13th Annual

National Human Services Training Evaluation Symposium

2010

May 26-27, 2010
University of California, Berkeley

Proceedings

Sponsored by:
the California Social Work Education Center (CalSWEC)
in conjunction with
the California Department of Social Services *and*
the National Staff Development and Training Association of
the American Public Human Services Association

CalSWEC

This page intentionally left blank

Proceedings
of the
13th Annual
National Human Services
Training Evaluation Symposium

The material in this volume is based on presentations and discussions at the 13th Annual National Human Services Training Evaluation Symposium, held May 26–27, 2010, at the University of California, Berkeley.
This volume is also available online at
http://calswec.berkeley.edu/CalSWEC/Publications_3.html

Co-sponsors
California Social Work Education Center (CalSWEC)
California Department of Social Services
National Staff Development and Training Association of the American Public Human Services Association, Washington, D.C.

Published by
California Social Work Education Center
University of California, Berkeley
School of Social Welfare
Marchant Building, Suite 420
6701 San Pablo
Berkeley, CA 94720–7420
http://calswec.berkeley.edu

Editors
Barrett Johnson (Principal Editor)
Eri Tsujii

Table of Contents

Acknowledgments

We are again grateful for all of the effort that went into this year's NHSTES, both from CalSWEC staff and from our partners. It is especially gratifying to continue the good work of the symposium even as we face fiscal constraints at our various organizations. I truly appreciate the participation of all that attend each year—over the years we have created together a rare opportunity for highly specialized, collegial discussion.

As in previous years, an active Steering Committee provides guidance on the symposium program, as well as other aspects of the event. This year's members included Jane Berdie, Julie McCrae, Nancy Chavkin, Michael Nunno, Sherrill Clark, Cindy Parry, Todd Franke, Doug Pratt, Michelle Graef, Salonje Rochell, Henry Ilian, Leslie Rozeff, Leslie W. Zeitler, Christine Mattos, and myself. The support provided by the ad hoc committee was also greatly appreciated. This year's members included Anita Barbee, Amanda Schreiber, Dale Curry, Christine Tappan, and Norma Harris.

I would especially like to thank Leslie Zeitler, who consistently provides excellent overall coordination of the symposium. Her organizational skills once again proved indispensable. Other CalSWEC staff that assisted in the planning and implementation of the symposium included Monica Asfura, Barbara Turner, Yemisi Ogbodo, Karen Ringuette, and Eri Tsujii. Eri and Karen provided great assistance in editing the *Proceedings.*

Of course, special thanks also goes out to our presenters and facilitators, whose thought and preparation stimulated all of the great discussions.

Barrett Johnson, *LCSW*
Director, Child Welfare In-Service Training Project
CalSWEC
Principal Editor of the *Proceedings*

Introduction

Our 13th year ushered in some significant changes for the NHSTES. The group of professionals dedicated to human services training evaluation has continued to develop and flourish, even as resources to support training and evaluation activities have come under pressure across the country. For the participants in the symposium, this environment set the stage for two big changes: The group defined itself as an entity broader than the symposium itself and decided (after facing a reduction in available fiscal support in California) to rotate the hosting of the symposium between UC Berkeley and other entities providing training evaluation.

Several activities at the symposium involved further developing the identity of the group of dedicated training evaluators that has met for the past 13 years to share knowledge, skills, and resources. Leslie Zeitler led the group in an activity on the first day to create a name for the group. After some deliberation, the group decided on the National Human Services Training Evaluation Collaborative (NHSTEC). One of the collaborative's first tasks is to develop a handbook on human services training evaluation; discussions on this topic were led by Anita Barbee and Todd Franke. As the group develops, so does the symposium, and moving the event to a different location every other year will further institutionalize the event, while spreading some of the logistical know-how to other entities.

Substantial Discussions

Of course, the main intent of the symposium remains to facilitate meaty discussions on training evaluation topics, and this year offered a great opportunity for the NHSTEC to share knowledge and build skills. Dale Curry, Michael Lawler, and Jann Donnenworth kicked off the symposium with a skill-building session on scale development and refinement to measure transfer of learning. Drawing upon their extensive work in

promoting and evaluating transfer, they led the group in a discussion about how to create and use brief scales to measure transfer effectively.

After this session, Julie McCrae and Helen Calahane facilitated a discussion with Michael Nunno on a very practical issue in training evaluation: how to assure that we can collect data, and that the data that we do collect is not biased by the composition of the (usually self-selected) group of respondents. This session was particularly useful for those that struggle with engaging both the agency leadership and staff in recognizing the importance of evaluating training, and participating accordingly. In the afternoon of day one, Norma Harris and Lazarina Topuzova presented some of the work associated with Dr. Topuzova's dissertation, which examined the training and education continuum in Utah.

The end of the first day and the beginning of the second day involved further refinement of the handbook discussed above, led by Dr. Barbee and Dr. Franke. The remainder of day two featured a facilitated discussion about formulating a model to evaluate trainer performance, with Leslie Zeitler and Leslie Roseff, as well as a skill-building session on Rasch modeling with a panel featuring Henry Ilian, Cynthia Parry, and James Coloma. Both of these discussions epitomized how the symposium distinguishes itself as a professional event—they allowed time for careful consideration and brainstorming, while also giving symposium participants some new skills to build upon back at their offices. The Rasch modeling discussion provided demonstrations and hands-on exercises to apply the model to real data.

New Venue in 2011

As the symposium ended, the group reflected on the learning this year, but also looked ahead to next year. I'm very pleased to report on the outcome of the discussions: we were able to secure an East Coast venue for the symposium in 2011—Cornell University in Ithaca, New York. We hope this will help to expand the audience for the symposium, and continue to build upon the work done over the years at Berkeley. On behalf of CalSWEC, I want to reassure everyone of our commitment to the work, and

our ongoing assistance to plan the event annually, hosting it in Berkeley every other year.

Barrett Johnson, *LCSW*
Director, Child Welfare In-Service Training Project
CalSWEC
Principal Editor of the *Proceedings*

2010 NHSTES
Proceedings

Scale Reduction: Developing User-Friendly Human Services Training and Development Assessment Instruments

Dale Curry *Ph.D., LSW, CYC-P;*
Jann Donnenwirth, *M.S; &* Michael J. Lawler, *M.S.W., Ph.D.*

Abstract

When developing human service training and development instruments, a primary concern is establishing evidence of validity for specific applications. When shortening (reducing the size/number of items) an instrument to promote greater ease of administration, the focus shifts to maintaining the validity of the original instrument. Although a vast amount of literature describes best practice methods for developing a new scale, far less pertains to standards for reducing an instrument. Despite the benefits of reducing a scale, evaluators will find few resources available to provide guidance in shortening scales. This article references two multi-county transfer of learning evaluation research projects involving large samples from 1) public child welfare workers in northeast Ohio and 2) public welfare (eligibility) workers in California. It explores implications for shortening the Transfer Potential Questionnaire, while maintaining its predictive validity.

Introduction

Training is used in many fields as an intervention to promote effective performance on the job. In human services, a growing literature on training and development reflects professionalization of the disciplines and practice areas, driven partially by major policy shifts in child welfare services (e.g., Adoption Assistance and Child Welfare Act of 1980) and public welfare (e.g., Personal Responsibility and Work Opportunity and Work Reconciliation

Act of 1996) (PWORA). Transfer of learning concepts and principles drawn from the field of human resources have been integrated into best practices within human services training and development (Curry, Lawler, & Bernotavicz, 2006; Wilcox, Greenwood, Lawler, Pendleton, & Shutz, 2002). Evaluation of transfer of learning to performance on the job in child welfare services has received increased attention in recent years (e.g., Curry, McCarragher, & Dellmann-Jenkins, 2005).

Yet there is limited literature on the work and training of public welfare staff (e.g., eligibility workers). Welfare reforms, notably PWORA, led some researchers to study public welfare services from the perspective of policy implementation at the front-line or "street level" (e.g., Brodkin, 1997; Sandfort, 2000; Meyers, Glaser, & MacDonald, 1998). Hagen and colleagues at SUNY Albany have studied public welfare staff for more than 20 years, examining the growing complexity of the work (Hagen, 1987; Hagen & Wang, 1993), unrecognized training needs (Hagen, 1990), and perspectives of frontline workers on welfare reform (Hagen & Owens-Manley, 2002). In Georgia, 250 workers participated in a carefully developed training program on policy and interpersonal skills (Morton & Lindsey, 1986) that was evaluated through independent rating of videotaped interactions (Lindsey, Yarborough, & Morton, 1987) and self-reports from supervisor, worker, and client perspectives (Lindsey, 1993; Lindsey, Kropf, & Carse-McLocklin, 1995).

In 2009 the Center for Human Services at the University of California, Davis, (UC Davis) undertook a study of transfer of learning among public welfare staff in California. The study adapted instruments and procedures used in a prior study by Curry (1997) of child protective services workers in northeast Ohio.

Primary research questions for the California study were:

1. What training and development instructional design factors contribute to transfer of learning from classroom to practice?
2. What organizational elements within an agency support transfer of learning?

3. What individual traits of public welfare staff support transfer of learning?

Additional research questions addressed methods of studying transfer of learning in human services:

4. Is the Transfer Potential Questionnaire a valid instrument for assessing transfer potential across human service populations?
5. Can an abbreviated (more user-friendly) version of the Transfer Potential Questionnaire be developed as a valid instrument for assessing transfer potential across human service populations?

This article will discuss some preliminary findings pertaining to the above questions but will primarily focus on the last research question. Data from both the Ohio child welfare (CW) study and the California public welfare (PW) study will be used to help select items to create a shortened version of the Transfer Potential Questionnaire.

Transfer of Learning Assessment Instruments

Transfer Potential Questionnaire (TPQ)

The TPQ is a 68-item questionnaire assessing individual, organizational and training design factors. Factor analysis from the Ohio child welfare sample indicates that the scale is comprised of 11 factors (see Table 1). Training design is conceptually represented by factors 1 (trainer adult learning and transfer strategies), 2 (training relevance and applicability), five (application planning), and 10 (training/organization congruence). Individual factors include factors 7 (pre-training motivation to attend) and 8 (prior experience with training and application). Organizational factors include factors 3 (supervisor support for training and application), 4 (top management/organization support), 9 (co-worker support for training and application), and 11 (pre-training preparation).

Table 1

Subscales Derived from Ohio Child Welfare Sample

1. Trainer Adult Learning and Transfer Strategies
2. Relevance & Applicability
3. Supervisor Support for Training/Transfer
4. Organizational/Top Management Support
5. Application Planning
6. Perceived Learning
7. Pre-training Motivation
8. Prior Experience with Training/Application
9. Co-worker Support
10. Training/Organization Congruence
11. Pre-training Preparation

For the present study, subscales were created for both samples for each of the factors (a sum of each of the items loading on the factor) based on the 11 factors that emerged from the Ohio study. Although confirmatory factor analysis using structural equation modeling has not yet been conducted, preliminary analysis of factor analysis results indicate that there is a significant amount of similarity in the factor analysis results from both samples *(see Table 2).* Similarly, a composite Transfer Potential Index (TPI) was developed (the overall mean of the 68 TPQ items).

Table 2

Item Correlation with Transfer and Factor Loadings for Ohio Child Welfare and California Public Welfare Samples

Item Description	Correlation with Transfer			Factor	Factor Loading	
	CW	PW	AVG	CW	CW	PW
1. … I substantially increased my knowledge…	.47	.40	.435	6	.59	.54
2. As a result of the training, I have developed new skills.	.50	.38	.44	6	.46	.54
3. The training has affected some of my attitudes….	.32	.31	.315	6	.55	.64
4. I have a better conceptualization of what I already do…	.50	.43	.465	6	.34	.52
5. I am motivated to put this training into practice …	.47	.49	.48	2	.52	.47 (.51 factor 6)
6. I will implement on the job at least one concept/skill...	.36	.42	.39	2	.52	.52
7. My supervisor will support me in implementing ...	.20	.21	.205	3	.77	.63
8. My supervisor is familiar with the content of this training.	.26	.19	.225	3	.72	.71
9. I will meet with my supervisor to discuss application …	.33	.36	.345	3	.60	.43
10. My supervisor values staff training.	.11	.24	.175	3	.79	.68
11. My supervisor views this training as a high priority.	.19*	.24	.215	3	.83	.68
12. My supervisor assures that my caseload is covered ..	.11*	.12*	.115	3	.58	.51
13. My supervisor expects me to use this training on the job.	.22	.28	.25	3	.78	.48

Table 2—continued

Item Description	Correlation with Transfer			Factor	Factor Loading	
	CW	PW	AVG	CW	CW	PW
14. In my organization, top management values staff training.	.20	.25	.225	4	.77	.82
15. … top management views this training as a high priority.	.25	.28	.265	4	.66	.76
16. Even if no one notices, I will use knowledge learned ...	.42	.39	.405	2	.52	.54
17. The training was relevant to my job duties.	.39	.31	.35	2	.74	.71
18. The trainer helped me to see how the training can be...	.54	.29	.415	1	.60	.47
19. The trainer used a variety of examples ...	.15	.35	.25	1	.71	.66
20. I felt as though the trainer knew the type of clients..	.37	.30	.335	1	.50	.53
21. My co-workers value training.	.21	.21	.21	9	.79	.41 (4 org factor)
22. My co-workers will support my attempts to use ...	.23	.26	.245	9	.82	.50 (4 org factor)
23. I have already made a plan with a co-worker...	.37	.29	.33	5	.67	.60 (preplanning)
24. There is at least one co-worker who will be supportive...	.30	.33	.315	9	.42	no factor
25. I will have sufficient opportunities to practice …	.39	.45	.42	2	.61	.60
26. My organization values training.	.19	.28	.235	4	.77	.84

Table 2—*continued*

Item Description	Correlation with Transfer			Factor	Factor Loading	
	CW	PW	AVG	CW	CW	PW
27. My organization will support my attempts to use ...	.20	.31	.255	4	.51	.79
28. My organization expects me to use the training ...	.26	.34	.30	4	.49	.67
29. The information… can definitely be used with my clients.	.45	.35	.40	2	.68	.71
30. My client(s) will cooperate ...	.31	.26	.285	5	.59	.68 (others factor)
31. If I have trouble … I can count on others...	.28	.24	.26	(no loading)		.52 (others factor)
32. In the past, I have found training to be useful.	.20	.31	.255	8	.79	.55
33. When I think back to other training..I have used….	.19	.35	.27	8	.79	.56
34. I have a plan to implement this training.	.45	.41	.43	5	.53	.56 (2 applicability)
35. I am very confident that I will use the training on the job.	.53	.44	.485	2	.64	.68
36. I actively participated ...	.34	.28	.31	no loading		.54 (active partic.)
37. I will have the time to review materials and.. plan.	.36	.30	.33	5	.50	.33 (preplanning)
38. This training addresses a high priority training need.	.41	.26	.335	2	.49	.38 (1 trainer strat)
39. I had input into the selection of this training.	.09*	.23	.16	7	.82	.58 (preplanning)

Table 2—*continued*

Item Description	Correlation with Transfer			Factor	Factor Loading	
	CW	PW	AVG	CW	CW	PW
40. I voluntarily attended this training.	.16	.23	.195	7	.87	.59 (motivation)
41. Prior to the workshop, I was motivated to attend.	.16	.23	.195	7	.68	.59 (motivation)
42. During the training, I was thinking of ways I could apply ...	.43	.37	.40	2	.47	.42
43. .. provided sufficient opportunities to practice ...	.44	.41	.425	1	.63	.61
44. I can think of specific cases/clients ..can be used.	.42	.36	.39	2	.63	.62
45. My supervisor helped to prepare me ….	.29	.31	.30	11	.45	.73 (preplanning)
46. The trainer … aware of underlying principles ...	.45	.26	.355	1	.56	.43
47. Prior ... I heard that this training was ..valuable.	.33	.20	.265	11	.76	.54 (preplanning)
48. Most training …by UC Davis is of the highest quality.	.21	.33	.27	8	.55	.34
49. The trainer is obviously an expert in this subject matter.	.30	.37	.335	1	.67	.68
50. The trainer provided some practical ideas ...	.49	.36	.425	1	.73	.75
51. I had fun during the training.	.40	.29	.345	1	.71	.57
52. The trainer .. a climate conducive to adult learning….	.40	.29	.345	1	.80	.70
53. There was sufficient time ...	.25	.16	.205	1	.56	.55
54. There was sufficient time ...	.25	.16	.205	1	.56	.55
55. There was sufficient time ...	.25	.16	.205	1	.56	.55

Table 2—*continued*

Item Description	Correlation with Transfer			Factor	Factor Loading	
	CW	PW	AVG	CW	CW	PW
56. There was sufficient time ...	.25	.16	.205	1	.56	.55
57. The other …participants contributed to an adult learning ..	.32	.21	.265	1	.63	.60
58. …trainer gave examples of when to use ..	.44	.35	.395	1	.74	.75
59. The trainer gave examples of when not to use..	.40	.35	.375	1	.61	.50
60. The trainer provided .. application aides…	.34	.36	.35	1	.64	.73
61. .. encouraged participants to learn from each other.	.31	.26	.285	1	.69	.59
62. The trainer helped motivate me to want to try out ..	.52	.35	.435	1	.66	.69
63. The … objectives were adequately addressed.	.48	.33	.405	1	.68	.72
64. … trainer … participants to write down ideas ...	.30	.27	.285	1	.41	.45 (also fac 10)
65. .. effectively plan for application of learning.	.44	.33	.385	1	.71	.53 (also fac 10)
66. .. content is consistent with my agency's mission…	.27	.33	.30	10	.50	.47 (3 org support)
67. … consistent with my agency's policies ….	.34	.34	.34	10	.50	.49
68. .. will help me to continue learning in this topic area	.48	.40.	.44	10	.43	.59 (trainer strat)
69. .. I will be a more effective worker.	.49	.50	.495	1	.49	.51
70. …can help make a difference with my client(s).	.52	.30	.41	2	.50	.54
71. Overall, I am very satisfied with this training.	.52	.37	.445	1	.72	.60

Note * not significant at .05 level.

Prior to its use in the California study the TPQ was found to have high internal reliability (Chronbach's alpha=.96) and significantly predict the transfer of learning of child protective social workers (r=.62). It was later found to have a positive relationship with long-term employee retention with the same group of child protective social workers (Curry, 1997; Curry & Chandler, 1999; Curry, McCarragher, & Dellmann-Jenkins, 2005).

Human Services Training Effectiveness Postcard (HSTEP)

The HSTEP is a brief questionnaire consisting of five items designed to incorporate the four major areas/levels of training evaluation in Kirkpatrick's (1975) model into a single assessment index. Kirkpatrick's widely accepted 4-level evaluation model includes 1) reaction/satisfaction, 2) learning, 3) behavior/transfer of learning, and 4) results of transfer of learning (Bell & Kerr, 1987; Birnbrauer, 1987; Krein and Weldon, 1994; Pine & Tingley, 1993). Table 3 lists items on the HSTEP and their association with the four levels. The last three HSTEP items were summed and used as a composite variable assessing transfer of learning.

Table 3

Relationship of Evaluation Postcard Items to the Four Levels of Kirkpatrick's Model

Level	HSTEP Item Number and Description
1. Reaction/satisfaction	#1 - Overall, I am very satisfied with the workshop.
2. Learning	#2 - During the workshop, I learned a substantial amount of information.
3. Behavior/transfer of learning	#3 - I have used the knowledge and skills I learned from the workshop on the job.
4. Results of transfer of learning	#4 - As a result of using the knowledge/skills from the workshop, I have observed client progress. #5 - As a result of the workshop, I am a more effective worker.

The HSTEP is also intended to provide acceptable return rates from training participants who may have high caseloads and little time for assisting in the evaluation of training programs. The return rate was 74% in the Ohio study and 57% in the California study. Internal reliability (Chronbach's alpha) for the Ohio study was .90 for all five HSTEP items and . 88 for the three transfer of learning items. Similarly, alpha=.87 in the California study for all five items and .85 for the three transfer items, indicating that the HSTEP has high internal reliability and can elicit adequate returns from participants (Curry, 1997).

In addition to providing quantitative outcome data, the instrument was designed to elicit qualitative information on the process of transfer of learning. The participants were requested to list factors which helped or hindered their application of learning.

Process and Samples

In both studies, the TPQ was administered to study participants at the conclusion of each training session. Three months after the conclusion of training, participants are mailed the HSTEP questionnaire.

In Ohio, study participants were line child protective services social workers who attended training at the Northeast Ohio Regional Training Center over a three-month period of time. The child protective services role (e.g., intake, ongoing, foster, adoptive, kinship care worker) was not identified. Ninety six percent (96%) of the training participants (598) representing 14 counties in northeast Ohio completed the TPQ and 441 returned the HSTEP.

The California study involved 459 public welfare staff in 23 northern and central California counties who attended training from March 23–June 30 2009. Eighty percent (80%) of the respondents self-identified as an eligibility worker, integrated case worker, or employment and training worker. Two hundred and fifty-nine (259) returned the HSTEP.

Findings

Key Differences between the Samples

- More public welfare staff were told they had to attend training (49% vs. 13 % of child welfare staff).
- The child welfare sample is more educated (66% with bachelor's degree compared to 35%).
- The Ohio child welfare staff had a larger percentage of African Americans (13% vs. 4%.)
- The California public welfare sample had a larger percentage of Hispanics (37% vs. 0%).

Variables/Factors Affecting Perceived Transfer of Learning

In the public welfare sample, there is a significant correlation between Transfer Potential Index (TPI) and perceived transfer three months later (predictive validity coefficient r=.53). While this is somewhat lower than child welfare sample (r=.62), it provides evidence that the TPQ is valid with both groups of human service professionals, suggesting the TPQ may be generalized across a range of human services.

Ten (10) of 11 subscales derived from the child welfare study significantly correlated with transfer ($p<.000$). Only the pre-training motivation subscale failed to significantly correlate with transfer of learning. In the public welfare sample, all 11 subscales significantly correlated with transfer of learning *(see Table 4).*

Table 4
Subscale Correlation with Transfer for CW and PW Samples

Subscale	Correlation with Transfer CW	PW
1. Trainer Adult Learning and Transfer Strategies	.53	.48
2. Relevance & Applicability	.57	.51
3. Supervisor Support for Training/Transfer	.30	.39
4. Organizational/Top Management Support	.26	.35
5. Application Planning	.49	.43
6. Perceived Learning	.49	.47
7. Pre-training Motivation	.04	.29
8. Prior Experience with Training/Application	.25	.41
9. Co-worker Support	.24	.32
10. Training/Organization Congruence	.49	.46
11. Pre-training Preparation	.37	.31

Note: Subscale 7 is not significant for the child welfare sample.

Only one TPQ item did not significantly correlate with transfer with a minimum $p<.05$ level (Item 12: "My supervisor assures my caseload is covered so I am not interrupted while I am in training"). This item was not significant in either study *(refer back to Table 2).*

Primary Research Questions—Training Public Welfare Staff

The primary research questions focused on training for public welfare staff. Key preliminary findings include:

- All training design subscales correlate with transfer: adult learning and transfer strategies; training relevance and applicability; application planning; and training/organization congruence.
- All organizational subscales correlated with transfer: supervisor support; top management/organization

- support; co-worker support for training and application; and pre-training preparation.
- Two individual participant subscales correlate with transfer: application planning and prior experience with training and application in the Ohio Child Welfare (CW) study. Pre-motivation to attend did not correlate with transfer. All subscales correlate with transfer in the California Public Welfare (PW) study.

Secondary Research Questions—Valid, Shortened Instrument for Human Services

Preliminary findings related to the secondary research questions show the potential for a transfer of learning instrument that can be used for the human services field. Specifically:

Is the Transfer Potential Questionnaire a valid instrument for assessing transfer potential across human service populations? Both populations show significant correlation between Transfer Potential Index and perceived transfer three months later—evidence of predictive validity. Internal reliability (Chronbach's alpha) results are similar between the two studies (.97 for the public welfare study and .96 for the child welfare study).

