Abstract Title Page
Not included in page count.

Title: Self-reporting across two reading interventions: Can it be the sole measure of adherence?

Authors and Affiliations: Lisa Hall Foster, Harvard University
James Kim, Harvard University

Abstract Body

Limit 4 pages single-spaced.

*****The majority of this report contains data and findings from the "W*hat works in gifted education*?" study conducted at the National Research Center on the Gifted and Talented at the University of Virginia. Preliminary findings of self-reporting from the *READS for Summer Learning* study being conducted at the Harvard Graduate School of Education will be introduced.

Background / Context:

Description of prior research and its intellectual context.

The need for fidelity of implementation (FOI) measurements and assessments during efficacy and effectiveness studies is becoming evident in the requirements of many educational research funding organizations. Fidelity is defined as the degree to which teachers and other program providers implement programs *as intended by the program developers* (Durlak & DuPre, 2008). The assessment of fidelity of implementation during efficacy and effectiveness studies is one way of increasing the validity and usefulness of research-based interventions that claim to be effective. The conditions under which programs or interventions are developed and tested are typically not be the same as those under which it is disseminated and implemented.

Although it is widely known that there are variations in educational programs and interventions when implemented (Blakely et al., 1987; Dusenbury, Brannigan, Falco, & Hansen, 2003), there is a point in variation that, when passed, could compromise the integrity of the program or intervention. Hence, fidelity measurements and assessments as part of research models facilitate the validity and proven effectiveness and help account for possible variance (Sanchez et al., 2007). A variety of methods and combination of methods are often used for measuring FOI during efficacy or effectiveness studies (Dusenbury et al., 2003). These methods can be direct or indirect. Direct assessments are typically observations. Indirect assessments include self-reports, interviews, and examination of permanent products.

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

The purpose of this study was to determine teachers' levels of FOI when implementing a reading curricular intervention and to investigate whether or not teacher self-report of implementation could be utilized as a reliable alternative to direct observations to measure FOI. Finding an alternative method for assessing fidelity would be beneficial to project funding agencies, researchers, and school districts (Fletcher et al., 2010) as direct observations or video recording are quite expensive.

Setting:

Description of the research location.

Data in this study were collected as part of a) the "What works in gifted education: Excellence and equity in education of gifted students" (WWIGE) study funded by a Javits grant and conducted by the National Research Center on the Gifted and Talented (NRC/GT) at the University of Virginia and b) the READS for Summer Learning study funded by an i3 grant conducted at Harvard University. In this five-year WWIGE study, teachers implemented a reading/language arts curriculum designed for 3rd grade gifted students using the *CLEAR* Curriculum Model (*CCM*). In the five-year READS study, teacher taught six scaffolded reading lessons to prepare students for the summer reading intervention.

Population / Participants / Subjects:
Description of the participants in the study: who, how many, key features, or characteristics.

Participants in were a) 30 third grade gifted teachers who voluntarily participated in the WWIGE study during the 2010-2011 school year and b) 29 teachers from three school districts who participated in the READS study during the 2011-2012 school year. All teachers were randomly assigned to a condition.

Intervention / Program / Practice:
Description of the intervention, program, or practice, including details of administration and duration.

Participants in WWIGE implemented two language arts/reading units, one on poetry and figurative language, the other on expository non-fiction text and research skills. The two units are a manifestation of the *CCM*, developed by researchers at the NRC/GT. The *CCM* is the assimilation and integration of three well-known models of instruction: the Differentiation of Instruction Model (Tomlinson, 1995; 1999), Depth and Complexity Model (Kaplan, 2005), and the Schoolwide Enrichment Model (Renzulli & Reis, 1985; 2000). The units emphasize conceptual thinking, real-world interdisciplinary inquiry, and problem-solving skills for encouraging student exploration with real-world utility. The poetry unit has 18 lessons and the research unit has 15 lessons. Each unit takes around 20 hours to implement. Teachers were sent the curriculum manuals and all supplies prior to implementation. They received professional development via webinars and constant support throughout the intervention. Teachers were at leisure to teach the units when scheduling allowed, as long as they taught both units during the one school year.

