Original Research Paper

Development and validation of the Diversity, Equity, and Inclusion Index (DEII) tool for assessing DEI in medical education lectures: a pilot study

Cortlyn Brown[1], Desiree Leverette[2], Christiana Agbonghae[3], Joseph Rigdon[4], and Edward Ip[5]

[1]MD, MCSO, Associate Professor of Emergency Medicine, Atrium Health Carolinas, Charlotte, United States
[2]MD, Department of Pediatrics, Yale School of Medicine, New Haven, United States
[3]MD, Department of Emergency Medicine, Atrium Health Wake Forest Baptist Hospital, Charlotte, United States
[4]PhD, Assistant Professor, Department of Biostatistics and Data Science, Wake Forest School of Medicine, Public Health Sciences, Winston Salem, United States
[5]PhD, Professor, Department of Biostatistics and Data Science, Wake Forest School of Medicine, Winston Salem, United States

Abstract

Background: Evaluating diversity, equity, and inclusion (DEI) in health education is crucial to ensure an inclusive and effective learning environment. In this pilot study, we aimed to create and validate the first tool, the Diversity, Equity, and Inclusion Index (DEII), to assess DEI within health education lectures. We hypothesized that the DEII would demonstrate strong face and content validity as well as acceptable to excellent interrater reliability. **Methods:** This pilot cross-sectional study was conducted from January 2022 to December 2023. We developed the Diversity, Equity, and Inclusion Index (DEII) and assessed its face and content validity. The study followed the COSMIN taxonomy guidelines. Interrater reliability was calculated using intraclass correlation coefficients from DEII scores of 50 YouTube lectures. **Setting:** The study was conducted at several academic medical centers in the United States of America, encompassing diverse health education settings. **Participants:** 10 expert-DEI reviewers as well as 10 non-expert DEI reviewers who represented the major sections of health professions including physicians, nurses, researchers, social workers, respiratory therapists, and physical therapists, as well as multiple institutions assessed face and content validity. Expert reviewers were defined based on specific criteria, including board certification, leadership positions in DEI, authorship of peer-reviewed DEI publications, national presentations on DEI topics, and implementation of DEI initiatives. Non-expert faculty were selected based on their involvement in health education but without specific DEI expertise. Participants were recruited through a convenience sample approach. All approached participants agreed to take part in the study. To assess interrater reliability, three researchers scored 50 health education lectures on YouTube whose audiences ranged from respiratory therapists, nurses, physicians, physician associates and those training to be each. **Results:** Interrater reliability analysis demonstrated that 16 out of 17 items met the threshold for acceptable reliability (ICC $\geq$ 0.50). Specifically, 7 items exhibited excellent reliability (ICC $\geq$ 0.90), 5 demonstrated good reliability ($0.70 \leq$ ICC $<$ 0.90), and 4 showed moderate reliability ($0.50 \leq$ ICC $<$ 0.70). At the domain level, three of the four domains achieved acceptable reliability (ICC $\geq$ 0.70), with one domain reaching the threshold for excellent reliability (ICC $\geq$ 0.90). **Discussion:** In this pilot study, the DEII is a reliable and valid tool for assessing diversity, equity, and inclusion in health education lectures. Its implementation may enhance the inclusivity of health education, ultimately leading to better-prepared healthcare professionals and improved patient outcomes. Further research is needed to refine the tool and explore its impact on educational practices and outcomes.

Keywords: medical education, diversity, lecture, equity, inclusion

Date submitted: 9-June-2025

Email: Cortlyn Brown (cortlyn.brown@atriumhealth.org)

Citation: Brown C, Leverette D, Agbonghae C, Rigdon J, and Ip E. Development and validation of the Diversity, Equity, and Inclusion Index (DEII) tool for assessing DEI in medical education lectures: a pilot study. Educ Health 2025;38:404-410

Online access: www.educationforhealthjournal.org
DOI: 10.62694/efh.2025.371

Published by The Network: Towards Unity for Health

Background:

