Comparing Traditional Versus Retrospective Pre-/Post-assessment in an Interdisciplinary Leadership Training Program

Erin Vinoski Thomas[1] · Rebecca Wells[2] · Stephanie D. Baumann[2] · Emily Graybill[2] · Andrew Roach[2] · Stephen D. Truscott[3] · Mark Crenshaw[2] · Daniel Crimmins[2]

Abstract

Objectives As the U.S. healthcare system shifts toward collaboration, demand for leaders with interdisciplinary skills increases. Leadership competencies guide interdisciplinary training programs; however, identifying cost-effective methods for evaluating leadership competencies is challenging, particularly when interdisciplinary trainees have different areas of expertise and professional goals. Traditional pre-/post-testing, a common method for evaluating leadership competencies, is subject to response-shift bias, which can occur when participants' understanding of a construct changes between pre- and post-test. As a result, participants may rate their knowledge of the construct lower at post-test. Retrospective pre-tests are one method thought to reduce response-shift bias in pre-/post-tests. The current study explores the use of a retrospective pre-test to control for response-shift bias in an interdisciplinary training program. *Methods* Over three cohort years, thirty-four trainees from an interdisciplinary leadership program completed a self-assessment aligned with MCH leadership competencies. The traditional pre-test self-assessment was completed at the beginning of the training program. The retrospective pre-/post-test self-assessment was completed at the end of the training program. *Results* Retrospective pre/post-test scores indicate significant self-reported increases in all 24 leadership areas (p ≤ .001). Furthermore, participants' self-ratings were significantly higher on the traditional pre-test for all 24 areas than on the retrospective pre-test (p ≤ .001). *Conclusions for Practice* Retrospective pre-tests appeared to control for response-shift bias and may be a cost-effective way to evaluate trainee change within an interdisciplinary leadership program. These findings suggest the methodology's usefulness in interdisciplinary training and its potential use in the broader world of community-based MCH training initiatives.

Keywords Workforce · Interdisciplinary training · Leadership development · Program evaluation · Measurement

✉ Emily Graybill
egraybill1@gsu.edu

Erin Vinoski Thomas
evinoski@uncc.edu

Rebecca Wells
rwells12@gsu.edu

Stephanie D. Baumann
sbaumann2@student.gsu.edu

Andrew Roach
aroach@gsu.edu

Stephen D. Truscott
sdt55@gsu.edu

Mark Crenshaw
mcrenshaw@gsu.edu

Daniel Crimmins
dcrimmins@gsu.edu

[1] Department of Public Health Sciences, University of North Carolina at Charlotte, Charlotte, NC, USA

[2] Center for Leadership in Disability, School of Public Health, Georgia State University, 75 Piedmont Ave. NE, Suite 514, Atlanta, GA 30303, USA

[3] Department of Counseling & Psychological Services, College of Education and Human Development, Georgia State University, Atlanta, GA 30302, USA

Significance

Interdisciplinary training programs can equip leaders with necessary skills to work effectively in the changing American healthcare system. Measuring the impact of interdisciplinary training can be difficult, however, due to the diversity of participants. Retrospective pre-tests have been used widely in education to measure learner change in knowledge and skills. This study replicates the use of the retrospective pre-test methodology and applies it to an interdisciplinary training program to demonstrate the methodology's potential for use in public health, MCH, and interdisciplinary settings. This study's findings have potential to inform the ways in which similar interdisciplinary training programs are evaluated.

Objectives

The importance of interdisciplinary approaches is widely recognized in healthcare and public health (Interprofessional Education Collaborative Expert Panel 2011; Reyes-Akinbileje 2013). The American healthcare system is transitioning towards more collaborative care models as evidenced by mandates in the Affordable Care Act, funding for the Patient-Centered Outcomes Research Institute, and shifts towards patient-centered models of care. These changes require healthcare professionals to understand themselves in relation to other healthcare professionals; to work effectively within interdisciplinary teams; and to collaborate with patients to support optimal patient care and outcomes all while recognizing where they fit within the larger healthcare system (Interprofessional Education Collaborative 2016).

Standard professional, academic, and continuing education programs and traditional leadership training programs seldom provide in-depth training in many of the skills needed by healthcare professionals in this new clinical age (Institute of Medicine [IOM] 2003). To fill that gap, the Maternal & Child Health Bureau (MCHB) funds a diverse training portfolio across the country, which includes programs for the current MCH workforce and for graduate students and future leaders (Health Resources & Services Administration 2017). These training programs seek to equip MCH leaders with the knowledge, attitudes, and skills necessary to influence necessary positive change in the American healthcare system. Leadership Education in Neurodevelopmental Disabilities (LEND) programs are one of the interdisciplinary leadership training programs funded by MCHB for future leaders in disability-serving disciplines. LEND program graduates are expected to transfer the knowledge and skills gained through the program into their professional and personal interactions with individuals with disabilities and their families.

