SPRING 2023
ISSUE

University of Connecticut
Neag School of Education

UCONN | NEAG SCHOOL OF EDUCATION

Journal

UCONN
NEAG SCHOOL OF EDUCATION

TABLE OF CONTENTS

MISSION STATEMENT

The Neag School of Education Journal is an editor-reviewed, open-access, annual journal founded and run by graduate students and published through the Neag School of Education at the University of Connecticut. Its primary purpose is to offer a platform for graduate students to share their research and knowledge with academic communities, to broaden and deepen the literature of education, as written and experienced by graduate and doctoral students, as well as early-career scholars.

The Neag School of Education Journal highlights the strongest, most robust student and early-career work from a broad range of disciplines such as educational psychology, curriculum and instruction, teacher education, as well as educational leadership.

Of particular interest are pieces providing reflection on student experience with their research processes and manuscripts showcasing research in the preliminary stages. The journal offers students and early-career scholars the chance to publish work through diverse types of academic writing including, but not limited to, research articles (e.g., qualitative and quantitative research), essays, literature review, as well as personal experience and reflective pieces.

Aligning with the mission of its governing institution, the Neag School of Education Journal places significance on pieces seeking to improve education and social systems in order to facilitate increasingly effective, equitable, and socially just practices for educators and practitioners from a variety of fields, perspectives, and theoretical lenses as they serve their local communities. Reflections are also sought after to foster relations and collaboration amongst graduate students and their colleagues, to pass along wisdom, innovation, and creativity in pursuit of fostering a graduate community rooted in rigorous and intentional research design and practices. The journal's open access ensures it as a source for current and future practitioners.

FOUNDING BOARD MEMBERS

CAITLIN BLACKSMITH
Research Methods, Measurement, and Evaluation Ph.D. Student

LUIS ORIONE FERRERIA
Giftedness, Creativity, and Talent Development Ph.D. Student

BRENNA FITZMAURICE
School Psychology Ph.D. Student

GINA NORMAN
School Psychology Ph.D. Student

SANDRA SEARS
Educational Psychology/Special Education Ph.D. Student

VAISHNAVI SIVAPRASAD
Educational Psychology Ph.D. Student

ANI TERTERIAN
Higher Education and Student Affairs Master's Student

SIERRA TRUDEL
MA, SYC, NCSP, LPC: School Psychology Ph.D. Student

JENNIE WEINER
Neag School of Education Faculty Advisor

EMILY WINTER
PhD, NCSP: School Psychology Ph.D. Alumna

LETTERS FROM THE BOARD

It is with great pleasure to present the inaugural edition of the Neag School of Education Journal. This project has been in the works for close to two years, beginning with conversations amongst graduate students during the pandemic who were in search of ways to gain experience in research and editing, as well as foster connection with others in graduate school whilst enduring remote learning. Drawing inspiration from various other graduate student-led journals in the field, the Neag School of Education Journal seeks to contribute to the development of graduate-student research and scholarship. Our publishing perspective was designed to prioritize novel ideas in research, tackling common problems in the review process, seeking to provide tailored and quick feedback to our contributors. Our goal is simple: to foster growth in research, whether that be for the authors of the five inaugural pieces or for the editorial board's development of their editing and feedback skills.

In our initial conceptualization of the Journal we wanted to appeal to research and the development of scholarship through soliciting manuscripts ranging from research articles, essays, literature reviews, and reflective pieces. The Journal was designed for graduate students, with graduate students in mind, so we conceptualized novel copyrighting policies to allow for just that: development. We hope that authors take our feedback, and feedback from consumers of their publication, to peer reviewed journals in the future. We hope to see our prior publications (our alumni!) in top peer-reviewed journals, in part thanks to the editor feedback and peer support received at Neag School of Education Journal. We are proudly an open-access journal, contributing to the idea that research should be free and easy to access for all, to bridge the gap between research and practice, and connect academia and tangible implementation.

Our featured pieces in this issue tie together themes aligning with the mission of the Journal and the Neag School of Education at the University of Connecticut. We sought to include articles that improve education and social systems, providing wisdom, innovation, and creativity. We are so excited for you to read the phenomenal work of graduate students across the country in our beautiful first issue.

For this inaugural issue we have many people to thank. To Dr. Jennie Weiner, thank you for your unwavering support and commitment to us and your graduate students. You are a wonderful role model as a scholar, academic, and all-around person. Thank you for all you have done for the team. To the dean of the College, Dr. Jason Irizarry, we are forever grateful for your enthusiasm and passion for our work. Thank you for believing in us and using your role at Neag to lift up the work of graduate students. Thank you to the design team at UConn who helped oversee the details of a beautiful first issue, we admire your creativity and collaboration. To the editors, thank you for your time and dedication to providing feedback to our submissions, we appreciate your thoughtfulness and willingness to learn. To the individuals who submitted to our inaugural journal, thank you for your interest. We look forward to continuing to uplift graduate students' work in years to come. To the board, thank you for the camaraderie, the passion and grit, and for being wonderful friends.

To learn more about our team and mission, please visit us at https://education.uconn.edu/neag-journal/

DISPUTES OF ASSESSMENT AND EVALUATION: AN ANALYSIS OF CONNECTICUT DUE PROCESS HEARINGS FROM 2020-2022

ELIZABETH ZAGATA
UNIVERSITY OF CONNECTICUT NEAG SCHOOL OF EDUCATION

ABSTRACT

Between January 2020 and March 2022, there were 20 fully adjudicated due process hearings in the state of Connecticut. Over half of those address disputes regarding assessment and evaluation. In this analysis, I provide an overview of the legal requirements for special education assessment and evaluation as well as look for trends in the decisions of these 11 hearings. The purpose of the analysis is to determine any potential patterns in these recent rulings to learn more about current issues in the state and make potential recommendations for districts desiring to reduce the likelihood of future disputes. Specifically, there were clusters of cases addressing the following topics: evaluator credibility when reviewing disputes about independent evaluations, movement between services received under Section 504 and special education, dismissal of complaints due to case circumstances, and economic-related geographic distribution of cases. Understanding these issues of current dispute provides opportunities for improved practice to subsequently reduce future conflict.

Keywords: Due process hearing, evaluation

Author Note
Elizabeth Zagata ORCID 0000-0003-1208-2691

Support was provided by Grant H325H190003 for Elizabeth Zagata, from the Office of Special Education Programs (OSEP). The author would like to acknowledge Jessica De Perio Wittman for her feedback on earlier versions of this manuscript.

In the United States, due process is a constitutionally protected right as per the Fourteenth Amendment (Lombardi & Ludlow, 2004). In the field of special education, the Continuum for Dispute Resolution Processes and Practices considers due process hearings to be among the most advanced stages of conflict (Center for Appropriate Dispute Resolution in Special Education [CADRE]), n.d.). In fact, CADRE describes due process hearings as "the most contentious and adversarial of required dispute resolution processes under IDEA (Individuals with Disabilities Education Act)" (CADRE, n.d.). As such, due process hearings can often serve as a bellwether for current issues in special education at the local, state, and even federal levels. Indeed, the Supreme Court of the United States will hear Perez v. Sturgis Public Schools during the 2022-2023 term, a case that centers on special education dispute resolution (Supreme Court of the United States, 2022).

Between January 2020 and March 2022, there were 20 fully adjudicated due process hearings in the state of Connecticut; over half (55%) of those addressed disputes regarding assessment and evaluation. In this analysis, I provide an overview of the legal requirements for special education assessment and evaluation as well as look for patterns in the decisions of these 11 hearings. Based on these findings, recommendations will be provided for districts interested in reducing the likelihood of future disputes.

IDEA REQUIREMENTS

To begin, I provide an overview of federal and state special education legal requirements for conducting dispute resolution and student evaluations. The Individuals with Disabilities Education Act (IDEA; 2004) is the primary federal law addressing special education. Chief among the many requirements included in IDEA is that public schools must provide a Free Appropriate Public Education (FAPE) to students with disabilities. FAPE encompasses special education and related services that are free of charge, meet the standards of the State Educational Agency (SEA) and IDEA, and delivered in a school setting under the context of an Individualized Education Plan (IEP). In order to receive special education services, a student must be identified as having a qualifying disability that requires specially designed instruction.

IDEA (2004) also addresses evaluations and eligibility determinations. If a student is determined in need of an evaluation for special education eligibility, parental consent must be obtained and a comprehensive evaluation must occur within 60 calendar days. If found eligible for special education services, a student must receive a re-evaluation no less than every three years to determine continued eligibility. As described in 20 USC § 1414b2-3, the evaluation should use a variety of technically sound assessment tools and instruments; investigate all areas of suspected disability; be administered by trained personnel in accordance with instructions provided; consider multiple data sources in determining eligibility; and consider input from parents and teachers as well as present levels of performance.

Parents of students with disabilities must be provided with a copy of their rights as well as prior written notice any time changes to the IEP are initiated (IDEA, 2004). Inevitably, families of special education students will not always agree with the recommendations made by a local education authority (LEA). Consequently, IDEA (2004) provides mechanisms for dispute resolution via Procedural Safeguards. Dispute resolution can occur through mediation or through an impartial due process hearing in which a hearing

officer considers the arguments and evidence presented by both sides, most often the parent/student and the LEA, before issuing a decision. While mediation is considered a less contentious avenue for dispute resolution (CADRE, n.d.), both parents and LEAs have the right under IDEA to elevate a dispute to a formal due process hearing. In these disputes, either the parent or the LEA makes an accusation of a procedural violation of IDEA's detailed procedural requirements or more commonly, a substantive violation that a particular child's IEP has not been reasonably calculated for that student to make appropriate progress (Lombardi & Ludlow, 2004). An example of the latter would be a parent disputing the results of a school's evaluation and subsequently requesting an independent educational evaluation by an outside evaluator at the LEA's expense. If the LEA believes that its evaluation was sound, then a due process hearing must be initiated to resolve the dispute (Musgrove, 2013). In all adjudicated due process hearings, both the parent and the LEA have the right to appeal the decision via an administrative review or a judicial review at a civil circuit court (Lombardi & Ludlow, 2004). In rare cases, the appeal process can continue all the way to the United States Supreme Court as happened most recently in Endrew F v. Douglas County School District (2017).

EVALUATIONS IN PRACTICE

As part of the Connecticut Department of Education Regulations for Children Requiring Special Education (2015), Connecticut General Statute § 10-76d addresses Conditions of Instruction, including referrals, evaluation, and eligibility. Specifically, parents can request the assessment and evaluation results three days prior to a planning and placement team (PPT) meeting in which eligibility decisions will be made. Connecticut General Statute Chapter 164 § 10-76fff (2015) echoes the federal statutes by requiring that assessments must be valid and reliable, administered by "trained and knowledgeable personnel… in accordance with any instructions provided by the producer of such tests." Once identified for special education, the assessment and evaluation results must be reflected in the Individualized Education Plan (IEP) in sections 4 and 5, known as Present Level of Academic Achievement and Functional Performance (Connecticut State Department of Education, 2021b).

34 C.F.R. §300.532(a) provides that any decision regarding placement may be subject to a request for a due process hearing. Connecticut General Statute § 10-76h (2015) fully outlines the procedures for due process hearings including hearing requests, mediation, scheduling, appointment of hearing officers, conduct and decisions. In Connecticut, the party who filed the due process "has the burden of going forward with the evidence" (Connecticut General Statute § 10-76h, 2015); however, the school district must prove the appropriateness of a given student's program or placement.

If parents disagree with an evaluation conducted by the public agency (i.e., school district), they have the right to request an independent education evaluation (IEE) under 34 C.F.R. § 300.502. If requested, the school district must either proceed with a due process complaint or provide public funding for the full cost of the IEE. The IEE is to be conducted by someone who is not affiliated with the school district. As stated in Connecticut Regulation 10- 76d-9, a parent is permitted to request an IEE per the provisions of IDEA. In Connecticut, a parent may proceed with obtaining an IEE and then request that the school district pay for it; however, the school district can refuse

if it determines that its own evaluation was appropriate or that the IEE did not meet criteria (Connecticut Department of Education, 2021a). In 2017, the Connecticut State Board of Education Task Force on the Implementation of IEEs, Observation and Related Matters was established by the Connecticut State Board of Education to review issues related to Connecticut Regulation 10-76d-9 (Connecticut State Department of Education, n.d.). While the meeting agendas for this task force are available, a final report of their recommendations is not.

SUMMARIES OF FULLY ADJUDICATED HEARINGS

Due process hearings serve as ever-current examples of special education case law, providing unique insight into the most recent issues of dispute between families and LEAs and the subsequent trends of how hearings officers are ruling. As an IEP is a highly personalized and complex document, the potential subjects of these disputes are seemingly infinite given the particular details of a given child's circumstances. Therefore, looking at recent decisions provides important understandings about current special education challenges facing both families and schools.

For the 2015-16 school year, Connecticut had the sixth highest rate of dispute resolution activity per ten thousand children out of all the states and entities served by the United States Department of Education (CADRE, 2017). Considering this fact that Connecticut is one of the most litigious states regarding special education, I examine the most recent due process hearing rulings to determine potential patterns, particularly in light of school disruptions due to the COVID-19 pandemic. In reviewing the 20 fully adjudicated due process hearings that have occurred in Connecticut since 2020, the most common topic for disputes was evaluation. In fact, over half (55%) of these 20 hearings addressed disputes about assessment and evaluation.

Subsequently, I decided that my analysis would focus on this area by limiting my review to fully adjudicated hearings in Connecticut from January 2020 to March 2022 that addressed disputes regarding evaluation and/or assessment. This time frame was initially chosen to determine potential impacts of the COVID-19 pandemic, although based on the timing of testimony and decisions, not all the cases decided overlapped with the pandemic. Non-adjudicated hearings were not included because they did not provide sufficient detail to determine the issue at hand. The purpose of the analysis is to determine any potential patterns in these recent rulings to learn more about current issues in the state and make potential recommendations for districts learning how to reduce the likelihood of future disputes. I hypothesize that these patterns will include disputes regarding pandemic-related interruptions to schooling and disputes regarding dyslexia as a result of the six state-level dyslexia laws passed in Connecticut since 2014. In sum, understanding these issues of current dispute provides opportunities for improved practice to subsequently reduce future conflict.

Table 1 presents a chronologically ordered summary of the 11 cases that meet the selection criteria of timeframe (January 2020- March 2022) and content (addressed issue of evaluation or assessment).

Table 1

Summary of selected fully adjudicated decisions

Case and year	Reasoning for dispute	Ruling
North Branford Board of Education v. Student, 2019	The family of a seventh-grade student identified under OHI-ADHD requested an IEE because they believed that the psychosocial and psychiatric evaluations conducted were not sufficiently comprehensive.	Citing the credentials of the evaluator and the content of the evaluation, the decision is ruled in favor of the school district.
Greenwich Board of Education v. Student, 2020	After the initial evaluation conducted of a fourth-grade student found her to be ineligible for special education services but eligible for services under Section 504, the parents disputed some of the evaluation's conclusions and interpretations and requested an IEE.	Citing the credentials of the evaluator and the content of the evaluation, the hearing officer ruled that the evaluation was appropriate and that the parents were not entitled to an IEE.
Student v. Meriden Board of Education, 2020	Both the parent and the school district filed separate due process hearings after the parent had refused consent for an initial evaluation. The parent was self- represented and did not communicate prior to the hearing or attend the hearing.	The ruling was in favor of the district proceeding with an initial evaluation without parental consent.
Watertown Board of Education v. Student, 2020	A sixth-grade student, classified under the disability of autism with a secondary disability of ADHD, was previously unable to be evaluated via a formal academic achievement assessment. However, the current special education teacher felt that the student was now able to engage in the assessment process with proper supports and accommodations and thus an academic achievement evaluation was conducted. The student's mother requested an IEE in reading based on lack of progress and requested that the special education teacher's testimony be found not credible.	Both parent requests were denied in the ruling.

Greenwich Board of Education v. Student, 2021	A triennial evaluation conducted of a fifth-grade student classified under SLD- dyslexia found the student to be no longer eligible for special education services but recommended services under Section 504 instead. Subsequently, the mother disagreed and requested an IEE.	The hearing officer ruled in favor of the mother, citing several reasons why the evaluation was not sufficient, including its failure to address the student's recent diagnosis ~~with~~ of Ehlers-Danos Syndrome.
East Hartford Board of Education and Student, 2021	A student requested an IEE and then became an adult who graduated from high school and subsequently was no longer eligible for special education services.	Case was dismissed because student was no longer eligible for special education services.
Enfield Board of Education v. Student, 2021	The mother of a seventh-grade student receiving services under a Section 504 Plan for ADHD made a PPT referral, and the student was found eligible for special education services under OHI-ADHD. Subsequently, the mother shared private evaluation results along with concerns about student performance, and the district conducted additional evaluations. The mother's primary disputes centered on the request for specific services as opposed to disputing the evaluation results.	Citing the credentials of the evaluators and the content of the evaluations, the hearing officer ruled that the evaluation was appropriate and that the mother was not entitled to an IEE.
Student v. Cheshire Board of Education, 2021	This case had multiple claims and issues, one of which was a request for an IEE. The student was homeschooled until sixth grade when she then attended a parochial school. At that time, her mother made a special education referral, and the district conducted an evaluation. She was found eligible under the category of specific learning disability but did not receive services in the parochial school. She subsequently returned to the public school and additional evaluations were conducted. After a period of average performance and additional evaluation	The hearing officer ruled in favor of the IEE request.

	results, the student was exited from special education services and found eligible for a 504 Plan. In the following year, the mother placed another referral to special education services, and in disagreement with the results of that evaluation, she requested an IEE. This was denied so the parents hired a private evaluator.	
Vernon Board of Education and Student, 2021	The parents disputed the appropriateness of the evaluations conducted for their preschool age son with autism. They did not submit any evidence to support their claim.	The hearing officer found that the assessments used were appropriate and technically sound and dismissed the request for an IEE.
Student v. Monroe Board of Education, 2021	A seventh-grade student's triennial re-evaluation were planned for February 2020, and subsequently, schools were closed for the pandemic. The student was identified under the category of SLD-dyslexia and attended a private school per a recent PPT decision. The speech/language evaluation was conducted prior to school closure. The neuropsychological and central auditory processing evaluations were also conducted successfully and after review of the assessments, adjustments were made to the IEP.	The dispute of the hearing had many issues, one of which was around the denial of FAPE because of failure to evaluate in a timely manner. This claim was denied.
Student v. Trumbull Board of Education, 2022	A middle school student had been unilaterally placed at an unapproved special education school. The parent requested that the home district contact the private placement for data and records in preparation for the annual review. The team did not review this information. The PPT recommended a diagnostic placement at the student's home school and denied the parent's "stay put" request for the diagnostic placement to be done at his current school. The evaluations were not completed during the 40-day diagnostic placement. The student began declining academically, emotionally, and behaviorally in the home school placement.	The hearing officer ruled on multiple points including the failure of the district to timely complete evaluations.

ANALYSIS

Next, I provide evidence of patterns that appeared in these eleven rulings. Specifically, there were clusters of cases addressing the following topics: evaluator credibility when reviewing IEE disputes, movement between services received under Section 504 and IDEA, dismissal of complaints due to case circumstances, and economic-related geographic distribution of cases.

EVALUATOR CREDIBILITY WHEN REVIEWING IEE DISPUTES

As previously stated, 55% of all the fully adjudicated cases held since January 2020 involved disputes regarding evaluation and assessments. Of that subset, about half of the cases involved an IEE request. For these cases, much of the findings of relevant fact relied on the specific credentials and experience of the evaluator as determined by 20 USC § 1414b2-3 which requires that special education evaluations use a variety of technically sound assessment tools and instruments that are administered by trained personnel in accordance with instructions provided. These same criteria are used by school districts in deeming the credibility of an IEE obtained by a parent (Zirkel, 2009). In fact, the regulations outline the steps for IEE disputes as such: the parent disagrees with an evaluation the school has conducted, the school district files without unnecessary delay and subsequently shows that its evaluation was appropriate and that the IEE is not appropriate (34 C.F.R. § 300.148(b)-(e), 2008). For these Connecticut cases, the hearing officers often relied on the background and training of the evaluators in determining that appropriateness of the school evaluations. In other words, in following federal guidance issued from the Office of Special Education Programs (OSEP; 1995), if the evaluator was knowledgeable and trained, then their evaluation was deemed appropriate. Consequently, the hearing officers in North Branford Board of Education v. Student (State of Connecticut Department of Education, 2019), Greenwich Board of Education v. Student (State of Connecticut Department of Education, 2020), Enfield Board of Education v. Student (State of Connecticut Department of Education, 2021), and Vernon Board of Education and Student (State of Connecticut Department of Education, 2021) deemed the evaluations in dispute to be appropriate based on the credentials of the evaluators and the content of the evaluations.

MOVEMENT BETWEEN SERVICES RECEIVED UNDER SECTION 504 AND IDEA

Over a third of the cases analyzed involved movement from ineligibility in special education to eligibility for services under Section 504 of the Disability Act and vice versa. Services received under Section 504 are typically accommodations provided in the general education setting as opposed to the specially designed instructed provided through special education under IDEA. Specifically, the disputes in Greenwich Board of Education v. Student (State of Connecticut Department of Education, 2021), Enfield Board of Education v. Student (State of Connecticut Department of Education, 2021), and Student v. Cheshire Board of Education (State of Connecticut Department of Education, 2021) all involved movement between services provided under IDEA (i.e., special education) and Section 504. These decisions were based partially on evaluation results in addition to student performance. Section 504 and IDEA both provide mechanisms

for serving students with disabilities; however, not all students who meet the definition of disability under Section 504 are eligible for services under IDEA (Yell, 2016). Indeed, IDEA services encompass specially designed instruction which is considered more robust and comprehensive than the accommodations provided by Section 504. Thus, it is not surprising that parents who are looking for additional supports for their children would seek eligibility under IDEA instead of Section 504 and subsequently use evaluation results to support their argument.

DISMISSAL OF COMPLAINTS DUE TO CASE CIRCUMSTANCES

Two of the cases had complaints dismissed outright due to lack of adherence to IDEA requirements (Gilsbach, 2015). The dispute in East Hartford Board of Education and Student, (State of Connecticut Department of Education, 2021) was considered under an improper party provision because the student in question had aged out of IDEA services. The argument in Vernon Board of Education and Student (State of Connecticut Department of Education, 2021) was dismissed due to insufficient complaint as the parents failed to submit any evidentiary support.

ECONOMIC-RELATED GEOGRAPHIC DISTRIBUTION OF CASES

A final pattern that emerged was related to geographic wealth distribution. Out of the eight counties in Connecticut, Fairfield County has the highest median household income (Connecticut Department of Economic and Community Development, 2022), and over one-third of the cases originated there. Securing legal representation in a due process hearing typically requires financial resources, and families with limited means often cannot afford to hire private counsel. However, parents represented by attorneys have more favorable outcomes than the pro se alternative (Hoagland-Hanson, 2014). Further, parents who have the financial means to secure an IEE also have more favorable outcomes in due process hearings (Hoagland-Hanson, 2014). Considering that the cost for neuropsychological evaluations can start at $5,000 (Carr, 2022), clearly families with ample income or access to wealth are better suited to obtain outside evaluations. Parental networking and knowledge-sharing may also lead to increased special education disputes; interestingly, two-thirds of Connecticut's Special Education Parent Teacher Associations are located in Fairfield County (Connecticut PTA, 2022).

COVID-19 PANDEMIC

The COVID-19 pandemic shut down Connecticut schools starting in March 2020, and thus I expected it to be a frequent factor in the cases analyzed. In reality, the pandemic disruption to schooling was only a factor in Student v. Monroe Board of Education (State of Connecticut Department of Education, 2021). In this case, the parent argument of failure to timely evaluate was denied since the school district had already conducted part of the evaluations prior to the shutdown and was able to successfully complete the remaining portions afterward. However, future hearings may more greatly reflect disputes over

pandemic-related education disruptions, including compensatory services and education (Zirkel, 2021). A study of future rulings would likely show a greater pandemic impact.

CONNECTICUT DYSLEXIA LEGISLATION

Since 2014, Connecticut has passed a series of six laws related to dyslexia. Although dyslexia was mentioned in a few of the cases, I did not discern any influence of the state legislation in these specific decisions. That said, recent research has shown a dramatic national increase in court cases related to dyslexia-specific instruction (Sayeski & Zirkel, 2021) and so it remains to be seen whether this trend will emerge in Connecticut.