Can an abbreviated TPQ be developed as a valid instrument for assessing transfer potential across human service population? A confirmatory factor analysis on the California sample needs to be done, but internal reliability results are similar—and high—between the two studies. The very high internal reliability suggests a certain degree of redundancy of items in the measurement of factors, so a shortened version may be indicated.

Developing an Abbreviated TPQ

The TPQ is an instrument that appears to be a valid predictor of transfer of learning for workers across human service populations. Given the difficulty of retrieving transfer of learning information from busy professionals once they leave the training room and are working on the job, capturing valid transfer of learning information in the session prior to leaving can serve a variety of useful purposes. Results from the tool can suggest to training administrators individual, organizational and training

design transfer-of-learning interventions. Since there is an established relationship between the TPQ and later transfer, the TPQ can possibly be used as a proxy for later transfer, becoming an outcome indicator itself.

However, completing 68 items (not including demographic items) can appear to be a daunting task after training, when workers are anxious to return home or back to the job to deal with pressing client needs. In addition, the very high internal reliability of Chronbach's alpha coefficients from both studies seem to indicate that shortening the scale is warranted. A shortened scale could perhaps be used after every training in addition to or replacing current reaction evaluations, providing useful information regarding transfer potential and suggestions for promoting more effective transfer of learning. This information could also be fed back to trainers and other key personnel, encouraging trainers to more effectively train for transfer. The process for developing a user-friendly shortened version of the TPQ that maintains the validity and utility of the original version will be the focus of the remainder of this article.

Internal, External, and Judgmental Indicators

When developing a human service training and development instrument, a primary concern is establishing evidence of the instrument's validity for specific applications. When shortening an instrument to promote greater ease of administration, the focus shifts to preserving the validity of the original instrument. While the literature provides a substantial amount of guidance for constructing new scales, relatively little research addresses guidelines for shortening scales. As a result, there is much variability and less documentation regarding the process of scale reduction (Stanton, Sinar, Balzer, & Smith, 2002). Research conducted by Stanton et al. (2002) suggests using a combination of internal, external and judgmental criteria for shortening scales. The following describes the researchers' progress toward reducing the length of the TPQ using these three criteria.

Internal item qualities. Stanton et al. (2002) describe internal item qualities as self-referential; determined in reference to the scale itself without reference to indicators external to the scale such as measures assessing later transfer of learning. Internal data

used for shortening the TPQ include factor loadings and corrected item-total correlations from both the Ohio CW and California PW sample. An attempt was made to include items pertaining to each of the 11 factors extracted in the Ohio CW study, especially those items that loaded on similar factors in the California PW sample. Since more than half of the items on the TPQ loaded on the first three factors for both samples, consideration was given to include more of these items on the shortened scale relative to the other factors.

External item qualities. A focus on external item qualities examines the relation of an item with a measure or measures of a criterion or criteria that is external to the instrument (TPQ). Although this study included only one external measure (HSTEP transfer composite variable), results were obtained for two different sample populations (Ohio CW and California PW). Since correlations sometimes vary across different samples and contexts, correlation results for the individual item to transfer criterion measure as well as the average correlation between the two samples for each item were obtained *(see Table 2).*

Judgmental item qualities. In addition to examining the internal and external statistical relations, other criteria such as feedback from experts regarding the relevance of items to a particular population should be considered. For this study participants from the National Human Service Training Evaluation Symposium (NHSTES) were invited to provide feedback regarding the relevance of the items to overall transfer potential as well as the 11 transfer factors/constructs (face validity), possible redundancy, and clarity of items. Several participants agreed to participate in a subsequent rating of preliminary items selected for inclusion in abbreviated versions based on internal and external criteria. Participants of the NHSTES are considered experts in the field of human service training evaluation and attend the symposium by invitation only.

Also, there are potentially conflicts between the internal and external indicators. According to Stanton et al. (2002), "thoughtful compromises" are sometimes made when choosing items for a shortened scale. For example, choosing the top items with the highest correlation with the transfer criterion may exclude items

that tap into important transfer factors as well as risk lower validity values with other external criteria. Similarly, a strategy that chooses items from each construct with the highest factor loadings (a common practice by many in scale reduction) without considering the relationship to an external criterion may sacrifice concurrent validity.

Thoughtful Compromises—Combining Internal, External and Judgment Criteria

When examining Table 2 (providing a summary of both internal and external indicators), a few items immediately appear to be candidates for inclusion (rating highly on both internal and external criteria) or exclusion (low ratings on both). For example, item 35 ("I am very confident I will use training on the job.") has the second highest average validity coefficient (r=.485) and relatively high factor loadings on factor 2 (relevance and applicability), suggesting it is a good candidate for inclusion on the reduced scale. Similarly, an example of an item to exclude is item 12 ("My supervisor assures that my caseload is covered so I am not interrupted when I am in training."). This item fails to correlate significantly with either sample and has moderate to low factor loadings on the two samples (.58, .51) relative to the other items.

However, the goal of creating a much briefer scale that can be completed by busy human service training participants prior to leaving a training session challenged the researchers to select items for inclusion that were not clear choices based on both internal and external data. For example, the item with the highest average validity correlation (item 66) has modest factor loadings (.49, .51) on factor 1. The item with the third highest average validity correlation (item 5) cross-loads on two factors in the PW sample, having a higher loading on factor 6 (perceived learning) than factor 2 (relevance and applicability). This item loaded on factor 2 in the CW sample. Stanton et al. (2002) recommend giving preference to external criteria when clear-cut choices are not apparent. With this in mind, the researchers selected items with the highest correlations while also sampling at least one item from each of the eleven factors. The researchers also recognized that several reduced size versions could perhaps be constructed

with different advantages and disadvantages relative to the original 68-item version of the TPQ.

Statistics regarding one draft shortened version (33 items) of the TPQ renamed as the Application Potential of Professional Learning Inventory (APPLĪ 33) *(see Appendix)* are provided below.

Validity and Reliability of the APPLĪ 33

After creating a reduced-length scale, the next step according to Stanton et al. (2002) is to assess the validity correlations between the shortened scale and its full-length scale. This is followed by assessing the internal consistency to ensure that it falls within an acceptable range, accounting for some shrinkage due to the smaller number of items.

The validity correlation between the APPLĪ 33 (average of 33 items) and the transfer criterion for the California PW sample is .57. The correlation with the full scale TPQ is .98. Chronbach's alpha for the 33 items is .95. For the Ohio CW sample, the correlation between the APPLĪ 33 and transfer is .60. The correlation with the full-length scale is .97. Internal consistency (Chronbach's alpha) is .95.

Reducing the TPQ by more than half of the items resulted in comparable validity coefficients and little shrinkage in internal consistency for both samples.

Factor analysis of the APPLĪ 33 for both samples resulted in the extraction of 5 and 6 factors for the PW and CW samples (eigenvalue>1). Further analysis of the factor structure of the APPLĪ 33 or other reduced-length versions is necessary prior to the development of subscales.

Discussion

The findings indicate that both the full-length TPQ and reduced version APPLĪ 33 is predictive of later transfer of learning for both human service worker populations. Although the full-length version can provide additional assessment information pertaining to its eleven factors, a shortened scale such as the APPLĪ 33 can provide practical information and evaluative data regarding the effectiveness of training. The APPLĪ 33 should be further field tested with another population since both samples were used in the selection of its items. Multi-group Structured Equation Modeling should be conducted to compare the scale-

level correlation matrices using a sample with the full-length scale and the cross-validation sample containing the reduced length scale (Stanton et al., 2002). The high internal consistency results of the APPLĪ 33 for both samples may indicate the possibility of further reduction of the scale while still maintaining acceptable validity and consistency standards. Additional feedback from NHSTES experts regarding the clarity and face validity of the items may help further refine the APPLĪ 33.

References

Bell, J.D., & Kerr, D.L. (1987). Measuring training results: Key to managerial commitment. *Training and Development Journal, 41*, 70–73.

Birnbrauer, H. (1987). Evaluation techniques that work. *Training and Development Journal, 41*, 53–55.

Brodkin, E.Z. (1997, March). Inside the welfare contract: Discretion and accountability in state welfare administration. *Social Service Review*, 1–33.

Curry, D.H. (1997). Factors affecting the perceived transfer of learning of child protection social workers. Unpublished doctoral dissertation, Kent State University, Kent, Ohio.

Curry, D., & Chandler, T. (1999). The Human Services Training Effectiveness Postcard (HSTEP): A tool for research and evaluation of human services training. *Professional Development: The International Journal of Continuing Social Work Education, 2*, 43–56.

Curry, D., Lawler, M., & Bernotavicz, F. (2006). The field of training and development in human services: An emerging profession? *Training and Development in Human Services, 3*, 5–11.

Curry, D., McCarragher, T., & Dellmann-Jenkins, M. (2005). Training, transfer, and turnover: Exploring the relationship among transfer of learning factors and staff retention in child welfare. *Children and Youth Services Review, 27*, 931–948.

Hagen, J. (1987). Income maintenance workers: Technicians or service providers? *Social Service Review, 61*, 261–271.

Hagen, J. (1990). Training income maintenance workers: A look at the empirical evidence. *Journal of Continuing Social Work Education, 5*(2), 3–9.

Hagen, J.L., & Owens-Manley, J. (2002). Issues in implementing TANF in New York: The perspective of frontline workers. *Social Work, 47*(2), 171–184.

Hagen, J.L., & Wang, L. (1993). Roles and functions of public welfare workers. Administration in *Social Work, 17*(2), 81–103.

Kirkpatrick, R. (1975). *Evaluating Training Programs*. American Society for Training and Development. Madison, WI.

Krein, T.J., & Weldon, K.C. (1994). Making a play for training evaluation. *Training & Development, 48,* 62–67.

Lindsey, E., Yarborough, D., & T. Morton. (1987). Evaluating interpersonal skills training for public welfare staff. *Social Service Review,* 61, 623–635.

Lindsey, E.W. (1993). Training Georgia's eligibility workers. *Public Welfare, 51*(2), 35–41.

Lindsey, E.W., Kropf, N.P., & Carse-McLocklin, S. (1995). Training public assistance workers in policy and interpersonal helping skills. *Research on Social Work Practice, 5*(1), 20–35.

Pines, J. , & Tingley, J.C. (1993). ROI of soft skills training. *Training, 30*, 55–60.

Meyers, M.K., Glaser, B., & MacDonald, K. (1998). On the front lines of welfare delivery: Are workers implementing policy reforms? *Journal of Policy Analysis and Management, 17*(1), 1–22.

Morton, T.D., & Lindsey, E.W. (1986). Toward a model for interpersonal helping skills use and training in public welfare practice. *Journal of Continuing Social Work Education, 4*(1), 18–24.

Sandfort, J.R. (2000). Moving beyond discretion and outcomes: Examining public management from the front lines of the welfare system. *Journal of Public Administration Research and Theory,* 10, 729–756.

Stanton, J.M., Sinar, E.F., Balzer, W.K., & Smith, P.C. (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology, 55*, 167–194.

Wilcox, M.R., Greenwood, B., Lawler, J.J., Pendleton, D., & Schutz, H. (2002). Best teaching practices in continuing professional education: An interdisciplinary study. *Training and Development in Human Services, 2*, 108–113.

Appendix
Items for Shortened Version of the TPQ: Application Potential of Professional Learning Inventory (APPLĪ 33)

1. As a result of the training, I substantially increased my knowledge on this topic.
2. As a result of the training, I have developed new skills.
3. The training has affected some of my attitudes concerning this topic area.
4. As a result of this training, I have a better conceptualization of what I already do on the job.
5. I am motivated to put this training into practice on the job.
6. I will meet with my supervisor to discuss application of this training on the job.
7. My supervisor expects me to use this training on the job.
8. Even if no one notices, I will use knowledge learned from this training on the job.
9. The trainer helped me to see how the training can be applied on the job.
10. The information I received from this training can definitely be used with my clients.
11. I have already made a plan with a co-worker to use this training.
12. There is at least one co-worker who will be supportive of my application attempts.
13. I will have sufficient opportunities to practice the new ideas/skills/techniques on the job.
14. My organization expects me to use the training on the job.
15. When I think back to other training I have attended, I can say that I have used the training on the job. I can even think of specific application examples.
16. I have a plan to implement this training.
17. I am very confident that I will use the training on the job.
18. I will have the time to review materials and make an implementation plan.
19. Prior to the workshop, I was motivated to attend.
20. During the training, I was thinking of ways I could apply the training content to the job.
21. The trainer/training provided sufficient opportunities to practice new information/skills.
22. I can think of specific cases/clients to which (with whom) this training can be used.
23. My supervisor helped to prepare me for this training by discussing my learning needs and potential applications.
24. The trainer provided some practical ideas that can be used on the job.

25. The trainer gave examples of when to use ideas/skills/strategies on the job.
26. The trainer helped motivate me to want to try out training ideas on the job.
27. The workshop objectives were adequately addressed.
28. This training content is consistent with my agency's mission, philosophy and goals.
29. This training content is consistent with my agency's policies and my individual responsibilities.
30. This training will help me to continue learning in this topic area.
31. As a result of the training, I will be a more effective worker.
32. The information I learned today can help make a difference with my client(s).
33. Overall, I am very satisfied with this training.

2010 NHSTES Proceedings

Scale Reduction: Developing User-Friendly Human Service Training and Development and Assessment Instruments

Discussion Summary

Presenters: Dale Curry, *Ph.D., Kent State University;* Jann Donnenwirth, *M.S., University of California, Davis; &* Michael Lawler, *M.S.W., Ph.D., University of California, Davis*

Wednesday, May 26, 2010
10:00-11:30 am

Today's Discussion Topics

This presentation and facilitated discussion centered on a research study and scale reduction. The presenters discussed three main issues: 1) internal correlation, 2) external correlation, and 3) professional judgment and face validity. All of these were framed in terms of evidence-based practice. People often use the term "evidence-based practice" without understanding that it includes testing, examining external comparisons and contexts, and professional judgment.

Background of the Study

- The process for this study was based on a scale that Dr. Curry had developed for a similar project in Ohio with child welfare social workers.
- Public child welfare work has shifted since the mid-1990s—most notably the work has changed from eligibility work and filling out documents, to interacting with clients.

- The study utilized 2 instruments adapted by Dr. Curry: the Transfer Potential Questionnaire (which has 68 items); and the Follow-Up Postcard (with 5 additional questions). The initial questionnaire was filled out at the end of the training session, and the postcard was mailed out three months later.

Research Questions

The presenters outlined the research questions; which included three primary questions and two supplemental questions.

1. What are the training and development instructional factors that contribute to transfer of learning?
2. What organizational elements contribute to transfer of learning?
3. What are the individual traits of public welfare staff that contribute to transfer of learning?
4. (Supplemental question) Is the questionnaire a valid instrument
5. (Supplemental question) Can we make the questionnaire smaller to use it more frequently as a measure of the value of training itself?

Findings

The presenters then discussed the findings for each question.

- Question 1: The study found a strong correlation with four factors associated with instructional design:
  1) adult learning and training strategies that are included in the training;
  2) the perceived training relevance and applicability;
  3) the actual application within the training session; and
  4) the congruence of training with the organization.
- Question 2: A correlation was found between public welfare staff and child welfare staff in terms of:
  1) supervisor support;
  2) top management and organizational support;
  3) coworker support for training and application; and
  4) pre-training preparation.
- Question 3: The findings for the third question were more mixed. Little difference was found between the individual traits, and the difference that was found was mainly in the area of pre-motivation to attend the training itself.

- Questions 4 and 5: For the two supplemental questions, no definitive and comprehensive answers could be formed, but results indicate that the internal validity is "very similar and very high" between this study and the previous study in Ohio.

Basics of Developing Specific Instruments

- Developing specific instruments in human service training is important. Many of the assessment tools that have been used for years have not had basic analysis done on them.
- Several instruments are currently under development, including those in Kentucky, at UCLA, and at the University of Toronto.
- It may be more helpful to look at specific abilities rather than abstract competencies.
- Other important considerations are reliability, validity, and user-friendliness (including brevity) of the tools that are being developed.

Developing a Transfer of Learning (TOL) Scale

- One must consider the intent of the transfer potential questionnaire as well as what happens before, during, and after training that might impact (either by facilitating or inhibiting) the transfer of learning process.
- Many people besides the trainer are important players in the transfer of learning process, including supervisors, coworkers, clients. These players are involved in the learning process before, during, and after training
- Several factors influence TOL before a trainings, such as prior knowledge, prior experience with training on a previous job, past behavior, and reputation of the training. For example, if the supervisor explains the training prior to attending, then this is a significant pre-training factor.
- Similarly, other factors impact TOL during the training, such as: emergencies that remove trainees from the training; fidelity of the curriculum; the voluntary/involuntary nature of the training; and whether or not trainings are conducted face-to-face or over long distances.

Discussion on Shortening a Scale

- The presenters discussed how to cut out items and shorten scales and still maintain validity in the results. Questions to consider include:

  - What is the conceptual framework?
  - What would be the potential to use this?
  - What do you want to know about your training population?
  - What could help you intervene?
- Some of the factors that can help influence transfer are not interesting to administrators or bureaucrats, but are still important to capture.
- One participant suggested using factor analysis to tease apart similar items. The presenter agreed that it might be helpful in figuring out what is important for the trainer, but not necessarily in shortening the scale.
- The presenters discussed a 10 step approach to shorten scales:
  - Step 1: Generate item-level indices of externalizing qualities
  - Step 2: Look at internal correlation, external correlation, and professional judgment to see if we can get the best out of all three and combine them.
  - Step 3: Have experts rate face validity.
  - Step 4: Sort items by item quality; prioritizing items.
  - Step 5: Use professional judgment to evaluate the item quality scores.
  - Step 6: Reduce the scale from the top items.
  - Step 7: Assess the internal consistency of the reduced scale.
  - Step 8: Field test the reduced-length scale.
  - Step 9: Re-check item-level performance.
  - Step 10: Conduct multi-group structured equation. modeling to compare the scale-level correlation matrices using a sample of the full-length scale and a cross-validation of the reduced length scale.

See page 5 for the paper submitted by the presenters.

2010 NHSTES
Proceedings

Is This Research Mandated?: Response and Non-response Bias in Child Welfare Training Evaluation and Research

Julie McCrae, *Ph.D.*, & Helen Cahalane, *Ph.D.*, *LCSW*, *University of Pittsburgh*, *School of Social Work*

Abstract

This brief article and discussion review the challenges of engaging public child welfare staff in training evaluation and data collection. Poor return rates can affect the generalizability of findings and can create a biased sample. Possible solutions to these problems were explored by the symposium participants.

Introduction

"Hard to reach" groups in human services research are often those who may be homeless, have cognitive challenges, or be otherwise difficult to locate or engage in research. Child welfare workers, while not typically as personally challenged, are challenging to engage in research because the work that they do leaves them little time in the office, and the demands of high caseloads leave them little cognitive resources for participating in anything beyond daily work tasks. Qualities of child welfare agencies, as historically compliance-driven and hierarchical, also contribute to recruitment challenges because workers distrust the research process. Pennsylvania has implemented several strategies to engage and retain workers in research, with disappointing results. The purpose of this discussion is

to describe Pennsylvania's experiences with response and non-response bias and to gain feedback from others about engaging child welfare workers in research and obtaining valid data.

Background

Research in training, transfer of learning, organizational effectiveness, and service is only as valid as the quality and representativeness of the data that is collected. In Pennsylvania (PA), three training-related studies and three service-focused studies have raised concerns about response and non-response bias. In a study of transfer of learning, workers who completed follow-up surveys were significantly younger, more likely to work in on-going compared with intake units, and be graduates of Title IV-E traineeships compared with workers who did not complete follow-up surveys. Response rates were 84% at baseline and just 42% of participants at follow-up. Web-based surveys sent to 63 supervisors in another effort yielded 19% participation. A user-friendly, report-producing database that helps agencies meet state and federal policy requirements around young children's developmental needs has yielded 65% participation after 7 months of engagement strategies and incentives. These studies suggest serious challenges with representativeness because of the extent of non-response bias.

A second issue concerns social desirability response bias. For example, informal conversations with agencies indicate that resources prevent some agencies from using the database just described; yet scaled response items during a research interview show that 80% of agencies report that they "have the resources necessary to support implementing the project." In another telephone survey, workers were reluctant to discuss any situation in which they missed a home visit and endorsed nearly 100% compliance with

mandated visitation of youth in federally defined foster care. State-level data, however, indicate visitation rates of 81%.

Discussion and Recommendations

Child welfare caseworkers function within a compliance-driven, hierarchical system that contributes to mistrust of research. While some caseworkers endorse research that can inform practice, others fear that results will be used against them and are reticent to disclose less than optimal information. In an environment of high caseloads and numerous "unfunded mandates," participation in research activities fares low on the priority list of the average caseworker, supervisor or administrator. The child welfare system in Pennsylvania is state-supervised and county-administered, further complicating challenges related to bias in response, non-response, and social desirability, particularly when research or evaluation is funded by the state. Researchers are in a triangular relationship with the state and county child welfare agencies, and often find themselves attempting to convince a reticent child welfare staff to provide an actual account of what is happening in practice while clarifying that the purpose is not to collect compliance data for the state. Additionally, researchers are not able to routinely use randomization or select "ideal" organizations for study (Cooke & Lafferty, 1994; Hemmelgarn, Glisson & James, 2006) which complicates reliable data collection.

Discussion

The group offered Drs. McCrae and Cahalane several suggestions for addressing response and non-response bias, including:

- Increase the stakes to move the field to where it should be by using the compliance-driven dynamic in child welfare as an advantage.
- Weigh project goals and agency culture (emphasis on compliance versus participation) to determine the best strategy for implementing the research. For example, more participatory styles of research and evaluation may work well in some agencies but not in others.

- Allow the state to be the enforcer and back out of conflicts that arise with compliance issues related to research participation.
- Reference work regarding *Evaluability Assessment* (e.g., Office of Juvenile Justice and Delinquency Prevention, Juvenile Justice Evaluation Center, 2003; Trevisan & Huang, 2003) to determine whether particular agencies are ready to embark on an evaluation.
- Use existing meetings to recruit individual participants as opposed to asking them to participate in research and evaluation activities on their own time.
- Provide compensation to participants whenever possible.
- Access case record data to compare with self-report data.
- Carefully review and augment written information for self-reports or introductory scripts for telephone surveys in order to explain projects thoroughly and prompt accurate feedback.
- Add social desirability scale to measures and determine whether it influences main findings.
- Frame the study issue and the concern with compliance rates, and then ask respondents to report on their colleagues instead of themselves.

Conclusion

The current emphasis on evidence-based practice, translational research, organizational performance, practice-based research, and outcome-based training evaluation requires evaluators and researchers to use strategies that produce high levels of participation and commitment by subjects. Basic reliability, validity and generalizability are compromised without this participation and commitment. The nature of child welfare work, the interrelated issues of turnover, caseload size and quality supervision, and the atmosphere within child welfare agencies pose challenges to research participation. The feedback and group discussion offered some viable solutions and reinforced the importance of engaging in applied research despite these challenges.

References

Cooke, R.A., & Lafferty, J.C. (1994). *Organizational culture inventory*. Plymouth, MI: Human Synergistic International.

Hemmelgarn, A.L., Glisson, C., & James, L.R. (2006). Organizational culture and climate: Implications for services and interventions research. *Clinical Psychology*, *13*(1), 73-89.

Juvenile Justice Evaluation Center (2003, May). *Evaluability Assessment: Examining the readiness of a program for evaluation*. Office of Juvenile Justice and Delinquency Prevention. Washington, D.C.