The potential public health impact of interdisciplinary MCHB training programs like LEND is substantial; however, evaluating these programs is challenging due to the inherent differences that exist between interdisciplinary trainees. Interdisciplinary training and leadership development programs are often guided by competencies that comprise a set of knowledge domains, skills, and dispositions carefully constructed to guide training; provide a framework for student and program assessment; and give direction for the immediate and long-term development of leadership trainees (Department of Human Resources for Health 2010; McDougal et al. 2005). Although these competencies are often well articulated and have a clear evolutionary lineage, there exists a critical need to ensure valid assessment of interdisciplinary trainees' baseline skills in the competencies, assess progress as they acquire new skills, and identify specific areas for additional attention. Such data will facilitate the move from conceptual models and hypothetical constructs to testable data-based evidence that can determine the added value of MCH leadership programs to professional training.

Many interdisciplinary training and leadership development programs rely on trainee self-report of knowledge, skill, and behavior change at baseline and post-training to assess trainee growth in program competencies (Piwowar and Thiel 2014). Self-report is the default evaluation method due to its feasibility and low cost (Moore and Tananis 2009). Potential problems with self-report methodology have been noted frequently in the literature (Coulter 2012; Goedhart and Hoogstraten 1992; Rohs 1999). For example, to compare baseline and post-training scores using self-report data, one must assume that a trainee's standards for measurement of the competencies remain constant throughout the course of the training program. This assumption is problematic because most training programs seek to change trainees' conceptualizations of the targeted program competencies (Piwowar and Thiel 2014); as a result, comparisons of baseline with post-training self-assessments may be confounded by response-shift bias. Response-shift bias occurs when participants' frame of reference about a topic changes as the result of receiving instruction (Howard et al. 1979; Nimon et al. 2011). Participants often realize that they think differently about a topic than they did at the beginning of the program, resulting in self-reports that can overestimate pre-test scores and underestimate post-test scores (Pratt et al. 2000).

Retrospective Pre-tests

The purpose of this article is to explore the use of retrospective pre-testing as an alternative to traditional pre-/

post-testing to assess trainees' perceptions of knowledge and skill change after participating in a MCH interdisciplinary training program. Retrospective pretests were introduced by Campbell and Stanley (1963) and empirically tested by Howard et al. (1979) as a compromise to traditional self-report pre-/post-tests. The methodology has since undergone rigorous examination. In a retrospective pre-test, individuals rate their level of baseline functioning (e.g., knowledge and skills) simultaneously with their post-training functioning. Early research on retrospective pretests as a way to improve internal validity and counter response-shift bias concluded that when trainees in an educational program did not have enough information to rate their initial level of knowledge and skills (that is, trainees did not yet know what they did not know), the retrospective pretest provided a more accurate baseline measure (Howard et al. 1979). Since then, this methodology has been successfully employed across a variety of settings, for example, to evaluate student growth with respect to program competencies in university programs (Coulter 2012; Drennan and Hyde 2008; Miller and Elder Hinshaw 2012; Moore and Tananis 2009), to assess patients' self-reported outcomes in healthcare settings (Kreulen et al. 2002), and to understand website visitors' responses to online content (Mueller 2015). Having trainees simultaneously rate their baseline and post-training knowledge/skills establishes a common metric for pre-/post-test responses and may provide a more accurate measure of subjective growth. Evidence has shown that retrospective pretest responses correlate highly with responses on objective measures of knowledge or behavior change, suggesting that the retrospective pretest may be a more valid measure of pre/post change than traditional pretests (Martineau 2004). Finally, retrospective pretests allow researchers to obtain complete datasets due to only collecting pre/post data at one time point, and often, on one form.

Context of the Current Study

Leadership Education in Neurodevelopmental Disabilities Programs

The present research examined trainees participating in a LEND program. The LEND program brings together professionals, graduate students, representatives from organizations, disability self-advocates, and family members of individuals with neurodevelopmental disabilities to engage in didactic coursework and training experiences related to the care and support of persons with neurodevelopmental disabilities and their families. MCH Leadership Competencies guide the LEND activities and curriculum (MCH Leadership Competencies Workgroup 2009).

The MCH Leadership Competencies

Table 1 lists the 12 MCH Leadership Competencies. The MCH Leadership Competencies frame training objectives for the MCH Leadership Training Programs (e.g., LEND), provide benchmarks for evaluating participants' progress, guide assessment of competencies within the existing MCH workforce, and promote and sustain MCH leadership in the healthcare system (MCH Leadership Competencies Workgroup 2009). A detailed review of the 12 MCH Leadership Competencies is described elsewhere (MCH Leadership Competencies Workgroup 2009). Mouradian and Huebner (2007) reported results of a 2004 conference of MCH stakeholders (MCH Working Conference: The Future of Maternal and Child Health Leadership Training). They recommended self-assessments as one method to assess trainees' competency development; however, additional guidance about methods to assess leadership competencies directly was not provided.