DISCUSSION

In this study, I aim to discern topical patterns in recent due process hearings in Connecticut; one goal in doing so is to suggest improvements to practice that may prevent future conflicts. Given the potential breadth of topics that these disputes can address, frequent analysis of recent rulings is essential to determine the issues currently impacting the field. Indeed, I find that over half of the cases since 2020 addressed issues around assessment.

Given the proportion of cases in this analysis that addressed disputes involving evaluation, LEAs should prioritize continued professional development on assessment for their special education staff. Indeed, teacher preparation programs and school districts should ensure that both future and current special education personnel are well-trained in administering and interpreting assessments. On-going professional development will also be important to stay abreast of instrument updates as test publishers often release revised editions.

Further, both parents and LEAs need to have a clear understanding of the eligibility criteria for special education services under IDEA and Section 504. To provide parents greater clarity regarding specific eligibility requirements under IDEA versus Section 504, LEAs may consider sharing information with parents through print or online resources or in-person workshops. Critically, districts should consider offering informational sessions for families to delineate these service provisions. Doing such would be a proactive step towards reducing confusion and potential future disputes.

Districts should increase parent communication and collaboration in an effort to maximize parent input and involvement prior to disputes (Otte, 2022). In one national analysis, Connecticut was among the five states with the highest number of due process hearings per capita (Mueller & Carranza, 2011). Cope-Kasten (2013) described the due process hearing system as unfair and advocates for the use of mediation instead. Connecticut stakeholders may consider using the resources and support provided by organizations such as the Center for Appropriate Dispute Resolution in Special Education (CADRE) to strengthen special education services, improve parent-school district relations and ultimately reduce the number of special education due process hearings.

Because due process hearing decisions are an ever-occurring source of case law, continued analyses such as these are helpful to discern current issues and patterns. Indeed, a future analysis using a wider time period may detect a greater impact of pandemic-related school interruptions and would also allow for a greater understanding of potential equity concerns than this limited analysis.

CONCLUSION

Disputes between LEAs and parents in special education will always exist. But the field can work towards reducing the need for extreme conflict resolution measures such as due process hearings. Indeed, careful examinations of due process rulings provides a unique opportunity to understand current issues in order to improve future practice – and hopefully decrease future disputes.

REFERENCES

34 C.F.R. § 300.148(b)-(e) (2008).

CADRE. (2017). IDEA dispute resolution data summary for Connecticut: 2004-05 to 2015-16. https://www.cadreworks.org/sites/default/files/resources/2015-16%20DR%20Data%20Summary%20-%20Connecticut%20rev.pdf

Carr, S. (2022, March 1). Want your child to receive better reading help in public school? It might cost $7,500. Hechinger Report. https://hechingerreport.org/an-independent-neuropsych-evaluation-is-critical-for-getting-access-to-special-education-services/.

The Center for Appropriate Dispute Resolution in Special Education, CADRE continuum. https://www.cadreworks.org/

Chapter 164 Educational Opportunities, Connecticut General Statute Chapter § 10-76fff. (2015). https://ctser.specialedreference.com/?s=1034699

Children Requiring Special Education, Connecticut General Statute § 10-76d. (2015). https://portal.ct.gov/-/media/SDE/Special-Education/Regulations-of-Connecticut-State-Agencies-Department-of-Education-Children-Requiring-Special-Educati.pdf?la=en

Connecticut Department of Economic and Community Development.,(2022). Housing and income data. https://portal.ct.gov/DECD/Content/About_DECD/Research-and-Publications/01_Access-Research/Exports-and-Housing-and-Income-Data

Connecticut Department of Education. (2021a). A parent's guide to special education in Connecticut. Bureau of Special Education.

Connecticut Department of Education. (2021b). IEP manual and forms. Bureau of Special Education. https://portal.ct.gov/-/media/SDE/Special-Education/IEP-Manual-REVISED-December-2021-PDF-ver.pdf

Connecticut PTA.(2022). Special Education Parent Teacher Association. https://www.ctpta.org/special- needs

Connecticut State Department of Education (n.d.). Independent educational evaluations (IEEs) Task Force. https://portal.ct.gov/SDE/Special-Education/IEETaskForce/Independent- Educational-Evaluations-IEEs-Task-Force

Cope-Kasten, C. (2013). Bidding (fair) well to due process: The need for a fairer final stage in special education dispute resolution. Journal of Law & Education, 42, 501-540. https://digitalcommons.macalester.edu/poli_honors/30

East Hartford Board of Education and Student, 21-0416. (State of Connecticut Department of Education, 2021).

Endrew F., a Minor, by and Through His Parents and Next Friends, Joseph F. et al. v. Douglas County School District RE–1, 64 IDELR 38, (D., Co. 2014), 580 U.S. _____ (2017).

Enfield Board of Education v. Student, 21-0380, (State of Connecticut Department of Education, 2021).

Gilsbach, T. E. (2015). Special education due process hearing requests under IDEA: A hearing should not always be required. Brigham Young University Education & Law Journal, (1) 187–201.

Greenwich Board of Education v. Student, 20-0314, (State of Connecticut Department of Education, 2020).

Greenwich Board of Education v. Student, 21-0212, (State of Connecticut Department of Education, 2021).

Hoagland-Hanson, K. (2014). Getting their due (process): Parents and lawyers in special education due process hearings in Pennsylvania. University of Pennsylvania. Law Review, 163, 1805–1842.

Individuals with Disabilities Education Act, 20 U.S.C. § 1400 (2004).

Lombardi, T. P., & Ludlow, B. L. (2004). A short guide to special education due process. Phi Delta Kappa Fastbacks, 523, 7–48.

Miguel Luna Perez, Petitioner v. Sturgis Public Schools, et al., No. 21-887. https://www.supremecourt.gov/DocketPDF/21/21887/246735/20221116172456338_certificate.%20motion.%2021-887.pdf

Musgrove, M. (2013, July 23). Dispute resolution procedures under Part B of the Individuals with Disabilities Education Act (Part B) [Memorandum]. United States Department of Education, Office of Special Education and Rehabilitative Services. https://sites.ed.gov/idea/files/policy_speced_guid_idea_memosdcltrs_acccombinedosersdisputeresolutionqafinalmemo-7-23-13.pdf

Mueller, T. G., & Carranza, F. (2011). An examination of special education due process hearings. Journal of Disability Policy Studies, 22(3), 131–139.

North Branford Board of Education v. Student, 20-0021, (State of Connecticut Department of Education, 2019).

OSEP Policy Letter, 22IDELR 563 (OSEP 1995).

Otte, E. (2022, April 14). Independent Review Supports Parent Complaints About Special Education in Colchester. The Connecticut Examiner, https://ctexaminer.com/2022/04/14/independent-review-supports-parent-complaints- about-special-education-in-colchester/

Sayeski, K. L., & Zirkel, P. A. (2021). Orton-Gillingham and the IDEA: analysis of the frequency and outcomes of case law. Annals of Dyslexia, 71(3), 483–500.

Student v. Cheshire Board of Education, 21-0334, (State of Connecticut Department of Education, 2021).

Student v. Meriden Board of Education, 20-0186 and 20-0233, (State of Connecticut Department of Education, 2020).

Student v. Monroe Board of Education, 21-0208, (State of Connecticut Department of Education, 2021).

Student v. Trumbull Board of Education, 21-0410, (State of Connecticut Department of Education, 2022).

Vernon Board of Education and Student, 22-0041, (State of Connecticut Department of Education, 2021).

Watertown Board of Education v. Student, 21-0181, (State of Connecticut Department of Education, 2020).

Yell, M. L. (2016). The law and special education. Merrill/Prentice-Hall, Inc., 200 Old Tappan Road, Old Tappan, NJ 07675.

Zirkel, P. A. (2009). Independent educational evaluations at district expense under the Individuals with Disabilities Education Act. Journal of Law and Education, 38, 223–244.

Zirkel. (2021). COVID-19 Confusion: Compensatory Services and Compensatory Education. Southern California Review of Law and Social Justice, 3, 391–414.

GAME-BASED ETHICAL INSTRUCTION IN UNDERGRADUATE ENGINEERING

COLTER MOOS
UNIVERSITY OF CONNECTICUT NEAG SCHOOL OF EDUCATION

LAUREN P. DOUGHER
UNIVERSITY OF CONNECTICUT NEAG SCHOOL OF EDUCATION

LANDON BASSETT
UNIVERSITY OF CONNECTICUT SCHOOL OF ENGINEERING

DR. MICHAEL F. YOUNG
UNIVERSITY OF CONNECTICUT NEAG SCHOOL OF EDUCATION

DR. DANIEL BURKEY
UNIVERSITY OF CONNECTICUT SCHOOL OF ENGINEERING

ANALYSIS

Ethics is a topic in undergraduate engineering curricula believed to frequently fall short in terms of professional application. In this paper we examine game-based learning approaches to engineering ethics in first year undergraduate students evaluated through the Engineering Ethical Reasoning Instrument (EERI). The EERI builds on the standard engineering ethics instrument, the DIT-2, and is focused on moral reasoning, while the game-based learning is based on situated decision making. Results showed no statistically significant interaction between participants who had game-based learning and lecture-based instruction over the time of the intervention, though ethical reasoning improved across all treatments. One conclusion we draw is that the lack of statistically significant interaction suggests that the EERI may not be the correct measure for these engineering ethics games, and further studies should research the creation of new instruments to incorporate this type of ethics instruction.

Keywords: Engineering, ethics, game-based learning, situated decision making, moral reasoning, assessment

GAME-BASED ETHICAL INSTRUCTION IN UNDERGRADUATE ENGINEERING

When visiting a typical undergraduate engineering classroom, one may expect to find a professor lecturing at the front of the room about formulae, equations, and theories in front of a lecture hall full of students. While these concepts are an important part of future engineers' education, so too is ethics, especially when you consider decisions that many engineers will be faced with when working in the field. In the United States, the Accreditation Board for Engineering and Technology (2022) guidelines for 2022-2023 require that students have: "an ability to recognize ethical and professional responsibilities in engineering situations and make informed judgments, which must consider the impact of engineering solutions in global, economic, environmental, and societal contexts" (Criterion 3: Student outcomes). In other words, engineers not only need to understand how to create within their field, but they must also be able to consider the implications of their work.

MORAL REASONING AND SITUATED DECISION MAKING

In this paper, we focus on two key terms in relation to ethics: moral reasoning and situated decision making. Moral reasoning is based on how people think about what is right and wrong, such as whether they prioritize self-interest, or look toward justice for all people (Kohlberg & Hersh, 1977). We define situated decision making as grounded in how contextual factors, such as authentic details, and people's past experiences in dealing with ethical problems, influence their current decision making (Bagdasarov et al., 2013). People respond and act differently when immersed in a situation as opposed to when presented with more "disengaged" scenarios. In the context of this study, we predict that an engineer will react and respond differently to ethical dilemmas at their workplace than to a predetermined scenario in compliance training or a classroom. Their moral reasoning, or how they think about what is right and wrong, may be the same in both contexts, but the decision they make may differ as a result of the aforementioned contextual factors.

ETHICS AND ENGINEERING

Ethics can be an obscure topic to teach and difficult for many students to learn. As a result, some undergraduate engineering programs have employed game-based learning toward enhancing student interest and achievement. While game-based learning can be defined in a variety of ways, most definitions agree that game-based learning is a type of gameplay with defined learning outcomes (Shaffer et al., 2005, as cited in Plass et al., 2015). Studies on game-based learning have yielded varying results, including significant improvement in learning outcomes for participants who underwent the learning intervention (Chee & Tan, 2012; Franciosi, 2017) as well as non-significant differences between groups' learning outcomes, specifically in content knowledge and critical thinking skills (Cicchino, 2015). Despite these mixed results, all of these authors see the utility of game-based learning and call on others in the field to continue these efforts and associated research.

In doing so, we follow the recommendations of Marklund and Taylor (2016) who call

on researchers to ensure that the game-based learning being utilized aligns with best practices and is implemented by those with expertise in this technique. Specifically, the professor was trained to utilize a variety of targeted games they constructed collaboratively with co-authors of this study who have expertise in game-based learning. As such, we could place greater focus on the learning outcomes from the game-based learning, rather than their implementation. Additionally, this approach allowed us to view game-based learning in the niche, yet expansive, topic of engineering ethics for insight into this game-based approach. If successful, this approach would have great implications for engineering schools and, in turn, engineering practice across the nation.

Our study was guided by the following research question: How did various arrangements of three engineering ethics games and modalities of play across three years differ from lecture-based instruction in their effect on students' ethics?

With a better understanding of the key terms used in this study and our research question, we next review some broader topics of interest, including engineering ethics and game-based learning. Next, our methods explore the participants, the three games used in the intervention, the lectures, as well as the outcome tool, the Engineering Ethical Response Inventory, or EERI. Using the EERI output, we will discuss the analysis and results of the two-way mixed-design ANOVAs across each year. We conclude with a discussion of our results and final conclusions regarding the role and importance of assessment in game-based ethical instruction.

LITERATURE REVIEW

To set the context for this study, we begin with a broad discussion of ethics education. We then focus on what ethics looks like in engineering education to better understand how lecture and game-based instruction are leveraged in this study. Next, we will consider "games" throughout the literature and an examination of what other studies have found when using games in the engineering ethics space.

In Western culture, many undergraduate universities and colleges offer philosophy courses focused on teaching philosophical ethics. These courses provide a theoretical background of topics such as virtue ethics, feminist ethics, and consequentialist ethics, among others (Paulson & Kretz, 2018). As Callahan (1980) explains, such courses should be included in undergraduate curriculum because "morality is part of any reflective personal life, and because ethical perspectives and specific moral rules are part of any cultural and civic life… ethical problems are inescapable…it is difficult to think of any aspect of personal or public life that will not be determined or conditioned by moral values" (p. 62). For students, such courses, at the very least, can provide clear evidence that there are ethical problems in all aspects of life, and how one understands and responds to these problems can make a difference in the lives of others, both positively and negatively (Callahan, 1980).

Ethics education at this stage is appropriate because ethical problems can happen at any time in a person's life and are also a part of jobs and professions (Callahan, 1980). For this reason, not only do we see introductory philosophy/ethics education as important in

undergraduate studies, but also as an integral part of many different disciplines' education. Some examples of these fields include counseling (Lamb, 1991), psychology (Plante & Pistoresi, 2017), and medicine (Wong et al., 2022). Engineering is yet another example of a field heavily intertwined with ethics, especially with the reach of its many branches including chemical, civil, electrical, industrial, mechanical, and so on. Considering a slice of ethics education, via the engineering discipline, may help improve ethics and engineering education broadly.

History shows us what can happen when ethical problems are encountered but poorly attended to, such as with the case Challenger disaster in 1986 where seven space shuttle members died due to known flaws in the spacecraft ("Space Shuttle Challenger Disaster", n.d.). But in engineering there is still a "disconnect between the ethics education of contemporary engineering students…and the ethics realities of contemporary engineering practice" (McGinn, 2018, p. 3). Some students and educators may even consider ethics as a simple box to check off rather than an essential part of the engineering design process (Lloyd & van de Poel, 2008).

To help curb these deficiencies some institutions have turned to game-based learning for their engineering ethics education, with Barab and Dede (2007) noting that, over the past few decades, game-based learning methodologies emerged as a type of curricula in science education. Lau et al. (2012) provided one example of game implementation in the engineering classroom. In this study, students were tasked with creating a design using colored paper in accordance with a list of constraints. Each group of students was set up with roles and could not talk to one another about their constraints. They were then given a "briefing" shortly before the end of the session that posed an ethical dilemma. The different roles within each team were then pressured in different ways to either move forward with the product or halt it due to concerns regarding its ability to function properly. The scenario was created in such a way as to mimic the Challenger disaster and the decisions/ lack of communication that led to it (Lau et al., 2012). Between the implementation of the game and viewing of the real Challenger disaster video, students' opinions on the importance of ethical statements in engineering practice increased for all but one statement (Lau et al., 2012).

The Challenger disaster case study is also often used in traditional engineering ethics instruction during lectures. Some methods used in this teaching format include providing students with "do's and don'ts" lists related to ethical engineering practice, having students use basic scenarios to apply said principles, and the use of case studies where students must analyze and provide an approach for resolving the case (Alfred & Chung, 2012). The latter is the most effective approach used in classroom lectures (Whitbeck, 1996, as cited in Alfred & Chung, 2012), but Drew (2011) described how engineering courses taught through lectures often fail to interest students, leading to reduced engagement and shallow learning. Therefore, other approaches, like games, have been introduced to increase the usefulness and engagement of engineering ethics instruction.

The example above describes Lau et al.'s (2012) study as implementing a "game," but there are additional terms with more distinct meanings used in education. Games, gamification, and game-based learning are terms that are frequently used interchangeably in educational settings. While there is some disagreement on exact definitions, Plass

(2017) differentiated between them. Specifically, game-based learning is distinct in that the original learning task is transformed into a game with a design grounded in discipline-specific applications (Plass, 2017). Bodnar et al. (2016) went into further detail, defining gamification as "the application of game design elements to nongame scenarios" (p.148). A common and simple form of gamification is the use of points, badges, and leaderboards where participants earn points for completing tasks (Bodnar et al., 2016). Further, they continued to explain that game-based learning has many benefits including that it provides immediate feedback, informs participants they are making progress, and motivates them (Bodnar et al., 2016). Bodnar et al.'s (2016) systematic review of the games meant to teach undergraduate engineering students, provided support for the conclusion that the implementation of games in undergraduate engineering classrooms improved student learning and attitudes.

For this paper, we define game-based learning as it aligns with Plass (2017) and Shaffer et al. (2005) in that an original learning task has been transformed into a game with learning outcomes. In the original lecture-based instruction scenario at the study site, students would engage in discussions of historic engineering and philosophical ethics problems, with the learning task being that they would contribute to the discussion about, and listen to, the problems. For the larger National Science Foundation (NSF) study from which this paper is derived, three games were created to transform the role of the student in different ways, as an individual (1) voting on an ethical response to a problem, (2) choosing a potentially ethically dubious card option due to the nature of the situation, or (3) ordering other engineers' views on ethical responses to issues in the field. All these games stemmed from the same situated engineering ethics problems of the lecture, but had the player take on a role with richer context and details, more agency, and from a different perspective on the ethical problem, leading to the varied learning outcomes from each game.

Slota and Young (2014) described the importance of implementing game-based learning beyond simply taking a lesson and adding game elements like rewards or points. Rather, incorporating principles where games can change to sustain player interest, have game narratives, and include opportunities for players to explore, expand, or build within the game, can result in a more effective game-based learning environment (Slota & Young, 2014). These game elements lead to more richly situated game-based learning environments, and the need for similarly richly situated assessments of engineering ethics. It is with this idea of richly detailed, and highly contextual, games in mind that we circle back to the current study, where we aim to examine whether there are differences between such engineering ethics games and more traditional scenario-based lectures.

METHODS

As part of a larger NSF-funded study on the use of researcher-created engineering ethics games on situated decision making, data was collected over the course of three years from an undergraduate Introduction to Engineering course at a public northeastern state research university. Throughout this study, various games were employed through different modalities across each of the three years of the study. The games played by the experimental groups included Mars: An Ethical Expedition (MAEE), Cards Against

Engineering Ethics (CAEE), and Toxic Workplaces (TW), with more detailed descriptions of each game below. Some key commonalities during the three-year period: the course always ran during spring semester and had no other connection to ethics, every year the Tuesday section had game-based learning and the Thursday section had lecture-based classroom instruction. These lectures occurred during the two full days set aside for CAEE, and later TW, in the middle of the semester. The lecture-based group had no substitute for MAEE, only receiving two lectures; first on the Challenger disaster discussed earlier, and second on the classic philosophical ethics trolley problem. In this problem, a bystander observes that a trolley will move forward and kill 5 people unless the bystander switches the track away, the result being that one person on the other track is killed. During each year of this study, participants took the Engineering Ethical Reasoning Instrument (EERI), our outcome measure, pre- and post- intervention.

Outside of these commonalities, the implementation differed across the years, both in terms of which games students played and via which modality (see Table 1 for further explanation). In 2019, MAEE and CAEE were run in person, with CAEE played for two weeks in the middle of the semester, while TW was not played. For 2020, MAEE was run half in person for the first six weeks, and half online for the last six weeks due to the COVID-19 pandemic. CAEE was not played in 2020, instead, for the first week, the game-based group received the same lecture as the lecture-based group. For what would be the second CAEE session, the game-based group instead discussed the prompts for TW used to generate ranked ethical choices for the responses to case studies in the game as will be discussed later. All three games were played online in 2021, with the COVID-19 pandemic still interfering with in person instruction.

PARTICIPANTS

Each year participants were students enrolled in a shared Introduction to Engineering class at a public northeastern state research university in the spring semester of their first year, a course required for all pursuing a bachelor's in engineering. The following guidelines were used as exclusion criteria for analysis: did not provide consent, did not answer all the questions for either the pre or post EERI, failure to complete either a pre or post EERI, or if a student reported they switched conditions (game vs. lecture) between the pre and post. Additionally, anyone who completed the EERI under 10 minutes was excluded as they were believed to have not taken enough time to review and respond to questions based on the EERI length. Last, if a participant submitted multiple times only their first complete submission was used and the rest excluded.

As seen in Table 2, after these factors for exclusion were implemented, there were roughly similar sample sizes across all conditions except the lecture-based instruction in 2021. We believe this difference was not due to the elimination criterion above, as the individuals removed were roughly equal across the classes, but instead due to participant self-selection. As the lecture-based instruction class was always held on Thursday, and the class of 2021 was fully online due to the COVID-19 pandemic, we believe that fewer people enrolled in that class for unknown personal reasons, resulting in an uneven sample size which we account for later.

Table 2 highlights the percentages of students across gender identity and race/ethnicity. Across all years, individuals who identified as male were predominant within each condition, while those identifying as Caucasian followed by Asian American or Pacific Islander were the highest and second-highest percentages across every condition, respectively, aligning with the larger demographics of engineering undergraduates (American Society for Engineering Education, 2022). From this, we felt comfortable comparing the participants across conditions within their respective years as there were no statistically significant differences.

GAMES

As part of the aforementioned NSF funded study, the research team designed the following three engineering ethics games with multiple rounds of playtesting and feedback before incorporation into the study, to teach students about engineering ethics through gameplay. Mars: An Ethical Expedition and Cards Against Engineering Ethics were created by some of the authors on this paper, while Toxic Workplaces was created by researchers at another institution working under this NSF study and supplemented and tested by some authors on this paper.

MARS: AN ETHICAL EXPEDITION

In this multiplayer choose-your-own-adventure game, the students collaborated as a class to make ethical decisions to survive on a Mars Colony. At the beginning of the semester the students were given a narrative introduction to the overall game, and then each week consisted of a specific ethical scenario. For each weekly scenario, the students would have a choice of two or three decisions they could choose from in handling the ethical situation. The students then voted on which decision they wanted to make, and the choice with the most votes was taken. Based upon their choice the class received a different situation prompt the next week, continuing for 12 weeks with branching events based on their choices. At the end of the 12 weeks, the students would learn whether they were successful in surviving/saving the Mars Colony.

The learning outcome from MAEE is for students to be able to take various perspectives in making situated ethical engineering decisions. Through playing the game they needed to vote and make an ethical decision. The goal is for them to make this decision with consideration of the impact their decisions would have on the Mars Colony from various characters' perspectives and determine not only whether those impacts and decisions are ethical, but which they would advocate for through their vote.

When played in person, the professor would stand at the front of the class and share their screen with the narrative introduction and weekly scenario as appropriate, while the students would use polling software on an electronic device to vote on the decision they personally would make. When played online the professor would similarly share their screen in the video call and collect votes from the polling software using the same method. In both cases, after the votes were collected, the professor would share their screen with the results of the vote and then continue with the rest of the lesson for that week unrelated to the gameplay.

Implementation of Games and Lectures Across Years

Table 1

Implementation of Games and Lectures Across Years

	2019		2020		2021	
	Games	Lecture	Games	Lecture	Games	Lecture
Lecture Modality	N/A	In Person	Online	Online	N/A	Online
Number of Lectures	0	2	1	2	0	2
MAEE Modality	In Person	N/A	Hybrid	N/A	Online	N/A
CAEE Modality	In Person	N/A	N/A	N/A	Online	N/A
TW Modality	N/A	N/A	Prompt Online	N/A	Online	N/A

Note. For TW Modality in 2020 "Prompt - Online" is used because while the students did not play Toxic Workplaces, they did discuss the prompts online to generate the ranked ethical choices, which differs from the Lecture condition, and the play in 2021.