Diversity, equity, and inclusion (DEI) refer to the presence of varied identities and perspectives (diversity), the fair removal of barriers to access and opportunity (equity), and the intentional creation of learning environments where all individuals feel welcomed, respected, and valued (inclusion). A diverse healthcare provider workforce in the medical field is associated with increased access and utilization of the healthcare system, high quality health care with improved outcomes and patient experience, and higher fiscal margins for hospitals.[1–4] In order for diverse individuals to thrive in medicine, however, it is important that health education is inclusive and accessible to all. Focus groups of medical students, however, revealed that students perceived a significant lack of diversity and awareness in their health education and, therefore, they did not feel prepared to appropriately treat a diverse patient population.[5] Medical students also reported feelings and experiences of exclusion in their health education, including facing stereotypes and prejudices.[5] Similar concerns have been raised in nursing, where students perceived inadequate integration of cultural competence into the curriculum and reported challenges in understanding professional culture and inclusivity.[6,7] These findings are not limited to a single discipline—interprofessional studies also highlight persistent barriers to diversity, equity, inclusion, and accessibility in health professions education.[8]

To evolve and meet the changing patient and medical provider demographics and needs, it is necessary to have a validated tool to assess the degree of diversity, equity, and inclusion (DEI) in health education lectures given to all medical professionals. In this pilot study, we present the Diversity, Equity, and Inclusion Index (DEII), the first validated scoring tool and to assess DEI within health education lectures for all medical professionals in real-time.

To our knowledge, there are three existing tools that address constructs of DEI in health education. The DEII is a validated tool that was formulated to measure both qualitative and quantitative aspects related to diversity within nursing school lectures.[9] While this tool is validated, it is specific to nursing lectures and was not tested or validated for other medical professionals. Neither of the two remaining tools, The Upstate Bias Checklist and the Byrne Guide for Inclusionary Cultural Content, have been confirmed to be validated and neither serves as a scoring tool. The Upstate Bias Checklist primarily aims to guide the lecturer as they are creating educational materials such as lecture slides/notes, clinical vignettes, multiple choice questions, standardized patient encounters among others for all health professions.[10] In contrast, the Byrne Guide is specifically tailored for nursing education and provides guidelines and examples to help nurse educators create and evaluate materials such as textbooks, syllabi, computer software, and examinations.[11]

The objective of this pilot study was to develop and validate the Diversity, Equity, and Inclusion Index (DEII), the first tool designed to assess DEI within health education lectures across all health professions. Existing resources such as the FRDC, Byrne Guide, and Upstate Bias Checklist are either limited to nursing or lack validation and a scoring framework. We hypothesized that the DEII would demonstrate strong face and content validity and acceptable to excellent interrater reliability.

Methods

After review of existing instruments[9–11], we drafted the preliminary DEII. Using the COSMIN taxonomy as our guide, we used this draft to assess for face validity and content validity.[12] To evaluate face validity, we asked 10 expert-DEI reviewers as well as 10 non-expert DEI reviewers if each of the domains (representation, equity/inequity, linguistic bias, and accessibility) measured the construct at hand (yes/no- with optional comments) as well as if the overall survey measured the constructs at hand (DEI) (yes/no- with optional comments). The following criteria was used to define expert-DEI reviewers.

1. Board-certified provider (physician, nursing OR resident in good standing with ACGME) AND,
2. Leadership position (local, regional, national) within the last 10 years focused on DEI AND/OR,
3. Author on greater than or equal to 3 accepted peer reviewed publications related to DEI in the last 10 years AND/OR
4. Speaker for greater than or equal to 2 national presentations on DEI topics in the last 10 years AND/OR

5. Practitioners who have successfully implemented at least 1 DEI initiative locally, regionally, or nationally within the last 10 years

Next, we evaluated content validity by asking those same reviewers to evaluate the survey content for clarity, accuracy, and relevance as well as if the DEII reflects all of our defined constructs (representation, equity/inequity, linguistic bias, and accessibility). We completed 3 rounds to evaluate face validity and 3 rounds for content validity. We then made changes to the survey based on the feedback.