Trevisan, M.S., & Huang, Y.M. (2003). Evaluability assessment: A primer. *Practical Assessment, Research & Evaluation*, *8*(20). Retrieved from http://PAREonline.net/getvn.asp?v=8&n=20

2010 NHSTES Proceedings

Is This Research Mandated?: Response and Non-response Bias in Child Welfare Training and Evaluation Research

Presentation & Facilitated Discussion

Presenters: Julie McCrae, *MSW, Ph.D., &* Helen Cahalane, *M.S.W., Ph.D., University of Pittsburgh*
Facilitator: Michael Nunno, *D.S.W., Cornell University*

Wednesday, May 26, 2010
12:45-1:55 pm

This facilitated discussion and presentation was about the challenges of conducting evaluations of child welfare trainings. The presenters briefly discussed the complex relationship between evaluators and workers. Training evaluation is often difficult because the child welfare work force is typically hard to reach, both at the individual and organizational level. The presenters provided an overview of the characteristics of child welfare organizations; explained response rates and socially desirable responses; and explored how to address these issues.

Characteristics of Child Welfare in Pennsylvania

- Pennsylvania has a state-supervised, county-administered child welfare system. It includes 67 diverse counties that consist of some urban, but mostly rural counties.
- The variety in county characteristics makes it difficult to conduct research and rigorous evaluation in the state.

General challenges faced in Pennsylvania in moving toward evidence-based practice and using outcomes to drive decision-making

- Challenges to improved practice include a stressed and strained workforce; expectations to do more with less funding (such as through unfunded mandates); and budget constraints which led to furloughs and hiring freezes.
- The challenges that the presenters focused on were socially desirable response/response bias and non-response bias.

Challenge #1: Socially Desirable Response/Response Bias

- This influences validity and generalizability of findings to agencies.
- Example 1: Casework visitation project
  - Initial visitation compliance rate was 81%; federal fiscal penalties will be enforced if they do not reach 90% compliance within 18 months.
  - A telephone survey of 2 caseworkers per county was conducted to establish a baseline. Survey questions included "What were the policies and practice of visitation?" and "Who were the caseworkers meeting with?" Respondents reported that they were 100% compliant, but federal data showed that compliance was actually at around 78-81%
- Example 2: Developmental and socio-emotional screening of very young children
  - A baseline survey was conducted, which asked about policies, procedures, and resources. A total of 80% of counties reported that they were doing everything they were mandated to do. However, actual compliance rate (after analyzing data) was 68%.
- The presenters stated that there may have been greater buy-in if people were engaged earlier in the process rather than mid-stream. Out-stationing many of the training staff all over the state helped to facilitate buy-

in, since the training staff built connections with county staff.

Challenge #2: Non-Response Bias

- When trying to get people to engage in a study or evaluation, people often ask, "Is this mandated?" If it is not, they often choose not to participate.
- Differences in characteristics of those who respond can bias samples. Those in ongoing service units participate more than intake workers. Younger workers complete follow-up surveys more than older workers (this was highly significant). Title IV-E graduates complete follow-up surveys more than non-Title IV-E graduates.
- Insufficient response rates are also an area of concern. Not all voices are being heard through the surveys.

Strategies to Address Non-Response Bias

- Attempts were made to increase survey participation. Initially, web-based surveys were used, but this was abandoned since responses were low, even after several attempts.
- Follow-up surveys (60-day follow-up and 6-month follow-up) were on a variety of topics (to gauge how much trainees had learned from the training), but the one with the lowest response rate was transfer of learning.
- Recently, some funding was acquired to pay people to complete survey, so this might improve response rates. However, it still might not result in an accurate picture. Suggested forms of payment included payment in the form of food and money, a gift card drawing, and compensation based on staff position (i.e., all who participate in 60-day will be compensated: $10 for caseworkers, $20 for supervisors).
- Direct observations might be more informative than self-reports, which are currently being used in the surveys.

Training Evaluation—Challenge #3: Triangulation

- The presenters explored how they can work with the agencies to develop rigorous studies and evaluations

that will be useful to them. There is an inherent distrust in the relationship that makes collaboration and engagement with agencies difficult because research findings will be reported to the state.

- The Title IV-E program evaluation with web-based survey had high a response rate (80-90%). This included a survey of current students, short-term grads, and long-term grads (at least 2–3 years after graduation) from 16 Pennsylvania schools, as well as county directors. Presenters attributed this high response rate to the pre-existing relationship that evaluators have with the Title IV-E programs. For instance, evaluators go to quarterly administrative meetings and continue to be physically present.
- Stress in the front-lines may be contributing to low response rates (55%).

Suggestions from Participants in the Symposium

- Evaluability assessments can help researchers see what difficulties might come up before administering the evaluation, and help them to think about how to overcome those difficulties. One audience member suggested the Vera Foundation as a source for evaluability assessments.
- It might be useful to tie survey completion to obtaining training credits.
- Another option would be to use an "opt-out" consent, in which participants would have to select "opt out" if they did not want their survey responses to be used for data purposes. Previously, participants had to "opt in" for their responses to be used in research.
- Symposium participants discussed how evaluation studies typically do not go through human subjects review committees, but research studies do. Thus, different rules apply for research studies that do not apply for evaluation studies.
- Trainers are in a unique position because they both train and conduct evaluations. This conflict can be addressed by discussing the concept of socially

desirable responses with participants before they complete the survey.

- A symposium participant mentioned that a social desirability scale is used in Undoing Racism workshop.
- There are 2 areas of social desirability: 1) fear of retaliation, and 2) being inclined to give a nice answer. The presenters also noted that there are intentional and unintentional forms of response bias.
- Symposium participants suggested providing respondents with general data about compliance rates and reasons for the research. It might also be useful to normalize non-compliance so that people would be free to talk about it.
- Symposium participants suggested that directors of organizations personally send out the survey. The presenters stated that this was already being done in Pennsylvania.

See page 31 for the paper submitted by the presenters.

2010 NHSTES Proceedings

Evaluating Utah's Training Continuum

Presentation & Facilitated Discussion

Presenters: Norma Harris, *Ph.D., University of Utah, &* Lazarina Topuzova, *Ph.D., Gonzaga University*
Facilitator: Sherrill Clark, *Ph.D., CalSWEC, University of California, Berkeley*

Wednesday, May 26, 2010
2:05-3:15 pm

This session was a facilitated discussion about the challenges of evaluating Utah's child welfare training and education system. Utah has a state-administered child welfare system and has developed set competencies that the state agency considers important for entry-level workers to master. The agency has requested an evaluation of the training system but this has not yet occurred. The presenters discussed how to evaluate whether practice principles are being incorporated into actual practice and training. Dr. Topuzova's dissertation findings examined the role of academic curricula by looking at Council on Social Work Education (CSWE) accreditation and the baccalaureate curriculum.

Background of Partnership Between Utah Division of Child and Family Services (DCFS) and School of Social Work

- The partnership between the Utah DCFS and the School of Social Work began in 1995. Last year marked the first year that they were open to working together on evaluation.

- The School of Social Work is not responsible for training, but it does train BSW and MSW students who are sponsored by the agency.
- Initially, the agency was reluctant to examine the training modules because the division was under court settlement at the time.
- First steps of the evaluation included establishing what was important to the agency in terms of entry level and advanced competencies. Dr. Topuzova then compared what the agency thought would support these competencies with what the BSW curriculum offered. She made recommendations of what could be improved in their training, as well as what the school could implement in its BSW program.
- Balancing the diverse missions of the organizations presented a challenge: DCFS is a practice organization that is concerned with practice behaviors, and the university has an academic mission and the associated accreditation requirements.
- Dr. Topuzova created a new list of competencies by looking at the research and talking to workers, administrators, and trainers. Around the same time, CSWE created a list of specific outcomes that they wanted for social welfare students.

Does Schooling Accurately Assess Master of Competency?

- Dr. Harris explained the accreditation process for CSWE and discussed how grade inflation at the university makes it difficult to assess mastery of competencies. A "B" grade is considered mastery of competency.
- A participant commented that grade inflation occurs at their institution, but the institution addresses this by giving students a standardized, critical thinking test upon entry and graduation. Faculty rate students on a scale that is tied to some of the competencies and some of the critical thinking measures. Results are used to inform instruction and curriculum. This is done at the BSW and the MSW level. The MSW program is evaluated on 4 domains.
- A student exam behind a 2-way mirror may be a better way of assessing mastery of competencies.
- Differences exist between practicum coordinators with regards to how students are evaluated. Low interrater reliability

among practicum coordinators may impact assessment of student mastery of competency.

Utah Division of Child and Family Services (DCFS) Training

- In Utah, a practice model drives the training, which consists of a 9-week pre-service training program followed by 3-month mentoring program.
- The presenters reviewed the history of the practice model, which began when the state agency was sued and the court mandated reform. An external agency was brought in to help develop practice model and training, and this led to decision to do a comprehensive child welfare reform. Community members, agencies, parents, and consumers participated in the reform efforts and eight principles were developed for the practice model. The practice model does not include a theory of change. Core training and skill-building trainings were developed based on the practice model.
- Worker retention is a big problem. Two-thirds of Utah's child welfare staff are bachelor's level, not BSW, so maximizing resources and training is critical.

Development of Competencies

- Dr. Topuzova began her process by talking with people from CalSWEC and the Institute for Human Services, Ohio, about what knowledge, skills, and competencies they believed were most important for child welfare workers. She also talked with trainers and regional directors, and conducted a survey with supervisors and workers.
- She asked supervisors what they expect workers to know right after completing a training, and one year later. She asked workers what they thought after the training and if their skills improved a year later.
- Forty-five (45) people responded out of an estimated 98 supervisors. Workers' response rate was slightly higher than supervisors. Since the survey was done through the DCFS, Dr. Topuzova did not have access to workers or supervisors and it was difficult to find out how many people received the survey.
- Common themes of the survey findings included:

  - Respondents were not interested in institutional racism and did not believe that it was important for people to know or understand this concept;
  - Supervision was deemed very important;
  - Workers need to be more knowledgeable about resources in the area;
  - Secondary traumatic stress and self-care are important;
  - It takes at least a year to get used to the job and feel comfortable.
- Dr. Topuzova reviewed the curriculum (including objectives, content items, assigned readings, and assignments) to ascertain that the content matched the desired competencies from the survey.
- She found that secondary traumatic stress and self-care were not being addressed well in the curriculum. The curriculum also did not adequately cover general information that was not specific to child welfare, such as the use of the data and information systems.

Discussion of Findings

- One goal of examining the training system was to reduce redundancy in training programs. Many BSW students have already been exposed to training content through their schooling, and an opt-out option might be useful for those who demonstrate competency.
- If the trainees come in well-prepared from their schooling and score high on the pretest and they do not show growth in posttest, that is okay because they already have demonstrated that they know the material.
- The next step of the process will examine the compatibility of DCFS and CSWE competencies. This will involve utilizing a similar process with the MSW competencies. Trainings will also be evaluated. BSW and MSW program directors are supportive of these efforts.
- Curriculum fidelity presents problems. Trainers are able to vary curriculum as they want, so there is no fidelity in the training delivery.
- The extent of evaluation of trainings up until now has been satisfaction surveys. There has not been any evaluation about

if and how the training information was helpful in actual practice.

Suggestions from Participants in the Symposium:

- Incorporating a theory of change in the practice model may help to get buy-in for these efforts. This provides an explanation of "if this happens, then this will happen" so that people will see how the different components are connected with outcomes.
- Symposium participants engaged in a discussion about the importance of behaviorally anchored statements about competencies, as well as the importance of being able to assess mastery of competency beyond a "B" grade.
- Schools, CSWE, state agency personnel, and researchers must collaborate in developing a grading rubric.
- "Backward design" may be a useful approach to developing a grading rubric or designing a course; start with what an instructor would like students to know by the end of the course, and then work backwards from there.
- Another participant suggested looking at some work done in Texas in the 1980s that showed the differences in competencies for different levels, but cautioned that it would need to be updated.

Follow-up Session, Part 1: Steps in Developing a Handbook of Child Welfare Training Evaluation

Facilitated Discussion

Facilitator: Anita Barbee, *M.S.S.W., Ph.D., University of Louisville*

Wednesday, May 26, 2010
3:25-4:25 pm

This facilitated discussion centered on the development of a handbook of child welfare training evaluation. Dr. Barbee reviewed the progress made on this project since the 2009 NHSTES when the handbook was first proposed. In an effort to gather input from those who were unable to attend this year's symposium, two Webinars™ were conducted prior to this session. About a dozen individuals participated. The bulk of this session was used as a brainstorming session to flesh out the purpose, contents, and target audience of the proposed handbook. Next steps were discussed at the end of the session.

Transfer of Learning

- Training evaluation is often required by legislation. However, due to a lack of accountability, it is usually not carried out. Further, most evaluations that are conducted are program evaluations, not training evaluations.
- Training evaluation is important because the child welfare workforce has a huge impact on the lives of children who are vulnerable.

- There is a need to fix the problems with social work education and selection of workers that resulted from de-professionalization.
- Despite the progress that has been made over the years about measuring learning and transfer of learning, many people in the field still have very basic questions about transfer of learning. People often use the jargon, but do not understand the full meaning of the terms or how they connect to the work. Overall, an understanding about the importance of transfer of learning has not been institutionalized beyond the special groups tasked with addressing transfer of learning. High staff turnover contributes to this.
- The judicial system is becoming interested in this and has begun to evaluate their trainings.

Handbook Discussion

- Symposium participants identified potential audiences, including child welfare administrators and training directors, evaluation personnel, and academics. The audience should include public, private, and non-profit agencies. A symposium participant wondered if the target audience needs to be more narrowly defined.
- Decisions need to be made about whether the handbook will be practice or academic oriented, and also about the titles of the sections. It would also be helpful to know if the handbook will focus on one model.
- A connection should be made between theories and evaluation. Symposium participants discussed the possible inclusion of constructivist theory and incorporating theory of change into the evaluation model to clarify the connection between training, practice, and outcomes.
- Writers for each section should be experts on that topic.
- A suggestion was made to change "Learning Theory in Training and Development" to "Adult Learning Theory…." Symposium participants also discussed the use of the term "training" versus "professional development." "Professional development" sounds more professional and respectful of the workers, whereas "training" sounds more pejorative.

What to Include for Administrators

- A chapter on the history of child welfare training should be included.
- The preface should include information about what the child welfare director is faced with and which of these issues is going to be addressed by the handbook. For instance, how do administrators motivate their staff to do a better job? Or, how can administrators evaluate how helpful the trainings are for staff? Ultimately, this section should explain how the trainings can help administrators.
- The handbook should explain what a training is, and what it is not. An explanation of the functions of trainings would be helpful. It should also be clear how trainings fit into the overall structure of the organization—*Where is training appropriate in an organization and where is it not?*
- Next, an explanation of the levels of evaluation should be included. Important considerations for administrators include an exploration of the administrator's goals and an explanation of the possibilities and limitations of evaluations and the data they produce.
- It may be helpful to include a chapter on different models, including Kirkpatrick, and how these models are relevant. Examples of how various models are being used across the country should be incorporated.
- The handbook should include information about how to develop a training organization that can handle and manage data.

Important Chapters to include

- Important topics include: gaps in knowledge about training and evaluation; training development; skill development and attitudinal change; and information on how to build, expand, sustain, and evaluate training capacity. It would also be useful to include a section on how to build a test from a specific curriculum.
- Training should be connected to the curriculum, and current performance should be linked to past trainings. A question was raised about whether or not the trainings should be explicitly linked to outcomes. A symposium participant mentioned that the *American Journal of Evaluation* has an

interesting article that talks about this. Researchers can help evaluators link the training to outcomes.

- Evaluation is a way to ensure mission compliance.
- The partnership between the agency, researchers, and evaluators should be addressed. This would include information on how to build relationships, establish a welcoming environment, and coordinate work with each other.
- A section on the differences in pre-service trainings across the country and between MSW and BSW programs should be included.
- The handbook should provide information on the quality and qualifications of trainers, and the expertise of evaluators. Technology issues should also be addressed.
- A section on how participant reaction can be measured and used should be included.
- Level Four of the handbook should include Kirkpatrick's organizational change, but it may also be helpful to add change in client outcome and retention.
- Level Five of the handbook should include return on investment.
- Quality of evaluation design and method depends on what the agency is willing and able to pay for. Important considerations are the practicality and cost-effectiveness of different evaluation models. Randomized trials are rarely done. Pre/post test designs are more often used.
- The effects of organizational culture and climate on training evaluation, and how to evaluate organizational culture and climate should be included.

Format of Handbook

- Mallon & Hess' "Child Welfare for the 21st Century: A Handbook of Practices, Policies, and Program" is an example of what a handbook in this field might look like.
- One possible format for the handbook would be to break it down into multiple volumes.
- The handbook could have a problem-solution theme. Discussion of training evaluation would start with a problem

and then proceed to gathering data, contracting to evaluator, implementing the program, and evaluating the program.
- Each chapter should include case examples.
- The appendix should include examples of measures, as well as examples of good knowledge questions, different question formats, and situational judgment formats.

Next Steps

- Dr. Barbee assigned homework to some participants so that certain sections could be developed further before Part II of the discussion.

Follow-up Session, Part 2: Steps in Developing a Handbook of Child Welfare Training Evaluation

Facilitated Discussion

Facilitator: Todd Franke, *Ph.D., University of California, Los Angeles*

Thursday, May 27, 2010
9:45-10:55 am

This session was a continuation of the handbook discussion led by Dr. Anita Barbee on Day 1 of the symposium. The main purpose of this discussion was to flesh out the content of the handbook. Symposium participants determined the format and basic content of the handbook, explored options for the selection process for writers, and briefly discussed the editing and publishing process.

Level of Detail for the Outline

- Symposium participants discussed the optimal level of detail for an outline. There needs to be enough detail to ensure that writers address desired topics, but it also needs to be general enough for writers to exercise some discretion over the content of their chapters.
- For those writers who have written about the topic in other publications, symposium participants decided to ask the writers directly if they would like to write a similarly themed chapter in this handbook.

Potential Audience

- The potential audience for the handbook includes evaluation researchers, training administrators and managers, program trainers and evaluators, educators, and students.

Chapters to Include in the Handbook:

- Currently, a total of about 42 chapter ideas have been brainstormed (pared down from an original count of 50 chapters).
- Preface
  - This section should include a brief historical background and a perspective of child welfare from the training and professional development director.
- Overview
  - Theories and models of training evaluation should be provided.
  - This section will also explore the current state of the field of training evaluation.
- Section 1: Getting ready to conduct the training evaluation
  - This section will begin by first explaining training, professional development, and training evaluation.
  - This section will also explore how to build organizational capacity for evaluation. This includes assessing agency culture, organizational readiness, skills and staffing patterns, technology, learning management systems, and the agency budget. It is also important to consider who will conduct the evaluation, who wants the evaluation completed, and how each player needs to educate the others and manage expectations. The roles and competencies of a training evaluator should also be explored.
- Section 2: Designing a training evaluation
  - A critical part of designing training evaluation is helping agencies use their administrative data. Part of this includes explaining evaluation methodology, data collection, and purpose of the evaluation. It also includes linking the data together, and linking the data to outcomes.
  - Assessment of curriculum compliance and trainer's quality and qualifications should be addressed in this section.

  - Agencies might consider using a system of levels for evaluation. This means that key decisions points exist for each level. This section should also explain how the levels are interrelated and how they are not. One symposium participant suggested looking at the work done by Holton and by Kirkpatrick.
  - In order to discuss transfer of learning, the handbook needs to define transfer (both positive and negative kinds of transfer) and how it can be measured. The role of supervisors and field instructors in transfer of learning should also be discussed. An exploration of factors that may further impact the "what" and "how" of transfer of learning should be included.
  - In addition, it may be useful to point out the consequences of a lack of transfer (such as loss of investment). This section will include information about the role of evidence-based practice in linking transfer to outcomes. A discussion of participatory evaluation may also be useful in this section.
- Section 3: Organizational issues in training
  - In order to reach a larger audience, both "professional development" and "training" should be used in the handbook.
  - This section might include a brief look at resources provided by the National Staff Development and Training Association (NSDTA), such as the NSDTA Organizational Effectiveness Handbook and a description of NSDTA roles.
  - Structure is important for effective training. Agencies should add structure to organizational culture and climate, or add structure to barriers and facilitators.
  - The chain of evidence should be explained in this section so that agencies are clear about what training evaluation can and cannot do (i.e., limitations of connecting test results to outcomes). The role of training and reinforcement of trainings should be explored.

Selection Process for Writers

Symposium participants agreed that everyone who is interested in writing should submit an outline. Two options for selecting writers were discussed.

- Selection Process Option 1:
  - A request will be sent to potential writers asking them to submit an outline for the chapter they would like to write. The request should be sent out to the listserv.
  - A committee reviews the submitted outlines (without being aware of the identities of the authors), selects the best outlines, and makes edits to the outline as needed.
  - There is also an option to take ideas from the outline that was not selected and see if the writers would be willing to incorporate those ideas into the official chapter. If they agree, they must cite the original writer.
- Selection Process Option 2:
  - A committee will select a document that was written for the Proceedings in a previous year and determine if that writer would be appropriate for a handbook chapter.
  - The writer will need to make updates to his/her original work.
  - A symposium participant asked about whether or not copyright laws prohibit the use of previously copyrighted materials. It was clarified that Proceedings are copyrighted, but that they are open and free to the public.

Chapter Format

- The handbook should reflect the language of training administrators, not the language of academics.
- Each chapter should be 12 pages in the handbook. This comes out to about 20 standard pages.
- The book as a whole should cover relevant, updated literature, but every chapter does not have to be a literature review.
- The format of the handbook will be similar to that of a textbook. This means that it will include helpful resources, tools, or websites (including examples of effective evaluations) as well as possible action steps throughout the chapter.
- Advanced organizers that outline the content and objectives of the chapter will be located at the beginning of each chapter.
- Each chapter will include a reflection piece.

- Some chapters will require a case example.
- A list of important questions should be available at the end of each chapter.
- Information from within and outside of the field of child welfare should be included in each chapter
- Symposium participants discussed whether a minimum number of references should be required for each chapter.

Publishing and editing

- Dr. Barbee and Dr. Franke will edit the handbook, along with section editors, and an editor from the publisher.
- A symposium participant suggested that it might be useful to reach out to NSDTA for their endorsement of the handbook.
- A suggestion was made by a symposium participant to try and get CWLA's endorsement, but the group decided not to pursue this option.

2010 NHSTES Proceedings

Developing a Statewide Model of Trainer Evaluation

Facilitated Discussion

Presenter: Leslie Zeitler, *LCSW, Training & Evaluation Specialist, CalSWEC, University of California, Berkeley*
Facilitator: Leslie Rozeff, *M.S.S.W., University of Maryland, Baltimore*

Thursday, May 27, 2010
11:05 am-12:15 pm

This was a facilitated discussion about the various trainerevaluation models that have been utilized across the country. Ms. Zeitler explained that California's strategic plan for child welfare training evaluation called for a statewide model for training evaluation. Data collection under the first strategic plan (which ended in 2009) occurred at the regional and county levels and included information such as trainee demographics, trainee satisfaction, knowledge, skill, and application of knowledge. The second strategic plan involves collecting additional data on people's attitudes and values in training via a stereotype threat study with the Public Child Welfare Training Academy. Data will also be used to evaluate fidelity to the training model, quality assurance, and the level of support for the professional development of trainers. This discussion centered on comparing different models for evaluating trainers, and outlined important considerations when developing such a model.

Existing Trainer Evaluation Models

Ms. Zeitler and symposium participants provided several examples of trainer evaluation models that have been used across the country.