CARDS AGAINST ENGINEERING ETHICS

This game is based on the similar party games Apples to Apples and Cards Against Humanity, in which players match a card with a word or phrase on it in their hand to a shared prompt. If a player's card is chosen as the best fit by the judge of that round, the player earns points. In this version, cards are focused on situations that can occur as an engineer, with some prompts and phrase cards based on real-world engineering ethics situations. Students played this game in groups of four, with the judge rotating between the players throughout their play. After 30 to 45 minutes the groups stopped playing and the whole class discussed some of the real-world ethical situations on the cards and considered their goals in playing the game. After this discussion, gameplay resumed for another 30 to 45 minutes before ending with a debrief on the decisions players made and a discussion of what they would do in similar real-life circumstances.

The learning outcome from CAEE is for players to recognize how the context of a situation modifies their situated decision making. Through both playing combinations of cards that are unethical and discussing why the context of playing the game allowed them to play such unethical combinations, students would be able to better recognize the importance of context in modifying their situated decision making. In this case, the learning came not from the playing of CAEE directly, but from the discussions and analysis of how the students were playing the game.

When played in person, the players would use physical playing cards and gather in groups with those sitting near them. When played online, the players would be randomly grouped together with others in their class through video conference breakout groups to play a digital version of the game. The virtual card deck contains nearly all the cards from the physical version, with some omissions due to copyright purposes. Therefore, the largest changes between the in person and virtual implementations were the use of slightly fewer cards in the digital version, the randomization of groups, and using video conferencing to communicate in which not all players would turn on their cameras or

DEMOGRAPHICS BREAKDOWN BY YEAR AND CONDITION

Table 2

Demographics Breakdown by Year and Condition

Demographics	2019		2020		2021	
	Games	Lecture	Games	Lecture	Games	Lecture
Sample Size	128	138	125	120	108	52
			Gender Identity			
% Male	65.6	58.7	71.2	65	53.7	53.8
% Female	34.4	40.6	28.8	35	45.4	46.2
% Other	0	0.8	0	0	0	0
% Prefer Not To Answer	0	0	0	0	0.9	0
			Race/Ethnicity			
% African American	3.9	4.3	3.2	3.3	6.5	3.8
% AAPI	10.2	18.8	11.2	15	16.7	11.5
% Hispanic	3.9	2.2	10.4	9.2	5.6	3.8
% Native American	0	0	0	0	0	0
% Caucasian	71.9	62.3	64	64.2	57.4	61.5
% Other	0.8	1.4	1.6	0.8	0	0
% Prefer Not to Answer	0	0	0	0	1.9	7.7
% Mixed	9.4	10.9	9.6	7.5	12	11.5

Note. The percentages in this table are all relative to their condition within their year. For example, the 2019 Games column has a sample size of 128 students, of which 65.6 identify as male, and 34.4 identify as female. This same interpretation applies to race/ethnicity.

keep their microphones open. While it is possible that those seated near each other are strangers, there is a higher chance that players would know at least one or two others in their group of four, while in the online random groups it is less likely that the players were familiar with each other.

TOXIC WORKPLACES

Toxic Workplaces uses a Family Feud style of play to discuss situated decision making. In this version however, instead of students blindly guessing what responses others have made, they are presented with responses to an ethical case study and have to rank order those responses based on their belief of how many people would choose that course of action. The response prompts were created by students from another university who voted both on how ethical each response was for each prompt, and how many of them would choose that course of action in the case study. This survey data was collected in 2020 and used to create the correct rank order for play in 2021. The gameplay occurred instead of the second week of CAEE.

The learning outcome from Toxic Workplaces is for players to be able to recognize how others within the engineering community perceive engineering ethics. This is achieved through not only the students' ranked ordering, but also through their scores which would be higher the more accurately their order aligned with the rank order of their peers from the previous year. Through this, they were directly able to get feedback on their alignment with how ethical others within the engineering community perceive various choices.

As this game was only played online, there was one implementation method used which utilized video conferencing and breakout groups, and Google Slides as the basis for playing the game. As mentioned before, after reviewing the case study players would discuss amongst themselves how to rank order the responses by dragging the response slides into the order from what they thought was most popular to least. After the responses were ordered, everyone in the group would take a screenshot of the order and record their choices for the end of the game. Groups would then be formed randomly two more times for a total of three rounds of play before scoring would occur. For every response in the correct rank order, the group would score a point. Individuals won if their teams scored the most points correctly throughout the three rounds.

LECTURES

Students in the lecture-based course received two lectures on ethics, one each over the course of two classes. In the first lecture, the professor would discuss the Challenger Shuttle disaster, situating the ethical problem for students in the engineers' knowledge about the potential explosion, and their decisions around moving forward with the shuttle launch anyway.

After going through the context of the ethical problem, the professor then led a discussion about alternative approaches the engineers could have taken and the ethics behind them, as well as the potential impacts of those decisions. The learning objective from this lecture was to have students take the perspective of an engineer in that ethical situation and consider what they might have done differently.

For the second lecture, the professor discussed the philosophical trolley problem described earlier. After setting this premise, the lecturer opened the class up for discussion about what they would do and the ethical implications of these actions. As the students presented different solutions, the professor lectured on the impact of different choices on others. The learning objective was to have students take the perspective of someone who had only bad options and determine what they would choose to do in an ethically complicated situation.

OUTCOME TOOL: ENGINEERING ETHICAL REASONINGS INSTRUMENT

Educators at Purdue University and the Illinois Institute of Technology recognized the gap between engineering ethics education and students' ability to apply these principles in practice. Together, they created an instrument for "individual ethical decision-making in a project-based design" (e.g., team-based, problem-solving) called the Engineering Ethical Reasoning Instrument (EERI) based off the leading ethics assessment of the time, the Defining Issues Test-2 (DIT-2) (Zhu et al., 2014). This instrument was chosen for the

present study due to the DIT-2 being the standard ethics assessment in the engineering ethics literature, and the EERI improving upon the DIT-2 through situating the scenarios in engineering.

In the EERI the participants were given six ethically complex scenarios, asked to decide on performing an ethically questionable action, and were then assessed on how important each factor was to them, and which factors were most important, in making their decision. In one such scenario, the participant's student project team is tasked with rating the overall quality of buildings in an impoverished section of the community to improve the quality of housing, some of which are thought to be unsafe to live in. The participant has good reason to believe that the data will be used to raze the least safe of those buildings, forcing residents to relocate with some residents stating that they may end up homeless if their residence is demolished. Thus, the ethical dilemma is to 1) report the unsafe dwellings so that people aren't exposed to the danger of the building but possibly leaving them unhoused or 2) not report the dwellings accurately, allowing residents to stay in their unsafe homes but off the streets. The participant must then decide whether to rate the homes, and then rate the importance of factors that could influence the decision. Some factors included whether the participant was friends with an affected resident, whether it could help the participant's future career, whether the unintended social ramifications of this work should be a concern, etc. Then after the importance of these factors were chosen, the participant would rank the four factors that were most important to them from most important to fourth most important.

From these scenarios, the EERI produces two scores: P and N2. The P score represents the extent to which participants' four most important factors were based on looking beyond self-interest (e.g., whether the unintended social ramifications of this work should be a concern). The N2 score is statistically derived from the P score, while also factoring in whether participants' rated factors based on self-interest (e.g., whether it could help the participant's future career) as less important than factors based on looking beyond self-interest. For the purposes of this paper, we are focusing on the N2 score, as it is the more comprehensive and interpretable score from the EERI, and is derived from the P score.

ANALYSIS

For participants who remained after exclusion criteria, we distilled their EERI results into pre and post P, and pre and post N2 scores. We had already split these individuals into groups based on their year of taking the course and whether they were in the game- or lecture-based instruction version of the course. Given our question's focus on determining if the conditions significantly differ in each year, we conducted an interaction analysis to determine how they may differ from pre to post, and if there are other underlying factors such as their groups, or time, that lead to perceived significant pre-post changes. We used two-way mixed-design ANOVAs for each year independently to compare the pre and post EERI N2 scores.

Before conducting the two-way mixed-design ANOVA we also analyzed whether there were statistically significant differences first between the pre EERI N2 scores for each group in their respective year to ensure the groups were comparable at the outset. We

found that, within each year, there was no statistically significant difference between the pre EERI N2 scores for the groups. In running the multiple two-way mixed-design ANOVAs not all assumption tests on normality indicated a normal distribution. The smallest sample size of a potentially non-normal distribution was 108 which can cause the assumption tests to be less accurate, as was the case here upon visual inspection of the distributions. For the assumption tests on the homogeneity of variance, there were multiple instances where the variances were below 0.05, but above 0.01. Thus, we adopted 0.01 as our threshold for these tests, as in all cases where the samples were identical or nearly identical, Hartley's Test resulted in a variance ratio under 1:2. The assumption tests identified a few outliers, however, removing them did not significantly change anything so they were kept in the model to reflect the data more accurately. Last, there was no assumption of sphericity as the within-subject factor of time only had two levels, pre, and post. From the above, we believe the two-way mixed-design ANOVAs used to analyze the EERI N2 data can be interpreted normally.

RESULTS

2019 RESULTS

In 2019, there were no statistically significant effects for the N2 score, though multiple effects neared significance. Specifically, our main interest was in the lack of statistically significant interaction effect for the N2 score ($F(1, 264) = .358$, $p = .550$, $\eta^2 = .001$), suggesting that students experienced equivalent changes in their moral reasoning regardless of treatment group. Figure 1 shows how the slopes of the lines do not significantly differ, supporting the non-significant interaction between group and time.

Figure 1

Estimated Marginal Means of EERI N2 for 2019

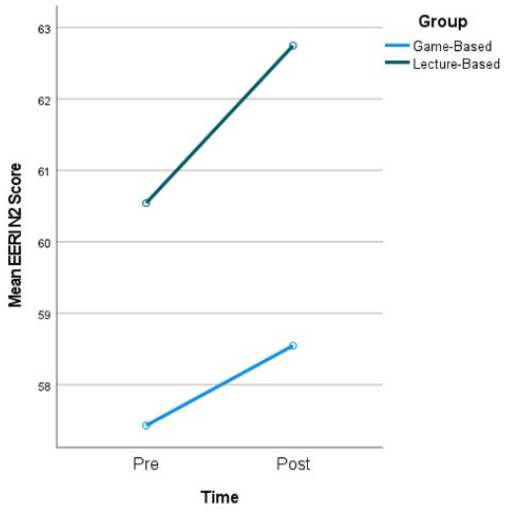
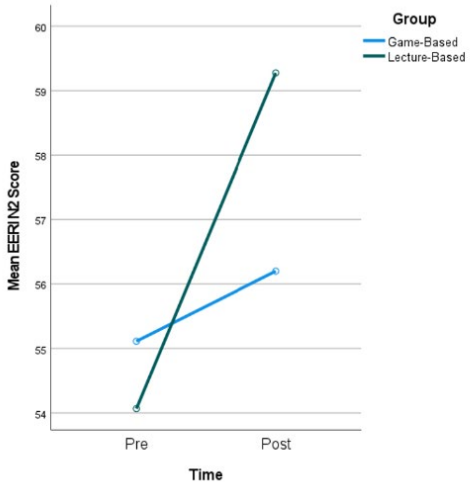


Figure 2

Estimated Marginal Means of EERI N2 for 2020



2020 RESULTS

The results of the two-way mixed-design ANOVA for 2020 indicate that the only finding of statistical significance was the main effect difference between the pre and post-EERI N2 score ($F(1, 243) = 8.453$, $p = .004$, $\eta^2 = .034$). Thus, in this case, we can understand students in either condition ended their time in the course with a significantly higher EERI N2 score. Figure 2 highlights the direction of this change, as in both groups there was a positive difference

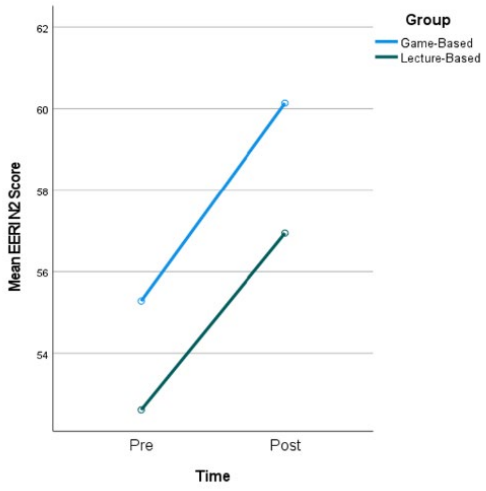
in their EERI N2 score from pre to post. We again see that the interaction between time and group was not statistically significant for the N2 score ($F(1, 243) = 3.620$, $p = .058$, $\eta^2 = .015$). This continues the trend that, regardless of the condition participants were in, there was an increase in N2 score. Figure 2 also highlights how the slopes of the lines do not significantly differ with the proportional difference over time in EERI N2 score being similar across groups.

2021 RESULTS

Again, the only finding of statistical significance was for the main effect difference between the pre and post N2 EERI score ($F(1, 158) = 11.727$, $p = .001$, $\eta^2 = .069$). Figure 3 highlights this difference in EERI N2 scores across time in both groups. As well, for the third time, the interaction between time and group was not statistically significant for the N2 score ($F(1, 158) = .040$, $p = .842$, $\eta^2 = .000$), such that students' scores were proportional over time regardless of treatment. Figure 3 also highlights how the slopes of the lines do not significantly differ, supporting the non-significant interaction between group and time.

Figure 3

Estimated Marginal Means of EERI N2 for 2021



DISCUSSION

From the results above it is clear that, as assessed through the EERI, there is a lack of evidence that within any year, the game-based learning condition significantly differed from the lecture-based instruction. This is a potentially encouraging finding in its own right, as a lack of significant difference between conditions, but significant increases in N2 scores over time in 2020 and 2021 may indicate that game-based learning is as impactful as traditional lectures. While other analyses should be performed to support this claim, there are other possible explanations for these results. For instance, it is possible that other shared elements of this engineering class led to increases over time, despite there being no other direct discussion or assessment of ethics in the course. Regardless of this positive change in moral reasoning over time, the focus of this study was on whether game-based learning and lecture-based instruction differed in their effect on ethics. Given this focus, we are more interested in how such drastically different methodologies resulted in no statistically significant difference.

There are multiple possible explanations for this lack of significant difference between the ethics games and traditional lectures. One explanation is that both interventions were beneficial themselves and simply did not differ significantly in how effective they were. Another reason for these findings could be that the lectures were particularly excellent, and that other types of ethics lectures by another professor would not have these same results. We can also speculate that the games may require more time to engage students

than anticipated, so the games may have been less effective than theoretically possible due to constraints such as class length, time to get students settled and focused, and other classroom priorities. Lastly, it is possible that switching the modalities (in person vs. online) and games played between years, and the effects of COVID-19, did not allow the instructors and students to engage with the games to the best extent possible. All of these possible explanations are worth further investigation, however, we believe there is another explanation that we view as the most compelling.

We believe that the best possible explanation for no statistically significant interaction terms lies with a potential misalignment between the EERI's measure of ethics through moral reasoning and the ethics games focus on exploring students' situated decision making. The EERI is focused solely on the underlying moral reasons, defined as their thoughts on right and wrong. The ethics games were focused on situated decision making, which is grounded in contextual factors, authentic details, and people's past experiences. This explanation is not rooted in data from the EERI, but in recognizing how the games were designed based on situated decision making, in contrast to the moral reasoning design of the EERI. These findings of no significant difference between the forms of pedagogy indicate to us that the difference between the games' focus on situated decision making as distinct from the lectures, was not captured by the EERI's measure of moral reasoning.

As the EERI is focused solely on the underlying moral reasons that students find most important (e.g,. self-interest versus justice), games seeking to change this underlying moral reasoning would potentially show some difference from traditional lectures as assessed through the EERI. For games that differ from lectures in their authentic, specific, and contextually rich scenarios, like the games used in this study, the EERI is unlikely to detect differences between the lecture and game-based pedagogies due to the EERI's format. The EERI's scenarios are antithetical to those from our games and therefore cannot accurately detect the changes in students' ethics. For these types of games, such as the ones in this study, research needs to be conducted to create new instruments that are targeted to assess these principles, rather than try to apply existing ethics measures that are not aligned.

CONCLUSION

Given this believed misalignment between the EERI's measurement of moral reasoning and the situated decision-making design of the engineering ethics games used in this study, some questions remain; how should assessments of ethics be designed to better capture situated decision making, and how can these assessments of ethics be better aligned with the interventions on ethics instruction? These are questions for both further designers of assessment and for researchers in choosing which assessments are the most appropriate for your research questions when studying situated decision making, or moral reasoning, in engineering instruction. There is likely no single answer to any of these, as various assessments are built from different moral and ethical philosophies and frameworks, and the various playful and other interventions being studied may be designed incongruously. This result occurred within this study, as the EERI was the best measure we could choose, being based on the standard ethics assessment in the literature

and further focused on engineering, but was not the correct measure for our engineering ethics games. Recognizing the importance of varied assessments of situated decision making and moral reasoning in engineering, how the design of ethics games aligns with these theories, and the impact alignment has on measuring the impact of ethics games on situated decision making, can generate further research on how ethics games may help shape the ethics of future engineers.

LIMITATIONS

There are several limitations to this study, most of which have been discussed already: the uneven sample size in 2021, and the COVID-19 pandemic occurring in the middle of 2020. The uneven sample sizes in 2021 were a minor limitation in the analysis of the EERI P and N2 scores for that year. While the analysis was conducted due to the use of assumption tests that account for uneven sample sizes, it is still limiting that one group had slightly over twice the number of participants as the other. It is impossible to control which section students sign up to take and whether they properly finish and take seriously the EERI, there could be better incentive structures increasing retention so that more students complete the EERI both times.

The other limitation of note is the COVID-19 pandemic that occurred halfway through 2020 and fundamentally changed the instruction in both conditions for 2020 and 2021, also causing a change in the interpretations. It is impossible to control major pandemics and their impact on students, all that can be done in these cases is to acknowledge their presence, and how they may impact interpretations. In this case, it is possible that some of the non-significant results such as interactions could have been the impact of the pandemic. While the findings are interpreted as the result of a misalignment between the EERI and games, it is possible that for 2020 and 2021, there were effects of the pandemic that impacted these results, unrelated to the problem of alignment.

ACKNOWLEDGMENTS

This research is being funded by the National Science Foundation, “Collaborative Research: Learning Engineering Ethics Through High-Impact Collaborative and Competitive Scenarios” (IUSE – 1934702).

REFERENCES

Accreditation Board for Engineering and Technology. (2022). Criteria for accrediting engineering programs, 2022-2023. ABET. https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2022-2023/

Alfred, M., & Chung, C. A. (2012). Design, development, and evaluation of a second generation interactive Simulator for Engineering Ethics Education (SEEE2). Science and engineering ethics, 18(4), 689-697.

American Society for Engineering Education. (2022). Profiles of Engineering and Engineering Technology, 2021. https://ira.asee.org/wp-content/uploads/2022/11/Engineering-and-Engineering-Technology-by-the-Numbers-2021.pdf

Bagdasarov, Z., Thiel, C. E., Johnson, J. F., Connelly, S., Harkrider, L. N., Devenport, L. D., & Mumford, M. D. (2013). Case-based ethics instruction: The influence of contextual and individual factors in case content on ethical decision-making. Science and Engineering Ethics, 19(3), 1305-1322. https://doi.org/10.1007/s11948-012-9414-3

Barab, S., & Dede, C. (2007). Games and Immersive Participatory Simulations for Science Education: An Emerging Type of Curricula. Journal of Science Education and Technology, 16(1), 1–3. https://doi.org/10.1007/s10956-007-9043-9

Bodnar, C. A., Anastasio, D., Enszer, J. A., & Burkey, D. D. (2016). Engineers at play: Games as teaching tools for undergraduate engineering students. Journal of Engineering Education, 105(1), 147-200.

Callahan, D. (1980). Goals in the teaching of ethics. In: Callahan, D., Bok, S. (Eds.) Ethics Teaching in Higher Education. (pp. 61-80). Springer, Boston, MA. https://doi.org/10.1007/978-1-4613-3138-4_2

Chee, Y. S., & Tan, K. C. (2012). Becoming chemists through game-based inquiry learning: The case of legends of Alkhimia. Electronic Journal of e-Learning, 10(2), 185-198.

Cicchino, M. I. (2015). Using game-based learning to foster critical thinking in student discourse. Interdisciplinary Journal of Problem-Based Learning, 9(2). http://dx.doi.org/10.7771/1541-5015.1481

Drew, C. (2011, November 4). Why science majors change their minds. New York Times. http://www.nytimes.com/2011/11/06/education/edlife/why-science-majors-change-theirmind-its-just-so-darn-hard.html

Franciosi, S. J. (2017). The effect of computer game-based learning on FL vocabulary transferability. Educational Technology & Society, 20(1), 123-133.

Kohlberg, L., & Hersh, R. H. (1977). Moral development: A review of the theory. Theory Into Practice, 16(2), 53–59. https://doi.org/10.1080/00405847709542675

Lamb, C. S. (1991). Teaching professional ethics to undergraduate counseling students.

Psychological Reports, 69(3_suppl), 1215–1223. https://doi.org/10.2466/pr0.1991.69.3f.1215

Lau, S. W., Tan, T. P. L., & Goh, S. M. (2012). Teaching engineering ethics using BLOCKS game. Science and Engineering Ethics, 19(3), 1357–1373. https://doi.org/10.1007/s11948-012-9406-3

Lloyd, P., & van de Poel, I. (2008). Designing games to teach ethics. Science and Engineering Ethics, (14), 433–447. https://doi.org/10.1007/s11948-008-9077-2

Marklund, B. B., & Taylor, A. A. (2016). Educational Games in Practice: The challenges involved in conducting a game-based curriculum. The Electronic Journal of e-Learning, 14(2), 122-135.

McGinn, R. (2018). The ethical engineer: Contemporary concepts and cases. Princeton University Press.

Paulson, J., & Kretz, L. (2018). Exploring the potential contributions of mindfulness and compassion-based practices for enhancing the teaching of undergraduate ethics courses in philosophy. The Social Science Journal, 55(3), 323-331. https://doi.org/10.1016/j.soscij.2017.12.003

Plante, T., & Pistoresi, S. (2017). A survey of ethics training in undergraduate psychology programs at Jesuit universities. Pastoral Psychology, 66(3), 353–358. https://doi-org.ezproxy.lib.uconn.edu/10.1007/s11089-017-0755-3

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. Educational psychologist, 50(4), 258-283.

Plass, J. L. (2017, March 15). Gamification v. Game-based Learning v. Playful Learning. LinkedIn. https://www.linkedin.com/pulse/gamification-v-game-based-learning-playful-jan-l-plass/

Shaffer, D. W., Squire, K. R., Halverson, R., & Gee, J. P. (2005). Video games and the future of learning. Phi delta kappan, 87(2), 105-111.

Slota, S. T., Young, M. F. (2014). Think games on the fly, not gamify: Issues in game-based learning research. Journal of Graduate Medical Education, 6(4), 628-630. https://doi-org.ezproxy.lib.uconn.edu/10.4300%2FJGME-D-14-00483.1

Space Shuttle Challenger Disaster. (2022, September 22). In Wikipedia. https://en.wikipedia.org/w/index.php?title=Space_Shuttle_Challenger_disaster&oldid=1111721759

Whitbeck, C. (1996). Ethics as design: Doing justice to moral problems, The Hastings Center Report, 26(3), 9-16. https://doi.org/10.2307/3527925

Wong, M. K., Hong, D. Z. H., Wu, J., Ting, J. J. Q., Goh, J. L., Ong, Z. Y., Toh, R. Q. E., Chiang, C. L. L., Ng, C. W. H., Ng, J. C. K., Cheong, C. W. S., Tay, K. T., Tan, L. H. S., Ong, Y. T., Chiam, M., Chin, A. M. C., Mason, S., & Radha Krishna, L. K. (2022). A systematic scoping review of undergraduate medical ethics education programs from 1990 to 2020. Medical Teacher, 44(2), 167-186. https://doi.org/10.1080/0142159X.2021.1970729

Zhu, Q., Zoltowski, C. B., Feister, M. K., Buzzanell, P. M., Oakes, W. C., & Mead, A. D. (2014,

June), The Development of an Instrument for Assessing Individual Ethical Decisionmaking in Project-based design teams: Integrating quantitative and qualitative methods Paper presented at 2014 ASEE Annual Conference & Exposition, Indianapolis, Indiana. https://doi.org/10.18260/1-2--23130

INITIAL DEVELOPMENT AND VALIDATION OF THE STUDENT SELF-DETERMINATION OPPORTUNITY SURVEY: TEACHER REPORT

SARAH ROSATI
UNIVERSITY OF CONNECTICUT NEAG SCHOOL OF EDUCATION

ABSTRACT

Self–determination instruction is associated with a variety of important outcomes for students with disabilities. A new instrument, the Student Self–Determination Opportunity Survey: Teacher Report version (SSOS–TR), utilizes Causal Agency Theory as a framework to measure teachers' perceptions of the extent to which their instructional practices facilitate this area of skill development. This paper outlines and describes the item development and content validation processes utilized in developing the SSOS–TR. A sample of 151 special education teachers spanning grades K-12 participated in this SSOS-TR instrument pilot. In alignment with the three essential characteristics of self-determined action identified in Causal Agency Theory, the SSOS-TR factor analyses suggested a three-factor extraction. The three resultant subscales produced alpha coefficients ranging from .87-.91, evidencing internal consistency. Implications, limitations, and suggestions for future research are described.

