we onrechtvaardige selectieresultaten als *ongeldig*. Dit is een concreet voorbeeld van hoe HPE stelling kan nemen bij het kiezen van hoe validiteitsargumenten worden geconceptualiseerd om aan te sluiten bij de waarden en overtuigingen van ons vakgebied. Als HPE echt waarde hecht aan inclusiviteit, is validiteit een instrument om te komen tot meer rechtvaardige beoordelings- en selectiebeslissingen.

## Hoofdstuk 8

Mijn samenwerking met Dr. Tim Dyster van MedEd Models (<https://www.mededmodels.com/>) heeft geleid tot de ontwikkeling van visuele infographics die de concepten van validiteitsargumenten en argumentatietheorie beschrijven. Deze infographics zijn ontworpen om gemakkelijk te delen (bijv. sociale media) en voor lezers die geen experts zijn in de validiteitstheorie.

## Hoofdstuk 9

Hoofdstuk 9 vat het onderzoek in dit proefschrift samen en stelt voor dat het omarmen van argumenten en argumentatie in HPE-validatie de validiteit zou kunnen transformeren door deze meer community-based te maken (bijv. door studenten en patiënten te betrekken), participatief (d.w.z. de co-creatie van validiteitsargumenten), inclusief (d.w.z. inclusiviteit een kernonderdeel van validiteit maken), holistisch (d.w.z. inclusief vele soorten bewijs en argumenten), en activistisch (d.w.z. HPE in staat stellen om te kiezen hoe validiteitsargumenten worden geoperationaliseerd om in lijn te zijn met zijn idealen en waarden). Ik stel voor dat toekomstig onderzoek het gebruik van implementatiewetenschap en realistische evaluatie zou kunnen omvatten om te begrijpen hoe een meer op argumentatie gebaseerde benadering van validiteit in de echte wereld zou kunnen worden geoperationaliseerd

## Verwijzingen

1. Newton PE, Baird J-A. Het grote validiteitsdebat: Taylor & Francis; 2016.
2. Lingard L. Wat we wel en niet zien als we kijken naar 'competentie': aantekeningen over een godsterm. Vooruitgang in het onderwijs in gezondheidswetenschappen. 2009; 14(5):625-628.
3. St-Onge C, Young M, Eva KW, Hodges B. Geldigheid: één woord met een veelvoud aan betekenissen. Adv Gezondheid Sci Educ Theorie Praktijk. 2017; 22(4):853-867.
4. von Bergmann H, Childs RA. Als ik zeg... geldigheid argument. Medisch onderwijs. 2018; 52(10):1003-1004.
5. Cronbach LJ. Vijf perspectieven op validiteitsargumenten. Validiteit van de test. 1988:3-17.
6. Messick S. Geldigheid van testinterpretatie en -gebruik: onderzoeksrapport voor de Educational Testing Service; 1990.
7. Kane M. De op argumenten gebaseerde benadering van validatie. Beoordeling van schoolpsychologie. 2013; 42(4):448-457.
8. Bachman LF. Het construeren van een argument voor beoordelingsgebruik en ondersteunende beweringen over interacties tussen testpersoon en beoordelingstaak in evidence-centered assessment design. METING-LAWRENCE ERLBAUM ASSOCIATES-. 2003;1:63-65.
9. Van Eemeren FH, Grootendorst R, Johnson RH, Plantin C, Willard CA. Grondbeginselen van de argumentatietheorie: een handboek met historische achtergronden en hedendaagse ontwikkelingen: Routledge; 2013.
10. Messick S. geldigheid in. R. Linn (Ed.) Educatieve meting (13-103): New York: Macmillan Publishing; 1989.
11. Boole G. De wetten van het denken: Dover, New York (oorspronkelijke uitgave 1854); 1957.
12. Barth EM, Krabbe EC. Formele dialectiek: instrumenten voor het oplossen van conflicten over geuite meningen. Spektator 7: 307. 1978;341.
13. Van Eemeren FH, Grootendorst R. Argumentatie, communicatie en drogredenen: een pragma-dialectisch perspectief: Routledge; 2016.
14. Walton D. Informele logica: een pragmatische benadering: Cambridge University Press; 2008.
15. Perelman C, Olbrechts-Tyteca L. De nieuwe retoriek: een verhandeling over argumentatie, vert. John Wilkinson en Purcell Weaver (Notre Dame, IN: University of Notre Dame Press, 1969). 1969;19.
16. Laclau E, Mouffe C. Hegemonie en socialistische strategie: naar een radicale democratische politiek. Deel 8: Verso Boeken; 2014.
17. Fairclough N. Analyse van het discours: tekstuele analyse voor sociaal onderzoek: Psychology Press; 2003.
18. Jeetje, JP. Hoe discoursanalyse te doen: een toolkit: New York en Londen: Routledge; 2011.
19. Hodge S. De oorsprong van competentiegerichte training. Australisch tijdschrift voor volwasseneneducatie. 2007; 47(2):179-209.
20. Holmboe ES, Sherbino J, Englander R, Snell L, Frank JR, Medewerkers I. Een oproep tot actie: de controverse van en de grondgedachte voor competentiegericht medisch onderwijs. Med onderwijs. 2017; 39(6):574-581.
21. Lucey CR, Hauer KE, Boatright D, Fernandez A. Het slechte probleem van medisch onderwijs: het bereiken van gelijkheid in beoordeling voor medische studenten. Academische geneeskunde. 2020; 95 (12S): S98-S108.
22. Cook DA, Lineberry M. Gevolgen Validiteitsbewijs: evaluatie van de impact van educatieve beoordelingen. Acad Med. 2016; 91(6):785-795.

