2020

Identifying Meaningful Individual-Level Change in Educational Experiences: Adding to Our Methodological Toolkit

Elizabeth Niehaus

Gudrun Nyunt

Identifying Meaningful Individual-Level Change in Educational Experiences: Adding to Our Methodological Toolkit

Elizabeth Niehaus & Gudrun Nyunt

Elizabeth Niehaus is Associate Professor of Educational Administration at the University of Nebraska–Lincoln.
Gudrun Nyunt is Clinical Assistant Professor of Counseling and Higher Education at Northern Illinois University.

Correspondence — Elizabeth Niehaus eniehaus@unl.edu

In recent years, improving the quantitative methods used to assess the effect of college, and particular college experiences, on student outcomes has received increased attention (*e.g.,* Mayhew et al., 2016). In *How College Affects Students*, Mayhew et al. (2016) highlighted the importance of issues of practical vs. statistical significance, self-selection into college (and by extension, self-selection into particular experiences), and direct and indirect effects, among other methodological challenges in identifying the relationships between college experiences and student learning and success. One particularly difficult challenge is identifying the conditional effects of experiences on student outcomes. Who benefits, or who does not, from particular experiences? There is growing evidence that the effects of educational experiences may differ among students, and in some cases, effects that may be positive for some students are negative for others (*e.g.,* Mayhew et al., 2016; Seifert, Gillig, Hanson, Pascarella, & Blaich, 2014).

Published in *Journal of College Student Development,* Volume 61, Number 5, September-October 2020, pp. 637-643.
Published by Johns Hopkins University Press
DOI:10.1353/csd.2020.0060

The most common methods of assessing conditional effects rely on group-level analyses (*e.g.,* introducing interaction terms or conducting subgroup analyses). Yet, these methods do not provide a way to determine whether an experience has had a positive, negative, or neutral effect on an individual student. Examining individual-level change can help researchers and practitioners further understand the complexities of how educational experiences affect students. With this article we aim to build on the work on conditional effects in higher education (*e.g.,* Seifert et al., 2014) to provide a way to assess meaningful individual-level change. We provide a theoretical framework for understanding why educational experiences might lead to positive or negative outcomes; discuss the challenges in assessing individual-level change; describe one method of assessing individual-level change; provide an example of how researchers might use this method to consider positive and negative outcomes for individual students; and discuss how this consideration might change the way we view college experiences.

Theoretical Foundations

Experiential learning theories are the basis of many educational experiences in higher education; however, Dewey (1938) cautioned that experiences are not automatically educational and can even be "miseducative." Research indicates that educational experiences do not always lead to positive outcomes (*e.g.,* Kilgo, Sheets, & Pascarella, 2015) and may have the potential for negative outcomes. For example, Becker and Paul (2015) argued that service-learning experiences may reproduce color-blind racism by reinforcing racist stereotypes. Ogden (2006) highlighted that study abroad experiences can similarly reinforce stereotypes of another culture, even when the students perceive that they had a transformative, positive experience.

Despite the potential for experiences to be mis-educative, most research on educational experiences in college reports positive benefits. In a review of 35 studies on out-of-class activities, Simmons, Creamer, and Yu (2017) found that only 6 referred to negative or neutral outcomes. Even scholars who question the benefits of particular experiences often fall short of pointing out the potential for *negative*

outcomes. For example, Kilgo et al. (2015) found that capstone courses and service learning both had a negative effect on certain educational outcomes. The authors, however, concluded that these "surprising" findings may not mean that these experiences always lead to negative outcomes, but encouraged further study on their effectiveness (p. 522).

The scarcity of literature on negative or neutral effects of educational experiences might be due to a bias against publishing negative results or to a lack of longitudinal research in recent years (Mayhew et al., 2016) or to the fact that many experiences may actually lead to positive gains on average (*e.g.,* Varela, 2017). Averages, however, obscure individual differences and thus do not provide the full picture of the learning that results from an educational experience. For example, in a large-scale study, Vande Berg, Connor-Linton, and Paige (2009) found an overall positive effect of studying abroad on intercultural competence, despite the fact that over a third of the women in the study "showed statistically insignificant intercultural gains or actual decline" (p. 25). Although this is one of the few studies that highlights neutral or negative effects, the researchers conflated "statistically insignificant . . . gains" with "actual decline," and did not specify how they determined gains or declines at an individual level.

The Challenge : Measuring "Significant" Individual-Level Change

Researchers may not be examining individual-level differences for one other reason: the challenge of determining what counts as a "significant" or "meaningful" change at the individual level. Unlike analyses across an entire sample where paired-sample *t* tests can be used to determine significant differences in mean pretest and posttest scores on a given measure, it is less clear how to determine whether an observed difference in scores is meaningful for an individual student. The way researchers determine group-level changes does not necessarily transfer to determining meaningful individual-level changes (McHorney & Tarloy, 1995).