Keywords: self-determination, instrument validation, disability, Causal Agency Theory

My experiences teaching in a variety of educational positions and settings have sharpened my focus on elevating the voices of students with disabilities (SWD). In what I hope will be my research arc as a scholar, I aim to explore the extent to which SWD perceive they are afforded opportunities by their special education teachers to have their student voices heard. Self–determination curricula can provide a means to elevate student voices. Implementing self–determination instruction is associated with a variety of important outcomes for SWD, such as educational goal attainment (Shogren et al., 2012) and positive postsecondary outcomes (e.g., employment, quality of life, integration into community; Nota et al., 2007; Shogren et al., 2012; Shogren, Wehmeyer, Palmer, Rifenbark, & Little, 2015).

Students need opportunities at school to practice self–determination skills and such opportunities for practice are largely designed, scheduled, and implemented by teachers. While tools exist to capture teacher ratings of a student's present levels of self–determination, missing are tools that focus on teachers' perceptions regarding the extent to which their instructional practices facilitate this skill development. Specifically, with this survey, I hope to measure these opportunities via K–12 special education teacher self–reports as the population of interest. My proposed instrument is named the Student Self–Determination Opportunity Survey: Teacher Report version (SSOS–TR).

This paper outlines and describes the item development and content validation processes that I completed in developing the SSOS–TR. My primary research question is: To what extent do special educators perceive they provide opportunities for SWD to develop self–determination skills? As I continue to explore the perceptions of SWD with regards to their educational experiences and opportunities, the SSOS–TR will serve as an elucidating comparative body of information and potentially expose important gaps in perception between teacher and student populations.

BRIEF REVIEW OF THE LITERATURE

Origins of self–determination trace back to the initial stages of research on personality development in the mid–1900s, particularly by Angyal (1941), who described autonomous–determinism as self–caused action in contrast to actions governed by external laws. Decades later, Wehmeyer (1992) defined and refined (Wehmeyer et al., 1996) self–determination as "acting as the primary causal agent in one's life and making choices and decisions regarding one's quality of life free from undue external influence or interference" (p. 632).

There has been a wide application of this model in special education research across a variety of school types and grade levels. Cobb et al. (2009) summarized the results of self–determination curricula and outcomes for individuals with disabilities in a narrative meta–synthesis. They concluded that (a) multi–element interventions yield more positive outcomes than single–component interventions (e.g., works included in the meta–synthesis by Algozzine et al., 2001; Fowler et al., 2007; Konrad et al., 2007; Wood et al., 2005), (b) self–determination outcomes can be enhanced in targeted interventions for SWD, yet these interventions do not appear to be effective with regards to increasing academic achievement (e.g., works included in the meta–synthesis by Fowler et al., 2007;

Konrad et al., 2007), and (c) there are strong and positive correlations between school self–determination interventions and adult outcomes (e.g., works included in the meta-synthesis by Algozzine et al., 2001; Chambers et al., 2007; Malian & Nevin, 2002). One of the included reviews published by Chambers et al. (2007) is particularly relevant. Among their findings was the fact that teachers' reported ratings on a quantitative scale measuring the importance of self–determination did not necessarily translate to them teaching these skills for various reasons (e.g., lack of teacher preparation and lack of perceived impact on students with more severe disabilities).

CAUSAL AGENCY THEORY

A re–conceptualization of the 1990s definition of self–determination (Wehmeyer et al., 1996) was deemed necessary due to (a) the emerging field of positive psychology and its focus on personal well–being and self–determination as a motivational force (Ryan & Deci, 2000), (b) the changing understanding of disability towards a strengths–based approach (Shogren, 2013), and (c) shifts towards increasingly inclusive models of special education delivery for SWD alongside general education students who would also benefit from self–determination interventions. This new reconceptualization, called Causal Agency Theory, additionally explains how individuals become self–determined (Shogren, Wehmeyer, Palmer, Forber-Pratt et al., 2015). The definition of self–determination as defined within the Causal Agency Theory is as follows:

> … dispositional characteristic manifested as acting as the causal agent in one's life. Self-determined people (i.e., causal agents) act in service to freely chosen goals. Self-determined actions function to enable a person to be the causal agent in his or her life. (Shogren, Wehmeyer, Palmer, Forber-Pratt et al., 2015, p. 258)

Burke et al. (2020) conducted a recent meta–analysis to update the metanalytic literature on self–determination and also to align a synthesis of the self–determination literature with the reconceptualization put forth in the Causal Agency Theory. Burke et al. found 34 articles meeting inclusion criteria, such as implementation of an intervention, participant populations containing (but not limited to) SWD, and outcome variables pertaining to self–determination skill(s). All "suggested positive outcomes of intervention for overall self–determination or one or more skills associated with self-determined action" (p. 183). The number of combined participants in the research studies contained within this meta–analysis was over three times greater than that of the comparable meta–analysis published by Algozzine et al. (2001) almost 20 years prior; thus, the construct of self–determination is one of continued relevance to the field.

According to Causal Agency Theory, there are layers of human agency that span the continuum from meeting basic biological and psychological needs to acting as a fully agentic self. The agentic self is a person who is personally empowered, engaged in goal–directed and self–regulated action, and is continuously self–monitoring in response to environmental changes that can function as either opportunities/supports

or impediments/threats to their goals (Shogren, Wehmeyer, Palmer, Forber-Pratt et al., 2015). According the Causal Agency Theory, people engage in an iterative series of analyses whenever there is a discrepancy between their current status and their goal status. They prioritize actions and evaluate their problem–solving abilities until they find a match between challenge and capacity towards goal–attainment that they can execute by making strategic decisions and choices, or they refine/revise the original goal (Shogren, Wehmeyer, Palmer, Forber-Pratt et al., 2015).

The layers of human agency between the basic and agentic self are the three essential characteristics of self–determined action within the Causal Agency Theory. An individual's motivation to secure basic psychological and biological needs promotes their well–being, which in turn facilitates the optimal condition for developing the essential characteristics of a fully agentic self. I employed these three characteristics as my hypothesized factors for the SSOS–TR: volitional action, agentic action, and action-control beliefs. These hypothesized factors are conceptually defined by Shogren, Wehmeyer, Palmer, Forber-Pratt et al. (2015) as follows:

- Volitional Action: self-initiation based on preferences and conscious choices
- Agentic Action: self-regulated and self-directed towards chosen personal goals
- Action-Control Beliefs: the individual's positive awareness of their sills, knowledge, and capacity towards their goals

These factors also lead to causal agency, which is when "the individual acts with an eye toward causing an effect to accomplish a specific end or to cause or create change" (Shogren, Wehmeyer, Palmer, Forber-Pratt et al., 2015, p. 259). Causal agency is thus both an outcome and an influencer of the fully agentic self, who experiences even higher levels of well–being. I chose Causal Agency Theory as my framework for this survey given its comprehensive utility in both defining what self-determination is and in explaining how to create opportunities for SWD to become agentic selves.

EXISTING INSTRUMENTS

Two predominant tools utilized to measure self–determination were developed in alignment with Wehmeyer's (1992) original theory of self–determination: the American Institutes for Research (AIR) Self–Determination Scale (Wolman et al., 1994) and The Arc's Self–Determination Scale (Wehmeyer & Kelchner, 1995). The Arc's Self–Determination Scale is a student self–report instrument that was designed for use with adolescents labeled with cognitive disabilities to measure their current areas of strengths and limitations as well as the factors promoting or inhibiting their self–determination. The AIR Self–Determination Scale additionally includes the aim of identifying goals that can be adapted for inclusion in a student's Individualized Education Program (IEP; Wolman et al., 1994).

When the functional model of self–determination was re–conceptualized as the Causal Agency Theory (Shogren, Wehmeyer, Palmer, Forber-Pratt et al., 2015), a corresponding instrument was designed in alignment: the Self–Determination Inventory: Student Report version (SDI–SR; Shogren et al., 2017). The SDI–SR is a departure from the scales above in that designed for use by students regardless of disability status (Shogren et al., 2017).

PROBLEM STATEMENT

In reviewing existing instruments, I noted a lack of tools measuring special educators' perceptions of the opportunities they afford to SWD to develop self–determination skills. The closest items I could find in measuring this construct were in the AIR Self–Determination Scale. The teacher version of the AIR Self–Determination Scale has just six items to measure the teachers' perceptions of opportunities that students have to perform self-determined behaviors at school. An example item is: "Student has opportunities at school to explore, express, and feel good about own needs, interests, and abilities" (Wolman et al., 1994). As with this example, the educator form of this scale is exclusively comprised of item stems beginning with "Student."

In contrast, I aim to capture a more direct measure to promote increased accountability for special educators and a tool to prompt meaningful reflection on their individual classroom practices. If the example above was reworded with the special educator as the active agent, it would instead read: "I create opportunities for my students to explore, express, and feel good about their own needs, interests, and abilities." The SSOS–TR will move beyond the measurement of students' skills and towards a measurement of the extent to which teachers themselves are working to create opportunities for SWD to become self–determined.

ITEM DEVELOPMENT PROCESS

I drew influence from the AIR Self–Determination Scale educator form and the Self–Determination Inventory: Student Report version in drafting an initial pool of 37 items. The items were hypothesized to operationally define one of the three essential characteristics of self–determination (i.e., 1=Volitional Action, 2=Agentic Action, 3=Action–Control Beliefs). Prior to the content validation process, I made scale selection considerations. I decided to select a Likert agreement scale with no midpoint to encourage deeper levels of critical thinking and thereby prevent strong satisficing of participants (i.e., selecting a neutral or no–opinion midpoint to minimize the difficulty level of the question on oneself; Krosnick et al., 1996). I selected six response options as opposed to four because of current guidance to include indicators with more than five categories in order for them to be treated as continuous in statistical analyses (McCoach et al., 2013). I created a content validation form to seek feedback regarding my conceptual definitions, capture expert opinions about which survey item belonged within each factor, to what degree they were certain of this judgment, and to what extent they found the item to be relevant to the factors. Last, I created a tab to capture qualitative feedback using open–ended questions probing about topics such as the clarity of survey items, range of content covered, and any additional comments.

CONTENT VALIDITY PROCESSES

EXPERT REVIEW METHODS

I contacted five content specialists in the area of self–determination via email to serve as expert reviewers for content validation. My original email included a line asking experts to suggest additional reviewers. In this way, I was able to generate additional experts for a total of seven experts. The final seven experts included one graduate student and six university professors who all specialize and publish in the area of self–determination. All experts provided their feedback via email using Excel spreadsheets that I created for each individual reviewer.

EXPERT REVIEW RESULTS

I compiled all of the seven experts' content validity forms into one spreadsheet. This sample of experts all indicated that my conceptual definitions were sufficient; no changes to the definitions were recommended. I then calculated the percentage of expert reviewers for each drafted item who were in agreement with its hypothesized factor as well as the percentage of certainty for each reviewer in their factor designation (i.e., Very Uncertain, Moderately Uncertain, Moderately Certain, Very Certain) and the percentage of relevance to constructs that the reviewers attributed to each item (i.e., Totally Irrelevant, Not Very Relevant, Somewhat Relevant, Totally Irrelevant).

The retained items on the revised SSOS-TR were all deemed somewhat relevant or totally relevant. All of the reviewers affirmed that the items seemed to span the range of content for each category as defined. Also, this sample of experts agreed that a 6–point Likert agreement response scale was appropriate. In response to the final qualitative question to reviewers asking for any additional thoughts or comments, one expert shared, "this will be a wonderful tool to add to our knowledge of teachers' current perceptions and practices and then plan for teacher preparation and professional development." This feedback encapsulates my hope for the future utility of the SSOS–TR.

INSTRUMENT ITEM DELETIONS PRIOR TO PILOTING

After analyzing the expert reviewers' content validation forms and qualitative feedback, a total of 17 items were deleted from the initial item pool. Items with fewer than five out of seven experts in agreement with the hypothesized factor were inspected first. Given that these items were so inconsistently classified, there was a possibility that they would load onto more than one factor and would consequently trouble a factor analysis. This criterion resulted in the removal of 13 items.

Two items, "I have my students identify what they can do by themselves," and "I encourage my students to act on their own," not only had low percentages of reviewers in agreement with the hypothesized factors (57% and 29% agreement, respectively), but were also critiqued by expert reviewers as being irrelevant to, and inappropriate for, inclusion within a self–determination instrument. In the qualitative feedback section, one

expert asked, “Why is it necessary for students to identify what they can do by themselves when what is more important is that they know the supports they need to be successful?” Another reviewer similarly commented, “I would make sure there is not a focus on independence. No one is truly independent; we all use supports to engage in our own self-selected goals.” These items went beyond the scope of self–determination and into the realm of independence, which was not the intended construct.

I removed four additional items given qualitative feedback. “I encourage my students to make choices” was consistently identified as redundant with “I provide opportunities for my students to make their own choices.” This redundancy “can cause the association between two variables to be even stronger than can be explained by the underlying factor(s)” (Flora & Flake, 2017, p. 86), which would also be problematic in a factor analysis. Experts who identified these items as redundant unanimously identified the later version as preferable for retention.

Three items that included the concept of confidence were removed: “I know which students are confident in their abilities,” “I know which students are working on their self–confidence,” and “I provide opportunities for confidence–building.” These items had high percentages of overall agreement with hypothesized factors among the experts (71%, 86%, and 86% respectively); however, I decided to remove them after extensive email correspondence with one of my experts. This expert reviewer was critical of the self–determination theoretical framing from within the field. The expert stated that “self–determination is often framed from White, individualistic perspectives” and believed that “questions to teachers should not sound like promoting White, cultural perspectives and practices, especially when it comes to doing things independently and being confident.” The expert also spoke to the reality that “students from culturally and linguistically diverse backgrounds may not present as confident as White students.” This researcher urged me to refrain from including items related to confidence, as it is not the same as self–awareness and self–knowledge, which are represented in the third factor of action–control beliefs. I had already removed the items referring to independence, but this researcher’s feedback was to also remove these three items referring to confidence.

I would like to emphasize the importance of this expert reviewer’s qualitative feedback in supporting me to create a more culturally responsive instrument. As discussed by Sankofa (2021), expert panels do not exhaustively check researcher bias as “experts are plausibly colleagues empowered with similar privileged worldviews and biases as the researcher who developed the scale” (p. 2). My original framing of items was indeed biased towards a cultural perspective uplifting Whiteness due to its value placed on confidence and independence.

INSTRUMENT ITEM REVISIONS

Two items on the SSOS–TR tool were revised for language but not deleted from the item pool. Expert feedback towards the item originally stated as, “I give time for my students to deal with obstacles,” was that “deal with” sounded like the teacher was not supporting students at all. Thus, the item was revised to read, “I give time for my students to confront obstacles.” This rewording still captures the teacher’s self–perception of the opportunities

they afford to students to practice navigating challenges, but exposure to barriers is prioritized over independent resolution of challenges. Another item originally read, "I minimize my influence on students' choices." Experts pointed out that students' choices are largely constrained by external forces beyond the teacher (e.g., curricular, resource, and time constraints) and that it is often necessary for teachers to ensure that authentic choices are available to SWD. The suggested revision was to replace "choices" with the word "decisions" so that the item now reads, "I minimize my influence on students' decisions."

DATA COLLECTION AND METHODS TARGET POPULATION

My revised target population for the SSOS–TR instrument pilot was special education teachers serving students in grades K–12 across a variety of service models most common to special education (e.g., self–contained classrooms for students requiring intensive instruction in all subject areas provided by a special educator within a separate setting or inclusive classrooms co–taught with a general education teacher where students with and without disability labels learn alongside each other). My original target population was secondary (i.e., high school) special education teachers; however, with expert feedback, I was pushed to open the target population to include all special education teachers serving students across grades K-12. Expert rationale for this broadening was to inquire about grades within the demographic section, as differences might emerge between grade bands that could inform my decisions with this tool moving forward. This suggestion to make a single instrument for special educators across all grade bands would help to align the SSOS–TR with the AIR Self–Determination Scale from which it drew much influence, as this tool also spans K-12 in its assessment of self–determination skills.

Some experts questioned my target population restricted to special educators. When Causal Agency Theory (Shogren, Wehmeyer, Palmer, Forber-Pratt et al., 2015) replaced the functional model of self–determination, it did so in part to reflect shifts towards inclusive models of special education for SWD alongside general education students, who would also benefit from self–determination interventions. The corresponding measure updated to align with Causal Agency Theory, the SDI–SR, was designed for use with all students regardless of disability labels, as its creators cited evidence that supports the importance of enhanced self–determination for all students.

Despite these suggestions, I decided to keep the target population restricted to special education teachers. Opening up the target population would generate information on general education teachers' self–perceptions, of which I am not interested at present. Also, as I ultimately aim to research the perceptions of students labeled with disabilities with regards to the opportunities, they perceive they are afforded, I do not wish to capture information on students without disability labels at this time. I do not negate the utility of my SSOS–TR tool for use with all teachers reporting on all students. However, for the specific purposes of my research, I am keeping my target population limited to special educators at this time. The expert suggestion to include all teachers does introduce a new line of inquiry to compare the self–perceptions of special and general educator teachers, which could be investigated in the future.

PILOT SAMPLING METHOD

For the purposes of this study, I used non–probability convenience sampling. I gathered a sample from social media (i.e., Facebook, Twitter) by linking to my survey and tagging relevant special education professional networks. I also generated a generic tweet and Facebook post containing the most widely used special education hashtags. I shared my survey link with graduate student colleagues who are former special education teachers using a snowball sampling outreach strategy in which these colleagues were the "seed" individuals with desired networks. The mechanism behind this strategy is therefore "semi–self–directed, chain–referral," (Sadler et al., 2010) as I asked colleagues to share the recruitment messages to reach the approximate sample size of 200. This goal sample size of 200 satisfied the best practices suggested by Boateng et al. (2018) given that there are 19 items on the revised scale (i.e., a 10:1 ratio of respondents to items) and it also meets the 200–300 sample size range described by the authors as appropriate for running a factor analysis.

INSTRUMENT PILOT TIMELINE

I sent my recruitment messages out to my graduate colleagues and posted the messages to social media platforms of Twitter and Facebook on March 15, 2022. I proceeded with recruitment utilizing the snowball sampling outreach strategy described above. The survey was distributed via a link to Qualtrics, a platform for designing and sharing web–based surveys, and remained active until April 19, 2022, at which point I had received 212 responses (i.e., within the required 200-300 range).

DESCRIPTION OF PILOT SAMPLE PARTICIPANTS

Differing numbers of respondents answered the demographic questions at the end of the SSOS–TR in that some answered all questions, while others answered some or none of them. The mean respondent age for those who provided their age was 42 years old with a range of 22 to 65. With regards to years teaching (including the 2021–2022 year), the mean was 13 years (range: 1–40). This sample reasonably approximates the expected characteristics for special education teachers provided by the U.S. Department of Education. With regards to race, the sample was slightly higher at 91% than the expected 83% for percentage of White–identifying teachers and was higher at 90% than the expected 75% for female–identifying teachers (U.S. Department of Education, 2020).

ANALYTICAL METHODS

I used an exploratory factor analysis (EFA) to analyze my data. An EFA was appropriate to understand the dimensionality of the SSOS-TR given that I was in the initial stages of instrument design. While I hypothesized that the SSOS–TR had three factors, I did not impose this structure a priori. Rather, I wanted to analyze if three was the "smallest number of interpretable factors needed to explain the correlations among a set of items" (McCoach et al., 2013, p. 111) and running an EFA would allow me to explore this hypothesis.

I also elected to run a reliability analysis on my data. I defined reliability in keeping with DeVellis' (2003) definition of scale reliability as "the proportion of variance" that is "attributable to the true score of the latent variable" (p. 27). I focused on internal consistency reliability to analyze the SSOS-TR scores using Cronbach's alpha (Cronbach, 1951), which is a measure of inter-item consistency (Clifton, 2019). Cronbach's alpha assumes that defined subscales are unidimensional and composed of items that are linearly related to an individual's total score. Cronbach's alpha additionally assumes that item errors do not covary, subscale items are normally distributed, and each unidimensional factor's items have equal factor loadings (Cronbach, 1951). I interpreted the overall SSOS–TR's Cronbach alpha as well as the alpha estimates of explored subscales according to DeVellis' standards of acceptability (2003), which outline that which is "unacceptable (<.60), undesirable (.60< < .65), minimally acceptable (.65< < .70), respectable (.70< < .80), very good (.80< < .90), and unnecessarily high such that one should consider shortening one's scale (.90<)" (Clifton, 2019, p. 2).

DATA ANALYSIS AND RESULTS

Two hundred twelve surveys were exported from Qualtrics. To begin, I screened and cleaned my raw data for potential missing items. Although my survey received 212 responses, I encountered a large number of empty entries. In the downloaded spreadsheet of raw data, 61 responses for the SSOS–TR items were blank. The remaining 151 responses were complete, excluding demographic questions. Of note, this was no longer in adherence to the general guidelines for sample size adequacy and will be discussed as an important limitation in the conclusion of this report. The data set of 151 complete cases was then imported for analysis into RStudio, a web–based program for statistical computation and graphics. The data that support the findings of this study are available from the author upon reasonable request.

EXPLORATORY FACTOR ANALYSIS CRITERIA

I looked first to the Kaiser–Mayer–Olkin (KMO) test as computed from the raw correlation matrix, which ranges from 0 to 1. The KMO is an index for analyzing the sum of the partial correlations relative to the sum of the correlations (McCoach et al., 2013). The KMO for my data was 0.91, which is above the .90 cut off for what would be considered a "marvelous" KMO in deciding whether an EFA is appropriate (i.e., the factor analysis will produce factors that are reliable and distinct; Kaiser & Rice, 1974). Similarly, the Bartlett Test of Sphericity endorsed the use of an EFA. Bartlett's Test proposes a null hypothesis that the correlation matrix computed from raw data would be equivalent to an identity matrix with ones on the diagonals and zeros on the off–diagonals. Rejecting this null hypothesis would mean that the data's correlation matrix is statistically significantly different from an identity matrix, which was the case with my data. The 2 value was high at 2038.21 with 171 df and a p–value <.001, indicating that an EFA would be appropriate.

In contrast were the results of the Measure of Sampling Adequacy (MSA), which is an indicator of the strength of an item's correlation with other items in the correlation matrix

(McCoach et al., 2013). The MSA did not fully endorse the appropriateness for an EFA. Although items all had MSAs above .70 and off–diagonals on the anti–image matrix were small, the diagonals on the anti–image matrix were all below .50, ranging from .23 to .43.

Bolstered by the support of the KMO and the Bartlett's Test, I did ultimately decide to proceed despite this concerning MSA and an inadequate sample size. Collecting more responses prior to completing my analyses would have been preferrable; however, I conducted this research as coursework within a doctoral program of study during the spring semester of 2022 and, thus, was under strict time constraints. For these reasons, I proceeded with caution when interpreting results.

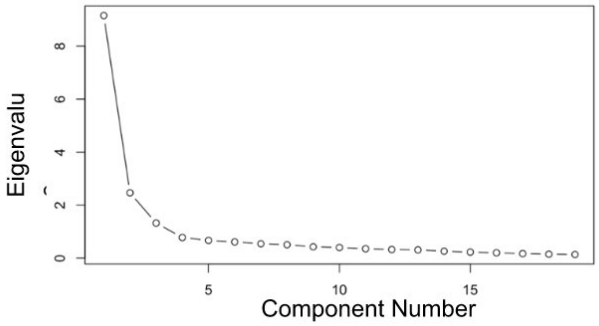
FACTOR EXTRACTION PROCESS

Next, I ensured that the means of the responses for each item fell within the scale options (i.e., 1–6), that the standard deviations demonstrated adequate variability (i.e., were not equal to 0), and that each item's histogram was mostly normally distributed. All of these criteria were met. Although there was indication of skewness on some of the item histograms, none were flagged as bimodal. Nothing stood out as an outlier in the data summaries and no coding errors were identified when examining item ranges. Upon examining inter-item correlations, I did not find any issues regarding low item correlations nearing zero or above 0.90. All the item correlations were positive, confirming that reverse coding would not be necessary moving forward in the EFA.