- Example #1:
  - One model measures a trainer's level of competency along two major areas of competence: 1) child welfare knowledge; 2) trainer skills.
  - This Level of Competency model is different from the National Staff Development and Training Association (NSDTA) model for trainer competencies.
- Example #2:
  - In this model, all trainers are certified in safety and risk assessment. The trainer evaluation is conducted by a training department.
- Example #3:
  - This model was developed by Dr. Dale Curry from a literature review of existing training evaluation practices. This training model evaluates skill and knowledge content. It was originally designed for the National Staff Development and Training Association (NSDTA). In this model, every instructor is evaluated at least twice a year. The evaluation involves gathering and analyzing information about the instructor, then providing the instructor with feedback from the data. The purpose of the evaluation is to see if instructors are having difficulties in a particular area of instruction.
- Example #4:
  - Colorado, which has a state-supervised child welfare system, uses a training evaluation model that looks at both the curriculum and the trainer. In this evaluation model, trainer liaisons ensure that all personnel attend the mandated trainings. Trainer liaisons are also trained to review the trainings. Quarterly meetings are held to provide feedback. A symposium participant noted that the purpose of the liaison was not fully appreciated at first, but over time the role of the liaison improved and the feedback they now provide is useful for trainers.

- Additional resources for developing a training evaluation model include the NSDTA competencies for trainers and the Institute for Human Services competencies for trainers.
- Public health models may be a useful resource due to their emphasis on using evidence-based curricula and fidelity measures. The trainer evaluation process at University of Maryland was developed from a public health model but uses many components that are similar to the NSDTA model.

Purpose of evaluating trainers

- Trainer evaluation is an integral part of ongoing professional development, and provides a meaningful way to measure and use data that can improve the quality of trainings. It can also lead to standardization of training delivery, which is one way to address issues of quality assurance and fidelity to the model. Further, the focus on the trainer rather than the curriculum can also impact personnel decisions, such as decisions to promote or terminate employment of a trainer.

Important Considerations for Measurement Instruments

- While the Likert Scale has a variety of uses, it also presents important limitations, particularly for those competencies that require a more nuanced evaluation of proficiency. For some competencies, it might not be enough to know if an individual's skills are "unacceptable," "needs improvement," or "shows mastery."
- It would be useful to have a set of specific anchors by which certain (training) skills/behaviors can be defined and identified.
- Additional considerations include deciding which skills to scale and how the data will be used.

Topics for Webinars™

- A national Webinar™ allows for a greater level of participation and may be a useful method of communication throughout the development process. Topics for future Webinars™ include identifying various methods of evaluation and sharing information from observations of existing training evaluation models.
- Symposium participants discussed the nuts and bolts of observations. Key questions were:
  - *Who does the observations?*

  - *Is it useful for peers to observe each other?*
  - *What about supervisors?*
  - *Does the trainer evaluation process include self-reflection?*
  - *Do we want more than satisfaction level for training feedback?*
- Symposium participants also expressed concern about evaluating trainers based on how well the trainees do. Doing so might negatively impact the quality of trainings.

Experiences with Assessment and Feedback

- Assessment and feedback are essential in the development of a trainer. Trainers often begin by feeling insecure about skills, then feeling hopeless about the fact that their trainees aren't picking up the material, and then recognizing the need for transfer.
- The frame with which trainer evaluation is viewed also affects how the trainer integrates the feedback. For example, if the agency supports trainer evaluation as part of a multi-level professional development process (as opposed to solely as a source of information for an annual performance evaluation), trainers are more likely to participate with an open and accepting attitude.
- Symposium participants discussed factors that impact the usefulness of feedback.
  - Trainers often prefer receiving feedback in the form of written comments provided by participants.
  - Timeliness of the feedback is important. If the feedback is given months from the original date of the training, the results may become irrelevant for the trainer.
  - Asking the trainees what they would have liked differently from the trainer (e.g., "What could the trainer have done to facilitate your learning?") might provide the trainer with useful data.
- Emerging technology also presents challenges for trainers. Issues included:
  - *How will skill qualifications change as we move more toward using technology?*
  - *Will there be more online courses?*
  - *Will it be useful to assess and provide feedback for homework assignments?*

In-Service Trainings

- Trainings are provided in different forms. Some trainings are done with co-trainers, while others are conducted by a single trainer. In California, most of the core trainings are delivered by a single person (no co-trainers). Some regions have transfer of learning specialists. Trainings also differ in how trainers are hired. In Pennsylvania, most trainers are contracted practitioners. In California, the Regional Training Academies (RTAs) are responsible for their own trainers and most are contract trainers. Some RTAs have full-time dedicated trainers on staff.

What Trainer Evaluation Looks Like Today

- Trainer evaluation typically occurs once or twice yearly.
- One example of an ongoing evaluation requires evaluations after the trainer's third training, seventh training, and tenth training. (Public Child Welfare Training Academy). An additional final evaluation serves as an annual satisfaction survey. These annual surveys are not in person. Results from satisfaction scores are discussed with the trainer. Scores from satisfaction surveys are often used to determine whether or not to call the trainer back for future trainings. If scores are low, technical assistance is provided. If scores still don't improve, then the trainer will not get called back to conduct a training.
- Sometimes, trainings are not well-received by trainees because the content is more challenging or difficult to train. Thus, some low satisfaction ratings might not be the fault of the trainer. Conversely, sometimes trainers receive high scores for being likeable, even if the trainee did not learn much. This highlights important ethical issues for trainers (i.e., some trainers might tell the trainees what will be on the test so that the trainer will get a higher satisfaction rating).

Additional Resources for Trainers

- NSDTA offers a national certification for trainers.

2010 NHSTES
Proceedings

Rasch Measurement for Child Welfare Training Evaluators

Henry Illian, *D.S.W.*, Cynthia Parry, *Ph.D.*, *& James Coloma*, *M.S.W.*

Abstract

This article describes the use of Rasch measurement in the analysis of data from tests and trainee reaction questionnaires ("reactionnaires"). Rasch measurement is an analytic approach in which individual abilities in carrying out specified tasks and the difficulty of those tasks can both be measured on the same scale. The logic underlying Rasch measurement in analyzing test and rating-scale data is discussed, and an example from the California Common Core Training Evaluation is presented to illustrate how this approach can be used to isolate the cause of anomalous results.

Introduction

In recent years, many child welfare training programs have begun to move toward testing the knowledge acquired by trainees through large-scale testing programs. Additionally, trainee reaction questionnaires ("reactionnaires") form an important part of the data for evaluating training programs. While the majority of evaluators are familiar with considerations of item quality, test reliability and test validity, several additional issues arise in these training programs related to the interpretation of data collected across varying trainee groups and over an extended period of time. Researchers and program administrators may be interested in questions such as "How do post-implementation scores of a revised curriculum compare with scores from previous years?" or "Do regions of the state with a higher concentration of MSWs score higher on the test than those where the most common degree is a BA or BS?" When the goal is to summarize or compare

scores over time, across regions, or on different versions of a test, a yardstick for measurement is needed that is not dependent on the exact composition of a given test or the ability of a particular trainee group. The Rasch model offers a way to make those types of comparisons.

The Rasch model has been used in a variety of evaluation settings, for example: validating or evaluating the psychometric properties of survey instruments (Green & Frantom, 2002) and those designed to measure complex problem-solving skills (Smith & Kulikowich, 2004; McCann & Stanley, 2006); health related concerns (Kelly, 2005; Chiu, Fritz, Light, & Velozo, 2006; Hahn, Cella, Bode, & Hanrahan, 2010) and social and cognitive growth (Brown & Unsworth, 2009); item banking for measuring a variety of constructs (Choppin, 1979; Stenner & Stone, 2003); computer adaptive testing (Smith, 1992); setting standards, cut-offs and cut-scores in educational and occupational measurements (Baghaei, 2007 and 2009); determining test-item discrimination (Kelley, Ebel, & Linacre, 2002); assessing the construct validity of test items (Linacre, 2004a); and evaluating the effectiveness of rater training (Mulqueen, Baker, & Dismukes, 2000).

A fundamental idea in Rasch measurement is that individual abilities in carrying out specified tasks and the difficulty of those tasks can be measured on the same scale. The analogy that is often given is performance on an athletic skill such as the high jump (see, for example, Wright & Mok, 2004). Competitors' ability is stated in terms of the how high they are able to jump, based on the highest they have jumped in the past. The difficulty of the jump is also stated in terms of height, so that both ability and difficulty are stated in the same metric. In the context of training evaluation, the tasks measured are not one-dimensional like height of a jump, but Rasch measurement nevertheless allows them to be expressed in terms of a continuum of difficulty, and participants can be assessed in terms of the difficulty of the skills they have mastered.

First, it is necessary to have items that can define an underlying ability variable. Items must be related to the ability and encompass a sufficient range of difficulty to allow accurate estimates to be made of what people know. If items have not been well chosen, the test will be a poor reflection of training content

and will not be valid for the intended purpose of testing. If items are not written with a sufficient range of difficulty, floor and ceiling effects may make it impossible to estimate how much trainees at the highest and lowest levels of ability know about the topic. If there are insufficient items in the range of proficiency where most test takers fall, estimates of ability will lack precision and have (Wright & Mok, 2004).

Second, one must have a valid response pattern that can locate a person's position on the underlying attitude or ability variable. The evaluator needs to be confident that the test score obtained reflects the trainee's knowledge as accurately as possible. In general, we expect that trainees will be more likely to answer easier items correctly and will be less successful as the item difficulties increase. Evaluators need a means of identifying response patterns that do not conform to expectations and that may not give a valid estimate of a person's ability, such as when someone is guessing or filling in random answers. Evaluators also need to be able to determine when a trainee who has not finished a test has not filled in enough answers to get a valid score.

Useful estimates of ability also must be stable and independent from conditions associated with the particular test administration. Estimates of a person's or group's ability obtained on one test should not change if measured with a different set of items that measure the same underlying ability. Similarly, estimates of an individual's ability should not vary based on the ability of the other test takers in the class. This independence is referred to as parameter separation within the Rasch measurement framework (Wright & Mok, 2004).

A fourth requirement is that measures be linear. This allows them to be used to study change over time and to make comparisons across groups. For these comparisons it is necessary to have a measure where a difference of a given number of points has the same meaning in terms of difference in underlying ability, regardless of where it occurs along the scale. In other words a five point change from pretest to posttest, or from one group to another, needs to mean the same thing whether it is from a score of 30 to a 35, a 50 to a 55, or a 90 to a 95. On a rating scale, a change from "strongly disagree" to "disagree" needs to represent

the same degree of opinion change as a change from "agree" to "strongly agree". Often, we treat test scores and ratings as if they form this type of equal interval scale. However, measurement in child welfare training, as well as other fields, seldom meets this requirement.

Measurement models have been developed to meet these requirements in different ways, with varying degrees of success. Rasch is one approach that can satisfy these measurement requirements.

Some Examples

In the first part of this article our goal is to discuss these measurement requirements in relation to situations commonly encountered in testing in child welfare training settings and to show how Rasch differs from classical approaches to common measurement problems. The remainder of the article will illustrate the applicability of Rasch measurement to training evaluation, with two brief examples: the New York City Child Protective Practice Core Training Reactionnaires; and the California Common Core Training Evaluation. The latter involves the construction of an item bank, conducting analyses across test forms, and investigating differential item functioning and differences in item difficulty associated with different trainers.

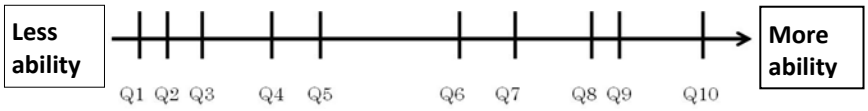
Description of Rasch Measurement

Defining an Ability Variable: A test is a way of turning observations of performance into measures of an underlying, or latent, ability. For example, trainees can be assumed to have some underlying knowledge about child and youth development. This underlying knowledge can be thought of as forming a continuum, from relatively little to a great deal of knowledge. The purpose of a knowledge test is to place trainees along that continuum by making an estimate of their ability. Since we can't see into their minds and do it directly, we need to do it indirectly by asking them to answer questions.

The questions that are asked (the test items) operationalize what is meant by "knowledge of child and youth development" and define a measure of the trainee's underlying ability. The items' positions on the continuum should reflect how much ability

someone needs to have to get them correct and that should not change. In effect the test items are a kind of yardstick (Wright & Stone, 1979) for measuring ability and that yardstick needs to remain constant if it is going to be useful. In other words, the 1 foot mark always has to come before the 2 foot mark which comes before the 3 foot mark, etc. and a foot long submarine sandwich always has to be the same length. Figure 1[1] illustrates this graphically for a ten item test. In this example, item 1 is the easiest and takes the least ability to answer, while item 2 is somewhat more difficult and item 10 takes the most ability to answer.

Figure 1



In classical test theory, item difficulties are estimated based on the proportion of people in a group who answer them incorrectly. Therefore their difficulty, or position on the line, changes, based on the overall ability of the group taking the test. When item difficulties change, the evaluator no longer knows what getting an item correct means in terms of underlying ability. In the yardstick analogy, we might suddenly have the 2 foot mark first, followed by the 3 foot mark, and then the 1 foot mark. Now being 1 foot long means something different than it did before.

A shifting scale is a problem for comparing scores for different groups of trainees and also for measuring change from pre to posttest. Under the Rasch model, calibration of test item difficulty is independent of the distribution of ability of the persons used to develop the difficulty estimates, so the yardstick remains constant.

Estimating Trainee Ability: Once a measure is defined, the next step is to determine how a score on that measure relates to a trainee's ability. Some types of measurement scales have an objective meaning. For example, if we say a man weighs 170 pounds, we know what that means (Wright & Stone, 1979). It is not necessary to know anything about the specific tool used to measure him. Within a small range of measurement error, a

[1] Figure adapted from Wright and Stone, 1979

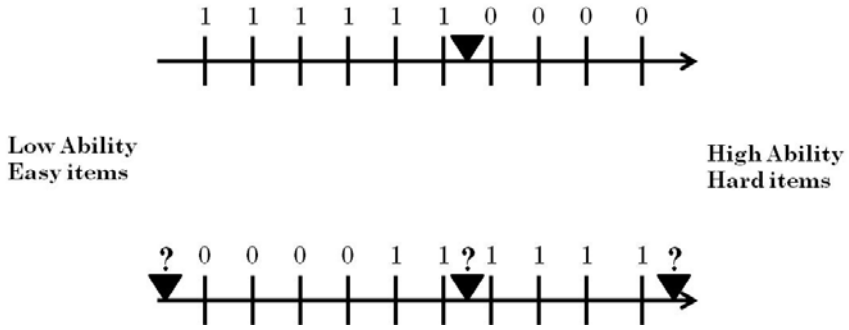
pound is always a pound. In educational testing the meaning of a score is less clear.

In classical measurement frameworks, test scores are based on the total number of correct answers. Like item difficulties, these scores need to be interpreted with reference to a known group and a specific test. For example, if we say that someone got a score of eight on a test, is that a good score or bad score? We only know how to interpret it if we know something about the test. If the person scored 8 out of a possible ten, this has a very different meaning in terms of ability than if he scored 8 out of 100. To understand what this score means in terms of ability we also need to know something about the other test takers. For example, was the mean score for the group 5 or 9? This need to place a score in context is not unique to raw scores. For example, if scores are expressed as percentile ranks, we might know that someone scored better than 80% of other test takers. However, we still need to know something about the normative group. Did the trainee score better than 80% of college graduates or better than 80% of MSWs? In Rasch a test score is based on the probability that the trainee will get items of a given difficulty level correct and it is interpreted independently of the scores of the rest of the group.

Response Patterns: In addition to knowing how many items a trainee answered correctly, making an estimate of trainee ability also requires knowing something about the pattern of correct answers that make up the test score. In Figure 2[2] below, the arrow represents the underlying ability variable being measured and the vertical lines represent test items. A "1" indicates that the item was answered correctly and a "0" indicates an incorrect answer. Trainees are represented by the inverted triangles. The top scale shows person one's pattern of responses. The bottom shows person two's pattern. Both people have a raw score of 6.

[2] Figure adapted from Wright and Stone, 1979

Figure 2



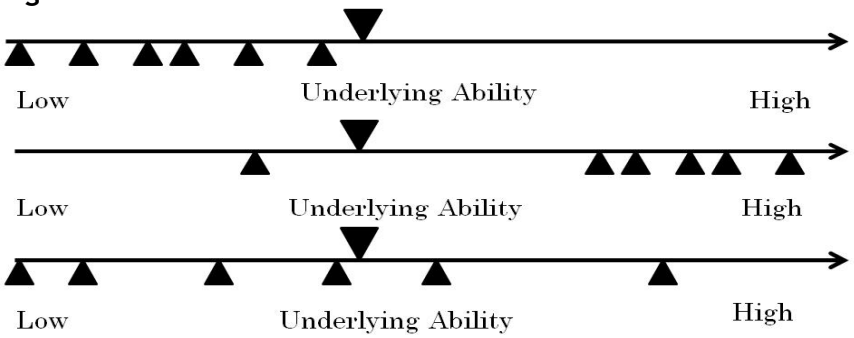
Based on the pattern of person one's responses, it seems clear that we would estimate his ability at the point where he stops answering correctly and begins to answer incorrectly. Placing person two on the continuum (estimating his ability) is more difficult. Do we put him at the bottom because he missed the first 4 easy items, at the top because he answered the hard items correctly, or somewhere in the middle? When scores are based on adding up the number of correct responses these people would look the same. However, the second person could have some gaps in knowledge that most of the group does not have, or his score may be invalid because he was marking answers at random (Wright & Stone, 1979).

Rasch is not additive. It does not rely on the number of items answered correctly, but takes into account which items have been answered correctly in order to make an estimate of ability. Because the model uses information about the pattern of responses in estimating a score, it is also able to identify unexpected, and potentially invalid, response patterns as part of the analysis.

Invariance: Estimates of a person's ability yielded by test scores should be stable and generalizable. We typically do not want to know only how a person or group did on a specific test at a specific point in time, but want the score to have consistent meaning in terms of the trainee's child welfare knowledge and how able he or she might be to do the job. Figure 3 below illustrates a situation where the same group of trainees takes three different tests of knowledge of child and youth development. The tests are based on the same content but have different items. In the

figure, the underlying continuum of ability is shown by the lines, with triangles below each line representing the test items. The three hypothetical tests are represented by the three different lines. The group's score is shown by the triangle above the line. This group's ability does not change, as shown by its location on the continuum of underlying ability in each of the three examples. However, its mean test score does change. Test one would give a mean score of four, test two would give a mean score of one, and test three would give a mean score of six. Did this group learn the content? One score (the first) would clearly indicate "yes", one (the second) would clearly indicate "no," and one (the third) is in the middle.

Figure 3[3]



Why is this important? Imagine that you want to monitor your program's effectiveness—to know if trainees' scores from year to year are similar, improving, or decreasing. Imagine too that the test is changing over time in response to curriculum changes. If scores go down did the program really do less well?

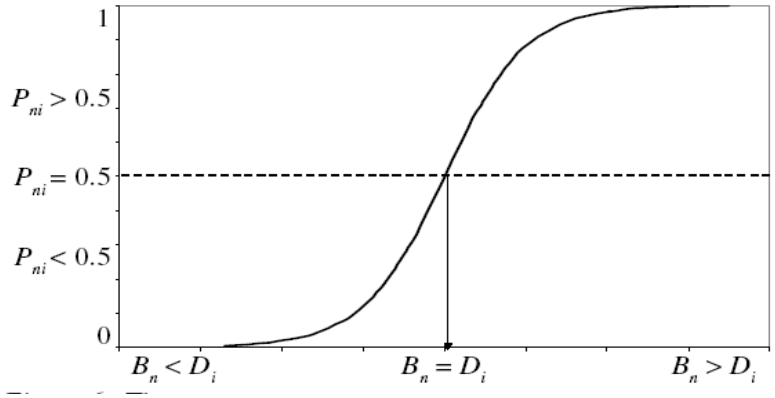
Most common test scores, such as raw scores or percents, are based on the assumption that the person's location on the ability continuum can be determined by the test score (Wright & Stone, 1979), but typical test scores cannot be generalized beyond the particular point in time, test takers, or materials. Classical test construction models have dealt with these issues by equating test forms or by judging ability in relation to a normative group.

[3] Figure adapted from Wright and Stone, 1979

However, how well this works depends heavily on how similar the group of interest is to the norming sample.

The Rasch Model: We have seen in the previous discussion how important it is for interpreting a test score to have a measurement tool that is invariant no matter when or with what group it is used, and how important it is that the interpretation of a person's score does not vary based on the performance of other test takers. To obtain measures that exhibited this invariance Rasch took a confirmatory approach rather than the descriptive approach taken by most measurement theorists (Wright & Stone, 1979). He identified the structure the data needed to exhibit in order to provide a useful measure, and then worked out a mathematical model using the logistic function (shown in figure 4[4] below) that produced measures with the desired properties.

Figure 4



The Rasch model assumes that a person's score should be the result of their ability (B), and an item's difficulty (D). Rasch allows for error in testing (e.g., from guessing or other factors) by making the model probabilistic; person ability and item difficulty do not completely determine the correct response, but do determine the probability of a correct response (Wright & Stone, 1979). This relationship is summarized in the curve in Figure 4. If a person's

[4] Note: from An Overview of the Family of Rasch measurement Models by Wright, B. and Mok, M. (2004) In Smith, E. and Smith, R. (Eds.) Introduction to Rasch Measurement title p14. Copyright (2004) Reprinted with permission?

ability is less than the item difficulty (lower left) the probability of a correct answer (shown on the Y axis) is less than 50/50. Probability of success is greater than 50/50 if the person's ability is greater than item difficulty. Unlike in classical measurement frameworks, a person's score is not represented by how many items he got correct, but by the odds that he would get items at a specific difficulty level correct.

The equation below (Figure 5) shows the Rasch model for dichotomous data; those where an item is either correct or incorrect.

Figure 5

$$\ln\left(\frac{P_{ni}}{(1-P_{ni})}\right) \equiv B_n - D_i$$

The probability of a response is governed by the difference between ability (B) and item difficulty (D). Because probability varies between 0 and 1, but the difference between ability and difficulty on the underlying variable can be infinite, a natural log transformation is used to allow for measure scores that range from minus to plus infinity. The mathematical model specified by Rasch allows the algebraic separation of "person ability estimates" and "item difficulty estimates" through conditional maximum likelihood estimation. For a fixed set of data and this underlying probability model, maximum likelihood picks the values of the model parameters that make the data "more likely" than any other values of the parameters would make them (Wright & Stone, 1979).

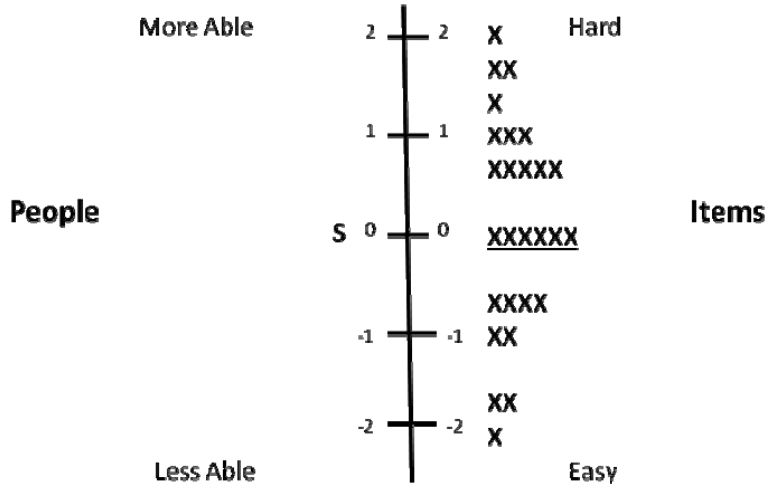
Rasch also developed a series of fit measures that are used to determine how well actual data fit the pre-specified model. If the fit is good, then the probability of a response is a function of person ability and item difficulty only. It is not dependent on the distribution of people taking the test or the distribution of difficulty of the specific items that make up the test.