- 23.** Coyle M, Bullen J, Poobalan A, Sandover S, Cleland J. Volg het beleid: Een actor netwerk theorie studie van het verbreden van deelname aan geneeskunde in twee landen. Medisch onderwijs. 2023.

## Scientific Impact

Validity is seen as a central concept in health professions education (HPE) assessment.<sup>1</sup> It is often seen as the standard by which assessment is deemed trustworthy, reliable, or meaningful. Within HPE, validity is most commonly understood as being an argument that supports the interpretations and uses of assessment data.<sup>2,3</sup> Rather than labeling an assessment score or instrument as *validated*, validity arguments focus on the decisions that result from assessment scores. Two frameworks are commonly used in HPE to organize evidence to support validity arguments. Messick's framework describes different types or sources of validity evidence that can be sought to support an argument.<sup>4</sup> Kane's framework lays out a series of inferences that are made between the moment of observation (i.e. scoring) to the moment of decision-making.<sup>5,6</sup> However, neither framework provides guidance on the specifics of argument and argumentation that undergird HPE validity. A useful analogy might be to imagine a courtroom. The lawyers need to have relevant evidence, and to have it organized in an understandable way, but they also need to understand the rules (both implicit and explicit) of the court itself. They need to know how their arguments should be structured, to whom they are making an argument (A jury? Only a judge? The other lawyer?) and by what standards their argument will be evaluated. Currently in HPE, there is a lack of understanding of how the validity 'courtroom' works and the rules therein. In other words, there is a lack of understanding of what argument means in HPE validity. The main objective of this body of research is to deeply explore how argument and argumentation are conceptualized and operationalized in HPE assessment validity.

## Relevance

This research is relevant to anyone who creates or uses assessment systems, as well as those who are impacted by assessment scores. Those most impacted by these research findings include medical trainees, patients, program leaders, and certification and accreditation bodies.

### *Impact on medical trainees and patients*

Validity experts in our study identified significant issues with current HPE validation practices. They noted that the most important audiences who should be able to review and evaluate validity arguments are learners and patients. Ostensibly, these groups are most impacted by assessment decisions. Learners have committed years of training and often huge amounts of money toward their careers; assessment decisions loom large in their lives and should therefore be defensible and valid. Patients rely on valid assessment decisions by HPE programs to ensure that clinicians are not graduated to unsupervised practice until they are truly ready. However, learners and patients are rarely (if ever) engaged in validity arguments. They do not get to scrutinize nor co-create the arguments supporting assessment decisions. One major implication of this research points to the need to address this gap: novel approaches for engaging with these audiences must be developed in order to include the most important stakeholders in validity argumentation. This will likely entail

the identification of interested representatives from each group (not every trainee nor patient would want to be involved). Such representatives would need education on validity and assessment theory, and mechanisms for longitudinal discourse with argument developers. For example, if a post-graduate medical education program uses a multifaceted assessment system to arrive at summative decisions about readiness for unsupervised practice, learner and patient representatives ought to weigh in on what type of validity evidence might be most important to support the overall argument of the validity of the ultimate summative decision. Perhaps a validity committee could exist in such a program (much like a competence committee) to engage in regular discourse, taking a dialectical approach (as we did in Chapter 2) to develop, iterate, and evaluate validity arguments. Such a committee could include learners, patients, and anyone else deemed an important stakeholder for assessment decisions. We began using this approach on a small scale in our competency-based, time-variable training pilot (TIMELESS) in which we educated our residents on our assessment system and allowed them to scrutinize the summative decisions resulting from synthesis of assessment data.<sup>7</sup> I learned that many trainees have meaningful input on the defensibility (i.e. validity) of assessment practices and their downstream decisions. For example, TIMELESS residents asked for improved group decision-making approaches in our clinical competency committee (CCC). In response, we reimaged our CCC processes to be more theory-informed in an attempt to improve defensibility.<sup>8</sup> Validity argumentation could be a mechanism to harness that input to improve assessment processes and decisions.