To calculate interrater reliability, we had three scorers who were members of the research team score the first 15 minutes of 50 lectures using the DEII. Time started once the actual material was introduced (ie. time at the beginning of the lecture when individuals are finding their seats was not counted). The lectures were broad in terms of topic, lecturer, and target audience and were found on YouTube. We did not apply formal inclusion or exclusion criteria when selecting lectures from YouTube, as our goal was to capture a broad and representative sample of health education content across professions and topics. To ensure consistency across lectures of varying length, we evaluated only the first 15 minutes of each recording, beginning when the educational content started (excluding introductory remarks or downtime), since prior research suggests that key concepts and representative teaching practices are typically introduced early in a lecture.[13,14] For each of the DEII questions, intraclass correlation (ICC) were calculated using linear mixed effects models. These models included the question response as the outcome, a fixed effect for rater, and a random effect for lecture.

In addition, ICCs were calculated for each domain (representation, equity/inequity, linguistic bias, and accessibility). Domain scores for each observation were calculated by taking the average question response per domain and multiplying by the number of questions in that domain. For example, if a rater's scores on the three linguistic bias questions for a particular lecture were one, three, and two (on a Likert scale), the average score would be two and the linguistic bias domain score would be two*three = six. Domain scores were set to missing if less than two questions were answered in that domain. For example, if the rater's scores for linguistic bias were one, NA, and NA, the linguistic bias domain score would be NA (missing). ICCs per domain were calculated using linear mixed effects models with the domain score as the outcome, a fixed effect for rater, and a random effect for lecture. All inter-rater reliability analyses were performed using R version 4.3.2.[15]

Results

Intraclass correlation coefficient (ICC) values below 0.50 are considered indicative of poor reliability; values between 0.50 and 0.69 represent moderate reliability; values between 0.70 and 0.89 reflect good reliability; and values of 0.90 or higher are indicative of excellent reliability.[16] Interrater reliability analysis showed that 16 out of 17 questions met the threshold for acceptable reliability (ICC of ≥0.50). Specifically, 7/17 had excellent reliability, 5/17 had good reliability, and 4/17 had moderate reliability (table 1). At the domain level, three of the four domains achieved acceptable reliability (table 1).

Discussion

In this pilot study, we developed and validated the Diversity, Equity, and Inclusion Index (DEII), the first structured scoring tool designed to evaluate DEI in health education lectures across multiple health professions. The DEII demonstrated strong face and content validity and acceptable to excellent interrater reliability, suggesting it can serve as a practical and psychometrically sound measure of inclusivity in health education.

Our findings build on and extend prior work in the field. The FRDC tool[9], while validated, is limited to nursing lectures and does not generalize to other health professions. The Upstate Bias Checklist and Byrne Guide provide useful frameworks for educators but remain unvalidated and lack a scoring system.[10,11] More broadly, studies have identified underrepresentation and bias in health education materials, including case vignettes, multiple-choice questions, and standardized patients, highlighting the need for structured tools to evaluate inclusivity.[17,18] The DEII addresses these gaps by offering a validated, quantitative instrument that can be used in real time across disciplines and by both expert and non-expert raters.

Beyond medicine, research across nursing, physical therapy, and social work has also documented gaps in DEI integration. Nursing

Table 1: Intraclass correlation (ICC) in 19 questions in DEI assessment. Per question, ICC estimated via linear mixed model with fixed effect for rater, and random effect for lecture (50 total).