The MCH Leadership Training Consortium at the University of North Carolina at Chapel Hill (UNC-CH) has developed and implemented a common leadership curriculum addressing the MCH Leadership Competencies across five MCH-funded programs (Dodds et al. 2010). Margolis et al. (2013) assessed graduates of the UNC-CH Leadership Training Consortium using a self-report measure and concluded that the program influenced participants' thinking and actions to be more interdisciplinary. Statistically significant differences between Leadership Training participants and non-participants were evident in some, but not all, cases. Participants who reported stronger agreement with the need for interdisciplinary practice were significantly more likely to report that they had worked to improve a program, organization, or partnership after graduation (Margolis et al. 2013).

Table 1 MCH leadership competencies (version 3.0)

MCH leadership competencies
Knowledge base
Self-reflection
Ethics & professionalism
Critical thinking
Communication
Negotiation & conflict resolution
Cultural competence
Family-centered care
Developing others through teaching & training
Interdisciplinary team-building
Working with communities & systems
Policy & advocacy

Purpose of the Current Study

The [*program name removed for blinded review*] program's evaluation is guided by an adapted version of the Kirkpatrick-Barr Model of Learner Outcomes (Barr et al. 1999). As part of our interdisciplinary training program's ongoing program evaluation, the present study investigates trainee level changes at Level 2 of our evaluation model: "Knowledge or Skills Related to Interdisciplinary Care and Disability Leadership." We build on the work of Fernandez et al. (2014) who found the retrospective pre-test method to be a valid tool in documenting perceived knowledge and skill change among trainees in an interdisciplinary MCH training program. To further substantiate the use of retrospective pre-test methodology in evaluating interdisciplinary training programs, we examined the traditional vs. retrospective pre-/post-test issue by comparing three cohorts' responses to a self-assessment tool in a traditional pre-/post-assessment in contrast to a retrospective pre-/post-assessment format. The present study was guided by the following research questions:

1) Does the response-shift phenomenon occur within a graduate-level interdisciplinary leadership program?
2) Are there meaningful differences in interdisciplinary trainees' change scores across the 12 MCH leadership competencies as measured through the use of the retrospective pre-test methodology?

Methods

Participants

After receiving ethical approval from our institutional review board, we recruited trainees from three separate training cohorts originating from one LEND program located in the southeastern United States ($N = 34$) to participate in the study. All trainees gave their informed consent prior to inclusion in the study. Each cohort lasted 1 year. To maintain confidentiality, specific years of training and training locations will not be provided. Trainees across cohorts were primarily female (87%) and diverse in terms of ethnicity (62% White, 32% Black, 2% Asian, 4% Hispanic). Trainees also represented a wide variety of disciplines, including Psychology, Law, Communication Disorders, Nutrition, Social Work, Disability Studies, Public Health, Rehabilitation Counseling, Special Education, and Physical Therapy. Several trainees also identified as self-advocates or as family members of an individual with a disability.

Measure

The *Early Career Professional: Self-Assessment Tool* (ECP; 26) is a 72-item web-based self-rating scale that yields scores for each of the 12 MCH Leadership Competencies (see Table 1). Each competency is represented by a series of from two to ten statements that serve as definitions of the key dimensions of that competency. The statements are categorized as representing either "basic" or "advanced" skills within each of the 12 competencies resulting in 24 possible scores. For example, the competency "Self-Reflection" is represented by three statements including: "Recognize that personal attitudes, beliefs, and experiences (successes and failures) influence one's leadership style," "Use self-reflection techniques effectively to enhance program development, scholarship, and interpersonal relationships," and "Identify a framework for productive feedback from peers and mentors." Trainees rate their proficiency on each of the 72 items using a 1 to 5 (low to high) Likert-type rating scale. Trainee responses are tabulated and a score report is generated that aggregates the trainee's responses into the identified categories and provides comparisons with other trainee groups. Graphs are also provided to visually represent individual trainee's responses in comparison to averages. The tool is traditionally administered online. The scale is generally intended as a means for increasing trainee awareness at the beginning of the LEND training period, but a number of LEND programs also use it as a pre-/post evaluation measure.

Procedures

At the beginning of the LEND training experience ($preT_1$), trainees rated their knowledge/skills in the 12 MCH Leadership Competencies via the ECP. One cohort completed their traditional pre-test on paper during their cohort orientation; the other two cohorts completed the ECP through the online tool prior to orientation and provided their results to one member of the research team. At the conclusion of the LEND experience each year, trainees simultaneously completed the traditional post-test ($postT_2$) and retrospective pre-test ($preT_2$) via a modified paper version of the online ECP tool. The retrospective pre-test utilized the same self-assessment of the 24 basic and advanced MCH leadership competencies, but respondents were asked to provide post-training ratings of what they *now* judged to be their *pre-training* level of performance—a retrospective pre-assessment at the end of their LEND training. Trainees did not have access to their Time 1 pre-test scores when completing the retrospective pre-test at Time 2.