The first criterion I used to extract factors was Kaiser's criterion (Kaiser, 1958), which explores eigenvalues (i.e., the roots in a correlation matrix) by running a principal components analysis. Kaiser's criterion suggests that all factors with eigenvalues greater than or equal to 1.0 be retained, as these factors are reasoned to account for more variance than a single item (McCoach et al., 2013). The output for my data produced three factors with eigenvalues greater than 1.0. The screen test is a visual form of analysis where eigenvalues on the y–axis are plotted against factor numbers on the x–axis. The plot is then analyzed by looking at the point at which the curve straightens out, colloquially referred to as the "elbow," and using the x–axis label just prior to the elbow as the number of factors to extract (McCoach et al., 2013). My judgment was that the recommended extraction from the scree plot (see Figure 1) was three factors.

Figure 1

Scree Plot from the EFA indicating Three Factors for the SSOS-TR



Another approach is to look to the proportion of variance explained by each factor in a preliminary principal components analysis. The biggest drop in explained variance was between one (48% of the variance) and two (13% of the variance), suggesting a one factor extraction.

As a departure from the non–statistical methods above, parallel analysis is a statistical method of factor extraction. Average eigenvalues are calculated from random data (i.e., a no-factor dataset) and compared to the eigenvalues in the actual dataset. When sample eigenvalues are greater than the random data average eigenvalues, the conclusion is that there exists a 'true' factor (Fabrigar & Wegener, 2012). With my dataset, the parallel analysis suggested the extraction of three factors. Another statistical approach to factor extraction is the Very Simple Structure (VSS) analysis, which provides the optimal number of factors to extract in correspondence with differing levels of complexity. The VSS complexity was highest (i.e., .95) with two factors.

The last statistical approach I explored was Velicer's minimum average partial (MAP) procedure. Velicer's MAP partials out common variance by extracting one component at a time, calculating the average squared partial correlations, creating a new residual matrix, and repeating the process until the smallest squared partial correlation is calculated and only unique variance remains (Velicer, 1976). I utilized Velicer's MAP procedure in accordance with the revision made by Velicer et al. (2000) to use average partial correlations raised to the fourth power instead of squared. With my data, the Velicer MAP achieved a minimum with three factors.

FACTOR EXTRACTION PROCESSES

In summary, the above recommendations for number of extracted factors were one factor (endorsed by the preliminary principal components analysis), two factors (endorsed by the VSS procedure), and three factors (endorsed by Kaiser's criterion, the scree plot elbow method, the parallel analysis procedure, and the Velicer MAP procedure). While the overwhelming endorsement was for three factors, I decided to be thorough by running separate EFAs using one–factor, two–factor, and three–factor models in order to ultimately

decide. For my models, I used an oblimin oblique rotation, as the factors would be correlated given that constructs composing self–determination are related to one another (McCoach et al., 2013).

ONE–FACTOR EXTRACTION MODEL

The one–factor EFA produced five communalities (h2) below .40; given that communalities are interpreted as the amount of variance in each item that is explained by the factor, having few items with low communalities indicates that the majority of the variance in the items is explained by this one factor. Items with low communalities were Items 1, 3, 6, 11, and 13. I flagged these items to attend to when investigating pattern coefficients to see whether they also ended up with low loadings. The output's "fit based upon off diagonal" of .93 was below the .95 cutoff, indicating that the one–factor model is statistically speaking not a good–fitting model.

The one-factor EFA model's sum of squares (SS) loading was 8.63, although a factor eigenvalue above one is to be expected in a one–factor model. At the item level, all items had factor loadings above .40, which is "the most popular cutoff for 'good' factor loadings onto a primary factor" and "it is recommended that satisfactory variables load onto their primary factor above 0.40" (Howard, 2016; p. 55). The factor loadings ranged from .513 to .778; thus, no items in the one–factor model were candidates for deletion based on this criterion. Next, I examined the residuals in the one–factor EFA model. One hundred eleven residuals were greater than the .05 cutoff, which indicated that 65% of items had large residuals. This proportion is above the 50% cutoff for concern. The mean residual was .12, which is above the .08-.10 cutoff for concern. By contrast, the histogram of residuals appeared approximately normally distributed with no outliers. Overall, a one–factor model was not a statistically good fitting model, nor is it conceptually meaningful considering that Causal Agency Theory defines three essential characteristics of self–determined individuals.

TWO–FACTOR EXTRACTION MODEL

The two–factor EFA produced no communalities below .40; all the variance in the items was explained by these two factors and thus, they are satisfactory indicators. The fit based upon off diagonal of .98 was above the .95 cutoff, indicating that the two–factor model is a statistically good–fitting model and a better fitting model than the one–factor model. Also, both of the two-factor EFA model's SS loadings were above one (6.44, 4.35).

All primary factor loadings for the two–factor EFA were above .40, ranging from .42 to .86. However, two of the items had loadings above 0.30 on more than one factor. McCoach et al. (2013) "recommend eliminating any item that has a loading of 0.30 or higher on more than one factor" (p. 143); therefore, in a two–factor EFA, items 16 and 17 would be potential candidates for elimination. Next, I examined the residuals. There were more large residuals in this two–factor model than in the one–factor model, as 141 residuals were greater than the .05 cutoff. Eighty-two percent of the residuals in the two-factor model were large, which is a proportion notably above the 50% cutoff for concern. The

mean residual of .19 was above the .08-.10 cutoff, meaning that a model with more factors might be preferred. Although the histogram of residuals appeared approximately normally distributed, there was some indication of multiple peaks. Therefore, the overall residual output in the two–factor model was of concern.

THREE–FACTOR EXTRACTION MODEL

The three–factor EFA produced no communalities below .40. Also, all of the three–factor EFA model's SS loadings were above one (4.33, 4.16, and 3.39, respectively). Its fit based upon off diagonal of .99 was above the .95 cutoff, indicating that the three–factor model was a statistically good–fitting model and also a better fitting model than the one– and two–factor models on this index. The three–factor model was what I hypothesized during my scale construction and content validation phases, so its primacy as the best fitting model was exciting.

The three–factor EFA had one item with a primary factor loading below than 0.40. This was item 15 (i.e., "I have my students identify the supports they need") and its primary factor loading was .371. Eight of the items had loadings above 0.30 on more than one factor. Next, I examined the residuals in the three–factor EFA model. There were 165 large residuals. Ninety-six percent of the residuals in the three–factor model were large, which is a proportion well above the 50% cutoff. Additionally, the mean residual was .22, which exceeded the .08-.10 cutoff. However, the histogram of residuals did appear approximately normally distributed.

FACTOR EXTRACTION DECISION

I made the decision to select the three–factor model for extraction based on statistical evidence and theory. The main evidence for this decision was that the majority of methods suggested a three–factor extraction. Four out of the six methods I employed (i.e., Kaiser's criterion, scree plot elbow method, parallel analysis, and MAP analysis) all suggested three factors. MAP analysis is largely considered the best method with the next best method being parallel analysis, as both of these are statistical in nature and therefore less subjective. That being the case, having both of the best statistical methods for extraction suggest three factors, in addition to having two other methods also in agreement, makes a strong case for my decision. Also, the three–factor had the best 'fit based upon off–diagonal' of all the three models at .99.

Aside from statistical evidence, the three–factor model conceptually makes the most sense. A three–factor model is in alignment with the Causal Agency Theory, which identifies three essential characteristics of self–determined action. When I compared the item clusters of the two–factor model to the item clusters of the three–factor model, the three–factor model clusters made more conceptual sense in their groupings. Interestingly, both the two– and three–factor models had the exact same items grouped in Factor 2: item 1 ("I ask my students what classroom activities they like"), item 2 ("I provide time for students to think about their goals"), item 3 ("I create opportunities for students to explore their interests"), item 11 ("I teach students to make choices they feel good about"),

item 12 ("I assist my students to align their goals with their interests"), and item 19 ("I provide opportunities for my students to make their own choices"). With that factor's items grouped equivalently across the two– and three–factor models, the other factor containing all remaining items in the two–factor model did not make conceptual sense. However, in the three–factor model, the remaining items were clustered in a way that made conceptual sense when analyzing the individual items.

The units of factor loadings in an EFA represent each item's association with the factor(s). While some items were only associated with a single factor, other items were associated with more than one factor (i.e., double or triple loadings). Given that constructs composing self–determination are related to one another, the factors in this EFA would be correlated. Therefore, an oblique oblimin rotation was the most appropriate rotation technique in this case, as it allows for intercorrelations among the factors (Fabrigar & Wegener, 2012). The factor pattern matrix from the selected three–factor EFA is below in Table 1 with conceptual labels of the subscales.

Table 1

Factor Pattern Matrix from for a Three-Factor Extraction of the SSOS-TR

Note. N=151. The exploratory factor analysis extraction method was principal axis factoring with an oblimin oblique rotation. Factor loadings above .40 are in bold. Factor pattern matrix sorted by size using the fa.sort() and fa.organize() functions in RStudio for ease of comprehension.

	SSOS-TR item	Factor loading		
		1	2	3
Factor 1: Acting on Goals				
Q10	I have my students identify paths to their goals.	**0.892**		
Q04	I teach my students how to monitor progress towards their chosen goal.	**0.714**	0.156	
Q08	I provide opportunities for my students to make their own plans.	**0.668**		0.122
Q13	I give time for my students to reflect on their skills.	**0.628**	-0.130	0.260
Q14	I encourage my students to find their own ways to meet goals.	**0.552**		0.384
Q09	I schedule time for my students to work on their goals.	**0.411**	0.264	0.272
Q05	I provide opportunities for my students to take action.	**0.409**	0.229	0.306
Q15	I have my students identify the supports they need.	0.371	0.189	0.342
Factor 2: Deciding on Actions				

	SSOS-TR item	Factor loading		
		1	2	3
Q19	I provide opportunities for my students to make their own choices.		**0.826**	0.161
Q01	I ask my students what classroom activities they like.		**0.789**	
Q03	I create opportunities for students to explore their interests.		**0.766**	
Q11	I teach students to make choices they feel good about.		**0.721**	0.236
Q12	I assist my students to align their goals with their interests.	0.320	**0.632**	
Q02	I provide time for students to think about their goals.	**0.460**	**0.587**	-0.208
Factor 3: Believing in Self				
Q18	I teach my students that meeting a goal requires effort.	0.168		**0.796**
Q17	I provide opportunities for students to practice self-regulation.	-0.122	0.387	**0.678**
Q07	I give time for my students to confront obstacles.		0.253	**0.642**
Q06	I teach my students that there is more than one path towards a goal.	**0.414**	-0.191	**0.553**
Q16	I provide opportunities for my students to work towards goals.	0.261	0.300	**0.447**

ITEM RETENTION DESCRIPTION

Item 15 had a primary factor loading of .371 and was flagged as a potential candidate for deletion given that this value was below the .40 cutoff. However, .371 is not very far off from .40 and, using another set of heuristics, it would not necessarily be flagged for deletion. Eight of the items in my selected three–factor model had loadings above 0.30 on more than one factor, as seen in the factor pattern matrix in Table 1. While these double and triple loadings were not ideal, I did not recommend any of these items for deletion. Seven of these eight items had primary factor loadings above .40, which serve as protective factors against item deletion. More importantly, I did not delete these items given my inadequate sample size and the caution with which I interpreted results from the EFA. As previously mentioned, the Measure of Sampling Adequacy (MSA) analysis indicated that an EFA was not fully appropriate. Therefore, it would not be wise to delete items from a piloted instrument with factors extracted from a potentially inappropriate exploratory factor analysis. All original 19 items on the SSOS-TR were retained. See Appendix for the final clean copy of the SSOS-TR information sheet and instrument.

SUBSCALES

SUBSCALE DESCRIPTIONS AND INTERPRETATIONS

Subscale 1, "Acting on Goals," contains items relating to how special educators perceive they provide opportunities for their students labeled with disabilities to be agents in their own lives. The items speak to how teachers perceive they facilitate students' abilities to pursue their goals through individual actions, including identifying how to achieve the goal, how to plan for achieving the goal, how to secure supports towards goal attainment, and how to measure goal progress. A high score on the Acting on Goals subscale means that the special educator perceives that they provide frequent, meaningful, and intentional opportunities for their students to be the agentic selves in their lives. High scorers perceive that they schedule time for students to plan, carry out, and monitor their actions towards their individual goals and that they provide students with the encouragement, instruction, and supports to do so. High scorers on the scale tend to strongly agree with items. A low score on the Acting on Goals subscale means that the special educator provides limited opportunities for their students to act as agentic selves in their own lives, likely planning pathways towards goal achievement for the students, offering teacher–identified and selected supports, and monitoring the students' progress for them. Low scorers on the scale tend to strongly disagree with items.

Factor 2, "Deciding on Actions," contains items that relate to how special educators perceive they provide opportunities for SWD to be the center of decision–making processes for their individualized actions based on their own interests. The items capture educator self–perceptions on the degree to which students' preferences drive decisions necessary to achieve goals, the degree to which decisions are the students' own, and the degree to which these decisions are ones that students feel good about making. A high score on the Deciding on Actions subscale indicates that the special educator has a minimized role on students' choices and that decisions are led by students as guided by the students' own interests. High scorers on the scale tend to strongly agree with items. A low score on the Deciding on Actions subscale indicates that the special educator makes decisions for students based on what they think the students would prefer and that the educator chooses goals they believe to be in the students' best interest. Low scorers on the scale tend to strongly disagree with items.

Factor 3, "Believing in Self," contains items that relate to how special educators perceive they provide opportunities for SWD to recognize that they have the capacity to develop skills and knowledge that will facilitate their goal attainment. The items capture how the educators foster the students' awareness that effort, obstacles, and forking paths are normal parts of the journey towards reaching goals and that they can overcome barriers and/or identify alternative routes through commitment, effort, and persistence. A high score on the Believing in Self subscale indicates that the special educator works to develop students' mindsets on how to persevere when faced with challenges. High scorers on the scale tend to strongly agree with items. A low score on the Believing in Self subscale indicates that the special educator devotes little time and attention towards fostering students' awareness of their own skills and have a more rigid, pass/fail approach to goal work. Low scorers might also be educators who set goals for students below their zone of proximal development (i.e., goals in areas where the student is already functioning independently) and generally tend to disagree with items.

These three names and verbal descriptions of subscales align well with the three essential characteristics of Causal Agency Theory. This finding affirms that my item generation and content validation processes prior to piloting the instrument were adequate. Recent communication with one of my expert reviewers alerted me to the fact that the three characteristics have been assigned more practitioner-friendly labels of Decide (formally Volitional Action), Act (formally Agentic Action), and Believe (formally Action–Control Beliefs; Bojanek et al., 2021). I drew from this knowledge when considering the clusters in my EFA but felt it useful to slightly expand upon the names in the case of the SSOS–TR instrument.

SUBSCALE INTERNAL CONSISTENCY RELIABILITIES

With the three–factor decision made, I ran the reliability analysis. The overall mean for the whole SSOS–TR instrument was 4.67 with a standard deviation of .76. I then ran the codes in order to obtain each subscale's Cronbach's alpha values, average inter–item correlations (IIC), and the standard deviations of these IICs for the subscales in Table 2.

Table 2

Reliability Analyses for Subscales 1-3 on SSOS-TR Instrument

	Subscale Name	# Items	Items	Subscale Alpha	Mean of IIC	SD of IIC
1	Acting on Goals	8	4, 5, 8-10, 13-15	.91 (.88, .93)	.60	.16
2	Deciding on Actions	6	1-3, 11, 12, 19	.90 (.87, .92)	.66	.16
3	Believing in Self	5	6, 7, 16-18	.87 (.83, .90)	.66	.19

Note. N=151. "SD" = Standard deviation. "Mean of IIC" = mean of the inter–item correlations for each scale. "SD of IIC" = standard deviation of the inter–item correlations for each scale.

The first subscale for Acting on Goals yielded a high alpha of 0.91 and a narrow 95% confidence interval of (0.88, 0.93). Although an alpha of .91 is above the .90 cutoff that indicates a subscale could potentially be shortened, I did not make any recommendations for deletions for the aforementioned reasons of inadequate sample size and failure to meet all criteria for determining that an EFA was appropriate. The second subscale for Deciding on Actions yielded an alpha of 0.90 with a narrow 95% confidence interval of

(0.87, 0.92); this alpha at 0.90 would be considered right on the line between very good and unnecessarily high in terms of inter–item consistency. The third subscale for Believing in Self yielded a very good alpha of 0.87 and a narrow 95% confidence interval of (0.83, 0.90).

The mean of the inter–item correlations for all subscales were all non–zero and below .90, which is a positive finding signifying that no single item's deletion would have significantly improved its subscale's alpha. Of note, a higher alpha would not be desired for Acting on Goals and Deciding on Actions given their already high to unnecessarily high Cronbach's alpha values. All corrected item–total correlations for items on each of the three subscales were similar to each other as desired and well above the .30 cutoff of concern.

Overall, the three subscale alpha coefficients were very good to high (.91, .90, and .87, respectively). Therefore, there is evidence that subscales on the SSOS–TR are internally consistent. The three subscales comprising the SSOS-TR had positive correlations that were moderate in strength, ranging from r=.32 to r=.49 (i.e., non-zero and below .90). These are good findings, as high correlations between factors (i.e., near or above .80) can signal issues with discriminant validity (Brown, 2015).

SUBSCALE DESCRIPTIVE STATISTICS

The subscale means and standard deviations are summarized in Table 3 below. All subscale ranges were as expected (i.e., from 1-6). While Factor 1 Acting on Goals and Factor 3 Believing in Self had approximately normally distributed histograms, the histogram for Factor 2 Deciding on Actions had some evidence of negative skewness.

Table 3
Scale Means and Standard Deviations for Each of the Subscales on SSOS–TR Instrument

	Subscale Name	Subscale Mean	Subscale SD
1	Acting on Goals	4.38	1.14
2	Deciding on Actions	4.81	1.07
3	Believing in Self	4.97	1.13

Note. N=151. SD= standard deviation.

CONCLUSIONS & IMPLICATIONS

Overall, the three–factor model that was endorsed by the majority of EFA extraction criteria was aligned with my hypothesized model, which was informed by the literature on Causal Agency Theory. This is a promising finding and provides some evidence for construct validity of the SSOS–TR. Although five of my written items did not end up clustering exactly as they were intended in the creation stage, fourteen of the items (i.e., 74%) did cluster together as hypothesized, which is promising. This high confirmation of cluster hypotheses is in large part due to the influence of reviewers. Recruiting and gathering feedback from content experts is not a step in the design process for which there are many clear and structured guidelines. In my experience, it was helpful to first send individual emails asking if the expert would be available to provide feedback and including a brief description of my personal research interests. I then provided individualized content validation forms to willing experts only after they agreed to volunteer their time. Another suggestion is to provide content experts with a specific date by which to return their feedback. More broadly, I recommend adopting the mindset that this expert feedback will lead to a better product, however different from one's original line of thinking. In my case, vulnerability to share drafted work and willingness to acknowledge errors to remediate potential harms to vulnerable, culturally diverse, and/or marginalized populations made for a tool in which I was more confident would benefit all students.

Another implication from my results can be found in the subscale means. My three subscales (i.e., Acting on Goals, Deciding on Actions, and Believing in Self) had item means of 4.38, 4.81, and 4.97. Given that the scale range was from one to six with no neutral middle option, these are all high means. New data would be needed to draw conclusions from subscale means of SSOS–TR scores, as these pilot data were collected during an instrument development process and are limited in their interpretation to factor and reliability analyses. With new data, implications could be made regarding the extent to which special educators perceive they are affording opportunities for SWD to become self–determined and, thus, act as causal agents in their own lives. High scores on items would imply that educators perceive they foster students' abilities to (1) decide on actions, (2) act on goals, and (3) believe in their abilities as causal agents in their own lives.

LIMITATIONS

There were several limitations that should be considered when evaluating the findings from this SSOS–TR instrument pilot. Most notably, the sample size was not adequate and fell below the recommendation for a 10:1 ratio of respondents to items (Boateng et al., 2018). With this logic, for a 19–item survey, the minimum acceptable number of respondents would have been 190; however, I only obtained a sample of 151 complete cases after the initial 212 respondent entries had been screened and cleaned for missingness. The sample size also falls below the general suggestion of securing an appropriate sample size in the 200-300 range in order to run an EFA (Boateng et al., 2018). The impact of this small sample size was that an EFA was not deemed fully appropriate by all of the criteria. For instance, the Measure of Sampling Adequacy (MSA) indicated that an EFA was not appropriate. As indicated previously, additional response

collection did not occur because this research was conducted as part of a doctoral course that was limited to one semester. I proceeded with the EFA because the sample size was sufficient for the software to produce output and because the Kaiser–Mayer–Olkin (KMO) test and the Bartlett Test of Sphericity both endorsed the use of an EFA with my data.

The 151 complete responses did not come from a random sample but rather came from non–probability, convenience sample using a snowball sampling outreach strategy and was thus not a representative sample of special educators. My colleagues who assisted in the snowball sampling were similar to me in terms of years of experience, meaning that the sample likely overrepresented the attitudes of more veteran teachers.

Prior to running the pilot, I only amassed the opinions of seven experts. Perhaps the SSOS–TR would have been composed of different or differently–worded items had there been more collective input from a larger sample of experts in the field. Also, I alone generated the SSOS–TR items on the initial tool sent out to experts for review without seeking input from special educators in focus groups or through cognitive interviews. Cognitive interviewing procedures would "add important qualitative information to ensure that each item on the survey is working as it should" (Ouimet et al., 2004; p. 234) and would enhance the content validity argument for the SSOS–TR.

Another limitation is with regards to the assumption that subscale items had a normal distribution. When screening my data, I found some indication of negative skewness in a few of the item histograms and, thus, a violation of the assumption underlying the use of Cronbach's alpha. Cronbach's alpha also assumes that items for each factor have equal factor loadings, which was not the case on the SSOS–TR. To test whether or not loadings could be constrained as equal at the subscale level would require a Confirmatory Factor Analysis. Therefore, interpreting subscale–level Cronbach's alpha values for the SSOS–TR should be done with caution.

SUGGESTIONS FOR FUTURE RESEARCH

If the SSOS–TR were rerun with an adequate sample size of at least 190 to achieve the minimum 10:1 ratio of respondents to items (Boateng et al., 2018), an EFA would be considered appropriate. The Measure of Sampling Adequacy (MSA) criterion with a sample of this size would also likely endorse the appropriateness of an EFA. Additionally, if reviewed by more experts, the SSOS–TR could make important contributions to the field of self–determination research specifically and to the field of special education more generally. An interesting question would be to compare these findings from the SSOS–TR to ratings of teacher behaviors measured through direct observation. Are educators accurate reporters of their own teaching behaviors?

I originally set out to design this instrument for students labeled with disabilities to measure their perceptions of the degree to which they are afforded opportunities by their teachers to become self–determined; unfortunately, given the IRB constraints under which the SSOS–TR was piloted, this was not possible. Moving forward, the SSOS–TR can serve as an important complement to a parallel student report version. Students labeled with disabilities are the experts on their own lives, so comparing the results of

the teacher and student versions would potentially lead to interesting findings on items that are discrepant. For instance, if teachers perceive that they are providing students with opportunities to make their own choices (i.e., item 19), but the students themselves perceive that they are not afforded the opportunity to make their own choices, what does this discrepancy in perceptions mean? How can exposing such discrepancies improve teacher pre– and in–service training and professional development?

Another area for future research would be to analyze findings of the SSOS–TR and, ideally, the parallel student version disaggregated by disability category to investigate whether opportunities to become self–determined are proportionally provided across disability categories. If not, why? Disaggregating by age of students would also be interesting. Are high school students labeled with disabilities afforded more self–determination opportunities than are elementary students? If so, why? Is having a family member labeled with a disability associated with higher teacher perceptions of the opportunities they afford to their own students labeled with disabilities? Answers to all of these questions would lead to important findings that could, ultimately, serve to inform pre-service teacher preparation programs and ongoing professional development programs. If differences emerge, for example, between disability categories and/or between age groups of students, these could be redressed through targeted interventions and supports for educators so that opportunities to become self–determined are afforded equitably to all students with disability labels.