#### *Impacts on education program leaders*

The work in this thesis suggests that validation ought to be conceptualized more as a social interaction (i.e. exchange of ideas) that can be audience-centric (i.e. specific stakeholders engaged). Bringing this conceptualization into being will likely require additional resources for HPE training programs. Currently, many HPE training programs have a dearth of educators with expertise in assessment generally, much less validity specifically. Validity can be an intimidating and confusing concept. Programs would need to recruit and/or develop local faculty with expertise and interest in validity and validation, as well as having protected time to develop validity arguments and engage with necessary audiences. Right now, validity is sequestered to the world of assessment research or psychometrics. This thesis suggests that validity argumentation should be made much more accessible to educators and audiences alike.

#### *Impacts on certifying and accrediting organizations*

I have been a post-graduate medical education associate program director or program director for nearly 10 years. I have never been asked to produce a validity argument for any organization that accredits my program or certifies my learners. In the United States, there is very little validity argumentation (if any) that occurs between those who are making summative decisions about graduation readiness (i.e., program leaders) and organizations who oversee learner and program quality via accreditation and certification. My program of

research suggests that there should be ongoing argumentation (i.e., claims, evidence, rebuttals, iteration) between education programs and regulatory bodies. Perhaps accrediting and certifying organizations could help create a process whereby the validity arguments individual programs develop could be reviewed on a regular basis. This review could include not simply an evaluative judgment of the validity argument, but also guidance on which types of evidence is most necessary for the organization to see, and help with how to gather such evidence (e.g. citing existing studies, conducting new research, etc). For example, the validity map from Chapter 2 could be shared with accrediting or certifying organizations, who could then provide an evaluation of the overall claim (our summative decisions are defensible) and feedback on the argument or evidence therein. As a program director, I would very much welcome such feedback.

### **Making validity arguments local**

Validity is situated and context dependent. The most important audiences for validity arguments are also context-specific (e.g. one program's learners, one specialty's patients). Therefore, validity arguments should be developed and shared by local education leaders. Validity arguments should be continually swirling within and around each program's assessment system, ensuring that the stakeholders most impacted by the validity arguments are involved in prioritizing argument claims and evaluating evidence provided in support of such claims. Validity should not be only the purview of national exam designers or oversight agencies. Validity argumentation should also have a local, program-level focus that simultaneously engages with national organizations. Keeping a local focus on validity could help bring meaning and defensibility to local assessment decisions, which are often the most consequential in determining when learners are deemed ready for unsupervised practice.

### **Evidence of Scientific Impact**

While validity is central to education, it is also a challenging topic to discuss. People have different understandings of what validity is, and validity theory can be dense and seem impenetrable. I have attempted to share my findings with a wide range of audiences including validity scholars, frontline educators, certifying and accrediting organizations, and learners. I have also become a collaborator with the Research Group in Pursuit of Validity, presenting my work at their webinars and colloquia. I have found that foregrounding the argument aspect of validity resonates with frontline educators even if they do not have significant experience with validity theory. Evidence of dissemination of my PhD work include:

- **9 peer-reviewed and invited presentations** at local, national, and international conferences including the Accreditation Council for Graduate Medical Education (ACGME) national meeting, the International Conference on Residency Education (ICRE), the American Board of Medical Specialties (ABMS) national meeting, and the International Competency-Based Health Professions Educators Collaborators international webinar series.

- **One podcast interview** with Medical Education<sup>9</sup> to discuss the paper that is presented in Chapter 3 of this thesis. <https://www.podbean.com/ew/pb-num46-1300ea7>
- **7 peer-reviewed publications** relating to validity in prominent HPE journals<sup>10-16</sup> such as Perspectives on Medical Education, Academic Medicine, Medical Teacher, and Medical Education.
- **1 book chapter** that explores validity in the context of entrustable professional activities<sup>17</sup>
- **A grant-funded year as a Visiting Scholar** with the American Board of Medical Specialties to examine the nature of validity argumentation<sup>18</sup>
- **Invited to be a collaborator** with the Research Group in Pursuit of Validity<sup>19</sup> as well as for discussions with multiple American Board of Medical Specialties member boards to explore how validation can be a more argumentation-based and engaging with stakeholders.

On an anecdotal level, I am now frequently consulted by local educators at my institution as well as education scholars from around the US who are seeking guidance on validation practices.