To our knowledge little if any work has been done to establish benchmarks for meaningful individual-level change in higher

education; however, researchers in the health fields have done a great deal of work to understand how to measure this type of individual-level change. Wyrwich and Wolinsky (2000) suggested the standard deviation or a measure of effect size (*e.g.,* Cohen's *d*) options for measuring meaningful "intra-individual" or individual-level change, but noted that because both are sample-dependent measures, benchmarks across many different samples are needed to generalize these measures as corresponding to meaningful individual-level change. Multiple health researchers thus identified the standard error of measurement (*SEM*) as the indicator with the strongest association with truly meaningful individual-level change (Wyrwich, Tierny, & Wolinsky, 1999; Wyrwich & Wolinsky, 2000).

The *SEM* is a measure of how much measurement error obscures the "true score" on a particular measurement (Wyrwich et al., 1999) and is calculated using the following formula:

$$SEM = \sigma_{xx}\sqrt{1 - r_{xx'}}$$

The standard deviation at the baseline (pretest) is multiplied by the square root of 1 minus the reliability of the measure, generally Cronbach's alpha (Wyrwich et al., 1999; Wyrwich & Wolinsky, 2000). The *SEM* is independent of sample size and is expressed in the same units as the measurement itself, facilitating interpretation. Multiple studies in the health fields have determined that a change in scores greater than ±1 *SEM* corresponds to meaningful individual-level change (Wyrwich et al., 1999; Wyrwich & Wolinsky, 2000).

Example: Individual-Level Change in Study Abroad

To demonstrate how researchers might use the *SEM* to examine individual-level change for participants in educational experiences—and how doing so might affect our understanding of these experiences—we provide an example using data from short-term study abroad participants at 7 different higher education institutions. For context, study abroad is an increasingly popular college experience: 332,727 US students studied abroad in 2016–17, an increase of 50% since 2005–06, with short-term study abroad programs, defined as a summer or 8

weeks or fewer, being the most popular type of study abroad (Institute of International Education, 2018). Although a full review of study abroad research is outside the scope of this brief article, there is substantial literature to back up the claim that study abroad can contribute to positive student outcomes (*e.g.,* Kilgo et al., 2015; Varela, 2017). Our findings may be particularly interesting to study abroad scholars and practitioners, but we discuss data analysis and results here to illustrate how and why researchers might consider individual-level change when examining educational outcomes more broadly.

Method

Sample

We sent survey invitations to all students in 50 faculty-led, short-term study abroad courses at 2 doctoral universities, 3 master's colleges/ universities, and 2 associate's colleges. The programs lasted between 7 and 43 days (M = 18 days) and were housed in various academic departments. A little less than half (46.0%) of the courses had European destinations; the others took place in Africa, Asia, Oceania, and the Americas. Of the 635 invited students, 398 completed the CQS pretest and posttest items (response rate: 62.7%). The majority of respondents (87.9%) were undergraduate students and over two thirds (69.8%) were women. Regarding race and ethnicity, 84.7% identified as White/ Caucasian (non-Hispanic), 6.0% as Asian / Pacific Islander, 5.8% as Hispanic, and 3.8% as African American / Black (non-Hispanic).

Data Collection

Data for this study came from pretest (before trips abroad) and posttest (after trips) surveys of short-term study abroad participants, including the Cultural Intelligence Scale (CQS; Van Dyne et al., 2012). For this example, we used students' pretest and posttest CQS scores to examine significant changes in cultural competence for the overall sample as well as meaningful changes on an individual level. The CQS measures individuals' capacity to effectively work and relate with

Table 1. Subdimensions of the Cultural Intelligence Scale (CQS)

Scales and Subscales	Description
Drive	
Intrinsic Motivation	Deriving enjoyment and sense of satisfaction from cross-cultural experiences
Extrinsic Motivation	Believing that one will benefit (*e.g.,* career advancement) from cross-cultural experiences
Self-Efficacy	Having the confidence that one can be effective in a different cultural setting
Knowledge	
Context General Knowledge	Having a macrolevel understanding of similarities and differences among cultures
Context-Specific Knowledge	Understanding how culture influences one's effectiveness in specific domains (*e.g.,* business, higher education)
Strategy	
Awareness	Cognizance of one's existing cultural knowledge
Planning	Strategizing before a culturally diverse encounter
Checking	Examining assumptions and adjusting mental maps when actual experiences differ from expectations
Action	
Verbal Behavior	Having and using a flexible range of culturally appropriate verbal behaviors (*e.g.,* accent, tone)
Nonverbal Behavior	Having and using a flexible range of culturally appropriate nonverbal behaviors (*e.g.,* body language, physical gestures, facial expressions)
Speech Acts	Modifying the manner and content of communication (*e.g.,* direct vs. indirect) to fit multicultural contexts