REFERENCES

Algozzine, B., Browder, D., Karvonen, M., Test, D. W., & Wood, W. (2001). Effects of interventions to promote self-determination for individuals with disabilities. Review of Educational Research, 71, 219–277. https://doi.org/10.3102/00346543071002219

Angyal, A. (1941). Foundations for a Science of Personality. The Commonwealth fund.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. Frontiers in Public Health, 6, 149–149. https://doi.org/10.3389/fpubh.2018.00149

Bojanek, E. K., Raley, S. K., Shogren, K. A., & Lane, K. L. (2021). Examining the impact of professional development on the Self-Determined Learning Model of Instruction for general and special educators. Inclusion, 9(2), 118-133. https://doi.org/10.1352/2326-6988-9.2.118

Brown, T. A. (2015). Confirmatory Factor Analysis for Applied Research (2nd ed.). New York, NY: Guilford Press.

Burke, K. M., Raley, S. K., Shogren, K. A., Hagiwara, M., Mumbardó-Adam, C., Uyanik, H., & Behrens, S. (2020). A meta-analysis of interventions to promote self-determination for students with disabilities. Remedial and Special Education, 41(3), 176–188. https://doi.org/10.1177/0741932518802274

Chambers, C. R., Wehmeyer, M. L., Saito, Y., Lida, K. M., Lee, Y., & Singh, V. (2007). Self-determination: What do we know? Where do we go? Exceptionality: The Official Journal of the Division for Research of the Council for Exceptional Children, 15(1), 3–15. https://doi.org/10.1080/09362830709336922

Clifton, J. D. (2019). Managing validity versus reliability trade-offs in scale-building decisions. Psychological Methods, 25(3), 259–270. https://doi.org/10.1037/met0000236

Cobb, B., Lehmann, J., Newman-Gonchar, R., & Alwell, M. (2009). Self-determination for students with disabilities: A narrative metasynthesis. Career Development for Exceptional Individuals, 32(2), 108–114. https://doi.org/10.1177/0885728809336654

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297–334. http://doi.org/10.1007/BF02310555

DeVellis, R. F. (2003). Scale Development: Theory and Applications (Vol. 26). London, UK: Sage.

Fabrigar, L. R., & Wegener, D. T. (2012). Exploratory Factor Analysis. New York: Oxford University Press.

Flora, D. B., & Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale development

and validation. Canadian Journal of Behavioural Science, 49(2), 78–88. https://doi.org/10.1037/cbs0000069

Fowler, C. H., Konrad, M., Walker, A. R., Test, D. W., & Wood, W. M. (2007). Self-determination interventions' effects on the academic performance of students with developmental disabilities. Education and Training in Developmental Disabilities, 42(3), 270–285.

Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? International Journal of Human-Computer Interaction, 32(1), 51–62. https://doi.org/10.1080/10447318.2015.1087664

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. Psychometrika, 23, 187–200. https://doi.org/10.1007/BF02289233

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark Iv. Educational and Psychological Measurement, 34(1), 111–117. https://doi.org/10.1177/001316447403400115

Konrad, M., Fowler, C. H., Walker, A. R., Test, D. W., & Wood, W. M. (2007). Effects of self-determination interventions on the academic skills of students with learning disabilities. Learning Disability Quarterly, 30, 89–113. https://doi.org/10.2307/30035545

Krosnick, J. A., Narayan, S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. New Directions for Evaluation, 1996(70), 29–44. https://doi.org/10.1002/ev.1033

Malian, I., & Nevin, A. (2002). A review of self-determination literature: Implications for practitioners. Remedial and Special Education, 23(2), 68–74. https://doi.org/10.1177/074193250202300202

McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). Instrument Development in the Affective Domain: School and Corporate Applications. (3rd ed.). Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-7135-6

Nota, L., Ferrari, L., Soresi, S., & Wehmeyer, M. (2007). Self-determination, social abilities and the quality of life of people with intellectual disability. Journal of Intellectual Disability Research, 51, 850–865. https://doi.org/10.1111/j.1365- 2788.2006.00939.x

Ouimet, J. A., Bunnage, J. C., Carini, R. M., Kuh, G. D., & Kennedy, J. (2004). Using focus groups, expert advice, and cognitive interviews to establish the validity of a college student survey. Research in Higher Education, 45(3), 233–250. https://doi.org/10.1023/B:RIHE.0000019588.05470.78

Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. The American Psychologist, 55(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sadler, G. R., Lee, H.-C., Lim, R. S.-H., & Fullerton, J. (2010). Recruitment of hard-to-reach population subgroups via adaptations of the snowball sampling strategy. Nursing & Health Sciences, 12(3), 369–374. https://doi.org/10.1111/j.1442-2018.2010.00541.x

Sankofa, N.L. (2021). Transformativist measurement development methodology: A mixed methods approach to scale construction. Journal of Mixed Methods Research, 1–21. https://

doi.org/10.1177/15586898211033698

Shogren, K. A. (2013). Positive psychology and disability: A historical analysis. In The Oxford Handbook of Positive Psychology and Disability. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195398786.013.013.0003

Shogren, K. A., Palmer, S. B., Wehmeyer, M. L., Williams-Diehm, K., & Little, T. D. (2012). Effect of intervention with the Self-Determined Learning Model of Instruction on access and goal attainment. Remedial and Special Education, 33, 320–330. https://doi.org/10.1177/0741932511410072

Shogren, K. A., Wehmeyer, M. L., Little, T. D., Forber-Pratt, A. J., Palmer, S. B., & Seo, H. (2017). Preliminary validity and reliability of scores on the Self-Determination Inventory: Student Report version. Career Development and Transition for Exceptional Individuals, 40(2), 92–103. https://doi.org/10.1177/2165143415594335

Shogren, K. A., Wehmeyer, M. L., Palmer, S. B., Forber-Pratt, A. J., Little, T. D., & Lopez, S. (2015). Causal agency theory: Reconceptualizing a functional model of self-determination. Education and Training in Autism and Developmental Disabilities, 50(3), 251–263.

Shogren, K. A., Wehmeyer, M. L., Palmer, S. B., Rifenbark, G. G., & Little, T. D. (2015). Relationships between self-determination and postschool outcomes for youth with disabilities. The Journal of Special Education, 48, 256–267. https://doi.org/10.1177/0022466913489733

Spearman, C. (1910). Correlation calculated from faulty data. British Journal of Psychology, 3, 271–295. http://doi.org/10.1111/j.2044- 8295.1910.tb00206.x

U.S. Department of Education, National Center for Education Statistics. (2020). National Teacher and Principal Survey (NTPS), "Public School Teacher Data File," 2017-18. Retrieved July 2022 from https://nces.ed.gov/programs/digest/d20/tables/dt20_209.50.asp

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. Psychometrika, 41(3), 321–327.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), Problems and Solutions in Human Assessment: Honoring Douglas Jackson at Seventy (pp. 41–71). Boston: Kluwer.

Wehmeyer, M. L. (1992). Self-determination and the education of students with mental retardation. Education and Training in Mental Retardation, 27, 302–314.

Wehmeyer, M. L. (1995). The Arc's Self-Determination Scale: Procedural Guidelines.

Wehmeyer, M. L., & Kelchner, K. (1995). The Arc's Self-Determination Scale. Arlington, TX: The Arc National Headquarters.

Wehmeyer, M. L., Kelchner, K., & Richards, S. (1996). Essential characteristics of self-determined behavior of individuals with mental retardation. American Journal of Mental Retardation, 100(6), 632–642.

Wolman, J., Campeau, P., Dubois, P., Mithaug, D., & Stolarski, V. (1994). AIR Self Determination Scale and User Guide. Palo Alto, CA: American Institute for Research.

Wood, W. M., Fowler, C. H., Uphold, N., & Test, D. W. (2005). A review of self-determination interventions with individuals with severe disabilities. Research & Practice for Persons with Severe Disabilities, 30(3), 121–146. https://doi.org/10.2511/rpsd.30.3.121

MIND THE GAP: CAN PODCASTS HELP BRIDGE THE DIVIDE BETWEEN EDUCATION RESEARCH AND CLASSROOM PRACTICE?

SARAH GILMORE
UNIVERSITY OF CONNECTICUT NEAG SCHOOL OF EDUCATION

ABSTRACT

There exists a significant and consequential divide between educational researchers and education practitioners. The purpose of this persuasive essay is to examine the reasons for this divide through the lenses of attitudes, access, and audience, and argue in favor of a podcast designed for an audience of teachers, based in the Neag School of Education. Drawing on adult learning theory, narrative theory, and comparative international research on teacher demographics and barriers to research engagement, there is evidence that such a podcast could be a valuable tool for synthesizing and communicating research in an accessible and audience-appropriate way, and could be used to support reciprocal and collaborative relationships between researchers and the wider teaching community. I am a former elementary teacher from Scotland in the United Kingdom with 15 years of teaching experience in primarily international contexts. After transitioning into the role of technology integration coach for teachers, I became interested in the difficulties my colleagues and I faced in accessing education research to improve our practice. I went on to co-found a podcast, and began to provide teacher professional development workshops to address a need I perceived for teachers to access and implement evidence-based practices. As such, my position in this essay is informed by my experiences as a teacher, teacher-educator, and podcast creator. While there is a growing body of research on the effectiveness of podcasts as a tool for professional learning in medical education, podcasts as a tool for teacher learning have yet to be studied extensively. This essay contributes to existing literature by synthesizing research on podcast best practices and barriers to teachers' research-engagement, and describing ways in which the Neag School of Education could both contribute to the wider education community, and benefit from the development of an education research podcast.

Keywords: technology-enhanced adult learning, teacher research, teacher-researcher collaboration, science communication

USING PODCASTS TO BRIDGE THE DIVIDE BETWEEN EDUCATION RESEARCH AND CLASSROOM PRACTICE

Teaching is a unique career: vocation and profession; science and art.

The science lies in what we know and understand about how children learn—academically, socially, and emotionally. The art can be seen in how teachers draw on experience, instincts, and relationships to apply that science in real-world situations with large groups of complex individuals who, in an environment that is challenging and dynamic, may feel and behave differently at 10 am than they do at 2 pm. Both teachers and researchers have valuable, complementary knowledge and experience to contribute to the shared goal of improving education for students and teachers. However, although research shows that teachers have generally positive attitudes towards education research and are willing to engage in and with it, most teachers don't regularly or meaningfully draw on research when making decisions about their practice (Nelson et al., 2017), dividing the critical interplay between science and art.

While many aspects of teaching are highly contextualized (e.g., school systems, socio-cultural or political influences, curricula, and much more), in some ways teachers around the world have a great deal in common. The FIT-Choice Scale, a tool for capturing teacher's motivations for entering the career, for example, was originally developed by Watt et al. (2012), and has been replicated and validated across numerous, diverse national contexts. The results consistently show that the majority are motivated to become teachers by their perception that teaching has an intrinsic value, as well as their desires to contribute to society and work with children or young people. For many, in short, teaching is both a vocation–a calling for those who care deeply about children and about education–and a profession that attracts those who value lifelong learning. It may seem surprising, therefore, that despite these motivations, teachers do not engage more actively in and with education research. So, why is this the case, and what can be done about it?

As in other professions, such as medicine or science, the landscape of educational knowledge and practice is always changing. Research is continually conducted, and new ideas are tested in context. However, education differs from these other professions in one very critical way: there is often a fundamental and practical divide between research and practice. In medicine and science, for example, research is often carried out by doctors and scientists, with findings published in papers that are intended to be read and implemented by fellow doctors and scientists. In contrast, the much of the research in education is carried out by researchers for an audience of fellow educational researchers, but the findings are intended to be implemented by a completely separate group of practitioners: teachers.

This divide between researchers and practitioners is extremely problematic. Researchers care deeply about the work they do, and they invest decades of their professional lives into work they hope will make a positive impact on education. When done effectively, education research has the power to guide and advance new, more effective educational approaches as well as to debunk out-of-date information and potentially harmful approaches. However, misconceptions such as the idea that individuals learn better when they receive information in their preferred learning style, or that learners can be 'right-brained' or 'left-brained,' that have been soundly debunked in literature for several

years continue to be endorsed by teachers (Hughes et al., 2021). Likewise, it is generally accepted that the vast majority of teachers care deeply about their students, and that they are committed to their vocation. Yet, in spite of the benefits education research can bring to classroom practice, relatively few engage in or with education research in a consistent or meaningful way (Nelson & Campbell, 2017).

If the application of research findings is viewed as a function of the relationship between communities of education researchers and teachers (Farley-Ripple et al., 2018), the research-practice gap will not be bridged by asking only why teachers don't engage more with education research, or what researchers are doing at the individual level to make their research accessible to teachers. Instead, it will require dialogue and understanding between communities about the needs and strengths each bring with them, and a concerted effort to build systems and structures that support the creation of sustainable and meaningful relationships between research and practice. If those on the research side are to take a proactive and supportive role in this process, we must seek to first understand and appreciate three key factors that influence teachers' research engagement, and then to play an active role in addressing them: attitudes, access, and audience.

ATTITUDES

How teachers and researchers feel about their own and each other's roles, and the nature of research itself, is of key importance to this issue their use of research in practice. The traditional model of education research is largely unidirectional and top down: academics conduct research in and on classrooms, then publish papers about their findings with the presumed expectation that teachers or educational leaders will read them, synthesize the findings, and apply them in the classroom context. This situates teachers in the role of technicians following instructions rather than as partners with a mutual goal and complementary knowledge and skills. While it is by no means the case that all or even most education researchers conceptualize teaching in this way, the debate about what teaching is and should be has continued for decades (Winch, 2004) with little consensus. At the same time, while some academics champion practice-generated teacher research, still others are dismissive of its validity and potential impact (Nelson & Campbell, 2019). Meanwhile, although many teachers report positive attitudes towards research and express willingness to engage in and with it (Nelson et al., 2017), not all are in agreement as to what constitutes authentic research (Shkedi, 1998), and many others are skeptical about its value and relevance in classroom practice (Vanderlinde & van Braak, 2010).

However, those teachers and researchers who agree on both the importance of education research and the need to close the research–practice gap suggest that bridges need to be built between researchers and practitioners to create a more cooperative approach to education research (Vanderlinde & van Braak, 2010). One approach to this is the concept of teacher–researcher partnerships in which teachers and researchers work together on topics of mutual interest to co–create solutions (Simmers, 2021). This approach has gained traction in recent years and with good cause. Teachers are the experts when it comes to their students, and we rely on them to exercise their judgment in adapting and applying research in each highly–individualized classroom context.

However, we can and should think further than individual partnerships between teachers and researchers and more systemically about the relationship between teaching and research. Teachers have a responsibility and a right to engage with, and be engaged in and by, education research. By failing to forge a reciprocal relationship with the teaching community, education research stands to lose a vital opportunity to further education and education research as a whole.

ACCESS

If teachers are to systematically and regularly read and implement education research, they need to be able to access it. There are many factors that make education research inaccessible for teachers, including cost, experience, and support (Rycroft-Smith, 2022).

The traditional model of research dissemination assumes that individual teachers are in a position to digest, translate, and implement a large quantity of highly specific and often conflicting findings into classroom practice. However, most teachers do not turn to journals for information about teaching (Vanderlinde & van Braak, 2010), and, for everyone outside academia, including the majority of teachers, research is expensive to read. Without access to an institutional subscription to journal resources, most published research is hidden behind paywalls and although some research can be found freely through Google Scholar or is published open-access, not all teachers have the knowledge and experience to identify which papers are methodologically sound and/or relevant. Research is also usually written in academic English, which not only excludes teachers from other language backgrounds, but also those who may find the academic writing style excessively difficult or time-consuming to read.

Additionally, in order to draw meaningful understandings from research, teachers need to read many papers, synthesize their findings, and translate those into classroom practice. Not only does this process require a significant investment of time, something that few teachers feel they have enough of (vide infra), but also expertise to identify reliable, high-quality research that is relevant to their specific instructional context.

AUDIENCE

If teachers are to be a regular and engaged audience for education research, it is vital that researchers have a deeper and more nuanced understanding of who comprises that audience and the unique needs and barriers they face. The demographics of teachers, internationally, are another way teachers are more similar than different. The average K-12 teacher in the U.S. is 43 and female (Tai et al., 2019), and this distribution is reflected globally: around 70% of all teachers of this age group are women, reaching 96% in the early education years and 82% in elementary (OECD, 2019). This is relevant because decades of research have shown that women are disproportionately affected by both relative and absolute time poverty. Additionally, around half of all teachers are between the ages of 30 and 49 are likely to be parents, which further drives time poverty (Conway et al., 2021). Further, almost 25% of teachers in the U.S. are considering leaving the profession with stress being the most common reason for doing so (Woo & Steiner, 2021).

Therefore, it is of particular importance that education research should be presented in a way that is easy to access and consume and directly applicable to practice.

Considering these factors as a whole, we are presented with an audience for education research that is chronically time-poor and stressed; who lack the resources, experience, and support to access published research; and who are underrepresented in the broader education research community and process. So, what can we do?

While education research should and can be more democratic in all regards, one major step we can take is to rethink how we communicate and engage with teachers about published research. The traditional model of publishing academic papers in academic journals is appropriate for an audience of academics but not necessarily for teachers. Rather than simply publishing individual papers and hoping that teachers will find, make meaning from, and implement them, teachers need us to take the additional steps of translating a body of research on a topic of teacher-determined interest into clear recommendations for practice and presented in an accessible, sharable format that works for people who are time-poor and stressed.

PODCASTS MAY PROVIDE THE ANSWER

Podcasts have gained popularity as a mode for both entertainment and education in a variety of fields in recent years. One of the reasons for the popularity and efficacy of podcasts as a medium for learning might be seen in the relationship between how successful podcasts communicate information and how humans learn. Narrative Theory suggests that storytelling is a fundamental aspect of how humans make meaning about the world and ourselves (Rossiter, 1999). Fisher (1984) went so far as to give humans the label homo narrans, so central did he consider the role of storytelling in our evolution, and theorists from Bruner (1990), to Gottschall (2013) continue to highlight the role of storytelling in human experiences and perception. In education, specifically, the value of narrative learning in adult learning has been argued by researchers who have suggested that adult learning is multidimensional and contextual. Presenting information in a narrative format can make it easier for adults to encode new ideas and make connections between new and prior knowledge and experiences (Clark, 2001; Clark & Rossiter, 2008; Merriam, 2008; Rossiter, 1999).

A host of highly-successful podcasts, such as This American Life, which uses narrative, long-form journalism, the ever–increasing slew of true crime shows, and investigative journalism podcasts such as Serial, leverage the power of storytelling to garner vast audiences of listeners seeking entertainment. However, there are growing genres of podcasts focused on adult learning and translational science. For example, following a dramatic increase in both popularity and ubiquity, medical podcasts have become the most commonly used asynchronous learning resource for medical residents (Zhang et al., 2022). In addition to supporting formal adult learning, many podcasts use narrative methods to support informal adult learning and translational science. Shows such as Science Vs, Radiolab, and The Infinite Monkey Cage, for example, share robustly researched and reliable information that synthesize published research on a variety of topics packaged in an engaging and entertaining format to attract a broad listener–base.

Several attributes can be associated with successful podcasts. The PodCred framework, for example, provides an analytical framework consisting of the indicators listeners take into account when they assess the credibility of a podcast including the podcast's content, context, host, and technical execution (Tsagkias et al., 2009). For listeners, the credibility of a podcast is strongly related to the consistency of its topical focus, the structure and duration of episodes, the fluency and style of the host's speech and way of addressing the audience, and production quality, among several other factors. A recent scoping review of medical education podcasts found that listeners identified similar attributes in the podcasts they favored and, in line with the narrative learning framework, tended to prefer a conversational tone with personal anecdotes and humor (Kelly et al., 2022).

The ubiquity and success of general entertainment and scientific translation podcasts, as well as the success story of medical education podcasts, demonstrates the potential for podcasts to provide a solution to education research-practice divide. Several education–research focused podcasts exist, including The Harvard EdCast, Education Research Reading Room, Meet the Education Researcher, The Evidence Based Education Podcast, and BOLD (Blog on Learning Development) Podcast. However, although interviews with researchers feature heavily in this genre, the tone and style of these podcasts is typically more didactic than truly conversational. Furthermore, the topic is often geared more towards researchers than teachers, usually focusing on the research itself over its practical implications and applications. Episodes usually center on a single researcher or published paper rather than seeking to synthesize multiple sources to make generalizable meaning at the practical level and, like traditionally presented research, are usually unidirectional, passing information from the host to the listener.

To meet the needs of the teaching community, a new type of podcast is needed that curates, collates, and communicates research findings with an emphasis on practice, in a more engaging tone and style, through multi-modal forms of communication (audio, blog, discussion threads) to establish bidirectional communication between the hosts and listeners. In response to this need, I co–created a podcast (Gilmore & Deos, 2019–present), Little Key Podcast, with the aim of trying to find evidence-informed answers to the questions I knew teachers had, e.g.

- Does homework actually improve student attainment and achievement?
- Is the time I spend creating beautiful, colorful display–boards helping my students or distracting them?
- As COVID-19 forces schools online, how worried should I be about screen time?

By organizing episodes around topics of interest to teachers rather than individual publications or researchers, we ensured that our episodes were relevant to the teaching community and translated research into real-world, transferable contexts.

For each episode, my co-host and I invested the time to find published research that appeared to be reliable and relevant, and derived contextualized meaning from them by

synthesizing researchers' findings and recommendations into key points that teachers could consider and trial in their classrooms. We did interview some researchers but, when we did so, it was to seek their input on a wider topic and its implications for practice rather than to discuss a specific paper or the research process. This translational approach created an experience that was accessible and efficient for an audience of teachers for whom the point is not the research itself but what that research means for them, in their context.

We recorded each episode as a conversation that was intended to be enjoyable. We shared our personalities and perspectives and tried (but mostly failed) to keep our episodes reasonably short, the length of a commute or the time it takes to cook dinner, to make it possible for our audience to engage with research while multitasking. Knowing the needs and barriers that the majority of teachers have, it was important that episodes should be both enjoyable and digestible; the last thing teachers want or need is a listening experience that feels like more work.

As a complement to each audio episode, we created a website that featured long form blog posts that shared a deep–dive into the research behind the episode, similar in scope and purpose to a review tutorial, but written in plain-English with teachers, rather than researchers, in mind. The posts were intended for listeners who wanted more information and included direct links to both peer–reviewed and non–academic sources. Teachers are often faced with situations in which it would be beneficial to be able to share and discuss evidence–informed practices such as in parent conferences or team meetings. To support them in doing this, we also created downloadable PDF resources that could be shared in print, digitally, or over social media and embedded them in our blog posts. To help ensure our podcast was accessible, each episode was fully transcribed in English for people with hearing difficulties, and we invested time and resources into creating the highest level of audio production quality possible.

While this podcast was aligned with the majority of the attributes and indicators of the PodCred framework (Tsagkias et al., 2009) and had topical relevance to teachers, it lacked a reciprocal element: a discussion board, forum, or other mechanism for teachers to propose episode topics, react, and share how their implementation of the research worked in their classrooms. This is important because, by providing a place where teachers can share the resources they create and the approaches they trial, both teachers and researchers can learn from their successes and challenges, make connections between the outcomes of implementation and context, and find active participants from the teaching community with whom researchers can collaborate.

In the fairly short time we were producing and releasing episodes, we gained hundreds of listeners with episodes having been listened to over 2,400 times. What this shows is that there is a need for this type of resource: teachers want to engage with research, they value research-informed classroom practice, and they both want and need a format, like this one, which makes that possible. Researchers also benefit in two main ways: first, greater interest and awareness from teachers in their areas of research can create opportunities for research partnerships at the classroom, school, or district level. Second, podcasts can create greater visibility for their published work beyond an academic audience, and increase the implementation of the initiatives and interventions they study.

However, just as it is unreasonable to ask teachers to be responsible for seeking out and synthesizing education research, it is also unreasonable to expect researchers to take responsibility for the translation of their academic work into other formats. Both roles have pressures and incentives that channel teachers' and researchers' time and attention away from closing the research–practice divide. Although this is beginning to change, partly in response to initiatives such as the National Science Foundation's key grant proposal evaluation criterion of Broader Impact in higher education (National Science Foundation, 2020), the merit process generally recognizes and rewards peer–reviewed publications and grants for research, but not its implementation or communication to a wider audience. Teachers also do not receive incentives in the form of either time or money to engage with or participate in research. Further, neither role is well–trained for this task: most teachers lack training in statistics or research methods, and researchers have trained to communicate primarily to a scientific or academic audience. Therefore, this proposed translational bridge must come from a third party for the benefit of both teachers and researchers.