## References

1. Downing SM. Validity: on the meaningful interpretation of assessment data. *Medical education*. 2003;37(9):830-837.
2. von Bergmann H, Childs RA. When I say... validity argument. *Medical Education*. 2018;52(10):1003-1004.
3. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ*. 2015;49(6):560-575.
4. Messick S. Validity of test interpretation and use: Research Report for the Educational Testing Service; 1990.
5. Kane MT. An argument-based approach to validity. *Psychological bulletin*. 1992;112(3):527.
6. Kane M. The argument-based approach to validation. *School Psychology Review*. 2013;42(4):448-457.
7. Kinnear B, Santen SA, Kelleher M, et al. How does TIMELESS training impact resident motivation for learning, assessment, and feedback? Evaluating a competency-based time-variable training pilot. *Academic Medicine*. 2023;98(7):828-835.
8. Kinnear B, Santen SA, Schumacher DJ, et al. Using Theory-Informed Group Processes to Make TIMELESS Promotion Decisions. *Academic Medicine*. 2024;99(1):28-34.
9. How argumentation theory can inform assessment validity: a critical review - Interview with Benjamin Kinnear. In: Eva K. *Medical Education Podcasts* 2022.
10. Touchie C, Kinnear B, Schumacher D, et al. On the validity of summative entrustment decisions. *Medical Teacher*. 2021:1-8.
11. Kinnear B, Kelleher M, May B, et al. Constructing a validity map for a workplace-based assessment system: Cross-walking Messick and Kane. *Academic Medicine*. 2021;96(7S):S64-S69.
12. Kinnear B, Schumacher DJ, Driessen EW, Varpio L. How argumentation theory can inform assessment validity: A critical review. *Medical Education*. 2022;56(11):1064-1075.
13. Kinnear B, Schumacher DJ. What the hell is water? Changing medical education's ideology through validity. *Medical Education*. 2024;58(3):274-276.
14. Kinnear B, Martini A, Varpio L, Driessen EW, Schumacher DJ. How do validity experts conceptualise argumentation? It's a rhetorical question. *Medical Education*. 2024.
15. Kinnear B, St-Onge C, Schumacher DJ, Marceau M, Naidu T. Validity in the Next Era of Assessment: Consequences, Social Impact, and Equity. *Perspectives on Medical Education*. 2024;13(1):452.
16. Kinnear B, Schumacher DJ, Varpio L, Driessen EW, Konopasky A. Legitimation Without Argumentation: An Empirical Discourse Analysis of 'Validity as an Argument' in Assessment. *Perspectives on Medical Education*. 2024;13(1):469.

17. ten Cate O, Burch V, Chen H, Chou FC, Hennis M. Entrustable Professional Activities and Entrustment Decision-Making in Health Professions Education: Ubiquity Press; 2024.
18. Kinnear named an ABMS Visiting Scholar for 2022-2023. 2022; <https://www.uc.edu/news/articles/2022/09/kinnear-named-abms-visiting-scholar.html>.
19. Research Group in Pursuit of Validity. <http://www.grqv-rgpv.ca/en/about-us/>.

## Curriculum Vitae

Benjamin was born on July 23, 1982 in St. Louis, Missouri, USA. He graduated summa cum laude from Saint Louis University and thereafter matriculated to the University of Missouri – Columbia School of Medicine. While at Mizzou, Benjamin was elected to and served as president of the local Alpha Omega Alpha honor society chapter. He also received multiple leadership and academic-based awards. After medical school, he entered residency training in Internal Medicine and Pediatrics (Med-Peds) at University of Cincinnati Medical Center and Cincinnati Children's Hospital Medical Center. After residency, he served as Med-Peds chief resident, subsequently joining as a faculty member in the Division of Hospital Medicine at Cincinnati Children's.

Benjamin currently is an associate professor of Internal Medicine and Pediatrics in the Division of Hospital Medicine at University of Cincinnati Medical Center and Cincinnati Children's Hospital Medical Center, caring for hospitalized children and adults. He served as Associate Program Director (APD) for the Med-Peds residency program from 2015-2024, and also served as APD for the Internal Medicine residency program from 2014-2024. As APD, he led the residency QI curriculum, developed and led the Medical Education Pathway, and led multiple education innovation projects. In 2024, he became Program Director for the Med-Peds residency program. He also became Program Director for the IMSTAR Medical Education Fellowship at University of Cincinnati.