Interpreting Test Scores Yielded by the Rasch Model: Rasch describes peoples' ability and item difficulty on the same underlying continuum of ability. Test scores are not based on adding up how many items a trainee answered correctly and

placing that in context of the test (e.g., 9 correct out of a possible 10, or 90%), or in the context of the rest of the group (e.g., Jane Doe scored 9 and the average for the group was 7). In Rasch, scores are expressed as logits. A person's logit score is the natural logarithm of their odds of success on items of a given difficulty level (Wright & Stone, 1979).

In the example shown in Figure 6 below The logit score Jane Doe gets on the Rasch variable (indicated by the "S" on the left side of the scale) means that she has a 50-50 chance of getting the underlined test items at the same point on the scale correct. As items get harder, her odds go down. The odds of a correct answer decrease to about 1 in 4 for items at 1 logit on the scale . As items get easier her odds go up. For items at the -1 level, her odds of being correct are a roughly 3 out of 4. It does not matter to our estimate of Jane's ability what the rest of the class does because we are not explaining her ability based on her position in the group.

Figure 6



It also doesn't matter if Jane doesn't respond to every item, since we are not adding up correct responses to arrive at a score. As long as the items are all placed on a common scale describing the underlying ability being measured, a person can respond to different subsets of the items and the estimate of his or her ability

will be the same[5]. This facilitates getting a meaningful estimate of a person's ability when there are missing data. It also facilitates building an item bank and gives us the ability to interpret the results from different tests, such as a pre and posttest, within a common framework.

Measuring Change with the Rasch Model: A key characteristic of Rasch is that it produces an equal interval scale. On the logit scale, a five point difference represents the same amount of change whether it is at the low end, the high end, or the middle of the scale. That is very helpful in interpreting score differences, as with pre to posttest change. We tend to treat all test scores as if they form an equal interval scale and we measure change as if this is true. In practice this is not the case. If test scores are normally distributed, a range of 5 points implies a larger change in ability at the extremes of a scale than in the middle. One practical implication of this phenomenon is that if a group of trainees has a high mean score at pretest, it may be harder to demonstrate change at posttest than if they come in with a lower mean pretest score.

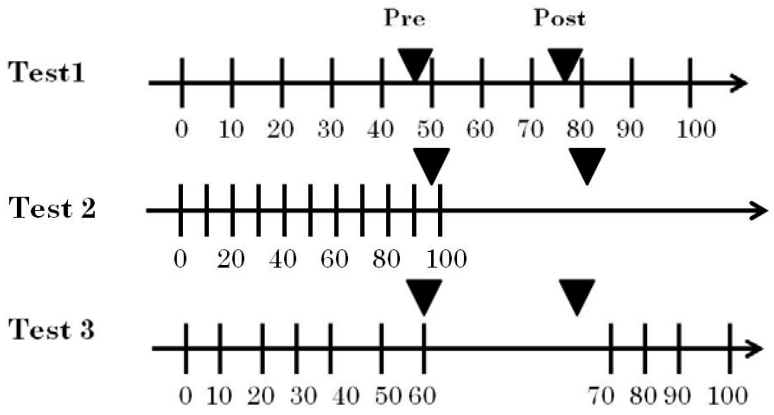
As discussed above, in classical approaches to testing, scores vary based on characteristics of the test and the scale it implies. Figure 7 below illustrates how the problem can be compounded when measuring change, as in a pre to posttest comparison.

Figure 7 shows pretest to posttest change based on three different tests of the same underlying ability. On each test the triangle on the left shows a groups' ability at pretest and the triangle on the right shows their ability at posttest. Suppose that the three tests are tests of knowledge related to a new worker training module on child and youth development. On test one the group showed a substantial improvement from a raw score of 45 to a raw score of 75; a 30 point difference. The next year the curriculum changed and test two was put into place. This time a group with the same ability would get a pretest score of 97 and a posttest score of 100; a change of 3 points. This test is obviously easier and there is a problem with a ceiling effect; but notice too

[5] Within a margin of measurement error and with the caveats that the different sets of items need to be well targeted to the ability level of the test takers and that the response data fit the model.

that the distance between points on the scale is much narrower. Thus, a change of 10 points does not mean the same thing here in terms of underlying knowledge as it did on test one. On test three there is a gap where no items fall on the underlying continuum. Here, the same raw score (60) would be obtained on both the pre and posttest.

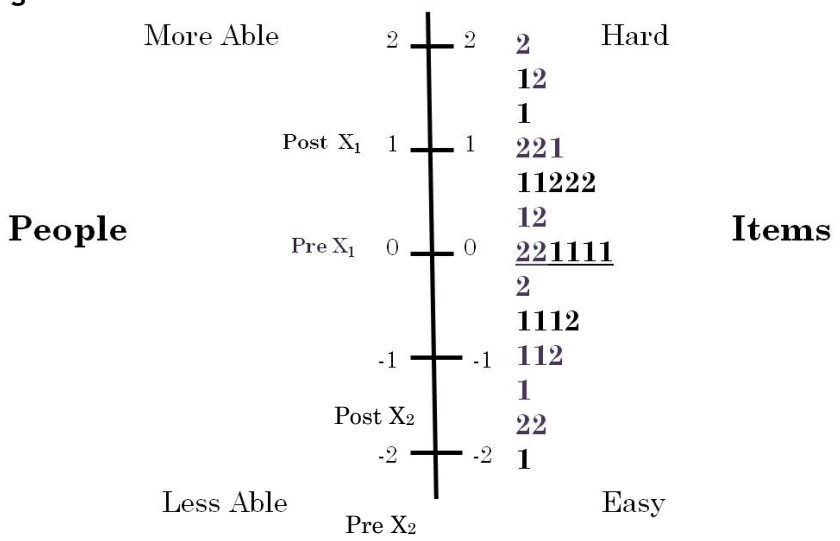
Figure 7



In classical approaches to measurement it is possible to transform scales statistically to scaled scores which have equal intervals. The scales used on common tests, such as the SAT and ACT, are scaled scores. There are also procedures for equating tests for difficulty. However, Rasch makes this somewhat simpler because the logit scale is already equal interval and no further transformations are necessary.

It is also not necessary for the distribution of item difficulty on two forms of the test to be exactly equal to compare pre and posttest scores. When a set of items have been scaled together and the data fit the specifications of the Rasch model, scores obtained from various subsets of those items will be interpretable within the same framework. In other words, if the difficulty values for individual items have been derived together and those difficulty values are used to anchor the pretest and posttest scales, the pretest and posttest can have different items and scores will still be directly comparable.

Figure 8



Suppose that the "1"s and "2"s in Figure 8 above represent a pool of items that were all scaled together. The 1s represent items that have been pulled out to make a pretest. The 2s are items that make up a posttest. On the pretest the mean score for a class is 0, meaning that on average the group would be expected to get items at the zero point on the scale correct 50% of the time. It does not matter if we mean 50% of all of them or 50% of the 1s or 50% of the 2s. This ability to derive scores from subsets of a group of items that have been scaled together greatly simplifies making comparisons across different test forms. The scale has not changed, so the pretest is being measured on the same scale as the posttest.

Moreover, the scale intervals have the same meaning. Suppose two classes take these same tests, labeled X_1 and X_2 in Figure 8. A change of one logit for group X_1 represents the same increase in ability as a change of one logit for group X_2, even though the group means are at different points on the scale. On average, group X_1 at pretest has a 50-50 change of getting items at the 0 point on the scale correct. At posttest, the odds of getting those items correct has increased to about 3 out of 4. On average, group X_2 has a 50-50 chance at pretest of getting items at the -2 level correct and by posttest that increases to 3 out of 4. Moreover,

because the scale is the same we can be confident in concluding that group X_1 has more ability than group X_2.

Rasch Measurement and Rating Scales

The use of Rasch measurement for developing and validating rating scales and for analyzing rating scale data is described in detail by Linacre (2004a) and by Bond and Fox (2007). Rasch measurement applied to rating scales is conceptually similar to its use with test data. What is being measured is still individuals' abilities on a set of progressively difficult tasks. In this case, the tasks are a set of items that are progressively more difficult to endorse, and the ability is an individual's ability to endorse them. The difference is that while a test taker can get a question either right or wrong—two possibilities—responding to a rating scale item involves indicating a degree of agreement or disagreement—several ordered possibilities. Nevertheless, Rasch measurement still looks for how much of something someone possesses: knowledge of a subject with a test, or ability to endorse increasingly difficult-to-endorse items at a high level of agreement. As with a test, an instrument that uses a set of scaled items embodies an underlying construct, e.g., an overall orientation toward the subject of the items.

One way of explaining how this works is through the example of the Guttman Scalogram, which is also a component of Rasch Analysis. The scalogram was developed in 1944 by Louis Guttmann (Guttman, 1944), a pioneer of American social psychology. Guttman, who was interested in measuring the strength of peoples' beliefs, used a set of statements with which people could either agree or disagree. The statements were developed to form a hierarchy from a mild to intense belief. Figure 9 is an example of a scalogram. Respondents are arrayed along a vertical axis, with those with the strongest belief at the bottom and those with the weakest on the top. Statements are arrayed from left to right, with those representing the weakest level of belief at the left and the strongest on the right. In the resulting diagram the 1s and 0s (agrees and disagrees) form a triangle, with 1s appearing more densely the lower and further to

the right you go, and 0s appearing more densely the higher and further to the left.

Figure 9

Person	*Item*				
	4	2	5	1	3
2	0	0	0	0	0
4	1	0	0	0	0
3	1	1	0	0	0
5	1	1	1	0	0
1	1	1	1	1	0
6	1	1	1	1	1

Source: ODL Open and Distance Learning 10/7/99

This example describes a set of yes/no questions, but it works the same way with degrees of agreement. Instead of 1s and 0s in the cells, there would be the numbers assigned to degrees of agreement or disagreement, 1 through 5 for example.

The Rasch model takes the scalogram one step further. The Scalogram is deterministic. It assumes that any person endorsing a hard item will endorse all of the easier ones. Reality follows this assumption in a general way, but rarely conforms perfectly. Instead, people will sometimes endorse a harder item while rejecting an easier one, and will sometimes reject an easier item but endorse a harder one. If the items are test questions, guessing will sometimes result in people answering harder questions after getting easier ones wrong. As with the analysis of test data, the Rasch model is probabilistic. As with tests, it predicts a probability, based on the pattern of answers. In this case, the probability is of a person endorsing a particular item at a particular level.

In the Guttman scale illustration, the hierarchy of difficulty among the items had been established before the items were administered. In most practical cases, this is not done, and one of the strengths of Rasch measurement is that it is able to do this

after data have been collected. In the example to be discussed, it was used to identify the relative difficulty among a set of tasks covered in Child Protective Services training.

Child Protective Services Tasks

The illustration involves trainee reaction questionnaire (reactionnaire) data from the Child Protective Services core (pre-service) training given by The James Satterwhite Academy, the training component of New York City's Administration for Children's Services. The reactionnaire is comprised of three sections. The first includes 25 items covering training content and delivery. It asks, essentially, how well the training was delivered. The second section, on which this discussion will concentrate, asks trainees to rate their own level of skill in 15 child protective services abilities that were trained (see Figure 10). The rating is on a Likert-type scale, running from 1 (very poor), to 5 (excellent), and anchored at all five points. To obtain a rough measure of learning, it uses a retrospective pretest format (Pohl, 1982). That is, it asks trainees to rate their post-training competence in these tasks and also to estimate their pre-training level of competence based on their current (post-training) understanding of these tasks,

Figure 10. The 15 CPS Tasks

1. Initiating a CPS investigation
2. Identifying and Addressing Risk Elements
3. Responding to domestic violence
4. Identifying Resources, formal/informal/societal
5. Interviewing the subject of a CPS report
6. Developing a Service Plan
7. Interviewing and assessing children
8. Documenting a CPS case
9. Interviewing collateral contacts
10. Making a determination of abuse/maltreatment
11. Identifying and Addressing Safety Factors
12. Managing removals and reducing the trauma of separation
13. Making safety decisions
14. Utilizing the Family Court to promote child safety and risk reduction
15. Developing and Implementing Safety Interventions

The Hierarchy of Tasks

Knowing the relative difficulty of the different tasks can indicate to training decision-makers which of the tasks covered in the training may require more effort to learn, and which are being learned effectively. The advantage of Rasch measurement in establishing the hierarchy is that it takes into account the differing levels of ability of the participants. Some have mastered at a high level only the tasks that turned out to be the easier ones. Others have mastered the full range of tasks at a high level. Simply averaging the ratings, which are ordinal in any case, gives an inconsistent picture. The average depends on the mix of abilities in any particular group. An unusually able group will make all tasks appear easier than they might otherwise be, and a less able group will make them appear harder. Because the results of Rasch measurement are independent of the make-up of any one group, it is possible to indentify a stable hierarchy of tasks without collecting the large amount of data it would take for differences in ability to cancel each other out.

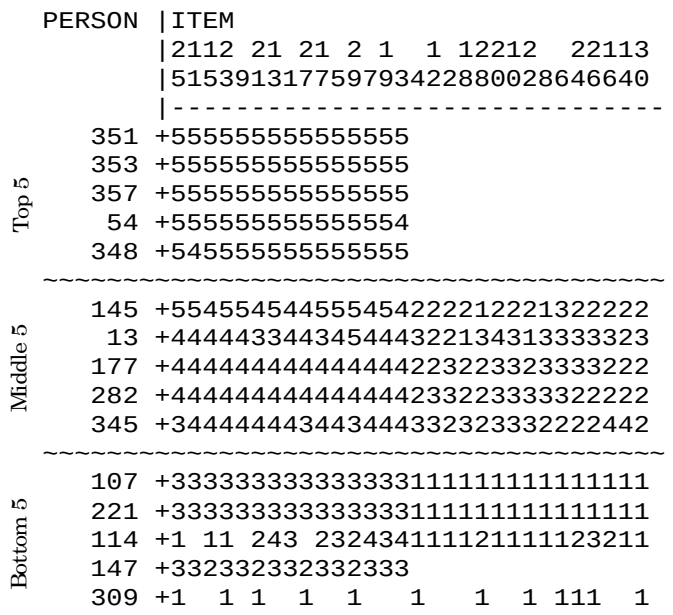
Figure 11 shows a Guttman scalogram of the 15 tasks rated by 366 participants. The scalogram has been truncated to show only those participants from the top, middle and bottom of the range of self-rated ability. The items are arrayed from left to right, with the easiest on the left and the hardest on the right. Because the 15 tasks were each rated twice, pre and post-training, they are numbered from 1 to 30. The odd-numbered items are post-training ratings, and group mostly toward the left. For example, the easiest to endorse at a high level was #25,"Making a determination of abuse/maltreatment." This was an ability that was heavily emphasized in the training.

The even-numbered tasks are pre-training, and group toward the right. The rightmost task is # 30, "Utilizing the Family Court to promote child safety and risk reduction." This task requires specialized knowledge that trainees likely would not have had prior to their training. The post-training version is #29, which still comes out as one of the harder ones.

All top five participants shown in the illustration rated all 15 of the tasks post-training as 5. The fact that none of them rated the items pre-training raises the question of whether they responded

to the rating task as intended; but for purposes of the illustration, it shows that they saw themselves as having mastered the tasks at the highest level. The group in the middle indicated that they mastered the tasks post-training at the 4 or 5 level, while they saw their ability as having been at the 1 or 2, or in some cases at a 3 level prior to training. The group at the bottom indicated that they had, for the most part not mastered the tasks post-training. They also saw themselves as at the 1, or in some cases 2, level pre-training.

Figure 11. Guttman Scalogram of Responses



Hierarchy of Difficulty

Figure 12 displays the CPS tasks arrayed in descending order of difficulty for only the post-training ratings—in the interest of saving space, the pre-training ratings were left out. The measure in the left column is the degree of difficulty expressed in logits. According to Linacre (2006), "The difficulty of an item is defined to be the point on the latent variable at which its high and low categories are equally probable." In this case, the degree of difficulty is equivalent to the level of mastery of the curriculum at

which a particular participant, who has just completed the training, would be equally likely to choose the highest or lowest categories (5 or 1) for the item. Those at a very low level of mastery would be equally likely to choose 5 or 1 for the easiest item, "Making a determination of abuse/maltreatment." Those at the highest level measured would be equally likely to choose 5 or 1 for "Responding to domestic violence." The values are all negative because all of the post-training items fall below the mean difficulty for all of the rated items, which includes the pre-training ratings as well.

Figure 12. CPS Tasks: Hierarchy of difficulty

MEASURE	CPS Task
-0.7	Responding to domestic violence
-1.18	Identifying Resources, formal/informal/societal
-1.37	Interviewing and assessing children
-1.39	Utilizing the Family Court to promote child safety and risk reduction
-1.4	Interviewing the subject of a CPS report
-1.42	Identifying and Addressing RAP and Risk Elements
-1.43	Managing removals and reducing the trauma of separation
-1.52	Initiating a CPS investigation
-1.7	Making safety decisions
-1.73	Developing a Service Plan
-1.75	Interviewing collateral contacts
-1.79	Documenting a CPS case
-1.92	Developing and Implementing Safety Interventions
-1.93	Identifying and Addressing Safety Factors
-2.03	Making a determination of abuse/maltreatment

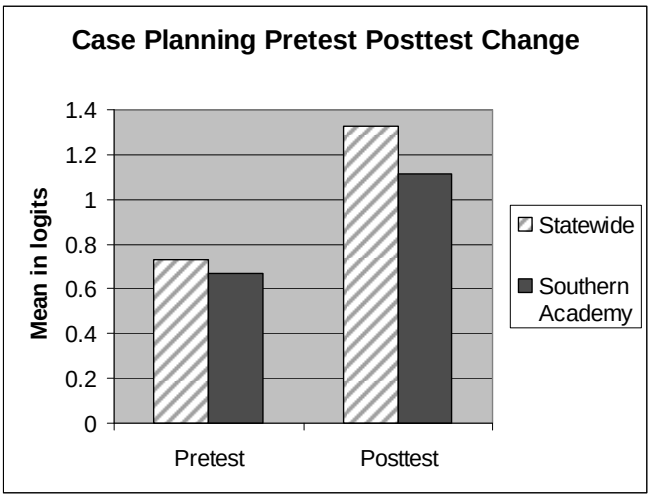
Knowing both the level of difficulty of different areas of training content—especially when measures of the difficulty are independent of a particular group—and individual participants' level of mastery makes it possible to distinguish whether

exceptionally high or exceptionally low ratings are due to the content or to the participants. This is especially so when ratings for one group deviate from ratings for other groups. The next example illustrates the use of Rasch measurement in making that distinction.

Rasch measurement and item difficulty differences associated with trainer. The second example comes from the California Common Core Training Evaluation. In California, there are four Regional Training Academies and the Inter-University Consortium that provide line worker core training to public child welfare workers. Seven modules of this line worker core are standardized curricula, meaning that the content, PowerPoint slides, and activities are provided to the trainer. Certain modules contain a pre/post-test consisting of multiple choice test items. While a module is a standardized curriculum, delivery of the curriculum may not be uniform. Each region uses a different pool of trainers to deliver these standardized curricula and depending on the expertise of each trainer and the make-up of the participants, trainers may emphasize certain knowledge/skills/tasks of the curriculum more than others.

During calendar year 2008 in one module, *Family Engagement in Case Planning and Case Management*, participants from the Southern Academy (the Public Child Welfare Training Academy) changed significantly less from pretest to posttest than participants in the other regions ($p<.01$). They also scored significantly lower at posttest ($p<.01$)., Statewide there were 743 complete pairs of pre/posttests for curriculum versions 1.25 and 1.26. There was an average gain of .60 logits from pretest to posttest, and a posttest mean of 1.3 logits. For the Southern Academy, there were 228 complete pairs of pre/posttests for curriculum versions 1.25 and 1.26, an average gain of .45 logits, and a posttest mean of 1.1 logits.

Figure 13: Case Planning Pretest Posttest Change



Since California's common core is standardized in curriculum, but not standardized in delivery, the Southern Academy wanted to know if there is a difference in the way the curriculum is delivered by the trainers in the southern region that can account for why the southern region's trainees changed significantly less from pretest to posttest than the rest of California. A report was produced that examined the differences in overall test scores and the difficulties of specific test items by trainer. The intent was to provide more information to trainers and curriculum developers to help fine-tune course content and delivery, and to use the report as part of discussions among training Academy staff, curriculum developers, and trainers.

Differences in scores, overall or on specific test items, are a function of many variables having to do with the trainee and the curriculum as well as the trainer. Moreover, when numbers are small, differences across trainers in item difficulty may be overly influenced by one or two trainees and may look different as more data become available. Thus, while there may be a training delivery-related reason for an item in a particular area to be more difficult (e.g. the learning point was not covered or was not covered in the same way), there also may not be any obvious link to how training was delivered, and discussion is necessary to

understand what may be operating to produce different patterns of scores.

Score differences by trainer. There were no statistically significant differences by trainer in either the overall posttest score or the amount of change from pretest to posttest. Which trainer conducts *Family Engagement in Case Planning and Case Management* did not appear to make an overall difference in terms of student learning.

Patterns of item difficulty by trainer. Although there were no overall differences in change from pretest score to posttest score, the possibility remained that different trainers cover the material differently and that these differences could cause individual items to be more difficult for the students of one trainer compared to another. To identify areas in which trainers may cover the material differently, Rasch was used to estimate item difficulties separately for each trainer. The goal of this analysis was to examine the consistency of training delivery. Comparing item difficulties across trainers allowed the identification of areas where a hard item for one trainer's students is an easy item for another trainer's students. Potential differences between these difficulty estimates are examined in two ways: through correlations of item difficulties across trainers and by graphing item difficulties to identify outliers or items that vary in difficulty based on which trainer delivers the curriculum.

High correlations indicate that items that are difficult for one trainer's students are also difficulty for another trainer's students and vice versa. Correlations of .80 and above indicate excellent agreement. Correlations from .70 to .80 indicate good agreement, and correlations of .60 to .70 indicate fair agreement. Correlations between pairs of trainers are shown in Figure 14. As shown, most correlations between trainers are in the good to excellent range.