Domain	Domain ICC[d]	Question Number	Question Prompt	Question ICC
Representation	0.88	1	The individuals in the pictures/graphics were diverse in terms of age, gender, skin color, ability status (e.g., requiring a wheelchair), and body habitus.	0.92
		2	In the pictures/graphics, people/body parts from non-majority groups were given the same priority in terms of size, composition, cropping, etc. as those from majority groups.	0.93
		3	The individuals in the case vignettes were diverse in terms of their identity characteristics (e.g., age, ability, gender, race/ethnicity, religion, sexual orientation, immigration status, primary language, and body habitus).	0.94
		4	The presenter clearly stated the inclusion and exclusion criteria for any research study presented.	0.78
		5	The statistics and data presented included individuals from non-majority groups. If they were not included, the presenter commented on their absence.	0.8
		6	In the figures and tables, statistics and data on individuals from non-majority groups were given the same priority as statistics and data on individuals from majority groups.	0.98
Equity/Inequity	0.93	7	The lecturer mentioned specific healthcare disparities (HDs) and/or social drivers of health (SDoH).	0.89
		8	The lecturer discussed the causes/background of HDs and/or SDoHs.	0.93
		9	The lecturer discussed potential solutions for HDs and/or SDoH.	0.83
Linguistic Bias	0.57	10	If diseases (e.g. HIV, substance use disorder, mental health disorders) and patient populations (ie. .LGBTQ+, unhoused) that are frequently stigmatized were presented, it was done to explore SDoH and highlight the disproportionate burden some diseases place on specific patient populations.	0.96
		11	The lecturer avoided Identity First Language	0.48
		12	The lecturer avoided Exclusive Language	0.6
Accessibility	0.76	13	Videos in the presentation had captions.	-[a]
		14	The general slide color palate was optimized for those with colorblindness	0.63
		15	Texture was used to help differentiate lines or points on a graph (see examples below).	0.6
		16	The font was large enough to be easily visible, left-justified, and not in all italics or underlined.	0.7
		17	The speaker avoided spatial language (words used to describe the relationship of objects in space)	0.62
		18	All audience members were able to participate equally in any activities	1.0[b]
		19	The speaker avoided slides with bold, flashing slide transitions and strobes or flashing GIFS.	-[c]

[a]40/50 lectures have no data; for remaining ones, there are eight lectures where only one rater gave a response, and then two lectures where all three raters responded, but all gave the same response; answers were always equal to "1"; model will not converge as variance-covariance matrix is not a positive definite matrix

[b]35/50 videos have no data; for remaining ones, there are five videos where only one rater gave a response, two videos where two raters responded but all gave same answer, and then eight videos where all three raters gave same answer; answers varied from "3", "4", and "5"; Model is nearly unidentifiable and may not have converged due to very large eigenvalue

[c]For all 50/50 lectures, all three raters gave answers that were always a "5"; model will not converge as variance-covariance matrix is not a positive definite matrix

[d]Domain-level ICCs were not obtained by averaging the item-level ICCs. Instead, domain scores were calculated for each rater and lecture, and separate mixed-effects models were used to estimate ICCs at the domain level. Because ICCs are derived from variance components, they cannot be computed as simple arithmetic means of item-level ICCs.

students report insufficient representation of diverse populations in lectures and clinical scenarios,[19] physical therapy educators note the absence of structured tools to evaluate cultural competence and accessibility in teaching,[20] and social work programs have long called for measurable DEI benchmarks in classroom and fieldwork education.[21] By positioning DEII as an interprofessional instrument, our study responds to these calls and provides a standardized way to assess inclusivity across the health professions.

The DEII showed particularly strong reliability in the domains of **representation** and **equity/inequity**, indicating that visible diversity and explicit discussion of health disparities are consistently identifiable by raters. These findings align with prior literature across disciplines demonstrating the importance of representation and equity content in educational outcomes.[5] In contrast, the lower reliability in **linguistic bias** and **accessibility** highlights areas that are more nuanced and subject to interpretation, consistent with broader literature on implicit bias in language and the variable adoption of universal design principles.[22] These results suggest that clearer operational definitions and additional rater training may further improve reliability in future iterations of the DEII.

By operationalizing constructs of representation, equity, bias, and accessibility into measurable items, the DEII makes a unique contribution to the broader body of evidence in health professions education. It can be used formatively to provide feedback to educators, longitudinally to monitor institutional progress, and in research to evaluate the effectiveness of DEI initiatives. Its applications extend beyond individual lectures to curriculum design, accreditation compliance, and institutional policy, aligning with calls across medicine, nursing, allied health, and social work for systematic evaluation of DEI integration in training.