Benjamin obtained his Master of Medical Education from University of Cincinnati in 2018 and completed a one-year research fellowship with the Education Research Scholars Program at Cincinnati Children's in 2020. In 2020 he was selected for the Macy Faculty Scholars Program, during which he piloted competency-based time-variable training in the UC Internal Medicine residency program. Benjamin has published over 100 peer-reviewed manuscripts (28 first author, 13 as senior author) and received grant support from the American Medical Association, the American Board of Medical Specialties, the American Board of Pediatrics, the Josiah Macy Jr. Foundation, and the University of Cincinnati Department of Medical Education. He currently serves as deputy editor for the Journal of Hospital Medicine and associate editor for Perspectives on Medical Education.

Benjamin spends most of his free time with his wife and two daughters hiking, playing board games, and travelling whenever possible. He is a St Louis Cardinals fanatic and strongly believes that mint-flavored ice cream is an abomination.

## Acknowledgements

Every reflection of gratitude I have begins and ends with my family. What started 23 years ago as an evening of roller skating and salsa dancing has grown into a marriage of 15 years and two amazing kiddos. Maja –you are the only reason I could even consider pursuing a PhD. Over the last 4 years there have been countless evenings, weekends, and vacations from which I have stolen time to do the work reflected in this thesis. Without your generosity and support, so much of my professional life would not be possible. Thank you for being the center of gravity for our family. Thank you for being the best partner I could ever hope for. Thank you for being my best friend. Thank you for making me a better human. To Adela and Mila – you are my happy thoughts on bad days. I long ago gave up the worry of caring what other people think of me, except for you. I carry with me the hope that you're proud of me, not because I obtained a PhD but because I chose to pursue a difficult goal just for the sake of curiosity and learning. Thank you for being my paranymphs, for making me laugh, for continually teaching me humility, and for being such amazing humans. I hope you're proud of me, but even moreso I hope you know how proud I am of you.

To my parents, who are no longer here - thank you for being the best mom and dad I could imagine. I never had to worry about whether I was loved, cared for, or safe. You showed me what kindness, grit, and humor look like. I wish you were here, and think about you all the time. To my brother Brad – thank you for being a role model of curiosity – of loving to learn new things. You're the person I look to when I need an aspirational figure for motivation. To Jeanne and Ron, my godparents – thank you for inviting me into your family, and for joining mine. Thank you for showing me what hard work and high standards look like, and for pushing me to do my best ever since kindergarten. Thank you for the love, time, and tears you have poured into me and my family. To Joe – thank you for being my brother from another mother. Thank you for generously donating your incredible artistic skills to this work. To Ron and Michael – I miss you. To Rick, Josh, and Carla – I'm so grateful that our families are now one and the same. Thank you for your support and encouragement, and for being such awesome role models to my kids. To Tomek and Ewa – jesteście dwoma najfajniejszymi, najzabawniejszymi ludźmi, jakich znam. Dziękuję, że pozwoliliście mi poślubić waszą córkę, że uczyniliście mnie częścią waszej rodziny i że pokazaliście mi, jak wygląda jednoczesne bycie kopającymi tyłki naukowcami i skromnymi, twardo stąpającymi po ziemi ludźmi. To Maggie, John, and Sophia – thank you for being a continual example of kindness and generosity for my kids.

To the educators at Cincinnati who have been my role models, collaborators, and supporters – I can't thank you enough. To Eric, Matt, Dana, and Dan – thank you for being the OG Voltron group who helped me fall in love with MedEd scholarship. Those early days of scholarship when we did not know what we were doing are still some of my favorite memories. To Jen and Caroline – thank you for being my Med-Peds role models, and for shepherding me in to the world of education. To Sally – thank you for being a superhuman catalyst for my career. I don't deserve your sponsorship or generosity, but am incredibly grateful for both. To the RADICAL lab

– thank you for providing a space for community, curiosity, debate, and friendship. Let’s change the world. Thank you to Samir, Pat, and Karen for supporting my pursuit of a PhD, and always finding solutions to barriers that arose. I’m so grateful to be part of the CCHMC HM division. To Leslie, Nicole, Charissa, and the Med-Peds residents – thank you for your support and patience while I pursued this PhD. I’m so grateful to be part of the best Med-Peds program in the world.

I also want to thank some HPE colleagues outside of Cincinnati, though the names are too numerous to completely list. To Meredith and Christina – thank you for your generosity and patience. Five years ago I arrived at a symposium on validity as one of the only non-PhDs there. I felt like an absolute fraud who was too ignorant to participate. You sensed my imposter syndrome, reached out, shared your wisdom, and became supportive friends and mentors. I would not be doing this PhD without you. To Holly – thank you for your friendship, your humor, and the thousands of commiseration texts as we shared challenges that inevitably are part of a PhD. To Justin and Hannah – thank you for sharing so many stimulating conversations and incredible ideas. I’m inspired by both of you, both as scholars and as people.