Data Analysis

Analyses were performed using IBM SPSS Statistical Software version 19. We used non-parametric analyses for tests in which at least one variable was not normally distributed as determined by Shapiro–Wilk tests. In the case that both test variables were normally distributed, paired samples t-tests were used. Related samples Wilcoxon signed rank tests were employed to compare data from $preT_1$ and $postT_2$ and from $preT_2$ and $postT_2$. Scores were then divided by the number of observations over the two time points to calculate effect size, and interpreted using Cohen (1988) criteria. To further examine scores, we conducted related samples Wilcoxon signed rank and paired samples t-tests, where appropriate, to compare the two sets of pre-test scores ($preT_1$ to $preT_2$). We used an alpha level of 0.05 for all statistical tests.

Results

Descriptive statistics for the traditional pre-test, the retrospective pre-test, and the post-test are presented in Table 2. Related samples Wilcoxon signed rank and paired samples t-tests were used to compare the means of the two different pre-tests and between the pre-tests and the post-test.

Results comparing traditional pre-/post-assessment scores ($preT_1$ to $postT_2$) revealed decreases for 17 of the 24 areas (Table 2) and statistically significant decreases in two of the 24 areas, communication—basic, $z = -2.34$, $p < .05$, with a small effect size ($r = .28$), and develop others/teaching & mentoring—basic, $z = -2.83$, $p < .05$, with a medium effect size ($r = .34$), as noted in Table 3. Thus, trainees rated themselves significantly lower in these two areas post-training.

Results comparing the retrospective pre-/post-assessment scores ($preT_2$ to $postT_2$), however, revealed statistically significant increases in all of the 24 areas with medium to large effect sizes, as depicted in Table 4.

We conducted additional analyses to compare the two sets of pre-test scores ($preT_1$ to $preT_2$). Results, depicted in Table 5, indicated that trainees rated themselves significantly higher on the traditional pre-assessment ($preT_1$) than they did on the retrospective pre-assessment ($preT_2$) for all 24 areas.

Conclusions for Practice

Evaluators of interdisciplinary training programs often include pre- and post- self-ratings to determine program effectiveness and establish value-added. Such ratings may be subject to response-shift bias. In the present study, we compared three cohorts' responses to a self-assessment tool administered in a traditional pre-/post-assessment and

Table 2 Descriptive statistics for traditional pre-, retrospective pre- and post-tests

Competency	Mean	SD
Knowledge base		
Basic		
Traditional	3.21	0.87
Retrospective	2.24	1.03
Post	3.40	0.44
Knowledge base		
Advanced		
Traditional	2.91	0.74
Retrospective	1.74	0.99
Post	3.21	0.65
Self-reflection		
Basic		
Traditional	4.09	0.83
Retrospective	2.93	0.89
Post	3.78	0.45
Self-reflection		
Advanced		
Traditional	3.47	0.92
Retrospective	2.38	0.82
Post	3.21	0.61
Ethics & professionalism		
Basic		
Traditional	3.57	0.87
Retrospective	2.35	1.05
Post	3.36	0.84
Ethics & professionalism		
Advanced		
Traditional	3.19	0.78
Retrospective	2.06	0.95
Post	3.20	0.93
Critical thinking		
Basic		
Traditional	3.10	0.97
Retrospective	2.19	0.89
Post	3.12	0.95
Critical thinking		
Advanced		
Traditional	3.27	0.92
Retrospective	2.18	0.77
Post	3.03	0.87
Communication		
Basic		
Traditional	3.86	0.69
Retrospective	2.86	0.61
Post	3.48	0.45
Communication		
Advanced		
Traditional	3.28	0.74
Retrospective	2.26	0.73

Table 2 (continued)

Competency	Mean	SD
Post	3.21	0.57
Negotiation & conflict resolution		
Basic		
Traditional	2.94	0.89
Retrospective	2.00	0.92
Post	2.88	0.73
Negotiation & conflict resolution		
Advanced		
Traditional	3.32	0.84
Retrospective	2.24	0.89
Post	3.00	0.74
Cultural competency		
Basic		
Traditional	3.42	0.78
Retrospective	2.22	0.78
Post	3.20	0.58
Cultural competency		
Advanced		
Traditional	3.13	1.00
Retrospective	2.07	0.92
Post	3.10	0.89
Individual/family-centered supports		
Basic		
Traditional	3.02	0.99
Retrospective	2.01	0.94
Post	3.29	0.79
Individual/family-centered supports		
Advanced		
Traditional	3.06	0.82
Retrospective	1.82	0.76
Post	3.11	0.79
Develop others/teaching & mentoring		
Basic		
Traditional	3.69	0.72
Retrospective	2.50	0.75
Post	3.25	0.58
Develop others/teaching & mentoring		
Advanced		
Traditional	3.38	1.01
Retrospective	2.37	0.92
Post	3.20	0.58
Interdisciplinary team building		
Basic		
Traditional	3.62	0.75
Retrospective	2.54	0.72
Post	3.28	0.66
Interdisciplinary team building		
Advanced		
Traditional	3.29	0.74
Retrospective	2.28	0.84