Figure 14. Correlation Table

		TRAINER01	TRAINER02	TRAINER03	TRAINER04	TRAINER05	TRAINER06
TRAINER01	Pearson Correlation		.809	.798	.786	.652	.791
	Sig. (2-tailed)		.000	.000	.000	.000	.000
	N		24	39	39	39	24
TRAINER02	Pearson Correlation	.809		.839	.919	.897	.805
	Sig. (2-tailed)	.000		.000	.000	.000	.000
	N	24		25	25	25	25
TRAINER03	Pearson Correlation	.798	.839		.840	.850	.704
	Sig. (2-tailed)	.000	.000		.000	.000	.000
	N	39	25		40	40	25
TRAINER04	Pearson Correlation	.786	.919	.840		.849	.876
	Sig. (2-tailed)	.000	.000	.000		.000	.000
	N	39	25	40		40	25
TRAINER05	Pearson Correlation	.652	.897	.850	.849		.764
	Sig. (2-tailed)	.000	.000	.000	.000		.000
	N	39	25	40	40		25
TRAINER06	Pearson Correlation	.791	.805	.704	.876	.764	
	Sig. (2-tailed)	.000	.000	.000	.000	.000	
	N	24	25	25	25	25	

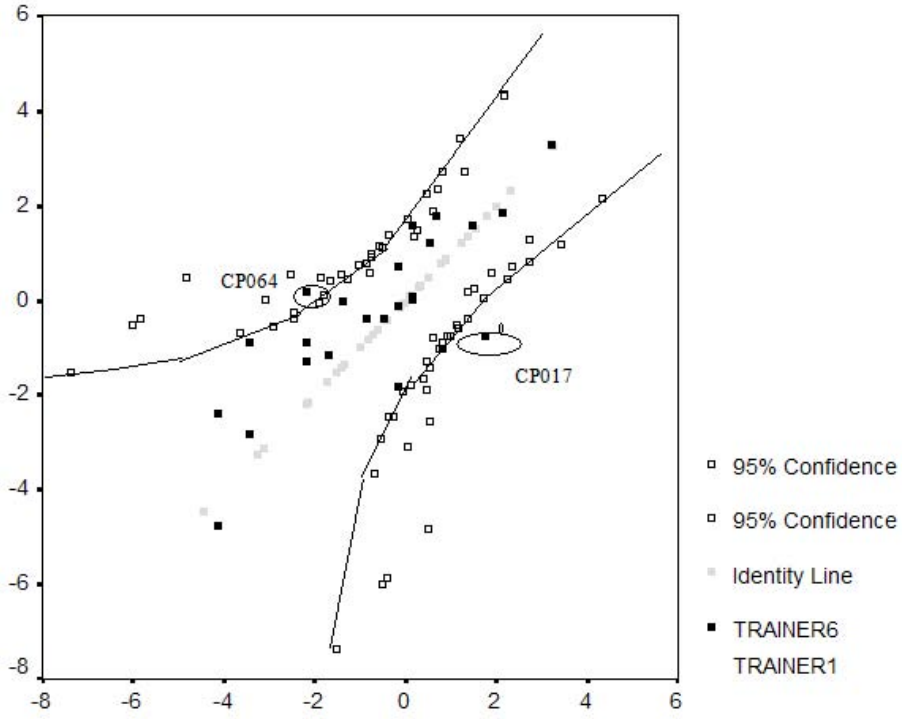
Estimates of Differential Functioning

Estimates of Differential Functioning

Graphing the item difficulty estimates of two trainers, we can identify items where the item difficulty estimate based on one trainer is significantly different from the item difficulty estimate based on another trainer. Figure 15 shows the item difficulty estimates between trainer 1 on the x-axis, and trainer 6 on the y-axis. The solid black squares are where the difficulty values of the two trainers intersect. The light gray squares are the desired state for each item - difficulty estimates that are identical. The curved lines, which approximate the hollow black squares, indicate a confidence interval of 95%. Any points that lie outside these lines represent items where the item difficulty based on one trainer is significantly different from the item difficulty based on another trainer.

In Figure 15, CP064 has an item difficulty value of -2.17 for Trainer 1 and 0.17 for Trainer 6. Since difficulty on the logit scale increases from negative to positive, this item was significantly more difficult for Trainer 6's students. CP017 has an item difficulty value of 1.75 for Trainer 1 and -0.75 for Trainer 6. This item was significantly more difficult for Trainer 1's students.

Figure 15. Graph of Item Difficulty for Trainer 1 & Trainer 6]



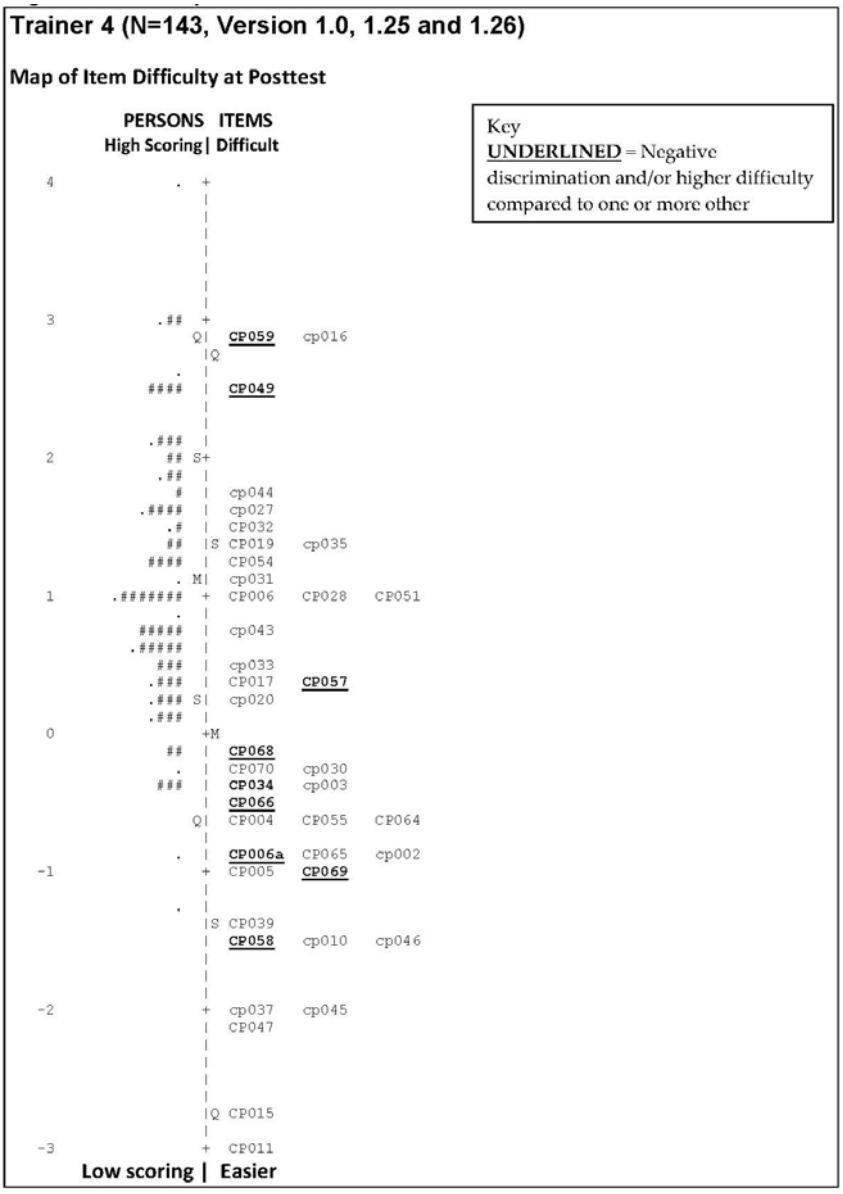
In Rasch differences in item difficulty are shown on item maps. In this example, for each trainer, maps were generated showing test items arranged along a scale from easiest to most difficult. The amount of space between the items indicates approximately how much they differ in difficulty. Items on the same line are approximately equal in difficulty.

The researchers identified items that were significantly different in difficulty depending on which trainer delivered the curriculum using graphs like the one in Figure 15. These items were then underlined on the maps produced by the Rasch software and used as part of the meetings with the trainers. Items

that were underlined received further scrutiny, either because they were more difficult for this trainer's students than one or more other trainer's students or because they showed negative discrimination (which means they were missed more often than expected by higher scoring people or were answered correctly more often than expected by lower scoring people).

Figure 16 shows a sample of an item map for Trainer 4. In this example, CP011 is an easy item for this trainer. CP059 and CP016 are hard items for this trainer. CP059 and CP016 are also equivalent in difficulty.

Figure 16. Item Map for Trainer 4



Since the item maps were run for each individual trainer and are difficult to compare, Figure 17 shows all of the items that differed significantly for at least one of the possible pairs of trainers. Items that are underlined either showed a negative discrimination and/or were more difficult for this trainer's

students than one or more other trainer's students. This table omits items which perform as expected.

In Figure 17, CP059 is significantly harder for students of Trainer 1, Trainer 4 and Trainer 6 compared to students of Trainers 2, 3 and 5. CP058 while on the easier end of items, is very easy for Trainers 1, 3, and 6 but slightly harder for Trainers 2, 4, and 5.

Figure 17. Item Difficulty Scores

Item	Trainer 1	Trainer 2	Trainer 3	Trainer 4	Trainer 5	Trainer 6
CP005	-1.71	-1.21	**-.28**	-1.03	-1.94	-1.14
CP006a		-1.70	-2.91	**-.93**	-1.44	**.78**
CP017	**1.75**	.35	.59	.42	-.31	-.75
CP019	**2.14**	1.66	**2.54**	1.39	1.51	1.87
CP028	.67	1.44	.78	1.00	**1.95**	**1.80**
CP032	.52	**2.00**	**1.73**	1.48	1.23	1.21
CP033	.31		1.24	.55	**2.25**	
CP034	-.15	**.89**	.28	-.40	-.31	-.13
CP037	**-.41**		-2.11	-2.02	-3.79	
CP045	-.04		-1.76	-2.01	**-2.51**	
CP049	1.47	1.34	1.16	**2.56**	2.02	1.60
CP051	.15	**1.55**	.49	.95	.43	**1.60**
CP057	-.15	**.41**	-.08	**.34**	**.11**	-1.82
CP058	-3.44	**-1.21**	-4.07	**-1.51**	**-1.31**	-2.81
CP059	**3.20**	1.88	1.62	**2.87**	1.01	**3.29**
CP064	-2.17	-.79	-.87	-.59	-.84	**.17**
CP065	**.15**	-1.21	-1.35	-.92	-1.18	**.10**
CP066	-3.44	**-.76**	**-.51**	**-.47**	**-.47**	**-.89**
CP068	-1.38	-.32	**.05**	**-.13**	-.74	**-.03**
CP069	-.48	**-.15**	-.87	-1.01	-1.46	-.38

NOTE:

- **Underlined values** = Negative discrimination and/or higher difficulty compared to one or more other trainers.
- Blank cells indicate items that were not included on any test versions administered by this trainer.

Discussion with Trainers

Differences in scores, overall or on specific test items, are a function of many variables, having to do with the trainee and the curriculum as well as the trainer. While there may be a training

delivery related reason that can account for the difficulty of an item in a particular area, there also may not be any obvious link to how training was delivered, and discussion is necessary to understand what may be operating to produce different patterns of scores.

The correlations matrix, the Graph of Item Difficulties, the Item Maps, and the Item Difficulty Scores were shared with each of the trainers that train *Family Engagement in Case Planning and Case Management* for the Southern Region. During this discussion with these trainers, differences in delivery that were related to specific items and their associated learning objectives emerged, which the following quotes illuminated:

- "I thought that everyone coming into this field knew this information and, thus, I didn't cover it in the training."
- "That content is covered primarily in the Jeopardy activity and I don't know how to play Jeopardy."

By engaging in the discussion with trainers, specific information was obtained which can contribute to increasing the fidelity of the curriculum delivery. Additionally, staff for the Southern Region was able to provide trainers with assistance or additional information that could be used by trainers when teaching this training.

Current Questions, Struggles, and Resolutions

From the foregoing it is clear that Rasch measurement presents a set of analytic tools with the potential to add to child welfare training evaluators' ability to understand the training process, and over time, these tools will undoubtedly find their place in the repertoire of evaluation methods and approaches.

The following are some considerations that may guide the use of these methods. One practical limitation is the large sample size needed to conduct the analyses. Some Rasch experts (B. Wright, personal communication, July 1995) suggest that sample sizes need not be as large as is commonly thought. However, substantial numbers are needed when we start to drill down to examine aspects of the data in detail. Another potential issue lies in interpreting the results of Rasch analysis to stakeholders who are unlikely to be familiar with the assumptions and terminology.

Additionally, Stacey and Steinle (2006) caution that Rasch measurement may not be applicable to studies of conceptual learning, and finally, comparisons between results obtained by classical testing theory and item response theory approaches, which include Rasch measurement, have found that results are, similar, with only some small exceptions (Fan, 1998; Nükhet, 2002). If this is the case, then Rasch measurement may be most useful for those tasks which are easier to accomplish using it or those which classical testing theory approaches do not handle well or at all. Two such tasks, which were not addressed in this article, are examining the ways in which response categories are used by respondents and correcting for rater bias where essays or interviews are scored.

Nevertheless, overall Rasch analysis is well suited to many of the practical issues confronting training evaluation in child welfare. It facilitates:

- longitudinal comparisons and tracking of trends when test versions and curricula are changing over time or when learner characteristics vary (e.g., when there is an influx of adult protection workers or private agency workers who have not previously been part of the training);
- cross-sectional comparisons across trainee groups that differ in education, experience or other background variables; and
- use of alternate test forms for pre and posttesting and the development of item banks.

It is important to remember, however, that Rasch models represent a theory of how useful measures can be constructed from data based on certain a priori requirements. Thus observed data must fit the Rasch model in order to take advantage of its properties.

Further Reading

Institute for Objective Measurement, Inc. - http://www.rasch.org/

Wright, B.D., & Stone, M.H. (1979). Best Test Design Rasch Measurement. Chicago, IL: Mesa Press. Available at https://mmm1406.sanjose14-verio.com/rascho/books.htm

Wright, B.D., & Masters, G. (1982). Rating Scale Analysis. Chicago, IL: Pluribus Press. Available at https://mmm1406.sanjose14-verio.com/rascho/books.htm

Smith, E., & Smith, R. (2004). Introduction to Rasch Measurement. Maple Grove, MN: JAM Press. http://www.jampress.org/

Bond, T.G., & Fox, C.M. (2007). Applying the Rasch Model. Fundamental Measurement in the Human Services. (2nd Ed.). London, UK: Routledge Academic.

References Cited

Baghaei, P. (2007). Applying the Rasch rating-scale model to set multiple cut-offs. *Rasch Measurement Transactions, 20*(4), 1075-1076. Retrieved from http://www.rasch.org/rmt/rmt204a.htm.

Baghaei, P. (2009). A Rasch-Informed Standard Setting Procedure. *Rasch Measurement Transactions, 23*(2), 1214. Retrieved from http://www.rasch.org/rmt/rmt232f.htm.

Bond, T.G., Fox, C.M. (2007). Applying the Rasch Model. Fundamental Measurement in the Human Services. (2nd Ed.). London, UK: Routledge Academic.

Brown, T., & Unsworth, C. (2009). Evaluating construct validity of the Slosson visual-motor performance test using the Rasch measurement model. *Perceptual and Motor Skills, 108*(2), 367. Retrieved from Alumni - Research Library

Chiu, Y-P., Fritz, S.L., Light, K.E, & Velozo, C.A. (2006). Use of item response analysis to investigate measurement properties and clinical validity of data for the dynamic gait index. *Physical Therapy, 86*(6), 778-87. Retrieved from Alumni - Research Library.

Choppin, B. (1979). Testing the questions – The Rasch model and item banking. MESA Research Memorandum Number 49. MESA Psychometric Laboratory. Retrieved from http://www.rasch.org/memo49.htm.

Fan, X. (1998). Item response theory and classical test theory: an empirical comparison of their item/person statistics. *Educational and Psychological Measurement. 58*(3).

Green, K.E., & Frantom, C.G. (2002). Survey Development and Validation with the Rasch Model. A paper presented at the International Conference on Questionnaire Development, Evaluation, and Testing. Charleston, SC, November 14-17.

http://www.jpsm.umd.edu/qdet/final_pdf_papers/green.pdf. Downloaded 8/17/10.

Guttman, L. (1944). A Basis for Scaling Qualitative Data. *American Sociological Review, 9*(2), 139-150.

Hahn, E., Cella, D., Bode, R., & Hanrahan, R. (2010). Measuring social well-being in people with chronic illness. *Social Indicators Research, 96*(3), 381-401. Retrieved from Alumni - ABI/INFORM Complete.

Kelley T., Ebel R., & Linacre, J.M. (2002). Item discrimination indices. *Rasch Measurement Transactions 16*(3), 883-884.

Kelly, C.W. (2005). Commitment to health scale. *Journal of Nursing Measurement, 13*(3), 219-29. Retrieved from Alumni - ProQuest Health Management.

Linacre, J.M (2004a). Test validity and Rasch measurement: construct, content, etc. *Rasch Measurement Transactions*. *18*(1), 970-971.

Linacre, J.M (2004b). Optimizing rating scale category effectiveness. *Introduction to Rasch Measurement.* E. Smith & R. Smith (Eds.). (pp. 258-278). Maple Grove, MN: JAM Press.

Linacre, J.M. (2006). WINSTEPS Rasch Measurement Software for persons and items. [computer software]. Available from http://www.winsteps.com

MacCann, R.G., & Stanley, G. (2006). The use of Rasch modeling to improve standard setting. *Practical assessment, research and evaluation. 11*(2). Retrieved from http://pareonline.net/getvn.asp?v=11&n=2

Mulqueen, C., Baker, D., & Dismukes, R.K. (2000). Using multifacet Rasch Analysis to examine the effectiveness of rater training. Poster paper presented at the 15th Annual Meeting of the 'society for Industrial/Organizational Psychology (SIOP) in New Orleans, LA.

Nükhet, Ç.D. (2002). A study of Raven Standard Progressive Matrices test's item measures under classic and item response models: an empirical comparison. Ankara University, *Journal of Faculty of Educational Sciences. 35*(1-2).

ODL Open and Distance Learning. (1999, October 7). Guttman's Scalogram. Retrieved from http://www.mathpsyc.uni-bonn.de/

Pohl, N. F. (1982). Using retrospective pre-ratings to counteract response-shift confounding. *The Journal of Experimental Education, 50*(4), 211-214. Retrieved from http://www.jstor.org/stable/20151462.

Smith, E.V. Jr.,& Kulikowich,J.M. (2004). An application of generalizability theory and many-facer Rasch measurement using a complex problem solving skills assessment. *Educational and Psychological Measurement. 64*, 617-639.

Smith, R. (1992). Computer adaptive testing. In Smith, R.(Ed.) *Applications of Rasch measurement.* (pp. 157- 172). Maple grove, MN: JAM Press.

Stacey, K., & Steinle, V. (2006). A case of the inapplicability of the Rasch Model: Mapping conceptual learning. *Mathematics Educational Research Journal. 18*(2), 77-92.

Stenner A.J., & Stone, M.H. (2003). Item Specification vs. Item Banking. *Rasch Measurement Transactions. 17*(3), 929-930. Retrieved from http://www.rasch.org/rmt/rmt173a.htm.

Wright, B.D., & Mok, M.M.C. (2004). An overview of the family of Rasch measurement models. In E. Smith & R. Smith. (Eds.). *Introduction to Rasch Measurement.* Maple Grove, MN: JAM Press.

Wright, B.D., & Stone, M.H. (1979). Best test design Rasch measurement. Chicago, IL: Mesa Press.

2010 NHSTES Proceedings

Rasch Measurement for Child Welfare Training Evaluators

Skill-Building Session

Presenters: Henry Ilian, *D.S.W., New York City Administration for Children's Services;* Cynthia Parry, *Ph.D., Consultant; &* James Coloma, *B.A., Public Child Welfare Training Academy, San Diego State University*

Thursday, May 27, 2010
1:15-2:45 pm

This skill-building session centered on Rasch measurement and its application in child welfare training evaluation. The discussion focused on presenting a brief history of Rasch measurement, highlighting the benefits of using this measurement, and providing an overview of how to use it. The session concluded with a practice session during which participants applied what they learned.

Rasch Measurement versus Measurement in Classical Testing

- Both types of measurement can calculate item difficulty through item analysis. However, with a classical testing framework, additional information must be collected about the characteristics of the test, the scale, and the distribution of the test takers in order to interpret the results. Otherwise, only the number of correct responses would be reported. Rasch measurement does not require all this information for analysis because it uses the item difficulties in its calculations. Rasch takes into account not only how many items a test-taker answered correctly, but also which items they answered correctly.

Overview of Rasch measurement

- The purpose of using Rasch measurement is to get a stable estimate of the person's ability that is not dependent on the type of test being taken.
- The presenters provided examples of Rasch scores and explained what the scores reveal about the test-taker.
- The Rasch measurement program is able to find unusual patterns in the person's responses (for instance, if the person was picking answers randomly, etc.). It can also assess whether or not a group's scores changed over time due to an actual change in ability or a change in test difficulty. Thus, changing test items does not impact Rasch's ability to measure a person's ability. Rasch accomplishes this by distributing the test items in a way that makes different tests equal.
- High-difficulty and low-difficulty items tend to produce more error and a greater confidence interval. Ideally, items should be targeted to the ability of the group.
- Rasch can also be used to assess if there are too many easy items or too many difficult items on a particular test. It allows for the removal of redundant test items, and can be used to make shorter tests in which the items reflect various levels of difficulty.
- Rather than analyzing test items in isolation, Rasch analyzes the probability of a correct response based on the person's ability and the item difficulty.
- Rasch can also be used for validating scales to see which items conform most closely to the construct. Those items that are not consistent with the construct can either be dropped or analyzed separately.
- The presenters gave a brief history of the development of Rasch measurement.

Basics of Data Analysis

- *Anchoring a scale* means using estimated item difficulty and keeping it consistent. The measurement must be constant so that comparisons can be made across tests. Doing so allows for item banking.
- It is important to have an equal interval scale.
- Change is relative and cannot be measured by simply looking at the raw test scores. An individual can begin with a low

level of knowledge, learn a great deal from the training, but still score lower than someone who started with a high level of knowledge but grew little from the training.
- The process of analyzing data includes removing unusual items (i.e., random responses, all "A" responses, etc.), removing invalid or redundant test items, and checking the distribution of easy and difficult items.

Guttman Scaleogram

- The presenters provided information on the Guttman Scaleogram (i.e., the Bogardus Social Distance Scale).
- The presenters used data from a 413-person data set to illustrate a scale. The sample rating scale ranged from 1 to 5. If people felt confident in their ability, they would self-rate themselves at a 5. If they were not confident, they would rate themselves at a 1.
- In the sample pretest, people were more confident in responding to domestic violence than utilizing family court. The curriculum did not contain much content on domestic violence. A posttest showed that the domestic violence component demonstrated the least amount of growth. The greatest amount of growth was in utilizing the family court. Equal interval scale allowed for a comparison of growth between the two subjects.
- Analysis also showed that the 5-point scale can be reduced to a 4-point or 3- point scale because people seem to use 1s and 2s and 3s interchangeably.

Rasch Resources

- There are several software programs that enable Rasch statistical analysis. One is called Winsteps and costs about $140. Another program that is available for purchase is called Facet. UC Berkeley offers a free software program called Construct Maps.
- The presenters advised participants to take some online statistical methods courses through statistics.com to learn more about Rasch measurement or to brush up on statistical measurement. There is also a course offered through the Institute for Objective Measurement.

Practice Session

- Participants formed groups of five or six and looked at some data that was provided through Rasch. They were instructed to look at the test items, try to determine what was going on with the test item, and provide any recommendations about the test item. Presenters walked participants through how to look at the tables and graphs to interpret the scores. Participants learned where to find information on how many people answered the question, the Rasch difficulty score, the standard error, and fit measures (if the person's ability matches the item difficulty).

See page 61 for the paper submitted by the presenters.

Wrap-up and Strategies for 2011

Barrett Johnson, *LCSW, Director, In-Service Child Welfare Training, &* Leslie Zeitler, *LCSW, Training & Evaluation Specialist, CalSWEC, University of California, Berkeley*

Thursday, May 27, 2010
3:00-4:00 pm

Bringing the symposium to a close, Barrett Johnson and Leslie Zeitler of CalSWEC facilitated a discussion about the Symposium, including next steps for future Symposia.

Positives about the Symposium

- Symposium participants stated that the highlights included the session on Rasch measurement and the skill-building components.
- Participants appreciated the flexibility of the presenters because it allowed them to actively participate in the presentations.
- They enjoyed the ability to speak freely without having to use a microphone to ask a question.
- Participants did not find the absence of a keynote speaker to be problematic.
- Overall, participants stated that they enjoyed hearing what other participants had to say and what they were doing.