Implications for Health education

The DEII has multiple applications across the medical education continuum. The DEII is not meant to penalize educators and students. Instead, it aims to assess improvement over time in terms of personal growth and monitoring. For lecturers and curriculum developers, it provides structured feedback that can guide the design and delivery of more inclusive lectures. For institutions and program directors, the DEII offers a way to monitor progress toward accreditation standards and institutional DEI commitments, complementing broader curricular evaluation efforts. For researchers, the tool enables standardized measurement of DEI content, allowing comparisons across institutions and longitudinal studies of interventions. At the policy level, the DEII could inform quality benchmarks and accountability measures as national organizations increasingly prioritize equity and inclusion in training standards. Finally, from a theoretical standpoint, the DEII operationalizes constructs of representation, equity, bias, and accessibility into measurable items, contributing to the growing body of scholarship on how DEI can be meaningfully embedded into health professions education.

Limitations and Future Directions

This study has several limitations. First, our sample of lectures was drawn from publicly available YouTube videos, which may not fully represent the range or quality of lectures delivered in academic medical centers. While this provided broad exposure to diverse topics and audiences, it also introduced potential selection bias. Second, our study relied on three raters who were members of the research team, which may limit generalizability; future work should include a larger pool of independent raters from multiple institutions. Third, although the DEII demonstrated strong reliability overall, the moderate agreement in the linguistic bias and accessibility domains suggests that some items may require clearer definitions or additional rater training. Fourth, while both expert and non-expert reviewers informed face and content validity, our sample size was relatively small, and reviewers were recruited through convenience sampling, which could limit diversity of perspectives. Finally, as a cross-sectional pilot study, we were unable to assess how use of the DEII may influence actual educational practices, learner experiences, or long-term outcomes.

Future research should address these limitations by testing the DEII in live lecture settings, expanding rater pools across disciplines and institutions, refining items with lower reliability, and evaluating the tool's impact on educational practices and learner preparedness to care for diverse patient populations.

Conclusions

In conclusion, the DEII is a reliable and valid tool for assessing DEI in health education lectures. Its implementation can enhance the inclusivity of health education, ultimately leading to better-prepared healthcare professionals and improved patient outcomes. Continued research and refinement of the DEII will help ensure its effectiveness and broad applicability in diverse educational settings.

References

1. Komaromy M, Grumbach K, Drake M, Vranizan K, Lurie N, Keane D, Bindman AB. The Role of Black and Hispanic Physicians in Providing Health Care for Underserved Populations. *New England Journal of Medicine*. 1996 May 16;334(20):1305–1310. https://doi.org/10.1056/NEJM199605163342006.

2. Gomez LE, Bernet P. Diversity improves performance and outcomes. *Journal of the National Medical Association*. 2019 Aug;111(4):383–392. https://doi.org/10.1016/j.jnma.2019.01.006.

3. Greenwood BN, Carnahan S, Huang L. Patient–physician gender concordance and increased mortality among female heart attack patients. *Proceedings of the National Academy of Sciences*. 2018 Aug 21;115(34):8569–8574. https://doi.org/10.1073/pnas.1800097115.

4. Saha S, Komaromy M, Koepsell TD, Bindman AB. Patient-Physician Racial Concordance and the Perceived Quality and Use of Health Care. *Archives of Internal Medicine*. 1999 May 10;159(9):997. https://doi.org/ 10.1001/archinte.159.9.997.

5. Verbree AR, Isik U, Janssen J, Dilaver G. Inclusion and diversity within medical education: a focus group study of students' experiences. *BMC Medical Education*. 2023 Jan 25;23(1):61. https://doi.org/10.1186/s12909-023-04036-3.

6. Sumpter DF, Carthon JMB. Lost in Translation: Student Perceptions of Cultural Competence in Undergraduate and Graduate Nursing Curricula. *Journal of Professional Nursing*. 2011 Jan;27(1):43–49. https://doi.org/10.1016/j.profnurs.2010.09.005.