people across different cultures and consists of 4 main scales (drive, knowledge, strategy, and awareness) that are divided up further into 11 subscales (see Table 1 for a brief description of each subscale). Participants responded to statements describing their cultural competence on a Likert-type scale from 1 (*strongly disagree*) to 7 (*strongly agree*), including a neutral option, 4 (*neither agree nor disagree*).

Analysis

We examined changes in students' CQS scores across all subdimensions on two levels: across the entire sample and for each individual student. Across the whole sample, we conducted a paired-samples *t* test to identify significant differences ($p < .05$) in mean scores of the

CQS subdimensions; we also calculated effect sizes using Cohen's d_{av}. Next, we used the *SEM* to identify meaningful individual-level change in students' CQS scores. We first calculated the *SEM* for each of the CQS subscales, then calculated each student's individual difference scores across each of the CQS subscales. We used ±1 *SEM* as the cut-off value for determining whether or not an individual difference score corresponded to "meaningful change," either positive or negative, and then counted the number of students who had meaningful increases or declines in scores between the pretest and posttest.

Limitations

This example has a few limitations. First, the *SEM* has been normed in the health fields, but to our knowledge no research on the use of the *SEM* in higher education exists. Future research might examine the validity of ±1 *SEM* representing meaningful individual-level change for higher education outcomes. Second, we did not examine predictors of individual-level change, only describing the individual-level change in our sample. Researchers should examine why we see such wide variation in students' learning in experiences like study abroad.

Results

While we found significant differences in students' cultural intelligence before and after studying abroad across the entire sample, our individual-level analysis tells a more nuanced and complex story (see Table 2). Specifically, across the entire sample, we found significant differences in 10 of the 11 CQS subdimensions, with effect sizes from very small (d = .12) to medium (d = .44). The largest effect sizes were in culture general knowledge and context-specific knowledge and the smallest in verbal behavior and checking. Only one subdimension showed no significant differences between the pretest and posttest: self-efficacy to adjust. At the individual level, however, we found meaningful increases *and* decreases (mean differences > ±1 *SEM*) between the pretest and posttest for all subdimensions of the CQS.

Table 2. Significant or Meaningful Differences at the Sample and Individual Levels (N = 398)

	Whole Sample (M)				Individuals With Pretest–Posttest Difference > ±1 SEM	
	Pretest	Posttest	ΔM	d_{av}	Decrease n (%N)	Increase n (%N)
Intrinsic Motivation	5.13	5.37	0.24**	.20	75 (19.1%)	146 (37.2%)
Extrinsic Motivation	4.88	5.19	0.30**	.24	91 (23.0%)	169 (42.8%)
Self-Efficacy	5.50	5.59	0.09	.08	109 (27.5%)	136 (34.3%)
Culture General	4.13	4.63	0.50**	.44	75 (19.3%)	181 (46.5%)
Context-Specific Knowledge	3.65	4.16	0.51**	.37	91 (23.2%)	218 (55.5%)
Planning	4.34	4.63	0.30**	.24	105 (26.4%)	168 (42.2%)
Awareness	5.35	5.55	0.20**	.18	90 (22.9%)	141 (35.9%)
Checking	5.32	5.46	0.14*	.13	94 (23.8%)	139 (35.2%)
Speech Acts	4.86	5.13	0.27**	.23	103 (26.1%)	155 (39.3%)
Verbal Behavior	4.59	4.74	0.15*	.12	120 (30.3%)	151 (38.1%)
Nonverbal Behavior	4.58	4.82	0.24**	.17	116 (29.1%)	166 (41.7%)

Total N ranges from 389 to 398 due to missing values for some items.
* Pretest–posttest difference significant at $p < .05$.
** Pretest–posttest difference significant at $p < .01$.