To my SUPERvisors – Lara, Erik, and Dan - you are the reason I have made it to this point. You have shown me nothing but patience, encouragement, wisdom, good humor, and kindness. I know not all PhD candidates have the joy of working with a group of supervisors like you, and I’ve been continually grateful for each of you. You have helped me push through imposter syndrome, discouragement, frustration, fatigue, and grief. You showed me grace, patience, and love when I lost my mom early in the course of my PhD. I would not have gotten back up and continued without your encouragement. Thank you for providing such a nurturing and safe environment to be curious, to fail, and to be silly. I have had so much fun on this PhD journey, and it’s largely because of you. Dan – thank you pushing me to first consider pursuing a PhD, and for being my go-to mentor and sponsor. You have opened countless doors for me, made me a better researcher, and given me innumerable pep talks. Thank you. Lara – thank you for taking a chance on me, for pushing me to take the more interesting but more difficult road of studying a field I knew very little about, and for helping me find my academic armor when facing critique. Erik – thank you for encouraging me to take chances, to be bold, and to stay true to my values in my research. All three of you are amazing supervisors, and even better humans. I’m grateful for your friendships. Abby – thank you for so generously donating your time, expertise, and energy toward my PhD journey, simply out of kindness and love of scholarship. I’m thankful to have gotten to know you.

I mention that my gratitude begins and ends with my family. Maja, Adela, and Mila – I love you so much. Thwip.

### **SHE Dissertation Series**

The SHE Dissertation Series publishes dissertations of PhD candidates from the School of Health Professions Education (SHE) who defended their PhD theses at Maastricht University. The most recent ones are listed below. For more information go to: <https://she.mumc.maastrichtuniversity.nl>

Bock, L. (25-04-2025) Exploring value-based healthcare decision-making in postgraduate medical education: A focus on medical residents' perspectives

Kolm, A. (24-03-2025) TRANSCENDING TIME, SPACE, AND CULTURE: International Online Collaboration Competencies in Health Professions Education

Suliman, S. (27-01-2025) Learners' Mental Health and Challenging Transitions: The Added Value of Co-Creation

Onan, E. (20-12-2024) Promoting the self-regulated use of desirable difficulties: The case of Interleaved Practice

Shimizu, I. (18-12-2024) Social interdependence in collaborative learning

McDonald, J. (28-11-2024) Collections for reflection: the transformative potential of portfolios in medical education

Barry, E. (20-11-2024) Leadership and Followership within Healthcare Teams: Exploring the roles and collaborative dynamics in interprofessional teams

Linden van der, J. (14-10-2024) Learning Beyond Assessment Compliance. The conundrum of self-regulated learning in the context of summative assessment

Smeets, H. (04-10-2024) Beyond the silos: Design guidelines for interprofessional assessment in higher healthcare education

Wisener, K. (24-09-2024) From Incentive Schemes to Feedback Processes: On the challenges of recruiting, supporting, and retaining clinical educators

Maulina, F. (06-09-2024) Developing Physician Leadership in Indonesia's Rural and Remote Contexts: Culture Matters

Wu, J. (24-06-2024) Sitting Smartly in Tertiary Education Exploring Strategies to Tackle Sedentary Behavior for Student's Cognitive and Physical Well-being

Bossen, J. (16-05-2024) Shared Decision Making in Orthopaedic Surgery: Implementation of tools and training in clinical practice

Klasen, J. (02-05-2024) Allowing Failure as an Unspoken Pedagogy in Residency Training

Behrens, C. (31-01-2024) Medical students' emotions in complex simulation

- Nicholson, J. (20-12-2023) Partnering for Success: Librarian Involvement in Assessment for Learning
- Barber, C. (20-12-2023) Aligning education to societal needs: Evaluating social accountability in health professions education
- Ilishkina, D. (19-12-2023) Self-Regulation of the Motivation to Learn
- Loosveld, L. (07-12-2023) Finding the merit of mentoring Mentors' personal knowledge and beliefs about mentoring in health professions education
- Puranitee, P. (02-11-2023) Alleviating burnout in medical school: increasing a sense of belonging and collegiality in the clinical workplace
- Nobel, M. (28-09-2023) The radiological report: a compromise between structured reporting and natural language processing
- Heijnens, L. (16-06-2023) Aspects of geometry, fixation and materials in total hip arthroplasty. What have we learned?
- Gillespie, H. (06-06-2023) Helping students become doctors: analysing tensions and releasing opportunities in clinical workplaces
- Khan, M. (16-05-2023) Overcoming barriers in the prevention of surgical site infections: a master plan employing task-based interprofessional training
- Tremblay, M-L. (08-05-2023) Lights. Camera. Action. Debrief. Designing Immersive Simulation for Novices to Promote Learning
- Bynum, W. (15-02-2023) Out of the Shadows: A qualitative exploration of shame in learners across the continuum of medical education
- Mordang, S. (07-02-2023) Challenges in high-value cost-conscious care training of residents: Exploring the various stakeholders' roles and attitudes
- Brouwer, E. (12-12-2022) Medical education without borders. The what, why and how of International Medical Programmes
- Kellar, J. (26-10-2022) Becoming Pharmacists: Professional identity struggles of a profession in transition
- Biwer, F. (08-07-2022) Supporting Students to Study Smart – a learning sciences perspective
- Bransen, D. (22-06-2022) Beyond the self: A network perspective on regulation of workplace learning
- Lee, J. (08-06-2022) The Medical Pause in Simulation Training
- Kruepunga, N. (17-05-2022) Development of the caudal part of the human embryo