Table 2 (continued)

Competency	Mean	SD
Post	3.10	0.71
Working with communities & systems		
Basic		
Traditional	3.25	0.80
Retrospective	2.15	0.96
Post	2.99	0.81
Working with communities & systems		
Advanced		
Traditional	2.91	0.90
Retrospective	2.04	0.80
Post	2.85	0.82
Policy & advocacy		
Basic		
Traditional	2.61	1.04
Retrospective	1.80	0.86
Post	2.85	0.79
Policy & advocacy		
Advanced		
Traditional	2.31	1.04
Retrospective	1.70	0.80
Post	2.62	0.86

$N = 34$ for all analyses

a retrospective pre-/post-assessment format to determine whether the response-shift phenomenon occurred and whether there were differences in interdisciplinary trainees' change scores across the relevant leadership competencies using the retrospective pre-test methodology.

Trainees rated their leadership skills at the beginning of the two-semester program and again at the end of the program. In these test conditions, we detected significant *decreases* in self-reported skill ratings from Time 1 ($preT_1$) to Time 2 ($postT_2$) in two of the 24 MCH leadership competency areas representing the 12 competencies. No significant increases were identified from Time 1 to Time 2. Thus, participants' traditional pre- and post-program self-reports indicated that they had gained *no* proficiency in most of the targeted areas and had decreased proficiency in two.

Trainees' pre-/post-program reports may have indeed been accurate representations, and there may have been no significant increase in perceived knowledge or skills during the yearlong interdisciplinary training program. To test this hypothesis, we compared the trainees' traditional pre-test self-ratings ($preT_1$) to their retrospective pre-test self-ratings ($preT_2$). We found that trainees had rated themselves as significantly more proficient in all 24 traditional pre-test ratings ($preT_1$; before any training) compared to the retrospective self-ratings of their proficiency at the beginning of the program ($preT_2$). To answer our first research question directly,

Table 3 Results of Wilcoxon analyses from traditional pre- and post-tests (preT$_1$ to postT$_2$)

Competency	Z	p-value	r
Knowledge base—basic	− 1.06	.29	.13
Knowledge base—advanced	− 1.88	.06	.23
Self-reflection—basic	− 1.71	.09	.21
Self-reflection—advanced	− 1.29	.20	.16
Ethics & professionalism—basic	− 0.53	.60	.06
Ethics & professionalism—advanced	− 0.79	.43	.10
Critical thinking—basic	− 0.35	.73	.04
Critical thinking—advanced	− 1.16	.25	.14
Communication—basic	− 2.34	.02*	.28
Communication—advanced	− 0.22	.82	.03
Negotiation & conflict resolution—basic	− 0.29	.77	.04
Negotiation & conflict resolution—advanced	− 1.90	.06	.23
Cultural competency—basic	− 1.73	.08	.21
Cultural competency—advanced	− 0.07	.95	.01
Individual/family-centered supports–basic	− 1.22	.22	.15
Individual/family-centered supports—advanced	− 0.25	.81	.03
Develop others/teaching & mentoring—basic	− 2.83	.01*	.34
Develop others/teaching & mentoring—advanced	− 0.96	.37	.12
Interdisciplinary team building—basic	− 1.91	.06	.23
Interdisciplinary team building—advanced	− 1.06	.29	.13
Working with communities & systems—basic	− 1.47	.29	.18
Working with communities & systems—advanced	− 0.45	.65	.05
Policy & advocacy—basic	− 1.25	.21	.15
Policy & advocacy—advanced	− 1.57	.12	.19

N = 34 for all analyses

*$p ≤ .05$

Table 4 Results of Wilcoxon analyses from retrospective pre- and post-tests (preT$_2$ to postT$_2$)