Comparing results at the individual and whole sample levels points to important implications for how we think about research on the effects of college experiences on student outcomes. First, findings of no significant differences between pretest and posttest scores or significant differences with a small effect size across a whole sample do not indicate that nothing happened at the student level. For example, despite verbal behavior having one of the smallest effect sizes across the entire sample (d = .12), we saw that 68.0% of students had meaningful changes in scores (38.1% positive, 30.3% negative). Similarly, there were no significant differences in self-efficacy to adjust across the entire sample, but 61.8% of students' scores changed in meaningful ways (34.3% positive, 27.5% negative). The numbers of students who experienced meaningful change in these two dimensions were similar to the number for culture general knowledge (65.8%), the dimension with the largest effect size across the sample (d = .44). In culture general knowledge more students (46.5%) showed increases in scores than for verbal behavior (38.1%) or self-efficacy to adjust (34.3%), and fewer students (19.3%) showed decreases (30.3% and 27.5%, respectively). Importantly, in both of these areas, more than

a third of students still demonstrated meaningful increases in scores, a finding we would have missed had we only looked at mean differences across the entire sample.

The second major finding is that significant increases in students' scores across the entire sample, even with medium effect sizes, did not necessarily correspond to universally positive changes at the individual level. In the two areas with the largest increases in students' scores across the entire sample, large numbers of students still showed meaningful decreases in scores between the pretest and posttest: 19.3% for culture general knowledge and 23.3% for context-specific knowledge. These decreases were balanced out, on average, by the larger number of students who had meaningful increases in scores, but we still saw about 1 out of every 5 students with declining scores for culture general or context-specific knowledge.

Conclusion

Our results highlight the need to examine individual-level change when examining learning outcomes of educational experiences. Based on the analysis of our entire sample, we would have concluded that short-term study abroad programs led to significant positive changes in cultural intelligence in almost all areas. Examining individual-level changes provided a more nuanced picture of the learning in these programs, both positive and negative. Gaining a more nuanced understanding of student learning will allow scholars to ask important questions about why an experience may lead to positive changes for some students but negative changes for others. Such information can help practitioners improve experiential learning initiatives to maximize the number of students who achieve the intended results of an educational experience.

References

Becker, S., & Paul, C. (2015). "It didn't seem like race mattered": Exploring the implications of service-learning pedagogy for reproducing or challenging color-blind racism. *Teaching Sociology, 43*, 184-200.

Dewey, J. (1938). *Experience and education.* New York, NY: Touchstone.

Institute of International Education. (2018). *Open doors report on international education exchange*. Washington, DC: Author. Retrieved from https://www.iie.org

Kilgo, C. A., Sheets, J. K. E., & Pascarella, E. T. (2015). The link between high-impact practices and student learning: Some longitudinal evidence. *Higher Education, 69*, 509-525. https://doi.org/10.1007/s10734-014-9788-z

Mayhew, M. J., Rockenbach, A. N., Bowman, N. A., Seifert, T. A., & Wolniak, G. C. (with Pascarella, E. T., & Terenzini, P. T.). (2016). *How college affects students (Vol. 3): 21st century evidence that higher education works.* San Francisco, CA: Jossey-Bass.

McHorney, C. A., & Tarlov, A. R. (1995). Individual-patient monitoring in clinical practice: Are available health status surveys adequate? *Quality of Life Research, 4,* 293-307.

Ogden, A. C. (2006). Ethnographic inquiry: Reframing the learning core of education abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *13*, 87-112.

Seifert, T. A., Gillig, B., Hanson, J. M., Pascarella, E. T., & Blaich, C. F. (2014). The conditional nature of high impact/good practices on student learning outcomes. *Journal of Higher Education, 85*, 531-564.

Simmons, D. R., Creamer, E. G., & Yu, R. (2017). Involvement in out-of-class activities: A mixed research synthesis examining outcomes with a focus on engineering students. *Journal of STEM Education: Innovations and Research, 18*(2), 10-16.

Van Dyne, L., Ang, S., Ng, K. Y., Rockstuhl, T., Tan, M. L., & Koh, C. (2012). Sub-dimensions of the four factor model of cultural intelligence: Expanding the conceptualization and measurement of cultural intelligence. *Social and Personality Psychology Compass, 6*(4), 294-313. https://doi.org/10.1111/j.1751-9004.2012.00429.x

Vande Berg, M., Connor-Linton, J., & Paige, R. M. (2009). The Georgetown Consortium Project: Interventions for student learning abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *18*, 1-75.

Varela, O. E. (2017). Learning outcomes of study-abroad programs: A meta-analysis. *Academy of Management Learning & Education, 16*, 531-561. https://doi.org/10.5465/amle.2015.0250

Wyrwich, K. W., Tierney, W. M., & Wolinsky, F. D. (1999). Further evidence supporting an *SEM*-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *Journal of Clinical Epidemiology, 52*, 861-873.

Wyrwich, K. W., & Wolinsky, F. D. (2000). Identifying meaningful intra-individual change standards for health-related quality of life measures. *Journal of Evaluation in Clinical Practice, 6*, 39-49.