At the Neag School of Education at the University of Connecticut, and perhaps at other schools of education across the globe, we have an opportunity to use our considerable resources and collective expertise to bridge that need. We can build on this existing proof of concept to create a professional–quality, translational education research podcast designed specifically for teachers, that showcases the incredible work happening both within our school and beyond it, and goes the extra step of translating theory into practice. To do this, we can draw on the expertise and interests of our graduate students and faculty to collate and synthesize research, interview researchers and in–service teachers, and learn out loud by asking questions on behalf of our listeners. We can actively involve the teaching community by seeking input and including interviews and contributions from our Teacher Education program students and alumni, inviting them to share their experiences of implementing evidence-based practices in their classrooms, and discussing their questions and hopes for research. We can also draw on the media presence of the school to enhance and extend this work through blog posts or a podcast magazine, videos, and social media to reach a wider audience.

It is time for the education research establishment to examine what we might learn from teachers about their needs and interests to make education research more accessible and relevant to them. Rather than expecting teachers to seek out researchers and their work, or researchers to reach out to teachers individually, it is time for the education research establishment to take one (or several) steps closer to teachers and work together to build a bridge over the divide between research and practice to accelerate the improvement of learning and teaching in schools.

REFERENCES

Bruner, J. (1990). Acts of meaning. The Jerusalem Harvard Lectures, 179. https://psycnet.apa.org/fulltext/1990-98641-000.pdf

Clark, M. C. (2001). Off the beaten path: Some creative approaches to adult learning. New Directions for Adult and Continuing Education, 2001(89), 83–92.

Clark, M. C., & Rossiter, M. (2008). Narrative learning in adulthood. New Directions for Adult and Continuing Education, 2008(119), 61–70.

Conway, K. M., Wladis, C., & Hachey, A. C. (2021). Time poverty and parenthood: Who has time for college? AERA Open, 7. https://doi.org/10.1177/23328584211011608

Farley-Ripple, E., May, H., Karpyn, A., Tilley, K., & McDonough, K. (2018). Rethinking connections between research and practice in education: A conceptual framework. Educational Researcher, 47(4), 235–245.

Fisher, W. R. (1984). Narration as a human communication paradigm: The case of public moral argument. Communication Monographs, 51(1), 1–22.

Gilmore, S., & Deos, K. (Hosts). (2019-present). Little Key Learning Podcast. Little Key Learning. https://littlekeylearning/com/podcast

Gottschall, J. (2013). The Storytelling Animal: How Stories Make Us Human (1st ed.). Mariner Books.

Hughes, B., Sullivan, K. A., & Gilmore, L. (2021). Neuromyths about learning: Future directions from a critical review of a decade of research in school education. Prospects. https://doi.org/10.1007/s11125-021-09567-5

Kelly, J. M., Perseghin, A., Dow, A. W., Trivedi, S. P., Rodman, A., & Berk, J. (2022). Learning through listening: A scoping review of podcast use in medical education. Academic Medicine: Journal of the Association of American Medical Colleges, 97(7), 1079–1085.

Merriam, S. B. (2008). Adult learning theory for the twenty-first century. New Directions for Adult and Continuing Education, 2008(119), 93–98.

National Science Foundation. (2020, June 1). Proposal and Award Policies and Procedures Guide. Chapter II - Proposal Preparation Instructions. Retrieved September 27, 2022, from https://www.nsf.gov/pubs/policydocs/pappg20_1/pappg_2.jsp#IIC2d

Nelson, J., & Campbell, C. (2017). Evidence-informed practice in education: meanings and applications. In Educational Research (Vol. 59, Issue 2, pp. 127–135). Routledge. https://doi.org/10.1080/00131881.2017.1314115

Nelson, J., & Campbell, C. (2019). Using evidence in education. In A. Boaz, H. Davies, F. Alec, & S. Nutley (Eds.), What Works Now? Evidence-informed policy and practice. (pp. 131–150). Policy Press.

Nelson, J., Mehta, P., Sharples, J., & Davey, C. (2017). Measuring Teachers' Research Engagement: Findings from a pilot study The Education Endowment Foundation (EEF).

OECD. (2019). TALIS 2018 results (volume I): Teachers and school leaders as lifelong learners. OECD. https://doi.org/10.1787/1d0bc92a-en

Rossiter, M. (1999). Understanding adult development as narrative. New Directions for Adult and Continuing Education, 1999(84), 77–85.

Rycroft-Smith, L. (2022). Knowledge brokering to bridge the research-practice gap in education: Where are we now? Review of Education, 10(1). https://doi.org/10.1002/rev3.3341

Shkedi, A. (1998). Teachers' attitudes towards research: A challenge for qualitative researchers. International Journal of Qualitative Studies in Education: QSE, 11(4), 559–577.

Simmers, K. (2021). Education Research Is Still Too Dense. We Need More Teacher-Researcher Partnerships. EdSurge. https://www.edsurge.com/news/2021-04-08-education-research-is-still-too-dense-we-need-more-teacher-researcher-partnerships

Tai, S., Goldring, R., & Spiegelman, M. (2019). Characteristics of Public and Private Elementary and Secondary Schools in the United States: Results From the 2017-18 National Teacher and Principal Survey. http://eric.ed.gov/?id=ED544178

Tsagkias, M., Larson, M., & Rijke, M., de. (2009). Predicting podcast preference: An analysis framework and its application. Journal of the American Society for Information Science and Technology. https://doi.org/10.1002/asi.21259

Vanderlinde, R., & van Braak, J. (2010). The gap between educational research and practice: views of teachers, school leaders, intermediaries and researchers. British Educational Research Journal, 36(2), 299–316.

Watt, H. M. G., Richardson, P. W., Klusmann, U., Kunter, M., Beyer, B., Trautwein, U. T., & Baumer, J. (2012). Motivations for choosing teaching as a career: An international comparison using the FIT-Choice scale. Teaching and Teacher Education, 28(6), 791–805.

Winch, C. (2004).What do teachers need to know about teaching? A critical examination of the occupational knowledge of teachers. British Journal of Educational Studies, 52(2), 180–196.

Woo, A., & Steiner, E. D. (2021). Job-Related Stress Threatens the Teacher Supply: Key Findings from the 2021 State of the U.S. Teacher Survey—Technical Appendixes. Rand Corporation. https://www.rand.org/pubs/research_reports/RRA1108-1.html

Zhang, E., Trad, N., Corty, R., Zohrob, D., Trivedi, S., & Rodman, A. (2022). How podcasts teach: A comprehensive analysis of the didactic methods of the top hundred medical podcasts. Medical Teacher, 1–5.

STONES FROM ANOTHER MOUNTAIN: A CRITICAL-CULTURAL COMPARISON OF CHINESE AND U.S. HIGH-IMPACT LEARNING PRACTICES

JINGYI XU
HARVARD GRADUATE SCHOOL OF EDUCATION

MITCHELL STRZEPEK
BOSTON COLLEGE LYNCH SCHOOL OF EDUCATION AND HUMAN DEVELOPMENT

ABSTRACT

The Association of American Colleges and Universities promotes 11 high-impact practices for undergraduate learning that have been widely adopted in the U.S. This study investigates the cross-cultural relevance of these practices in 18 Chinese universities, based on reports from 20 mid- and high-level higher education administrators employed in a varied set of universities throughout China. Findings indicate that the U.S. practices are either absent in Chinese higher education or are present in significantly different forms. Results also include distinctly Chinese high-impact practices. A critical-cultural "trans-positional" analysis focuses on how Chinese and U.S. colleges and universities might translate potentially useful practices across cultures according to each country's distinct sociocultural context and postsecondary goals.

Keywords: high-impact practices, Chinese higher education, critical-cultural analysis, trans-positional analysis

In 2005, the American Association of Colleges and Universities (AAC&U) published its Liberal Education and America's Promise (LEAP) initiative, prompted by increasing pressure on undergraduate liberal arts education in the United States to respond to growing social, economic, and political challenges posed by globalization (AAC&U, 2005, 2007). LEAP prescribes the following "high-impact practices" (HIPs) as the most effective means for colleges and universities to fulfill essential undergraduate learning outcomes: 1) first-year seminars and experiences; 2) common intellectual experiences; 3) learning communities; 4) writing-intensive courses; 5) collaborative assignments and projects; 6) undergraduate research; 7) diversity/global learning; 8) service-learning; 9) internships; 10) capstone courses and projects; and 11) ePortfolios (Eynon & Gambino, 2017; Kuh et al., 2017; NSSE, 2006). A robust body of empirical literature suggests that student engagement in these practices is associated with positive outcomes such as higher retention (Kuh, 2008; Provencher & Kassel, 2017; Zilvinskis, 2019), career attainment (Miller, et al., 2018; Zilvinskis, 2019), and compensatory benefits for historically underrepresented student populations (Kuh, 2008).

Although the AAC&U practices have been widely adopted in the U.S., research regarding their cross-cultural applicability is limited. Comparative research is imperative because of the tensions between indigenous and isomorphic forces in international higher education (DiMaggio & Powell, 1983; Hayward & Siaya, 2001; Siaya & Hayward, 2003). The case of China offers a key example of a large system of higher education that has intentionally imitated many Anglo-American practices but whose sociopolitical and cultural context differs sharply from the U.S. (Hayhoe & Bastid, 2017; Huang, 2019; Marginson & Yang, 2021). The current study aims to investigate the localization and applicability of the HIPs in Chinese higher education institutions. Firstly, a brief introduction of the culture and purpose of the Chinese higher education system is offered, as well as the analytical and theoretical frameworks involved in the analysis and discussion of the data. Then, research methodology and findings from the data are described, which are student essays written by higher education practitioners taking an Ed.D. course at Peking University. Finally, findings are discussed from a critical-cultural perspective.

THE CULTURE AND PURPOSE OF CHINESE HIGHER EDUCATION

The massification of postsecondary education began in the 1940s in the immediate postwar era with the United States taking the lead, enrolling about 30 percent of its suitable age cohort into its higher education system. Afterwards, similarly in response to the various demands of the modern society, many European, Asian, and African countries, especially those industrialized, also experienced a dramatic expansion of their higher education system in the second half of the twentieth century (Altbach, 1998). In an attempt for economic development and international competitiveness, China also began its massification of higher education in the 1950s, but such transition had been interrupted by multiple political and social events. The current landscape of the Chinese higher education system was shaped by the modernization goals proposed in the 1990s (Neubauer & Zhang, 2015). Specifically, the modernization of Chinese higher education was marked by accelerated massification and internationalization.

China's massification of higher education outpaced that of most developed countries,

hitting the conventional benchmark of 15% higher education participation rate in 2004 (Shan & Guo, 2014). Simultaneously, the pattern of the massification process was distinctive in ways other than its rapidity. Zha (2011) traces the U.S. origin of the massification of higher education, which was facilitated by the decentralization of control, the pluralism of institutional types, ideals, and goals, as well as the diversity of sources of funding, and compares it with the differential higher education development patterns around the world. He analogizes the evolution of the Chinese higher education system to the East Asian (or Confucian) model of higher education development, where academia is closely tied to state management and emphasizes central control (Marginson, 2011). In other words, in consistency with Confucian traditions, higher education has been deployed as an instrument for social development and global competitiveness (i.e., state instrumentalism) in China (Zha, 2011).

Nevertheless, the development of the Chinese higher education system is distinguished from that of other East Asian countries, such as Japan and Korea, for its intentional commitment to internationalization (Altbach, 1998; Zha, 2011), through dispatching students abroad, adopting foreign academic models, and forming partnerships with foreign institutions (Lin, 2019; Neubauer & Zhang, 2015). Particularly, numerous efforts to internationalize higher education have been subject to Western influences since the Chinese economic reform (Altbach, 1998; Neubauer & Zhang, 2015). Studying Chinese college students' experience with internationalization at home, Guo and colleagues (2021) found that students typically perceived internationalization as westernization. In addition, Yang (2014) characterizes Western influence on the Chinese higher education system as one of the "two cases in which foreign influences brought to Chinese culture had such a great impact that the host culture was fundamentally changed" (p. 59).

Furthermore, evidence suggests that the United States has been increasingly influential on Chinese higher education in the past several decades. For instance, in studying Chinese higher education institutions' adaption of globally held ideas about research, Yoder (2010) found that Peking University and Beijing Normal University, which are both prestigious universities in China, explicitly encouraged integration of U.S. faculty and curriculum. Additionally, Tsinghua University initiated the Student Research Training (SRT) program, one of the first undergraduate research programs in China, after visiting the Undergraduate Research Opportunities Program (UROP) program at MIT in 1995. Funded by the Ministry of Education in 2000, Peking University also visited multiple U.S. institutions (e.g., UCLA) to investigate their undergraduate education and subsequently developed their own undergraduate research program (Lu, 2000). Apart from these institutional changes, U.S. ideas and practices also seamlessly permeate the Chinese higher education system. For example, an increasing number of Chinese universities adopt a general education curriculum that is similar to the liberal arts education of U.S. institutions. Additionally, the U.S. has become a major destination of Chinese international students, many of which return to China for faculty or staff positions with what they have learned in the U.S. higher education system.

Massifying its higher education system with a strong orientation of westernization and even Americanization in an attempt to achieve national prosperity and central control, the development of the Chinese higher education system presents an intricate picture of the adoption of Western policies and practices. Previous scholars have made some

preliminary attempts to understand the adoption of Western and specifically American models in the Chinese higher education system. Mohrman (2010) discusses five aspects of the U.S. higher education system that Chinese universities should not learn from, including sole concentration on research and publications, overemphasis on ranking and size, and the misuse of financial aid, due to certain inherent problems and structural differences. On the other hand, Yang (2013) evaluates China's incorporation and indigenization of the Western conception of university from a cultural perspective. Specifically, he points out that overreliance on U.S. experience to reform the Chinese higher education system while overlooking the fundamental cultural and ideological differences produces an arbitrary separation of structure and substance. Moreover, increasing emphasis on socialist values on the government's political agenda (Zhu & Li, 2018) intensifies the tension between the internationalization and indigenization of Chinese higher education. While the call for "higher education with Chinese characteristics" signifies elevated attention to the consistency between social and cultural contexts and higher education policies, it is important to examine how some Western-centric practices have been implemented in Chinese institutions.

Given the apparent tension between the Confucian traditions of state instrumentalism and the intentional westernization, especially Americanization, of academic models in Chinese higher education since the 1990s, it is necessary to further the understanding of the use of Western-centric practices among Chinese universities and colleges. At the same time, provided the lack of scholarship on the adoption of practices on institutional level, compared with policies on national and provincial levels, the ability to study this particular comparison between the implementation of HIPs in two such different contexts are ideal for investigating HIPs in a global perspective. Hence, the purpose of this study is to determine whether and in what form the U.S. high-impact practices are in use in Chinese universities, to uncover any indigenous Chinese high-impact practices, and to consider these results in light of a cultural-critical framework.

ANALYTICAL FRAMEWORK

We employed Bray and Thomas' (1995) framework for comparative education analyses to examine the application of HIPs in Chinese higher education institutions from an international comparative perspective. Bray and Thomas' framework is illustrated as a cube composed of smaller cubes (Figure 1), where the three dimensions of the cube represent three foci of comparison: 1) geographic/locational levels, 2) aspects of education and of society, and 3) nonlocational demographic groups. Bray and Thomas' framework calls for "multifaceted and holistic analyses of educational phenomena" (Bray et al., 2007, p. 8), which makes it suitable for examining the cross-cultural applicability of such Western-centric practices as HIPs in a Chinese context. Specifically, a strength of this present framework is that the wide range of aspects of education and society, including curriculum, teaching methods, management structures, and so on, speaks to the comprehensiveness of HIPs. For instance, first-year seminars and experiences, common intellectual experiences, and writing-intensive courses elaborate the curriculum aspect, while service/community-based learning and internships correspond student experiences. At the same time, this framework also takes the political and social factors that influence

the educational practices in China and the U.S. in distinct ways, as discussed above, into consideration. Hence, Bray and Thomas' framework for comparative education analyses was adopted with a focus on cross-country (level 2 on the front face) comparison across multiple aspects of education and of society (side).

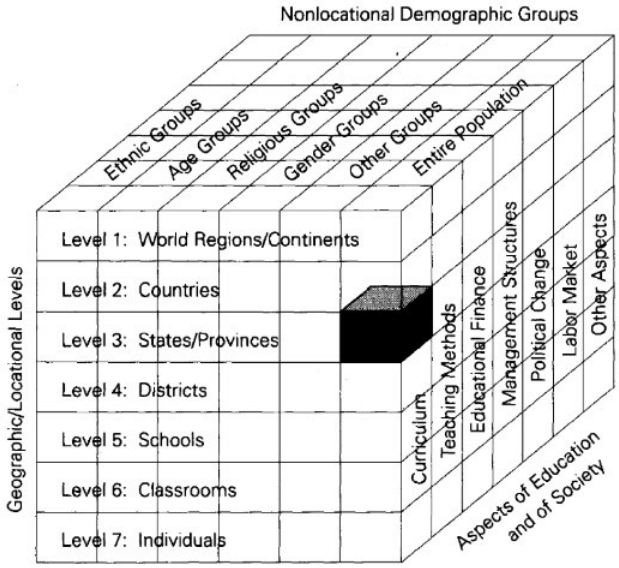


Figure 1

A Framework for Comparative Education Analyses

Source: Bray & Thomas, 1995, p. 475.

THEORETICAL FRAMEWORK

A critical-cultural theoretical framework is appropriate for investigating the effects of context on organizational practices in different national settings. Critical theorists foreground the role of structures, systems, and practices within social institutions such as universities (Abes et al., 2019). These social forces, in turn, emerge from and instantiate cultural norms, assumptions, values and beliefs (Guido et al., 2010; Patton et al., 2016). A critical-cultural framing attempts to make visible and question cultural norms and assumptions with the goal of critiquing and redressing inequitable power relations. Relevant to this study, unequal power relations have been evident in Chinese universities' adoption of English language scholarship and selected Anglo-American organizational practices (Hayhoe, 1989, 2017; Hayhoe & Bastid, 2017). The critical-cultural lens focuses description and interpretation on the differences between Chinese and U.S. cultural and socio-political foundations as these might influence the content and implementation of high-impact educational practices. In particular, the administration of higher education institutions by the central government and the Communist Party, which reflects the state instrumentalism of Confucian traditions discussed above, is distinct from the decentralized U.S. higher education system (Huang, 2019). Additionally, the collectivist orientation of Chinese culture, which emphasizes the interests and goals of groups than individuals, also differs from the individualist culture of the U.S. As a result, the elevation of Western-centric practices on Chinese university campuses brought contrasting cultural norms and created complex dynamics, the weaknesses, strengths, and opportunities of which are best examined through a critical cultural perspective.

The study investigated the cross-cultural relevance of U.S. high-impact practices in China by posing the following research questions:

1. What Association of American Colleges and Universities high-impact practices do administrators at Chinese universities identify as operating in their own institutions?

2. How do the same high impact practices differ in the US and China?

3. What high-impact practices are identified in Chinese higher education that are not part of the AAC&U HIPs?

METHODS, PARTICIPANTS AND DATA

The data source was from the final papers written by participants, who were students in an intensive summer course in the executive Ed.D. program of Peking University. The participants were a sample of 20 mid- and high-level Chinese higher education administrators employed in a varied set of 18 institutions around China. The course, co-taught by an American professor and a Chinese professor (see acknowledgements), included reading, lectures, and discussion on the AAC&U HIPs. As the assignment prompt shows (see Appendix B), participants were asked to compare HIPs in their own university to U.S. practices. An IRB review was required at neither Peking University nor Boston College for data collected as a part of a course assignment. However, participants signed a consent form giving permission for the authors to use their papers in future research (see Appendix A).

The characteristics of the participants' institutions were also gathered in a short survey (see Appendix A) and summarized in Table 1. The institutions were categorized based

Table 1

Frequency of Chinese Institutional Rankings Sample by Geographic Distribution of Sample

Geographic region	Institutional ranking	
	Double First-class	Non-Double First-class
North	4	5
East	2	2
Northeast	1	0
South Central	0	3
Total	7	10

*Non-Double First-class institutions may be a First-class institution in one or neither category.

on geographical location and Double First-class status. The Double First-class Initiative (i.e., the development of world-class universities and first-class disciplines) was launched by the Chinese government in 2007 to facilitate the internationalization and global competitiveness of Chinese higher education. Specifically, the Chinese government designated 42 higher education institutions as potential world-class institutions and 96 institutions to focus on building first-class disciplines. Double First-class institutions, which include a total of 42, refer to universities on both lists and represent the top-ranked institutions in China. Such a distinction undergirds the operations of Chinese universities and the present study on the adoption of HIPs because Double First-class universities typically receive greater financial and policy support (Liu et al., 2019).

DATA ANALYSIS

A team of Chinese and U.S. researchers, which included a Chinese and a U.S. professor, and a Chinese and a U.S. student assistant, conducted the analysis, beginning with translating the documents from Chinese to English. A frequency analysis of HIPs by institution type was then produced. The main part of the analysis was a thematic content analysis (Neuendorf, 2018) in which "'theme' can be described as the subjective meaning and cultural-contextual message of data" (Vaismoradi & Snelgrove, 2019, para. 3). In keeping with this tradition, the research team employed memoing as the main analytical strategy by engaging in extensive, iterative memo writing in an effort to conceptually connect raw data to abstractions (Birks et al., 2008), which, in this case, were cultural norms and assumptions on which HIPs were based. Such conceptual connections were vital to the present study given the nature of the raw data, which were student essays of varying degrees of subjectivity, criticality, and explicitness of underlying assumptions. The two student researchers each independently read the essays and wrote detailed memos on the themes of each essay. Then, the two sets of memos were compared, contrasted, and discussed with the professors so as to resolve inconsistent interpretations and reach agreement. Additional memos were continuously added about the sources of knowledge, which could be previous knowledge and experiences, used to examine the connections between raw data and cultural assumptions (Erlingsson & Brysiewicz, 2017). Keeping the issue of reflexivity in mind, the memos were validated on cultural assumptions that were not explicitly stated in the essay by referring to the literature that discussed the connections between explicit practices and underlying assumptions.

FINDINGS

Chinese higher education practitioners identified the high-impact practices that were present in Chinese universities and colleges. They also discussed the forms in which the practices were operated on Chinese campuses. This section summarizes significant findings derived from the practitioners' essays in response to the three research questions.

HIGH-IMPACT PRACTICES OPERATING IN CHINESE HIGHER EDUCATION INSTITUTIONS

All of the AAC&U high-impact practices were mentioned by Chinese higher education administrators in their essays. Undergraduate research, internships, and diversity and global learning were most frequently addressed by the administrators. They were discussed in 12, 11, and 8 essays, respectively. In contrast, most of the other practices were discussed by only a few administrators. In addition, most administrators also described the China-specific practices that they considered to have high impact on student learning in their own institutions, including academic competitions, ideological education, and administrative class (banjiti). Table 2 shows the frequency of mention for each of the AAC&U practices and the China-specific practices.

Although quite a few practices were not identified as operating in their institutions by the participating administrators, the cultural, social, and institutional factors underlying their rare presence varied from practice to practice across institutions. Institutional ranking, primarily determined by Double First-class status, was one important factor associated with the adoption of certain HIPs. Particularly, aside from a few top-ranked institutions that are experimenting with Anglo-American liberal arts models (Cheng, 2017), Chinese universities do not routinely offer writing-intensive courses, common intellectual experiences (in the form of general education), or service-learning programs. These results are consonant with the mainstream Chinese system of undergraduate specialization and lack of experiential learning (Huang, 2019). In contrast, liberal education involving HIPs has expanded to private and public institutions that were not conventionally considered liberal arts in the U.S., with the goal of promoting both intellectual and practical skills for a broader range of population, especially those historically marginalized (Kuh, 2008). Moreover, although collaborative assignments and projects are common in the U.S., respondents described this practice as just beginning in China. In both countries, students are graded and ranked individually. E-portfolios were described as entirely inapplicable by all of the three administrators addressing this practice, given the prior existence of a comprehensive individual dossier (dang'an) that systematically records personal, academic, and professional information, maintained by universities and government institutions.