Areas for Improvement

- Symposium participants discussed the need for better recruitment.
    - This could be an issue of budget or retirement, but it could also be that the group is too insular. One participant suggested incorporating icebreakers. However, this had been attempted before. Further, even when there was high

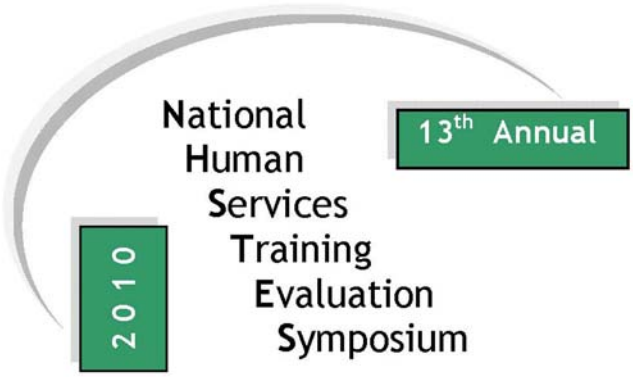
turnout, not all states were represented. A previously conducted survey revealed that budget was the main reason for non-attendance.
    - Scheduling may also be an issue. Some participants stated that Wednesday to Thursday was a better schedule than Thursday to Friday. The Memorial Day holiday may also impact attendance.
    - Outreach to a broader audience might be an option. This expanded audience could include after-school providers and those that provide technical assistance
- It was suggested that pre-symposium sessions should be made available to participants.
    - One proposal was for a three-hour or full-day pre-symposium session.
    - Another proposal was for a pre-symposium session that was no more than two hours. However, if it was part of the larger program, then three hours could be made available for a pre-symposium session, with skill-building taking up to two hours during the larger program.
- Online Attendance/Virtual Technology
    - While new technology does seem to hold some potential, the main issue is that much of the symposium is centered around discussions, not traditional lectures or presentations.
    - New technology might be better suited for a format that requires scheduling five presenters and putting them up on the same day as a Webinar™ (first, a brief presentation, followed by a virtual discussion). Immediately after the Webinar™, a follow-up email could be sent that includes an opportunity to share tools.
- Collaboration with the National Staff Development and Training Association (NSDTA) Conference could also be a useful option.
- Symposium participants would like to continue the yearly face-to-face conference. A suggestion was made to alternate every other year at a non-California location. This requires that participants rotate the work of finding a location and coordinating logistics so that CalSWEC is not always

responsible for a non-California location. This may result in fewer California participants attending, but may increase attendance from other states. Symposium participants would like to try this alternative. In order for this to be carried out, materials (i.e., logo, stationary, etc.) need to be consistent. The approval of funders will also be needed to see what the possibilities are for locations, scheduling, etc.

- One proposal for a topic at the next symposium was training as an intervention in an evidence-based model.

Possible new locations to explore for next Symposium

- Madison, Wisconsin
- Cornell University, New York
- Pittsburgh, Pennsylvania
- Baltimore, Maryland
- Columbia University, New York
- Denver, Colorado
- Others will be looked into.



May 26–27, 2010

International House
University of California, Berkeley

Synopses of Projects

Table of Contents

Current Training Evaluation Projects

Jane Berdie, MSW

1. **CalSWEC**
   Contacts: Barrett Johnson, barrettj@berkeley.edu, and Leslie Zeitler, lzeitler@berkeley.edu
   Working with the statewide training evaluation committee to design and implement the second strategic plan for evaluating Caseworker Common Core Training. Current evaluations include formative evaluation for curriculum design, knowledge tests, and embedded evaluations of skill.

2. **Denver University School of Social Work, Butler Institute**
   Contact: Robin Leaked Robin.Leake@du.edu
   Formative evaluation of the National Child Welfare Workforce Institute's Leadership Academy for Middle Managers training curriculum.

3. **Academy for Professional Excellence, San Diego State University School of Social Work** Contact: James Coloma, jcoloma@projects.sdsu.edu
   Working with team to develop embedded evaluations of training on adult protection.

4. **Bay Area Training Academy**
   Contact: Krista Brown, kbrown70@sfsu.edu
   Evaluation of adult protection training.

5. **Child Welfare Academy, University of Maryland School of Social Work**
   Contact: Leslie Rozeff , LROZEFF@ssw.umaryland.edu
   Consultation on revision of knowledge test for revised caseworker pre-service training.

Contact

Jane Berdie, MSW
Child Welfare Consultant
435 South Gaylord St.
Denver, CO 80209
Phone: 303-733-9532
jberdie@msn.com

Children's Interviewing Institute

The Children's Interviewing Institute is a week long training series that guides participants in developing a stronger knowledge base around conducting interviews with children as well as developing the participants' skills at conducting, documenting, and articulating the methods used in communicating with children. Gathering information from children is a fundamental part of the job of a child welfare professional or law enforcement investigator. The ability to communicate with children can be the difference between success and failure, safety and danger, or simply an accurate glimpse into the mind of a developing personality.

Evaluation Focus
Participants at this week long training institute are provided with the opportunity to engage in two mock interviews. These mock interviews are individually critiqued by expert field interviewers. The mock interviews are also video taped. In addition, participants are brought back six and twelve months after the initial training for a half-day refresher training. The refresher training also provides the participants another opportunity to engage in additional mock interviews.

Evaluation Design and Data Collection Plan
At the beginning of the institute, participants are provided with a self-assessment of their interviewing skills. During the week long institute, participants are taught the different stages of the interviewing process with a child. Participants are also taught how to effectively create open-ended questions which are known to elicit the most information from a child. As each interview is taped, a scoring rubric has been created to score the videos. Participants are tracked twelve months after the initial training to see if they have incorporated the knowledge and skills that were presented at the training.

Leaders in Action: Executive Development Program

Leaders In Action (LIA)
is a partnership between the Southern Area Consortium of Human

Services
(SACHS), San Diego State University and the Academy for Professional E xcellence. This
Executive Development Program originated out of an interest from the SACHS directors for an
executive talent pipeline to fill executive positions left vacant due to inc reased retirement rates.

This program is unique in that it is based on a thorough needs assessme nt and research process tailored to Southern California Human Services director needs. The program stresses the
development of "strategic thinking" skills using case material and other data available from the
collaborative practices study, as well as other sources. Participants are p repared for strategic
leadership in complex human service organizations with an emphasis on building collaboration, organizational change, interagency partnerships, performance measurements, and related
subjects.

Leaders In Action is being evaluated to assess the participant outcomes from participating in the
LIA Executive Development Training Program. Factors being evaluated a re the transfer of
learning of the curriculum to the work place setting, tracking of particip ants after they leave
the program, and satisfaction for participating in the program.

Contact
James Coloma, M.S.W.
Academy for Professional Excellence
San Diego State University
School of Social Work
6505 Alvarado Road, Suite 107
San Diego, CA 92120
Phone: 619-594-3546
FAX: 619- 594-1118
jcoloma@projects.sdsu.edu
http://theacademy.sdsu.edu

Evaluating Utah's Training Continuum

In Utah the Division of Child and Family Services (DCFS) offers core and advanced skill-level training to child welfare staff. The College of Social Work (CSW) offers the BSW and MSW programs. CSW provides Title IV-E stipends to both BSW and MSW students.

We will be evaluating the DCFS training programs. We are also evaluating the BSW and MSW training programs. Another step in the project is to determine the competencies that are necessary to deliver effective child welfare services. We will then evaluate both DCFS training and BSW/MSW training to determine if the competencies are addressed. We will then have an overall evaluation of Utah's training continuum.

Utah Division of Child and Family Services is a state-supervised/-administered system and DCFS offers nine weeks of core and advanced training (mandated).

Contact

Norma Harris, M.S.W., Ph.D.
Director, Social Research Institute
Research Professor/Lecturer
College of Social Work
University of Utah
395 S. 1500 E Rm. 111
Salt Lake City, UT 84112
Phone: 801-581-3822
Fax: 801-585-6865
norma.harris@socwk.utah.edu

Lazarina Topuzova, Ph.D.
Assistant Professor
Department of Organizational Leadership
Gonzaga University
502 E. Boone Ave., MSC 2616
Spokane, WA 99258-2616
Phone: 509-313-3620
Fax: 509-313-3566
topuzova@gonzaga.edu

Developing Usable Feedback to Stakeholders from Child Welfare Training Evaluation Data

The Assessment and Evaluation Department of the James Satterwhite Academy for Child Welfare Training (part of the New York City Administration for Children's Services) collects test and participant evaluation data from a variety of courses and from several thousand participants per year. Heretofore, we have provided reports to some, but not other, stakeholders, and many of those reports were in a format that was not easy for stakeholders to use to make changes in training or curriculum.

We have developed a three-part initiative to make training evaluation reporting more usable for stakeholders and thereby have a greater impact on the training.

1) We have automated our reporting system so that nearly all test scoring and reports can go through one application (developed for us by the agency's MIS), making the production of a variety of reports for stakeholders both fast and efficient.
2) We have developed a report that summarizes the main curriculum issue covered by each test question in one to two sentences and gives the proportion of a class that answered that question correctly, i.e., understood that curriculum issue. This report is produced in pre-test and post-test versions. The posttest version also gives the gain (or loss) from the pre-test. The pre-test report is given to trainers as soon as possible, usually before the lunch break, on the first day of training, so that they are able to adjust which areas they give the greatest emphasis in the training. The post-test report is particularly useful during the piloting phase of course development. This report was developed after the main report generator was nearly completed, and was automated very simply, using a combination of MS Excel and MS Access.
3) We have begun semi-annual presentations to the entire Training Academy, covering important findings from testing, participant evaluation surveys, and trainer assessment of participant skills. The first presentation, which will occur in March or April, will cover the core training and address such concerns as content that did and did not do well on the tests, how trainers rated participants on a variety of skills, components of the training that participants felt were strong and skills that they felt confident that they had learned.

Context

The New York State child welfare system is state supervised and county administered. A standardized core curriculum is mandated by the state. With state involvement and oversight, New York City developed a separate core curriculum from that used in the rest of the state. This curriculum, called the Child Protective Service Practice Core, integrates common core skills with child protective content.

Contact

Henry Ilian, D.S.W.
Assessment and Evaluation Department
James Satterwhite Academy for Child Welfare Training
ACS Children's Center, 5th Fl.
492 1st Ave.
New York, NY 10016
Phone: 646-935-1410
henry.ilian@dfa.state.ny.us

Evaluating Transfer of Learning In Public Welfare Training and Development:
Factors Affecting the Transfer of Learning of California Public Welfare Workers

The Center for Human Services, University of California at Davis

Goal

Examine transfer of learning for public welfare staff participating in the California Inter-County Training Consortium.

Primary Research Questions

6. What training and development instructional design factors contribute to transfer of learning from classroom to practice?
7. What organizational elements within an agency support transfer of learning?
8. What individual traits of public welfare staff support transfer of learning?

Additional/Secondary Research Questions

9. Is the Transfer Potential Questionnaire a valid instrument for assessing transfer potential across human service populations?
10. Can an abbreviated (more user-friendly) version of the Transfer Potential Questionnaire be developed as a valid instrument for assessing transfer potential across human service populations?

Sample

809 potential participants; 456 completed TPQ questionnaire (56% of 809); 259 completed second survey –HSTEP (57% of 459).

Methods

Training Transfer Assessment Instruments

Transfer Potential Questionnaire

The Transfer Potential Questionnaire (TPQ) is a 68-item questionnaire assessing individual, organizational, and training design factors. The TPQ was found to have high internal reliability (Chronbach's alpha=.96) and significantly predict the transfer of learning of child protective social workers (r=.62). It was later found to have a positive relationship with long-term employee retention with the same group of child protective social workers.

Human Services Training Effectiveness Postcard (HSTEP)

The HSTEP is a brief questionnaire consisting of five items designed to incorporate the four major areas/levels of Kirkpatrick's training evaluation model into a single assessment index. Kirkpatrick's widely accepted four-level evaluation model includes (1) reaction/satisfaction, (2) learning, (3) behavior/transfer of learning, and (4) results of transfer of learning. The HSTEP is also intended to provide acceptable return rates from training participants who may have high caseloads and little time for assisting in the evaluation of training programs. In addition to providing quantitative outcome data, the instrument was designed to elicit qualitative information on the process of transfer of learning (factors that helped or hindered transfer).

Process

The TPQ was administered to study participants at the conclusion of each training workshop. Three months after the conclusion of training, participants were mailed the HSTEP questionnaire.

Contact

Michael Lawler, M.S.W. (mjlawler@ucdavis.edu)
Jann Donnenwirth, M.S. (jdonnenwirth@ucdavis.edu)
Center for Human Services
University of California, Davis Extension
1632 Da Vinci Ct.
Davis, CA 95618-4860
Phone: 530-757-8643
Fax: 530-754-5104

Dale Curry, Ph.D., LSW, CYC-P
School of Family & Consumer Studies
Kent State University
Nixson Hall
P.O. Box 5190
Kent, OH 44242
Phone: 330-672-2998
Fax: 330-672-2194
dcurry@kent.edu

Engaging Child Welfare Workers in Research and Obtaining Valid Data

"Hard to reach" groups in human services research are often those who may be homeless, have cognitive challenges, or be otherwise difficult to locate or engage in research. Child welfare workers may be considered similarly challenging because the work that they do leaves them little time in the office, and the demands of high caseloads leave them little cognitive resources for participating in anything beyond daily work tasks. Qualities of child welfare agencies as historically compliance-driven and hierarchical likely contribute to workers' confusion and distrust about research. Pennsylvania has implemented several strategies to engage and retain workers in research, with disappointing results. The purpose of this discussion is to describe Pennsylvania's experiences and to gain feedback from others about engaging workers in research and obtaining valid data.

Background

Research in training, transfer of learning, organizational effectiveness, and service is only as valid as the quality and representativeness of the data that is collected. In Pennsylvania (PA), three training-related studies and three service-focused studies have raised concerns about response and non-response bias. In a study of transfer of learning, workers who completed follow-up surveys were significantly younger, more likely to work in on-going compared with intake units, and be graduates of Title IV-E traineeships compared with workers who did not complete follow-up surveys. Response rates were 84% of this agency at baseline and just 42% of participants at follow-up. Web-based surveys sent to 63 supervisors in another effort yielded 19% participation. A user-friendly, report-producing database that helps agencies meet state and federal policy requirements around young children's developmental needs has yielded 65% participation after 7 months of engagement strategies and incentives.

A second issue concerns social desirability response bias. For example, informal conversations with agencies indicate that resources prevent some agencies from using the database just described, yet scaled response items during a research interview show that 80% of agencies Agree or Strongly Agree that they "have the resources necessary to support implementing the project". In another telephone survey, workers were reluctant to discuss any situation in which they missed a home visit.

This project will be discussed at the symposium, along with questions to stimulate further thought and ideas on this topic.

Contact

Helen Cahalane, M.S.W., Ph.D
Principal Investigator
Child Welfare Education & Research Programs
Clinical Associate Professor
University of Pittsburgh
2328 Cathedral of Learning
Phone: 412-624-6386
hcupgh@pitt.edu

Julie McCrae, M.S.W., Ph.D.
Research Assistant Professor
Child Welfare Education & Research Programs
School of Social Work
University of Pittsburgh
2332 Cathedral of Learning
Pittsburgh, PA 15260
Phone: 412-383-5168
Fax: 412-624-1159
jsmccrae@pitt.edu

Children and Residential Experiences: Creating Conditions for Change (CARE): Evaluating a New Program Model for Residential Care
Whose Primary Implementation Strategy Is Training

Residential Child Care Project at Cornell University

This project implements and evaluates, within a quasi-experimental design, a new residential care program model that supports and maintains research based and best practices principles in 20 residential treatment facilities.

The CARE program model uses six principles as unifying themes that have the potential to create conditions for change in children's lives. The CARE principles are developmentally focused, family involved, relationship based, competency centered, trauma informed and ecologically oriented. These principles are grounded in theory, in evidence-based practices, and in practice wisdom.

The project uses training as a primary implementation strategy to introduce the CARE principles to all levels of the agency and to ensure on-the-job use. To implement and to sustain the CARE program model, the project incorporates supervisory, organizational self-assessment, participation-centered management strategies, and continuous quality improvement methods.

The evaluation measures learning post-training, assesses the immediate impact of training on intended and current practice, and determines appropriate implementation and sustainability strategies. The evaluation is embedded in a quasi-experimental design with each agency assigned either to an intervention or to a control group. Assessments at baseline and time 1 and 2 determine the impact on organizational congruence, culture, climate, and child outcomes. A qualitative methodology is employed that seeks to discover the organizational change dynamics necessary to implement and sustain the CARE program model.

Contact
Michael Nunno, D.S.W.
Family Life Development Center
College of Human Ecology, Cornell University
Phone: 607-254-5127
Phone: 315-209-1038
man2@cornell.edu

Current Training Evaluation and Human Resource Management Projects

University of Nebraska–Lincoln, Center on Children, Families and the Law

The UNL—Center on Children, Families and the Law (CCFL) provides pre-service training to all new Child and Family Services Specialist/Trainees (CPS worker title in Nebraska) employed by the Nebraska Department of Health and Human Services. This state-administered system requires new workers to complete a one-year training program, divided into two phases: pre-service and required in-service. Trainees assume limited case management responsibilities after the completion of the pre-service phase (at approximately 12 to 15 weeks, depending on upcoming job assignments), with a gradual increase to a full caseload by the 12th month of employment. After the first year, additional in-service training is offered on a wide variety of subjects.

Our team at CCFL is involved in a broad range of activities including: training evaluation, needs assessment, curriculum development, research, and consultation on human resource management issues to support the training. Past work with the agency has also focused on a complete redesign of the personnel selection system, including: job analysis, development of selection interviews and a writing skills assessment, designing and conducting test validation studies, advising on the implementation of a test of critical thinking skills, development of policy and procedures for use of selection tools, and training of systems users. Current training evaluation and human resource management projects include:

- **Evaluating Trainee Reactions**
  - Collecting, summarizing, and reporting trainees' written ratings and comments for each training unit
  - Conducting online post-training surveys to gather perceptions of training from recent trainees and their supervisors, both at the end of training and six months later

- **Evaluating Trainee Knowledge and Skills**
  - Creating new knowledge and skill assessments, and refining existing assessments to respond to curriculum changes

- Collecting and summarizing data from trainee knowledge and skill assessments, such as for testifying, documenting, and meeting with families

- **Evaluating Trainee Attitude and Behavior**
  - Collecting and summarizing trainers' behavioral ratings and comments regarding trainee attitudes and behavior in each training class (e.g., alertness, respectfulness, participation, etc.)

- **Reporting Trainee Progress**
  - Preparing routine trainee progress reports that summarize trainee assessment scores, performance feedback, and attitude and behavior evaluations; distributing reports to trainees and supervisors

- **Evaluating Job Performance and Turnover**
  - Providing in-service training to supervisors across the state on the use of our performance evaluation tool that assesses workers' task performance in 17 performance dimensions
  - Analyzing trainee and probationary employee job performance data
  - Completing a longitudinal study to determine predictors of employee job performance and turnover

Contact
Megan E. Paul, Ph.D. (mpaul2@unl.edu)
Kristin I. Saathoff, M.A.
Michelle I. Graef, Ph.D. (mgraef1@unl.edu)

Center on Children, Families and the Law
University of Nebraska–Lincoln
206 S. 13th Street, Suite 1000
Lincoln, NE 68588-0227
Phone: 402-472-3479

UNIVERSITY OF MARYLAND
SCHOOL OF SOCIAL WORK

525 W. Redwood St.

Baltimore, MD 21201-1705

410 706 7790

410 706 6046 *fax*

www.ssw.umaryland.edu

Training System Evaluation

The Child Welfare Academy (CWA) is located within the University of Maryland, School of Social Work, Ruth Young Center for Families and Children. The CWA operates in partnership with the Maryland Department of Human Resources, Social Services Administration in the development and delivery of competency-based training and evaluation for child welfare professionals, as well as foster and adoptive parents and providers.

Maryland is a state supervised, county administered system. Attendance at the six-week pre-service training for child welfare professionals is mandatory and continued employment requires passing of a knowledge test at the end of the training. Trainees have up to three opportunities to pass the exam over the course of six months.

Training System Evaluation
The CWA is currently engaged in several activities that are focused on assessment of the overall training system including:

- A comprehensive revision of the core pre-service curriculum
- Development and implementation of tools to assess trainer quality and ensure fidelity to the curriculum
- Examining effectiveness of embedded evaluation
- Surveys to assess the organizational culture/climate and impact on training effectiveness
- Training System Statewide Survey

Contact
Leslie Rozeff, LCSW-C

Director, Child Welfare Academy
University of Maryland School of Social Work
525 W. Redwood St.
Baltimore, MD 21201
Phone: 410-706-2059
Fax: 410-706-2423
lrozeff@ssw.umaryland.edu

Program

Wednesday, May 26

8:30 a.m.

Registration begins outside Homeroom.

9:30–9:40 a.m. **Homeroom**

Convene, Welcome, Introduction to the Symposium:
Barrett Johnson, *LCSW, Director, In-Service Child Welfare Training, CalSWEC, University of California, Berkeley*

9:40–10:00 a.m.

Ice-breaker
Leslie Zeitler, *LCSW, Training & Evaluation Specialist, CalSWEC, University of California, Berkeley*

10:00–11:30 a.m.

Skill-Building Session
Developing User-Friendly Human Service Training and Development and Assessment Instruments

- *Presenters*

– Dale Curry, *Ph.D., Kent State University*
– Jann Donnenwirth, *M.S.,* and
– Michael Lawler, *M.S.W., Ph.D., University of California, Davis*

11:30–12:45 p.m. | Lunch
Great Hall
Lunchtime Discussions

12:45–1:55 p.m.
Presentation & Facilitated Discussion
Is this research mandated?: Response and non-response bias in child welfare training and evaluation research
- *Presenters*

– Julie McCrae, *M.S.W, Ph.D.,* and
– Helen Cahalane, *M.S.W., Ph.D., University of Pittsburgh*
- *Facilitator*

– Michael Nunno, *D.S.W., Cornell University*

1:55–2:05 p.m. | Break

2:05–3:15 p.m.
Presentation & Facilitated Discussion
Evaluating Utah's training continuum
- *Presenters*

– Norma Harris, *Ph.D., University of Utah,* and
– Lazarina Topuzova, *Ph.D., Gonzaga University*
- *Facilitator*

– Sherrill Clark, *Ph.D., CalSWEC, University of California, Berkeley*

3:15–3:25 p.m. | Break

3:25–4:25 p.m.
Follow-up Session, Part 1
Steps in Developing a Handbook of Child Welfare Training Evaluation,
Part 1
- *Facilitator*

– Anita Barbee, *M.S.S.W., Ph.D., University of Louisville*

4:25–4:30 p.m.
Logistics for evening and morning, break for the evening
Leslie Zeitler, *LCSW, Training & Evaluation Specialist, CalSWEC,*
University of California, Berkeley

Thursday, May 27

9:30–9:45 a.m. **Homeroom**

Reconvene

Barrett Johnson, *LCSW, Director, In-Service Child Welfare Training, CalSWEC, University of California, Berkeley*

9:45–10:55 a.m.

Follow-up Session, Part 2

Steps in Developing a Handbook of Child Welfare Training Evaluation, Part 2

- *Facilitator*

– Todd Franke, *Ph.D., University of California, Los Angeles*

10:55–11:05 a.m. | Break

11:05 a.m.–12:15 p.m.

Facilitated Discussion

Workshop: Developing a statewide model of trainer evaluation

- *Presenters*

– Leslie W. Zeitler, *LCSW, CalSWEC, University of California, Berkeley*

- *Facilitator*

– Leslie Rozeff, *M.S.S.W., University of Maryland, Baltimore*

12:15–1:15 p.m. | Lunch
Great Hall
Lunchtime Discussions

1:15–2:45 p.m.