Cantillon, P. (28-04-2022) The Social construction of clinical education: being and becoming in clinical teams

Pieters, J. (01-04-2022) Let's talk about it: Palliative care education in undergraduate medical curricula

Jonge de, L. (28-03-2022) Is it all in the mind? Stakeholder conceptions on workplace based assessment

Beuken, J. (25-3-2022) Waves towards harmony: Learning to collaborate in healthcare across borders

Ilgen, J. (15-12-2021) Comfort with uncertainty in medical professionals. An exploration of how clinicians experience and manage dynamic problems in practice

Schut, S. (9-12-2021) The Burden of Proof - Agency and Accountability in Programmatic Assessment'

Hui, L. (6-12-2021) 'Fostering Self-Regulated Learning: the Role of Perceived Mental Effort'

Meeuwissen, S. (12-11-2021) 'Team learning at work. Getting the best out of interdisciplinary teacher teams and leaders'

Nguyen Thi, V.A. (02-11-2021) Motivating and educating health professionals to work in less attractive specialties: Findings from experiences of Vietnam

Martens, S. (15-10-2021) Building student-staff partnerships in higher education

Lestari, E. (05-10-2021) INTERPROFESSIONAL EDUCATION Lessons from Indonesia

Atherley, A. (27-09-2021) Beyond the struggles: Using social-developmental lenses on the transition to clinical training

Schillings, M. (06-07-2021) Talking about feedback: Face-to-face peer dialogue about written feedback

Wilbur, K. (05-07-2021) NO WHERE | NOW HERE: Context and Competency Expectations in Workplace-Based Training

Bendermacher, G. (02-07-2021) Navigating from Quality Management to Quality Culture

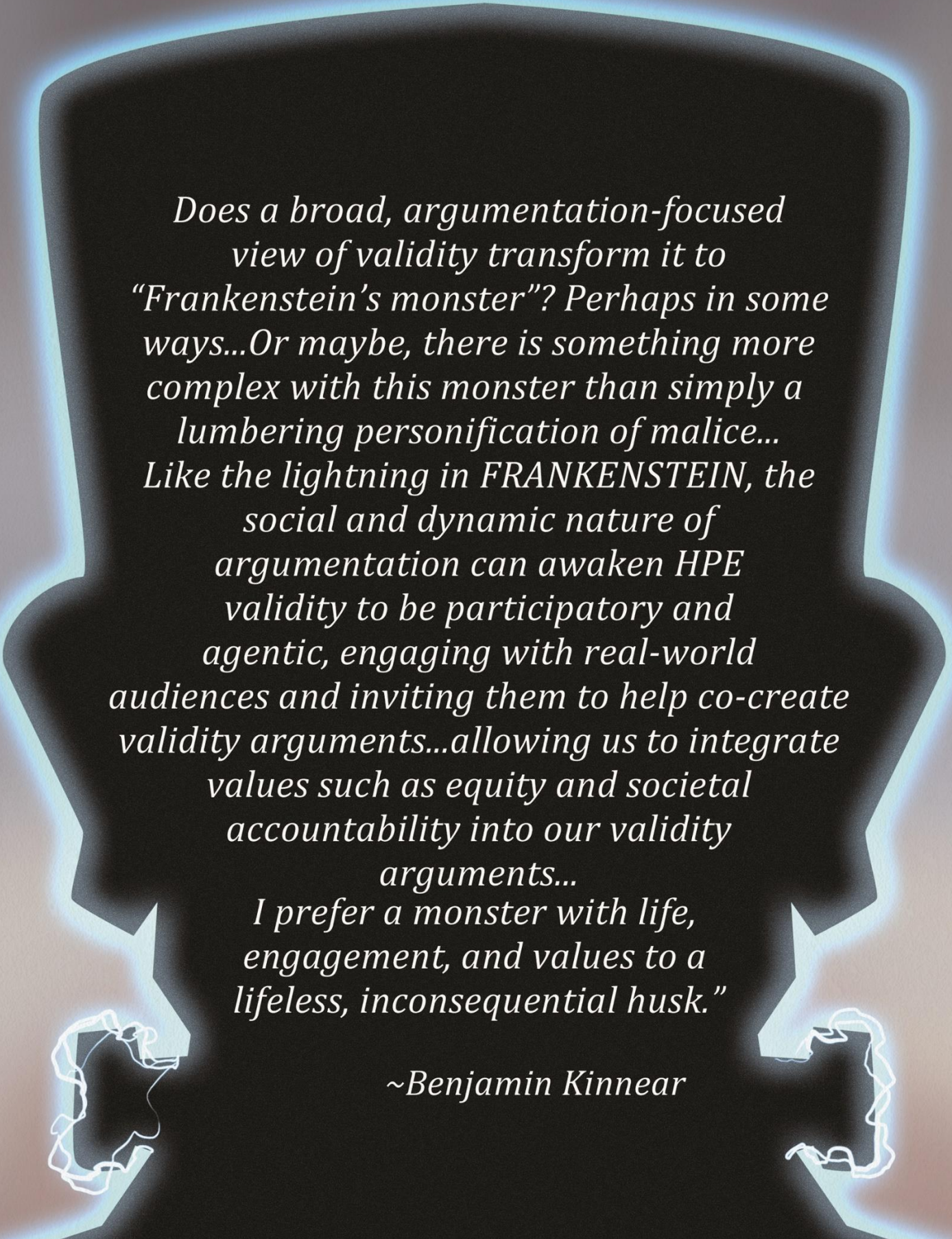
Ahmed Khan, R. (29-06-2021) Assessing curriculum viability in Undergraduate Medical Education