Competency	Z	p value	r
Knowledge base—basic	− 4.65	<.001**	.56
Knowledge base—advanced	− 4.89	<.001**	.59
Self-reflection—basic	− 4.24	<.001**	.51
Self-reflection—advanced	− 4.20	<.001**	.51
Ethics & professionalism—basic	− 4.60	<.001**	.56
Ethics & professionalism—advanced	− 4.80	<.001**	.58
Critical thinking—basic	− 3.76	<.001**	.46
Critical thinking—advanced	− 4.28	<.001**	.52
Communication—basic	− 4.57	<.001**	.55
Communication—advanced	− 4.78	<.001**	.58
Negotiation & conflict resolution—basic	− 4.31	<.001**	.52
Negotiation & conflict resolution—advanced	− 3.54	<.001**	.43
Cultural competency—basic	− 4.63	<.001**	.56
Cultural competency—advanced	− 4.52	<.001**	.55
Individual/family-centered supports—basic	− 4.73	<.001**	.57
Individual/family-centered supports—advanced	− 4.80	<.001**	.58
Develop others/teaching & mentoring—basic	− 4.34	<.001**	.53
Develop others/teaching & mentoring—advanced	− 4.48	<.001**	.54
Interdisciplinary team building—basic	− 4.58	<.001**	.56
Interdisciplinary team building—advanced	− 4.79	<.001**	.58
Working with communities & systems—basic	− 4.63	<.001**	.56
Working with communities & systems—advanced	− 4.77	<.001**	.58
Policy & advocacy—basic	− 4.75	<.001**	.58
Policy & advocacy—advanced	− 4.46	<.001**	.54

N = 34 for all analyses

**$p ≤ .001$

our data suggest that trainees in our study did exhibit a response-shift bias. Their pre-program self-ratings appear to be inflated and left little room for trainees to report increased competency at Time 2.

Our second research question referred to differences between results depending on the measurement method: traditional vs. retrospective pre-/post-test. In all cases, trainees rated themselves as more proficient before the program began (i.e., at Time 1) than they rated those same proficiencies retrospectively. Further, no Time 1 to Time 2 traditional rating indicated a statistically significant increase, yet all retrospective ratings indicated statistically significant increases in the targeted proficiencies. The retrospective rating method was superior in identifying participant's self-reported increases in proficiencies on the targeted skills.

As established by researchers in other fields (Pratt et al. 2000), our results demonstrate meaningful and statistically significant differences between the methods, with

the retrospective pre-test method appearing to minimize response-shift bias (Nimon et al. 2011) and providing a more sensitive measurement of self-reported growth. Our data suggest that participants may inflate their traditional pre-test response scores. In the language of training, trainees may "not know what they do not know" prior to the training experience and consequently be overconfident about their knowledge and skill levels. From a measurement standpoint, such inflated ratings leave little room to report increases in scores at the end of the program. This finding may explain our observed flat or decreased scores in the traditional format and provides justification for using the retrospective pre-/post-test format specifically for evaluating interdisciplinary training programs. Using this method, trainees may be better able to report the transformational nature of their experiences in an interdisciplinary training program, and program evaluators may obtain more accurate reflections of program results.

Table 5 Results of Wilcoxon/paired *t* test analyses from traditional pre-tests and retrospective pre-tests (preT$_1$ and preT$_2$)

Competency	t	Z	p value
Knowledge base—basic	4.85	–	<.001*
Knowledge base—advanced	–	−4.53	<.001*
Self-reflection—basic	–	−3.98	<.001*
Self-reflection—advanced	–	−3.55	<.001*
Ethics & professionalism—basic	–	−3.84	<.001*
Ethics & professionalism—advanced	5.34	–	<.001*
Critical thinking—basic	–	−3.51	<.001*
Critical thinking—advanced	–	−4.08	<.001*
Communication—basic	–	−4.30	<.001*
Communication—advanced	6.41	–	<.001*
Negotiation & conflict resolution—basic	–	−3.96	<.001*
Negotiation & conflict resolution—advanced	–	−3.97	<.001*
Cultural competency—basic	–	−4.44	<.001*
Cultural competency—advanced	–	−4.15	<.001*
Individual/family-centered supports—basic	–	−3.83	<.001*
Individual/family-centered supports—advanced	7.67	–	<.001*
Develop others/teaching & mentoring—basic	–	−4.51	<.001*
Develop others/teaching & mentoring—advanced	5.27	–	<.001*
Interdisciplinary team building—basic	–	−4.43	<.001*
Interdisciplinary team building—advanced	5.79	–	<.001*
Working with communities & systems—basic	6.38	–	<.001*
Working with communities & systems—advanced	4.97	–	<.001*
Policy & advocacy—basic	–	−3.68	<.001*
Policy & advocacy—advanced	3.53	–	.001*

N = 34 for all analyses

*$p \le .05$

Limitations

While our study makes a significant contribution to the literature on evaluating interdisciplinary training, it is not without limitations. First, our study used a relatively small sample size. Our LEND program is not offered as a credit-bearing course, thus it is difficult to ensure that *all* trainees attend *all* meetings as they manage academic and professional commitments throughout the training year. Not all trainees completed the survey at each of the three time points, resulting in a small sample of trainees for which we had complete data. While this missing data is a limitation of the current study, it may be an additional justification for using the retrospective format within interdisciplinary training programs. Traditional pre-/post-tests require attendance at two time points, whereas the retrospective format requires attendance at only one end-of-program time point.