7. Strouse SM, Radtke E. Nursing Students' Perceptions About the Culture of Nursing: A Focused Ethnography Study. *Nursing Education Perspectives*. 2024 Mar;45(2):71–76. https://doi.org/10.1097/01.NEP.0000000000001214.

8. Ellis AL, Pappadis MR, Li CY, Rojas JD, Washington JS. Interprofessional Perceptions of Diversity, Equity, Inclusion, Cultural Competence, and Humility Among Students and Faculty: A Mixed-Methods Study. *Journal of Allied Health*. 2023;52(2):89–96.

9. Scisney-Matlock M, McCloud PK, Barnard RM. Systematic assessment and evaluation of diversity content presented in classroom lectures: the FRDC tool. *Journal of Cultural Diversity*. 2001;8(3):85–93.

10. [CSL STYLE ERROR: reference with no printed form.]

11. Byrne MM, Weddle C, Davis E, McGinnis P. The Byrne Guide for Inclusionary Cultural Content. *Journal of Nursing Education*. 2003 Jun;42(6):277–281. https://doi.org/10.3928/0148-4834-20030601-11.

12. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, De Vet HCW. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*. 2010 Jul;63(7):737–745. https://doi.org/10.1016/j.jclinepi.2010.02.006.

13. Bradbury NA. Attention span during lectures: 8 seconds, 10 minutes, or more? *Advances in Physiology Education*. 2016 Dec 1;40(4):509–513. https://doi.org/10.1152/advan.00109.2016.

14. Wilson K, Korn JH. Attention during Lectures: Beyond Ten Minutes. *Teaching of Psychology*. 2007 Apr;34(2):85–89. https://doi.org/10.1080/00986280701291291.

15. [CSL STYLE ERROR: reference with no printed form.]

16. Liljequist D, Elfving B, Skavberg Roaldsen K. Intraclass correlation – A discussion and demonstration of basic features. Chiacchio F, editor. *PLOS ONE*. 2019 Jul 22;14(7):e0219854. https://doi.org/10.1371/journal.pone.0219854.

17. Bullock JL, Lockspeiser T, Del Pino-Jones A, Richards R, Teherani A, Hauer KE. They Don't See a Lot of People My Color: A Mixed Methods Study of Racial/Ethnic Stereotype Threat Among Medical Students on Core Clerkships. *Academic Medicine*. 2020 Nov;95(11S):S58–S66. https://doi.org/10.1097/ACM.0000000000003628.

18. Ross DA, Boatright D, Nunez-Smith M, Jordan A, Chekroud A, Moore EZ. Differences in words used to describe racial and gender groups in Medical Student Performance Evaluations. Gold JA, editor. *PLOS ONE*. 2017 Aug 9;12(8):e0181659. https://doi.org/10.1371/journal.pone.0181659.

19. Jeffreys MR. Teaching Cultural Competence in Nursing and Health Care: Inquiry, Action, and Innovation [Internet]. 3rd ed. New York, NY: Springer Publishing Company; 2016 [cited 2025 Aug 16]. https://doi.org/10.1891/9780826119971. Available from: http://connect.springerpub.com/lookup/doi/10.1891/9780826119971.

20. Cultural Diversity in the Practice of Physical Therapy: A Review of the Literature. *International Journal of Humanities, Social Sciences and Education* [Internet]. 2021 [cited 2025 Aug 16];8(1). https://doi.org/10.20431/2349-0381.0801005. Available from: https://arcjournals.org/pdfs/ijhsse/v8-i1/5.pdf.

21. Bibuss A, Boutte-Queen N. The ethics of inclusion: Developing culturally competent social work education. *Journal of Social Work Education*. 2019;55(2):221–234. https://doi.org/10.1080/10437797.2018.1526726.

22. Lie DA, Lee-Rey E, Gomez A, Bereknyei S, Braddock CH. Does cultural competency training of health professionals improve patient outcomes? A systematic review and proposed algorithm for future research. *Journal of General Internal Medicine*. 2011 Mar;26(3):317–325. https://doi.org/10.1007/s11606-010-1529-0.