Chim, H.Q. (30-03-2021) Physical Activity Behavior and Learning in Higher Education

Dominguez, L.C. (23-02-2021) Persistence in surgical training: The role of job crafting and leadership

Bindels, E. (22-02-2021) DOING WELL, GETTING BETTER; Facilitating physicians' reflection on their professional performance

- Iqbal, Z. (15-12-2020) All stakeholders matter in faculty development: Designing entrustable professional activities for small group facilitation
- Tran, QT. (09-12-2020) Nationwide implementation of medical skills training laboratories in a developing country: studies from Vietnam
- Pacifico, J. (30-11-2020) Making the Implicit Explicit: Uncovering the Role of the Conceptions of Teaching and Learning and the Perceptions of the Learning Climate in Postgraduate Medical Training.
- Nishigori, H. (17-11-2020) Why do doctors work for patients? Medical professionalism in the era of neoliberalism
- Oudkerk Pool, A. (06-11-2020) Competency-based portfolio assessment – Unraveling stakeholder perspectives and assessment practices
- Geel van, K. (05-11-2020) Lifelong learning in radiology: all eyes on visual expertise
- Stammen, L. (16-10-2020) Pursuing - High-Value, Cost-Conscious Care - The Role of Medical Education
- Meulen van der, M. (15-10-2020) Assessment of physicians' professional performance using questionnaire-based tools
- Matsuyama, Y. (05-10-2020) Contextual attributes fostering self-regulated learning in a teacher-centered culture: learner's professional identity formation is a trigger
- Rovers, S. (16-09-2020) Growing knowledge, supporting students' self-regulation in problem-based learning
- Bourgeois-Law, G. (03-09-2020) Conceptualizations of remediation for practicing physicians
- Giuliani, M. (19-05-2020) A Critical Review of Global Curriculum Development, Content and Implementation in Oncology
- Schreurs, S. (20-03-2020) Selection for medical school; the quest for validity
- Schumacher, D. (19-03-2020) Resident Sensitive Quality Measures: Defining the Future of Patient-Focused Assessment
- Sehlbach, C. (21-02-2020) To be continued.... Supporting physicians' lifelong learning



*Does a broad, argumentation-focused  
view of validity transform it to  
“Frankenstein’s monster”? Perhaps in some  
ways...Or maybe, there is something more  
complex with this monster than simply a  
lumbering personification of malice...  
Like the lightning in FRANKENSTEIN, the  
social and dynamic nature of  
argumentation can awaken HPE  
validity to be participatory and  
agentic, engaging with real-world  
audiences and inviting them to help co-create  
validity arguments...allowing us to integrate  
values such as equity and societal  
accountability into our validity  
arguments...*

*I prefer a monster with life,  
engagement, and values to a  
lifeless, inconsequential husk.”*

*~Benjamin Kinnear*