This feature may be particularly desirable when training community members and professionals who often balance significant and competing demands. Second, our sample was recruited from a single LEND program. LEND programs vary greatly in structure, size, pedagogical philosophies, setting (i.e., medical school vs. academic settings) and other factors; thus, our results may not be generalizable to all LEND programs.

A general limitation of our study is that retrospective pre-tests may still be affected by multiple biases. Because admission to LEND programs is competitive and many trainees use LEND training as a career springboard, trainees may experience self-enhancement bias—they may want to appear as if they have experienced large gains in skills (Taylor et al. 2009). Implicit theory of change (the expectation that pre-test scores *should* be lower than post-test scores) may also be a factor when using this methodology (Taylor et al. 2009). Trainees may feel a sense of completion at the culmination of their training year and may thus underreport or temporarily ignore the bad while over-reporting the good (Krosnick 1999). Further, this methodology is subject to acquiescence, which is the tendency to respond affirmatively to any question (Krosnick and Fabrigar 1998). Finally, retrospective pre-test methodology is subject to recall error, which can result in underreporting (Schwarz 2007). It is important to note that these biases are not specific to retrospective pre-test methodology; all self-ratings have the potential for bias (Taylor et al. 2009).

Conclusions

This study indicates that trainees do perceive meaningful differences in their leadership competency following a year-long, intensive interdisciplinary training program. While competency-based interdisciplinary training programs present a unique challenge to MCH program evaluators due to trainees' diverse areas of expertise and individual professional goals, this study presents an alternative to traditional pre/post self-report measures. Although traditional pre-post measures are often the most common method for evaluating trainee progress along program competencies, this method is subject to response-shift bias because interdisciplinary training programs may aim to change the way participants conceptualize program competencies through the training process. The present study offers a cost-effective, feasible alternative to traditional pre/post self-report measures that may help control response shift bias in public health, MCH, and interdisciplinary settings.

These findings confirm the usefulness of the retrospective pre-test methodology in a unique interdisciplinary training setting. Future research should explore the potential of this methodology in other community-based MCH training

settings including continuing education sessions, employee and staff in-service trainings, and health promotion programming. Furthermore, future research exploring the use of retrospective pre-test methodology within and beyond interdisciplinary training programs should focus on minimizing the other potential biases described herein. Future studies should also examine the use of retrospective pre-tests in conjunction with other methodologies for assessment of trainee development, perhaps by multiple raters (e.g., trainees and program leadership). This method may provide a more objective measure of pre-/post- training change and reduce the potential for bias.

References

Association of University Centers of Disabilities. *The early career professional self-assessment tool*. Retrieved from http://www.aucd.org/ecp/satool/.

Barr, H., Freeth, D., Hammick, M., Koppel, I., & Reeves, S. (1999). Evaluations of interprofessional education: Two systematic reviews for health and social care. *British Educational Research Journal, 25*, 533–544. https://doi.org/10.1080/0141192990250408.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research on teaching*. Chicago: RandMcNally.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn.). Hillsdale: Lawrence Earlbaum Associates.

Coulter, S. E. (2012). Using the retrospective pretest to get usable, indirect evidence of student learning. *Assessment & Evaluation in Higher Education, 37*(3), 321–334. https://doi.org/10.1080/02602938.2010.534761.

Department of Human Resources for Health. (2010). *Framework for action on interprofessional education & collaborative practice (WHO/HRH/HPN/10.3)*. Geneva: World Health Organization.

Dodds, J., Vann, W., Lee, J., Rosenberg, A., Rounds, K., Roth, M., et al. (2010). The UNC-CH MCH leadership training consortium: Building the capacity to develop interdisciplinary MCH leaders. *Maternal and Child Health Journal, 14*, 642–648.

Drennan, J., & Hyde, A. (2008). Controlling response shift bias: The use of the retrospective pre-test design in the evaluation of a master's programme. *Assessment & Evaluation in Higher Education, 33*(6), 699–709. https://doi.org/10.1080/02602930701773026.

Fernandez, C. S. P., Noble, C. C., Jensen, E., & Steffen, D. (2014). Moving the needle: A retrospective pre- and post-analysis of improving perceived abilities across 20 leadership skills. *Maternal and Child Health Journal, 19*(2), 343–352. https://doi.org/10.1007/s10995-014-1573-1.

Goedhart, H., & Hoogstraten, J. (1992). The retrospective pretest and the role of pretest information in evaluative studies. *Psychological Reports, 70*, 699–704.

Health Resources & Services Administration. (2017). *Workforce training*. Retrieved from https://mchb.hrsa.gov/maternal-child-health-initiatives/workforce-training.