Differences in High Impact Practices between the U.S. and China

Despite U.S. and Chinese higher education institutions large overlap in the adoption of such practices as undergraduate research, internships, and diversity and global learning, implementation differed significantly between the two countries. This section presents findings on the differences between the same high-impact practices in the U.S. and China.

UNDERGRADUATE RESEARCH

Participants indicated thesis research as a common form of undergraduate research that was shared by U.S. and Chinese higher education institutions. Thesis research is typically called "graduation design" and is more widely required for graduation in China. Students usually complete a research project based on their field experience and

conceptual work. Another form of undergraduate research in U.S. institutions is doing research with faculty members, where students typically work on faculty members' research projects. Nevertheless, several participants indicated that it is relatively rare for Chinese undergraduate students to work with faculty on their projects. Indeed, six of the 12 participants that wrote on undergraduate research, as well as two participants that separately designated competitions involving undergraduate research as a China-specific HIP, pointed out that most undergraduate research in China takes place in competitions, where students complete a research project individually or as a group with the advising of a faculty member. As elaborated by two participants who traced the development of undergraduate research practices in China, some top-ranked universities and new universities, in an attempt to learn from the undergraduate research programs in U.S. universities (e.g., the Undergraduate Research Opportunities Program at MIT) provide research training programs specifically for undergraduate students to gain research skills and experience. Overall, as explicitly named by another participant, undergraduate research was described as an "imported good," either explicitly or implicitly, by the administrators, while the "imported good" was adapted into the form of competitions and undergraduate research programs in accordance with the talent cultivation goals of different institutions with a focus on STEM fields.

INTERNSHIPS

Participants commented that both U.S. and Chinese undergraduates participate in internships to enhance their employment prospects. As a graduation requirement in many Chinese universities, internships are not mandatory for most U.S. academic majors. Most U.S. undergraduates acquire their own internships, sometimes with college-sponsored advising. A holistic analysis of participants' essays suggested that in China, the more selective the university, the more autonomy students have in the internship-searching process. Specifically, two participants from top-ranked Chinese universities suggested that career centers usually function as a platform for employment opportunities, alumni connections, and career advising. Nevertheless, based on the essays of the majority of the 11 participants discussing internships, more common forms of internships in Chinese institutions include university-business partnerships and business-owned universities that provide students with multiple channels to find career-related internships. Nevertheless, it is noteworthy that multiple participants pointed out that although internships serve to improve vocational outcomes in both the U.S. and China, the connection between college education and internships is weaker among Chinese institutions. One respondent that discussed the internship practices in three institutions of different types described internships as "employment-oriented, and basically detached from college education." She also compared internship experience to "the bargaining chips of the success rates of getting employed." Such characterization of internships deviates from the AAC&U ideal of internships as experiential learning.

GLOBAL AND DIVERSITY LEARNING

Global and diversity learning practices in the U.S. are relatively homogeneous. In

Table 2
Frequency of Participant-Identified HIPs by Chinese Institutional Ranking

AAC&U High Impact Practices	Double First-class	Non-Double First-class	Total
First-year seminars & experiences	2	0	2
Common intellectual experiences	1	0	1
Learning communities	3	1	4
Writing-intensive courses	2	0	2
Collaborative assignments & projects	2	2	4
Undergraduate inquiry & creative activity	6	6	12
Diversity/study away/global learning	4	4	8
Service learning/community-based learning	3	1	4
Internships	7	4	11
Capstone courses & projects	2	1	3
ePortfolios	2	1	3
China-specific High Impact Practices			
Academic competitions	5	3	8
Ideological education	3	1	4
Administrative class (*Banjiti*)	1	1	2

contrast, analogous practices in Chinese higher education institutions are much more varied. One important form of global learning is study-abroad experience that involves exchange or vising programs, which are similar in Chinese and U.S. institutions. At the same time, there is increasing popularity of short-term, self-funded visiting programs and internships, which may be paid or unpaid, among Chinese universities, while most U.S. undergraduate students study abroad in a university where regular tuition and financial aid apply. Another important practice of global learning among Chinese institutions is international conferences. On one hand, as pointed out by a respondent, in an attempt to increase international reputation, Chinese higher education institutions are holding more

academic conferences on their campuses. On the other hand, lots of funding support is available for students and scholars to travel and attend conferences abroad. Other forms of global learning practices in Chinese universities mentioned by respondents include international branch campuses, foreign partnership programs, and joint-degree programs, which involve, more or less, the presence of foreign faculty members on Chinese campuses.

An issue that is applicable to both Chinese and U.S. higher education institutions is foreign language fluency. Nevertheless, while U.S. students have the choice of attending an exchange program where the language of instruction is English, most Chinese students can only choose programs that are delivered in a foreign language other than Chinese. Chinese institutions may invite international faculty to teach discipline-specific materials in the faculty members' native languages. Such practice is widely adopted but problematic due to the separation of academic and administrative duties between international and local faculty members. In addition, several participants expressed the concern that the delivery of academic content in foreign languages and various pedagogical styles post challenges for students' foreign language skills and adaptability, the lack of which may undermine desired academic outcomes. In general, while multiple global learning programs are provided in Chinese higher education institutions, a variety of factors, as indicated by the administrators, can affect students' actual learning outcomes.

CHINA-SPECIFIC HIGH IMPACT PRACTICES

The participating administrators identified three practices that they regarded as high impact for student learning that are not included in the AAC&U practices: 1) academic competitions, 2) ideological education, and 3) administrative classes. This section discusses how the China-specific practices are operated and in what ways they are meaningful to student experience.

ACADEMIC COMPETITIONS

Twelve administrators addressed the prevalence of academic competitions as extracurricular activities in which Chinese undergraduates work on a research or entrepreneurial project individually or in a group. Some administrators separately described competitions in detail, while others weaved what they presented as this China-specific practice into the discussion of other AAC&U practices, such as undergraduate research and collaborative assignments and projects. Typically, as per participants' descriptions, students work on a scientific research project and submit deliverables, for example, research papers, for review. Excellent papers are ranked and awarded at university, city, and province levels. Competitions may involve various fields and disciplines, but most of the Chinese institutions from which the participants came were said to focus on natural sciences or entrepreneurship. One of the largest and most influential competitions in China is the College Student Innovation and Entrepreneurship Training Plan, which is carried out at the institutional, city, provincial, and national levels. According to the participating administrators, students have the opportunity to build their collaborative, problem-solving, and research skills by participating in competitions.

IDEOLOGICAL EDUCATION

Compulsory ideological education is integrated throughout Chinese education. In higher education, ideological education penetrates every student's college life from the beginning of their four-year journey. For example, per participants' explanations, most Chinese higher education institutions enact one-month military training and freshman education prior to the start of the first semester. The goals of these practices were described as getting familiar with the campus and the community, as well as the history and core values of communism and socialism. Participants frequently identified lectures and freshman concerts as typical activities involved in freshman education. In addition, when students start their college life, they are also required to take courses on Maoist and Marxist values. These ideas are also delivered in academically oriented courses. As a respondent wrote, "this action is called 'ideological and political theories teaching in all courses'; in this way, the impartation of knowledge and the guidance of values are combined." Other practices of ideological education are reflected in students' extracurricular activities, such as singing competitions.

ADMINISTRATIVE CLASSES (BANJITI)

A final China-specific practice, per the assignments, is the administrative class or banjiti (literally, the "class collective") in which students are grouped together based on year and major for their entire undergraduate career. As described by participants, students in the same administrative class usually take classes together. At the same time, they also share a lot of extracurricular activities, such as sports competitions, singing competitions, and so on. Therefore, they spend much time together on campus, which spontaneously increases their bonds with each other. Respondents viewed administrative classes as a representation of the collectivist culture. Specifically, a participating administrator wrote that the administrative class's "effects are mainly realized through the construction of the organizational system that is suitable for the Chinese collectivism." Furthermore, according to this respondent, a student typically leader is elected as the class monitor to help the classroom teacher with administrative duties. A U.S. counterpart of the Chinese administrative class is learning communities, where students engage in the same learning activities. Nevertheless, the Chinese administrative class is far more comprehensive than U.S. learning communities, and they do not share the same student leadership structures.

DISCUSSION

Identifying commonalities and distinct features of Anglo-American and Chinese higher education can lead to better understanding of one's own context (Marginson & Yang, 2021) and to adapting promising practices through cross-national "dialogue rather than domination" (Hayhoe & Liu, 2010, p. 92). In this spirit, the current study adds to the sparse comparative research on high-impact practices in student learning. Findings indicate that overall, U.S. HIPs are either absent or present in significantly different forms in Chinese higher education. The greatest similarity in HIPs occurs in top-ranked Chinese universities, which have been experimenting with the U.S. liberal arts model (Cheng, 2017; Pang et al., 2020) that the AAC&U explicitly endorses.

A critical cultural lens can help surface the cultural norms and assumptions underlying the similarities and differences between the HIPs practiced in Chinese and U.S. institutions. It is unsurprising that specific institutional practices in undergraduate education fail to translate in exact forms across countries with different societal ideologies, roles of higher education, and cultural norms and practices. The integration of Western, particularly American academic models and Confucian state instrumentalism of higher education that is featured in the development of the Chinese higher education system, is also evident in institution-level practices. In other words, in prioritizing indigenous forms of "higher education with Chinese characteristics" that highlight communist and socialist values (Zhu & Li, 2018, p. 1144) with continuing Western influences (Sporn & van der Wende, 2020), Chinese higher education operates Western-centric practices, such as the HIPs, in significantly different forms. On one hand, some study participants were able to explicitly identify the contextual influences that shaped the student learning practices in Chinese and U.S. institutions. For example, some participants attributed the inapplicability of e-portfolios in Chinese universities to the institutional practice of keeping personal dossiers by institutions and government agencies. They also related to the lack of credibility of portfolios created by students rather than faculty or staff members. At the same time, they also acknowledge the integrative and reflective value of e-portfolios. A few respondents attended to the cultural aspects that formulated certain practices. For instance, they pointed out that a collectivist culture determined the large scale and centralization of the activities involved in first-year seminars and experiences among Chinese universities.

On the other hand, although some respondents did not touch on the differences in norms and assumptions that led to the distinct forms of student learning practices, they were aware of the ideological and cultural foundations on which the practices were based. For example, while participants did not completely agree upon if diversity/global learning was not practiced or practiced differently in Chinese institutions, most of them were aware that diversity/global learning in the Chinese context focuses on international experiences or ethnic groups without referencing diversity in racial or gender identities that are common in Western discourses.

Similarly, participants connected the structural features of undergraduate research, including institutionalization, goal-orientations, and selection purposes, to the talent cultivation goals, which are typically a part of national development plans, of undergraduate research programs in Chinese institutions, without comparing them with the intellectual and interpersonal goals of undergraduate research originally designated by the AAC&U (Kuh, 2008). Nevertheless, a small number of participants assumed a culture-central perspective by placing certain Chinese or U.S. forms of HIPs in a superior position. Overall, a critical cultural perspective sheds light on the ways high impact is reflected in the significant benefits yielded by students in U.S. institutions (Kuh, 2008) and the realization of national development goals among Chinese universities and colleges (e.g., Wen et al., 2014; Xu et al., 2020).

More interesting is a trans-positional view (Sen, 2002) in which institutions consider how to translate potentially useful practices from another culture within their distinct sociocultural context and goals. China, for example, might design opportunities for undergraduate research and collaborative assignments that increase student engagement

toward collectivist, prosocial goals. The U.S. might address the separation of academics and student life by experimenting with undergraduate competitions or by adapting elements of the university-sponsored forms of student governance and intensive academically centered peer connections characteristic of the Chinese administrative class. Although the distinctly Chinese practice of compulsory ideological education might appear entirely inapplicable to an Anglo-American context, U.S. educators and state policymakers have begun to call for increased civic education in postsecondary schooling (Brennan, 2017; National Task Force, 2012).

In both China and the U.S. much more research is needed on the actual effects of presumed high-impact practices. Particularly important are studies that investigate practices from a student view, not just from an organizational or socio-cultural perspective. In carrying out this work, researchers should acknowledge and interrogate how educational practices reflect and promote cultural, ideological, and political norms and values.

A study respondent wrote that the Chinese have a long history of "using stones from another mountain to polish one's jade." Marginson and Yang (2021) echo this idiom by noting that Chinese scholars routinely make use of both Western and Chinese concepts and models but that few Anglo-American scholars draw on ideas from the Chinese context. As they write: "The possibility that more than one tradition can contribute to higher education studies is intellectually liberating" (p. 3). The present study suggests the importance of bi-directional learning about Chinese and US high-impact practices and the value of using this knowledge to consider adapting relevant HIPs in ways that are culturally appropriate for one's own context.

ACKNOWLEDGMENT

We would like to express our gratitude for the advice and support of Dr. Karen Arnold at Boston College Lynch School of Education and Human Development and Dr. Hong Zhu at Peking University Graduate School of Education. The data were collected in an Ed.D. course that they co-taught at Peking University. They also advised on the data analysis and writing of the paper.

REFERENCES

Abes, E. S., Jones, S. R., & Stewart, D. L. (Eds.). (2019). Rethinking college student development theory using critical frameworks. Stylus Publishing.

Altbach, P. G. (1998). Comparative higher education: Knowledge, the university, and development. Greenwood Publishing Group.

American Association of Colleges and Universities. (2005). College Learning for the New Global Century. American Association of Colleges and Universities.

American Association of Colleges and Universities. (2007). Liberal Education and America's Promise. American Association of Colleges and Universities.

Birks, M., Chapman, Y., & Francis, K. (2008). Memoing in qualitative research: Probing data and processes. Journal of Research in Nursing, 13(1), 68–75. https://doi.org/10.1177/1744987107081254

Bradburn, N. M., & Gilford, D. M. (1990). A framework and principles for international comparative studies in education. https://eric.ed.gov/?id=ED336417

Bray, M., Adamson, B., & Mason, M. (Eds.). (2007). Comparative education research: Approaches and methods. Springer.

Bray, M., & Thomas, R. M. (1995). Levels of comparison in educational studies: Different insights from different literatures and the value of multilevel analyses. Harvard Educational Review, 65(3), 472-490. https://doi.org/10.17763/haer.65.3.g3228437224v4877

Brennan, J. (2017). Higher education civic learning and engagement: A Massachusetts case study. Education Commission of the States.

Cheng, B. (2017). A comparative study of the liberal arts tradition and Confucian tradition in education. Asia Pacific Education Review, 18(4), 465-474. https://doi.org/10.1007/s12564-017-9505-6

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. American Sociological Review, 48(2), 147-160. https://doi.org/10.2307/2095101

Erlingsson, C., & Brysiewicz, P. (2017). A hands-on guide to doing content analysis. African Journal of Emergency Medicine, 7(3), 93-99. https://doi.org/10.1016/j.afjem.2017.08.001

Eynon, B., & Gambino, L. M. (2017). High-impact ePortfolio practice: A catalyst for student, faculty, and institutional learning. Stylus Publishing, LLC.

González, F., Moskowitz, A., & Castro-Gómez, S. (2001). Traditional vs. critical cultural theory. Cultural Critique, 49(Fall), 139–154. http://www.jstor.org/stable/1354706

Guido, F. M., Chávez, A. F., & Lincoln, Y. S. (2010). Underlying paradigms in student affairs

research and practice. Journal of Student Affairs Research and Practice, 47(1), 1-22. https://doi.org/10.2202/1949-6605.6017

Guo, Y., Guo, S., Yochim, L., & Liu, X. (2021). Internationalization of Chinese higher education: Is it Westernization? Journal of Studies in International Education, 26(4), 436-453. https://doi.org/10.1177/1028315321990745

Hayhoe, R. (1989). China's universities and Western academic models. Higher Education, 18(1), 49-85. https://doi.org/10.1007/BF00138961

Hayhoe, R. (1995). An Asian multiversity? Comparative reflections on the transition to mass higher education in East Asia. Comparative Education Review, 39(3), 299-321. https://doi.org/10.1086/447325

Hayhoe, R., & Bastid, M. (Eds.). (2017). China's education and the industrialised world: Studies in cultural transfer. Routledge.

Hayhoe, R., & Liu, J. (2010). China's universities, cross border education and the dialogue among civilizations. In D. Chapman, W. Cummings, & G. Postiglione (Eds.), Border crossing in East Asian higher education (pp. 77–102). Comparative Education Centre, University of Hong Kong/Springer Press. https://doi.org/10.1007/978-94-007-0446-6_4

Hayhoe, R. (2017). China's universities 1895–1995: A century of cultural conflict. Routledge. http://doi.org/10.1017/S0305741000043812

Hayward, F. M., & Siaya, L. M. (2001). Public experience, attitudes, and knowledge: A report on two national surveys about international education. American Council on Education, Washington.

Huang, F. (2010). Transnational higher education in Japan and China: A comparative study. In D. W. Chapman, W. K. Cummings, & G. A. Postiglione (Eds.), Crossing borders in East Asian higher educbation. CERC studies in comparative education (Vol. 27, pp. 265-282). Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0446-6_12

Huang, F. (2019). China's higher education system: 70 years of evolution, University World News. https://www.universityworldnews.com/post.php?story=20191001085233566

Kuh, G. D. (2008). High-impact educational practices: What they are, who has access to them, and why they matter. Association of American Colleges and Universities.

Kuh, G., O'Donnell, K., & Schneider, C. G. (2017). HIPs at ten. Change: Magazine of Higher Learning, 49(5), 8-16. https://doi.org/10.1080/00091383.2017.1366805

Lin, P. L. (2019). Internationalization of higher education in China: Challenges and opportunities. US-China Education Review B, 9(1), 1-12. http://dx.doi.org/10.17265/2161-6248/2019.01.001

Liu, Q., Turner, D., & Jing, X. (2019). The "Double First-class initiative" in China: Background, implementation, and potential problems. Beijing International Review of Education, 1(1), 92-108. https://doi.org/10.1163/25902547-00101009

Lu, X. (2000). 本科教育的重要组成部分 —— 伯克利加州大学本科生科研 [An important component

of undergraduate education – Undergraduate research at the University of California – Los Angeles]. *高等理科教育*, 5, 67-74.

Marginson, S. (2011). Higher education in East Asia and Singapore: Rise of the Confucian model. Higher Education 61, 587–611. https://doi.org/10.1007/s10734-010-9384-9

Marginson, S., & Yang, L. (2021). Individual and collective outcomes of higher education: a comparison of Anglo-American and Chinese approaches. Globalisation, Societies and Education, 1-31. https://doi.org/10.1080/14767724.2021.1932436

Miller, A.L., Rocconi, L.M. & Dumford, A.D. (2018). Focus on the finish line: Does high-impact practice participation influence career plans and early job attainment?. Higher Education 75, 489–506. https://doi.org/10.1007/s10734-017-0151-z

Mohrman, K. (2010). Educational exchanges: What China should not adopt from United States higher education. In D. W. Chapman, W. K. Cummings, & G. A. Postiglione (Eds.), Crossing borders in East Asian higher education. CERC studies in comparative education (Vol. 27, pp. 127-144). Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0446-6_6

National Task Force on Civic Learning and Democratic Engagement. (2012). A crucible moment: College learning and democracy's future. Association of American Colleges and Universities.

National Survey of Student Engagement. (2006). Engaged learning: Fostering success for all students. Annual report 2006. https://files.eric.ed.gov/fulltext/ED512619.pdf

Neubauer, D., & Zhang, J. (2015). The internationalization of Chinese higher education. CIQG Publication Series. Council for Higher Education Accreditation. https://eric.ed.gov/?id=ED587196

Neuendorf, K. A. (2018). Content analysis and thematic analysis. In P. Brough (Ed.), Advanced research methods for applied psychology (pp. 211-223). Routledge.

Pang, H., Cheng, M., Yu, J., & Wu, J. (2020). Suzhi education and general education in China. ECNU Review of Education, 3(2), 380-395. https://doi.org/10.1177/2096531120913171

Patton, L. D., Renn, K. A., Guido, F. M., & Quaye, S. J. (2016). Student development in college: Theory, research, and practice. John Wiley & Sons.

Provencher, A., & Kassel, R. (2017). High-impact practices and sophomore retention: Examining the effects of selection bias. Journal of College Student Retention: Research, Theory, & Practice, 21(2), 221-241. https://doi.org/10.1177/1521025117697728

Schein, E. (2010). Organizational culture and leadership. Jossey-Bass.

Sen, A. (2002). Rationality and freedom. Harvard University Press.

Shan, H., & Guo, S. (2014). Massification of Chinese higher education: Opportunities and challenges in a globalizing context. In A comparative analysis of higher education systems (pp. 9-23). Sense Publishers, Rotterdam. https://doi.org/10.1007/978-94-6209-533-5_2

Siaya, L., & Hayward, F. M. (2003). Mapping internationalization on US campuses.

American Council on Education.

Sporn, B., & van der Wende, M. (2020). The New Silk Road and the idea of the university. In M. van der Wende, W. C. Kirby, N. C. Liu, & Simon Marginson (Eds.), China and Europe on the New Silk Road: Connecting universities across Eurasia (pp. 331-360). Oxford University Press.

Vaismoradi, M., & Snelgrove, S. (2019). Theme in qualitative content analysis and thematic analysis. Forum: Qualitative Social Research, 20(3), Art. 23. https://dx.doi.org/10.17169/fqs-20.3.3376

Wen, W., Chu, J., & Shi, J. (2014). "9 8 5"高校高影响力教育活动初探 [On high-impact educational practices in Chinese "985" universities]. *高等教育研究,* , 35(8), 92-98.

Xu, D., Lv, L., & Fu, D. (2020). 中国研究型大学本科生高影响力教育活动特征探析 [On the characteristics of high-impact educational practices for undergraduates in China research universities]. *高等教育研究,* 41(2), 58-65.

Yang, R. (2013). Indigenizing the Western concept of university: The Chinese experience. Asia Pacific Education Review, 14(1), 85-92. https://doi.org/10.1007/s12564-013-9254-0

Yang, R. (2014). China's strategy for the internationalization of higher education: An overview. Frontiers of Education in China, 9(2), 151-162. https://doi.org/10.3868/s110-003-014-0014-x

Yoder, B. (2010). Adaptation of globally held ideas about research in China's universities. In D. W. Chapman, W. K. Cummings, & G. A. Postiglione (Eds.), Crossing borders in East Asian higher education. CERC studies in comparative education (Vol. 27, pp. 103-126). Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0446-6_5

Zha, Q. (2011). China's move to mass higher education in a comparative perspective. Compare: A Journal of Comparative and International Education, 41(6), 751-768. https://doi.org/10.1080/03057925.2011.590316

Zhu, X., & Li, J. (2018). Conceptualizing the ontology of higher education with Chinese characteristics. Educational Philosophy and Theory, 50(12), 1144-1156. https://doi.org/10.1080/00131857.2018.1504707

Zilvinskis, J. (2019). Measuring quality in high-impact practices. Higher Education, 78(4), 687-709. https://doi.org/10.1007/s10734-019-00365-9

APPENDIX A

PARTICIPANT CONSENT FORM AND SURVEY

Consent Form

[Do we have permission to cite your opinion and quote your class discussion and/or paper comments in a research article? We will not use your name or any information that could identify you individually.]

[Signature]:

Survey

Gender:

Age:

Current position

Title:

Area: Student affairs; Faculty; General university administration; Government; Business; Other

Years in current position:

Institution/company

Name of institution/company/government organization:

Location: Beijing/Tianjin/Hebei; Eastern China; Mid-China; Western China

Is it a higher education institution? Yes/No

Type: Research university; Teaching university; Applied university; Vocational college; Independent college; China-foreign partnership/foreign university

Classification: First-class A type; First-class B type; Non-first-class

Other work experience? Yes/No

If Yes, total number of years and title:

Total years of working in higher education industry:

APPENDIX B

ASSIGNMENT

This paper asks you to consider the extent to which the High-Impact Practices (HIPs) described in the U.S. professional literature apply to higher education in China. The evidence you provide should be from your own professional experience in higher education, informed by the class readings on high impact practices and the transcript of the class discussion. This is an analytic paper that is not based on empirical research beyond your own knowledge and professional experience.

[Based on your chosen practice, discuss whether or how the practice emerges in Chinese higher education. If the practice takes a different form, describe how the Chinese practice is different from the practice of Kuh's version. If there is not such a practice in China, discuss why the practice is not suitable for China. If you work in the government, departments making decisions, or businesses, consider the relevance of your chosen high-impact practice based on your role in higher education.]

UCONN

NEAG SCHOOL OF EDUCATION

education.uconn.edu/neag-journal