Skill-Building Session

Rasch Measurement for Child Welfare Training Evaluators

- *Presenters*

– Henry Ilian, *D.S.W., New York City Administration for Children's Services;*

– Cynthia Parry, *Ph.D., Consultant;* and

– James Coloma, *B.A., Public Child Welfare Training Academy, San Diego State University*

2:45–3:00 p.m.
Closing Remarks
Barrett Johnson, *LCSW, Director, In-Service Child Welfare Training, CalSWEC, University of California, Berkeley*

3:00–4:00 p.m.
Wrap-Up and Strategize for 2010
Barrett Johnson and Leslie Zeitler, *CalSWEC, University of California, Berkeley*

Acknowledgments

CalSWEC extends its gratitude to the members of the Steering Committee of the 13th Annual National Human Services Training Evaluation Symposium, who made this event possible.

Jane Berdie
Nancy Chavkin
Sherrill Clark
Todd Franke
Michelle Graef
Henry Ilian
Barrett Johnson
Christine Mattos
Julie McCrae
Michael Nunno
Cindy Parry
Doug Pratt
Salonie Rochell
Leslie Rozeff
Leslie W. Zeitler

Ad Hoc Committee

Anita Barbee
Dale Curry
Norma Harris
Amanda Schreiber
Christine Tappan

Directory of Presenters and Participants

Sevaughn Banks, *M.S.W., Ph.D.*
Training & Evidence Based-Practice Specialist
CalSWEC
School of Social Welfare
University of California, Berkeley
Marchant Bldg., Ste. 420
6701 San Pablo Ave.
Berkeley, CA 94720-7420
Phone: 510-643-0226
Fax: 510-642-8573
sevaughn@berkeley.edu

Sevaughn Banks' duties at CalSWEC include helping practice communities, research communities, and academia to collaborate on research-related projects to provide better outcomes for children and families in child welfare. She coordinates all evidence-based practice activities, forums, and projects. She is a trainer and lecturer in the California State University system and has connections to three of the four schools of social work in the Bay Area. Dr. Banks began her career in social work as a generalist practitioner, working in substance abuse, with at-risk children and their families, and in mental health.

Anita Barbee, *M.S.S.W., Ph.D.*
Professor and Distinguished University Scholar
Kent School of Social Work
University of Louisville
Oppenheimer Hall
Louisville, KY 40292
Phone: 502-852-0416
Fax: 502-852-0422

Anita Barbee was principal investigator or coprinicpal investigator on seven grants totaling $1.2 million for 2004–2005. She has spent the past 13 years evaluating child welfare training and outcomes for the Cabinet for Health and Family Services in Kentucky. She served as the evaluator for the

anita.barbee@louisville.edu	Children's Bureau Training and Technical Assistance Network of National Resource Centers for October 2004–September 2009.
Jane Berdie, *M.S.W.* Child Welfare Consultant 435 South Gaylord St. Denver, CO 80209 Phone: 303-733-9532 jberdie@msn.com	Jane Berdie is a trainer and training evaluation consultant in child welfare. She formerly was the training director for the Colorado Department of Human Services and the American Humane Association.
Susan Brooks, *M.S.W.* Director of Child Welfare Programs The Center for Human Services Northern California Training Academy UC Davis Extension 1632 Da Vinci Ct. Davis, CA 95616 Phone: 530-757-8643 Fax: 530-752-6910 sbrooks@unexmail.ucdavis.edu	Susan Brooks has nearly 20 years of experience in social services, with expertise in substance abuse, collaboration, team-building, and supervision. For seven years, Ms. Brooks supervised the multi-disciplinary team for children's services in San Mateo County. She was also the founder and executive director of the San Mateo Perinatal Council, a nonprofit community collaborative.
Helen Cahalane, *M.S.W., Ph.D.* Principal Investigator Child Welfare Education & Research Program Clinical Associate Professor University of Pittsburgh 2328 Cathedral of Learning Phone: 412-624-6386 hcupgh@pitt.edu	Helen Cahalane is the principal investigator of the Child Welfare Education and Research Programs, which include two federal IV-E professional education components and a statewide training program devoted to workforce development, practice improvement, and organizational effectiveness in

public child welfare. She is responsible for oversight of the evaluation and research division of the child welfare programs and management of the IV-E educational stipend program for public child welfare employees.

Soledad Caldera-Gammage, *M.S.W.*
Curriculum & Evaluation Specialist
Central CA Child Welfare Training Academy
California State University, Fresno
2743 E. Shaw Ave., Ste. #121
Fresno, CA 93710
Phone: 707-647-1655
Fax: 559-246-5581
scaldera@csufresno.edu

Soledad Caldera-Gammage coordinates and implements the evaluation efforts for the Central California Training Academy. The Academy's evaluation projects include the training evaluation of Child Welfare Social Worker and Supervisor Core Training, evaluation of the Field-Based Training Program, and trainer evaluations.

Sherrill Clark, *Ph.D., ACSW, LCSW*
Research Specialist
CalSWEC
School of Social Welfare
University of California, Berkeley
Marchant Bldg., Ste. 420
6701 San Pablo Ave.
Berkeley, CA 94720-7420
Phone: 510-642-4480
Fax: 510-642-8573
sjclark@berkeley.edu

Sherrill Clark, formerly curriculum specialist and executive director at CalSWEC, has taught policy, practice, and research methods to first- and second-year Master's students in the School of Social Welfare at the University of California, Berkeley. Her primary research responsibility is the evaluation of the CalSWEC project, including workforce preparation, retention of specially educated MSW child welfare workers, and child welfare education.

James Coloma, *M.S.W.*
Training & Evaluation Specialist
Public Child Welfare Training Academy
Academy for Professional Excellence
San Diego State University
6505 Alvarado Rd., Ste. 107
San Diego, CA 92120
Phone: 619-594-3219
Fax: 619-594-1118
jcoloma@projects.sdsu.edu

James Coloma focuses on training evaluation for the Public Child Welfare Training Academy; Tribal STAR, a training program to increase outcomes for Tribal foster youth; Leaders in Action, an executive development training program; Project MASTER, a training program for Adult Protective Service workers; and Behavioral Health Education & Training Academy.

Melissa Connelly, *M.S.W.*
Training & Curriculum Specialist CalSWEC
School of Social Welfare
University of California, Berkeley
Marchant Bldg., Ste. 420
6701 San Pablo Ave.
Berkeley, CA 94720-7420
Phone: 510-643-9067
Fax: 510-642-8573
mconnelly@berkeley.edu

Melissa Connelly's background includes 10 years of child welfare direct service as a social worker and supervisor. Since she joined CalSWEC in 2008, she has worked on the California Core Curricula and projects related to the California Child and Family Services Review.

John B. Cullen
Executive Director
CalSWEC
School of Social Welfare
University of California, Berkeley
Marchant Bldg., Ste. 420
6701 San Pablo Ave.
Berkeley, CA 94720-7420
Phone: 510-643-7904
Fax: 510-642-8573
johnbcullen@berkeley.edu

John B. Cullen retired as County Administrative Officer for Contra Costa County in 2008 where he was responsible for overseeing a $2 billion budget and 10,000 employees providing social services, health, law enforcement, public works, land use, and general government services. Prior to that, he served for several years as director of Employment and Human

Services and County Welfare director for Contra Costa County. He has also provided many years of service in Merced County as Human Services Agency director and as program manager for Social Services, Benefits Payments, and the Child Welfare Services Bureau. Mr. Cullen started his career with Merced County as a social worker following completion of his MSW at California State University, Fresno. Among the many professional organizations of which he is a member, he has served with distinction as past president of the CalSWEC Board of Directors and the Bay Area Social Services Consortium (fostered by UC Berkeley's School of Social Welfare's Center for Non-Profit Management). Mr. Cullen has served as past president of the National Association of County Human Service Administrators, the County Welfare Directors Association of California, and First Five of Contra Costa County.

Dale Curry, *Ph.D., LSW, CYC-P*
Assistant Professor
Kent State University
School of Family & Consumer Studies

Dale Curry recently served as the principal investigator of a validation study of the North American Certification Project. He has previously been involved with several statewide training

Nixson Hall
P.O. Box 5190
Kent, OH 44242
Phone: 330-672-2998
Fax: 330-672-2194
dcurry@kent.edu

evaluation projects in Ohio and Pennsylvania and a variety of evaluation studies at the Northeast Ohio Regional Training Center. He is on the Board of the National Staff Development and Training Association (NSDTA) and is editor of NSDTA's journal *Development in Human Services*.

Jann Donnenwirth, *M.S.*
Program Director
Center for Human Services
UC Davis Extension
University of California
1632 Da Vinci Ct.
Davis, CA 95618-4860
Phone: 530-757-8643
Fax: 530-754-5104
jdonnenwirth@ucde.ucdavis.edu

Todd M. Franke, *M.S.W., Ph.D.*
Professor
Department of Social Work
University California, Los Angeles
1100 Glendon Ave. Ste. 850
Los Angeles, CA 90024
Phone: 310-794-0999
Fax: 310-362-8673
tfranke@ucla.edu

Todd Franke has numerous years of experience in research, teaching, and evaluation in education and child welfare. He has considerable experience in designing and conducting cross-sectional and longitudinal research and evaluation in the areas of child abuse and adolescent violence. He is currently the principal investigator for the evaluation of Partnership for Families, a community- based child abuse and neglect prevention initiative in Los Angeles County. He is

also working with the Department of Children and Family Services on the redesign of the Core Academy Assessment and on evaluating the implementation and impact of training numerous modules including strength-based family-centered case planning and concurrent planning.

Frankie Freitas
Regional Training Coordinator
Central CA Child Welfare Training Academy
California State University, Fresno
2743 E. Shaw, Ste. 121
Fresno, CA 93740
Phone: 559-294-9767
Fax: 559-292-7371
frankief@csufresno.edu

Carrie Gibson, *B.A.*
Research & Evaluation Analyst
Public Child Welfare Training Academy
Academy for Professional Excellence
6505 Alvarado Rd., Ste. 107
San Diego, CA 92120
Phone: 619-594-6717
Fax: 619-594-1118
cgibson@projects.sdsu.edu

Before joining the Public Child Welfare Training Academy in May 2001, Carrie Gibson performed various research duties for other non-profit agencies in San Diego. She has experience in supervising, data collection, survey design, data entry and analysis, and reporting. At the Academy, Ms. Gibson conducts research activities for Parents and Children Together (PACT), Leaders in Action (LIA), the

	Southern Area Consortium of Human Services (SACHS), and the Network for Excellence (NEHS).
Elizabeth Gilman, *M.A., J.D.* Curriculum Specialist CalSWEC School of Social Welfare University of California, Berkeley Marchant Bldg., Ste. 420 6701 San Pablo Ave. Berkeley, CA 94720-7420 Phone: 510-642-9273 Fax: 510-642-8573 egilman@berkeley.edu	Elizabeth Gilman joined CalSWEC in March 2002. Her responsibilities include curriculum development, evaluation, and planning associated with the statewide Title IV-E MSW and BASW programs. She serves as staff to the CalSWEC Board Curriculum Committee. Prior to her return to California, Ms. Gilman was an associate research scientist in the Psychology Department at Yale University, where she served for 10 years as a policy analyst and instructor in child development and social policy. Her direct service experience includes work as a CPS social worker and as a group facilitator with troubled adolescents.
Erika Gonzalez, *B.A.* Training Specialist Bay Area Academy 2201 Broadway, Suite 100 Oakland, CA 94612 Phone: 510-419-3600 Fax: 707-643-3482	Erika Gonzalez is the training specialist for Alameda and San Mateo counties. She received a Bachelor's degree in psychology with a minor in dance from San Francisco State University. Her goal is to attain an MSW and continue working with the public. Ms. Gonzalez is also a volunteer with CASA (Court Appointed Special Advocates)

	and has a family-owned event planning business.
Michelle I. Graef, *Ph.D.* Associate Director Midwest Child Welfare Implementation Center Research Associate Professor Center on Children, Families & Law University of Nebraska—Lincoln 206 S. 13th St., Ste. 1000 Lincoln, NE 68588-0227 Phone: 402-472-3741 Fax: 402-472-8412 mgraef1@unl.edu	Under a contract with the Nebraska Health and Human Services, the Center on Children, Families & Law designs and implements all the training evaluation procedures for the pre-service training for child protection and juvenile parole workers in Nebraska. Dr. Michelle Graef's work includes research and consultation with human services agencies in Nebraska and other states on staff recruitment, selection, and retention issues. Most recently she is the co-principal investigator on a new cooperative agreement with the Children's Bureau for the Midwest Child Welfare Implementation Center. The Center's primary role is to support sustainable systems change to improve the quality and effectiveness of state and tribal child welfare services in a 10-state region.
Norma Harris, *M.S.W., Ph.D.* Director, Social Research Institute Research Professor/Lecturer College of Social Work University of Utah 395 S. 1500 E, Rm. 111 Salt Lake City, UT 84112	Norma Harris is currently responsible for the management of grants at SRI that exceed $8 million. Former director of community services in Montana, Dr. Harris has extensive experience in child welfare services. She is experienced in

Phone: 801-581-3822
Fax: 801-585-6865
norma.harris@socwk.utah.edu

evaluation of child protective and child welfare programs and was a member of the committee that assisted in the development of the U.S. Department of Health and Human Services' *Resource Guide: Rethinking Child Welfare Practice Under the Adoption and Safe Families Act of 1997.* Dr. Harris is a contributing author of the first edition and the principal author of the revised *Guidelines for a Model System of Protective Services for Abused and Neglected Children and Their Families.*

Darlene (Modesto) Hill, *M.S.W.*
Research and Evaluation Specialist
Bay Area Academy
2201 Broadway, Suite 100
Oakland, CA 94612
Phone: 510-419-3600
Fax: 510-663-5532
darlenemodesto@yahoo.com

Darlene (Modesto) Hill earned her MSW with a concentration in Administration & Planning from San Francisco State University. Ms. Hill attended Point Loma Nazarene University in San Diego for her undergraduate degree. She is a newlywed and enjoys spending her free time with her amazing husband.

Henry Ilian, *D.S.W.*
Training Evaluator
James Satterwhite Academy for Child Welfare Training
ACS Children's Center492 First Ave., 5th Floor
New York, NY 10016
Phone: 646-935-1410
Fax: 646-935-7282
henry.ilian@dfa.state.ny.us

Henry Ilian has been involved in research and evaluation at the New York City Administration for Children's Services (ACS), James Satterwhite Academy for Child Welfare Training since 1987. He has researched child protective worker turnover; conducted evaluations of training effectiveness and

training needs assessments; and has developed tests for a variety of courses offered by the Satterwhite Academy. Currently he is working on developing an evaluation package for a new initiative to train caseworkers who will staff an expanded preventive services component within the agency.

Susan E. Jacquet, *Ph.D.*
Research Specialist
CalSWEC
School of Social Welfare
University of California, Berkeley
Marchant Bldg., Ste. 420
6701 San Pablo
Berkeley, CA 94720-7420
Phone: 510-643-9846
510-642-8573
sjacquet@berkeley.edu

Susan E. Jacquet works with Dr. Sherrill Clark on CalSWEC's research component, including ongoing surveys of California's MSW students and CalSWEC Title IV-E MSW graduates who have completed payback work in child welfare, and the development of new research initiatives on outcomes for child welfare and the efficacy of the Title IV-E program. Dr. Jacquet manages the CalSWEC Student Information System (CSIS), the database of all CalSWEC Title IV-E students and graduates. She is also responsible for coordinating CalSWEC's funded research process from RFP through review of proposals.

Phyllis Jeroslow, *M.F.A., MFT*
Training & Curriculum Specialist
CalSWEC
School of Social Welfare
University of California,

Phyllis Jeroslow works on the development and revision of California's standardized Common Core Curriculum. She also produces the annual Symposium on Fairness & Equity Issues in Child Welfare

Berkeley
Marchant Bldg., Ste. 420
6701 San Pablo Ave.
Berkeley, CA 94720-7420
Phone: 510-643-5440
Fax: 510-642-8573
pjero@berkeley.edu

Training & Education. She is currently a doctoral student in the School of Social Welfare at UC Berkeley, studying the relationship of disparities across the public health, mental health, education, and child welfare systems.

Barrett Johnson, *M.S.W., LCSW*
Director, Child Welfare In-Service Training Project
CalSWEC
School of Social Welfare
University of California, Berkeley
Marchant Bldg., Ste. 420
6701 San Pablo Ave.
Berkeley, CA 94720-7420
Phone: 510-643-5484
Fax: 510-642-8573
barrettj@berkeley.edu

Barrett Johnson oversees CalSWEC's coordination of statewide training efforts, including development and implementation of a Common Core Curriculum and strategic planning for a statewide training evaluation system. He is currently involved in planning the training efforts for California's Outcomes and Accountability System. He has worked for many years with urban children and families, with an emphasis on intervention in cases of child sexual abuse.

Michael Lawler, *M.S.W., Ph.D.*
Director
Center for Human Services
University of Extension, UC Davis
1632 Da Vinci Ct.
Davis, CA 95616
Phone: 530-757-8643
Fax: 530-754-5104
mlawler@ucde.ucdavis.edu

Michael J. Lawler is the director of the Center for Human Services and co-director of the Center for Public Policy research at UC Davis. With nearly 30 years of experience, he holds a Ph.D. in Human Development from UC Davis and M.S.W. from UC Berkeley and is a licensed clinical social worker in California.

Chris Lee, *M.S.W.*
Graduate Student Researcher
CalSWEC
School of Social Welfare
University of California, Berkeley
Marchant Bldg., Ste. 420
6701 San Pablo Ave.
Berkeley, CA 94720-7420
Phone: 510-643-0223
Fax: 510-642-8573
clee07@berkeley.edu

Chris Lee received her M.S.W. from the University at Albany, State University of New York and is currently pursuing a Ph.D. in social welfare at the University of California, Berkeley. Her research interests include: vulnerable youth, especially those with fragile family connections; the integration of evidence into policy and practice "on the ground"; and holistic youth-serving systems of care.

Chris Mathias, *M.S.W.*
Director, Title IV-E Programs
CalSWEC
School of Social Welfare
University of California, Berkeley
Marchant Bldg., Ste. 420
6701 San Pablo Ave.
Berkeley, CA 94720-7420
Phone: 510-642-7490
Fax: 510-642-8573
cmathias@berkeley.edu

Chris Mathias joined CalSWEC in March 2000 as head of the Regional Training Academy Coordination Project. As the current director of the Title IV-E Programs, she leads the development and evaluation of the Title IV-E Stipend Program for public child welfare. For 14 years prior to joining CalSWEC, Ms. Mathias worked primarily in the private non-profit sector with children in out-of-home care. During that period, she did direct practice with children and families in the foster care system and developed curriculum, training, and quality assurance methods for direct care workers, clinicians, and administrators.

Christine Mattos
Manager
Resources Development &
Training Support Bureau
California Department of
Social Services
744 P. St., M.S. 8-11-83
Sacramento, CA 95814
Phone: 916-651-6076
Fax: 916-651-6239
christine.mattos@dss.ca.gov

Julie McCrae, *M.S.W., Ph.D.*
Research Assistant Professor
Child Welfare Education &
Research Program
School of Social Work
University of Pittsburgh
2332 Cathedral of Learning
Pittsburgh, PA 15260
Phone: 412-383-5168
Fax: 412-624-1159
jsmccrae@pitt.edu

Julie McCrae's work focuses on early intervention around children's emotional and developmental needs following contact with child welfare, and transfer of learning concerning child welfare-mental health interventions and strengths-based practice.

Michael Nunno, *D.S.W.*
Senior Extension Associate
Family Life Development
Center
College of Human Ecology
Cornell University
Beebe Hall
Ithaca, NY 14853-4401
Phone: 607-254-5127
Fax: 607-255-4837
man2@cornell.edu

Michael Nunno is the principal investigator of the Residential Child Care Project, coordinating and supervising the evaluation of the project's international training efforts in the prevention and de-escalation of aggression and violence in children's residential care programs. He also evaluates the impact of the Children and Residential Experience: Creating Conditions for Change curriculum for residential child

care personnel on participants' learning, and organizational congruence and climate.

Cindy Parry, *Ph.D.*
Consultant
C. F. Parry Associates
520 Monroe Ave.
Helena, MT 59601
Phone: 406-449-3263
Fax: 406-449-3263
cfparry@msn.com

Cindy Parry specializes in program evaluation in human services. She holds a Ph.D. in Educational Psychology with a strong emphasis on curriculum design and assessment. She has had over 25 years of experience in evaluation in child welfare, juvenile justice, and education. Dr. Parry is a former director of Program Analysis and Research at the American Humane Association. She is currently involved in evaluations of training effectiveness in Colorado, California, Nevada, and Washington, and has been a co-facilitator for a strategic planning process for training evaluation in California.

Leslie Rozeff, *M.S.S.W., LCSW-C*
Director, Child Welfare Academy
School of Social Work
University of Maryland
525 W. Redwood St.
Baltimore, MD 21201
Phone: 410-706-2059
Fax: 410-706-2423
lrozeff@ssw.umaryland.edu

In addition to overseeing specialized traininig development and delivery for Maryland's child welfare workforce, Leslie Rozeff has worked with local and state jurisdictions in areas such as reducing reliance on residential care for children in the foster care system, improving organizational effectiveness in systems serving children and families, and developing effective strategies for

translating child welfare policy into practice.

Lazarina Topuzova, *Ph.D*.
Assistant Professor
Department of Organizational Leadership
Gonzaga University
502 E. Boone Ave., MSC 2616
Spokane, WA 99258-2616
Phone: 509-313-3620
Fax: 509-313-3566
topuzova@gonzaga.edu

Lazarina Topuzova holds a Ph.D. in social work from the University of Utah; an M.A. in American Literature from Sofia University, Bulgaria; and a Master's in Peace and Conflict Resolution from the European Peace University, Austria. She spent time in Kosovo working on educational program development, capacity building, and conflict resolution skills training for youth from diverse ethnic backgrounds, as well as coexistence and art workshops. Her research interests and expertise are in working with youth, families, refugees, internationally displaced persons, as well as services targeted toward those populations. Dr. Topuzova is currently engaged in competency-based and online education and evaluation, as well as developing leadership capacity in human services professionals.

Jennifer Tucker-Tatlow, *M.S.W.*
Director
Academy for Professional Excellence
School of Social Welfare
San Diego State University

Jennifer Tucker-Tatlow has worked with San Diego County's Health and Human Services Agency on service integration efforts, and has managed program operations for adult and adolescent alcohol,

6505 Alvarado Rd., Suite 107
San Diego, CA 92120
Phone: 619-594-3917
Fax: 619-594-1118
jttatlow@projects.sdsu.edu

tobacco, and other drug prevention services in health care facilities. For the past 10 years, she has been with the Academy for Professional Excellence, and currently has oversight responsibility for all its programs and program support services. Ms. Tucker-Tatlow works closely with the directors of the Southern California region (via SACHS) to facilitate the exchange of ideas and information and strategic planning regarding issues pertaining to public human services.

Leslie W. Zeitler, *M.S.W., LCSW*
Training & Evaluation Specialist
CalSWEC
School of Social Welfare
University of California, Berkeley
Marchant Bldg., Ste. 420
6701 San Pablo Ave.
Berkeley, CA 94720-7420
Phone: 510-643-6400
Fax: 510-642-8573
lzeitler@berkeley.edu

Leslie Zeitler coordinates and implements CalSWEC's statewide training and evaluation efforts, and provides technical assistance to California's Regional Training Academies and counties. Prior to joining CalSWEC, Ms. Zeitler for six years provided direct social work services to low-income children and families in the San Francisco Bay Area. She is a graduate of the Coro Center for Civic Leadership's FellowshipProgram in Public Affairs, and has also taughtfamily assessment and risk management skills to Head Start direct services staff at its annual conferences.