Howard, G. S., Ralph, K. M., Gulanick, N. A., Maxwell, S. E., Nance, D., & Gerber, S. L. (1979). Internal invalidity in pretest-posttest self-report evaluations and a revaluation of retrospective pretests. *Applied Psychological Measurement, 3*, 1–23.

Institute of Medicine [IOM]. (2003). *Health professions education: A bridge to quality*. Washington, DC: The National Academies Press. https://doi.org/10.17226/10681.

Interprofessional Education Collaborative. (2016). *Core competencies for interprofessional collaborative practice: 2016 update*. Washington, DC: Interprofessional Education Collaborative. Retrieved from http://www.aacn.nche.edu/education-resources/IPEC-2016-Updated-Core-Competencies-Report.pdf.

Interprofessional Education Collaborative Expert Panel. (2011). *Core competencies for interprofessional collaborative practice: Report of an expert panel*. Washington, DC: Interprofessional Education Collaborative. Retrieved from http://www.aacn.nche.edu/education-resources/ipecreport.pdf.

Kreulen, G. J., Stommer, M., Gutek, B. A., Burns, L. R., & Braden, C. (2002). Utility of retrospective pretest ratings of patient satisfaction with health status. *Research in Nursing & Health, 25*(3), 233–241. https://doi.org/10.1002/nur.10031.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*(1), 537–567.

Krosnick, J. A., & Fabrigar, L. R. (1998). *Designing good questionnaires: Insights from psychology*. New York: Oxford University Press.

Margolis, L. H., Rosenberg, A., Umble, K., & Chewning, L. (2013). Effects of interdisciplinary training on MCH professionals, organizations and systems. *Maternal and Child Health Journal, 17*(5), 949–958. https://doi.org/10.1007/s10995-012-1078-8.

Martineau, J. (2004). *Evaluating leadership development programs: A professional guide*. Greensboro: Center for Creative Leadership.

McDougal, J. A., Brooks, C. M., & Albanese, M. (2005). Achieving consensus on leadership competencies and outcome measures: The Pediatric Pulmonary Centers' experience. *Evaluation & the Health Professions, 28*(4), 428–446. https://doi.org/10.1177/0163278705281076.

MCH Leadership Competencies Workgroup. (2009). *Maternal child health leadership competencies: Version 3.0*. Retrieved from http://devleadership.mchtraining.net/mchlc_docs/mch_leadership_comp_3-0.pdf.

Miller, M., & Elder Hinshaw, R. (2012). The retrospective pretest as a gauge of change. *Journal of Instructional Psychology, 39*(4), 251–258.

Moore, D., & Tananis, C. A. (2009). Measuring change in a short-term educational program using a retrospective pretest design. *American Journal of Evaluation, 30*(2), 189–202. https://doi.org/10.1177/1098214009334506.

Mouradian, W. E., & Huebner, C. E. (2007). Future directions in leadership training of MCH professionals: Cross-cutting MCH leadership competencies. *Maternal Child Health Journal, 11*(3), 211–218. https://doi.org/10.1007/s10995-006-0170-3.

Mueller, C. E. (2015). Evaluating the effectiveness of website content features using retrospective pretest methodology: An experimental test. *Evaluation Review, 39*(3), 283–307. https://doi.org/10.1177/0193841X15582142.

Nimon, K., Zigarmi, D., & Allen, J. (2011). Measures of program effectiveness based on retrospective pretest data: Are all created equal? *American Journal of Evaluation, 32*, 8–28. https://doi.org/10.1177/10982140103878354.

Piwowar, V., & Thiel, F. (2014). Evaluating response shift in training evaluation: Comparing the retrospective pretest with an adapted measurement invariance approach in a classroom management training program. *Evaluation Review, 38*(5), 420–448. https://doi.org/10.1177/0193841X14546932.

Pratt, C. C., McGuigan, W. M., & Katzev, A. R. (2000). Measuring program outcomes: Using retrospective pretest methodology. *American Journal of Evaluation, 21*(3), 341–349.

Reyes-Akinbileje, B. (2013). *Health workforce programs in title VII of the public health service act* (CRS Report No. R43177). Retrieved from Congressional Research Service website: https://www.hsdl.org/?view&did=744097.

Rohs, F. R. (1999). Response shift bias: A problem in evaluating leadership development with self-report pretest-posttest measures. *Journal of Agricultural Education, 40*(4), 28–37. https://doi.org/10.5032/jae.1999.04028.

Schwarz, N. (2007). Retrospective and concurrent self-reports: The rationale for real-time data capture. In A. S. Stone, A. Shiffman, A. A. Atienza & L. Nebeling (Eds.), *The science of real-time data capture* (pp. 11–26). New York: Oxford University Press.

Taylor, P. J., Russ-Eft, D. F., & Taylor, H. (2009). Transfer of management training from alternative perspectives. *Journal of Applied Psychology, 94*, 104–121.