Participants in READS taught six highly scripted and scaffolded reading lessons using Story Impressions and Main Ideas to prepare students for a summer reading intervention designed to reduce the effects of summer reading loss. Teachers modeled a before, during, and after comprehension routine using narrative and informational texts. In the narrative comprehension routine, teachers modeled how to (a) make a story guess using content drawn from text, (b) answer questions based on narrative text structures (e.g., characters, setting, plot), and (c) check similarities and differences between the story guess and actual story. In the expository comprehension routine, teachers modeled how to (a) answer questions about a topic using background knowledge about the text, (b) discuss questions based on expository text structures (e.g., description, sequence, problem-solution), and (c) check new content learned about the topic. The lessons required six consecutive days of implementation with two hours of in-person training conducted prior to implementation. Each lesson lasted approximately 45-60 minutes.

Research Design:
Description of the research design.

This study incorporated a mixed methods design (Creswell & Plano Clark, 2007; Tashakkori & Teddlie, 2003). To this end, data were collected via instruments and procedures that supported both qualitative and qualitative analysis of the extent to which self-reported fidelity correlated with directly observed fidelity and intended curriculum.

Data Collection and Analysis:
Description of the methods for collecting and analyzing data.

Qualitative sources of data included direct observations of teachers and follow-up interviews. Observations were recorded on a semi-structured protocol designed to capture

adherence to critical components. Following the observations, teachers self-reported adherence on a condensed self-report form reflective of the observation guide. Researchers then interviewed participants using a semi-structured interview protocol to gain an understanding of the fidelity and rationale for how the teacher implemented the lesson. The interviews were recorded and transcribed for later analysis to determine if a modification or omission had been made in accord with the *CLEAR* Curriculum's framework. Quantitative data sources included a scoring rubric for observed and self-reported adherence to the curriculum.

After identifying all of the critical components and design principles for each lesson, guiding principles for determination of acceptability of adaptations were developed to score adherence percentage to the critical components on observed and self-reported protocols. Inter-rater reliability was calculated for both measures. A Spearman's rho correlation was calculated between observed and teacher self-reported adherence scores because the data did not have normal distribution. The skewness and kurtosis values were not within acceptable limits (> 1.0) and highly negatively skewed.

Findings / Results:

Description of the main findings with specific details.

Overall, teachers implemented the lessons with moderate fidelity; however, with a mean of 83.27 (*SD* = 20.39), they were very close to having implemented the observed lessons with a high fidelity level according to the criteria that had been established (high fidelity ≥ 85% adherence) (see Table 1). The overall mean for self-reported adherence was 90.17 (*SD* = 13.46) (see Table 2). Large standard deviations were due possibly to the high variability, as there were no outliers. There was a strong positive correlation between the two variables, $r_s(28) = 0.820$, n = 30, $p < 0.000$. In checking for the explained variance, $r_p^2 = .6724$, meaning 67 percent of the variance in self-reported adherence was explained by observed adherence. The high correlation indicated that the teachers were able to report their implementation of the lesson at the same rank of fidelity as the researcher. As a result, self-report can be considered an acceptable alternative to direct observations in order to determine FOI levels.

Comparing the self-reported and observed adherence scores by fidelity group, the likelihood of a teacher reporting the same way as did the observer fell on a continuum of "very likely" to "not very likely." The higher the observed adherence, the more likely the teacher was to report implementation in a manner consistent with the observer; the lower the fidelity, the less likely.

The teachers in the high fidelity group averaged a 2.5 points difference from my scoring. Thirteen of the 18 teachers matched my score exactly, while three were seven to 15 points higher and two were six to seven points lower (see Table 3). Thirteen of the teachers reported scores within five percent of my rating. Three teachers rated their adherence within a 10% differential. Two of the high fidelity teachers reported scores difference of greater than 10%.

Teachers in the moderate fidelity group averaged a 12.8 points difference, with four of the five teachers reporting higher scores than observed data. One teacher reported her score with less than 5% differential and one with 10% differential. Three of the five had over 10% differential in reporting.

Finally, the lower fidelity group averaged a 20.6 points difference. Six of the seven teachers over-reported, meaning they reported higher percentages than observed data. One teacher reported in a manner that resulted in the same percentage score and one reported within a five percent differential. Five of the seven had greater than 10% differential.

Conclusions:

Description of conclusions, recommendations, and limitations based on findings.

Consistent with extant literature (Lillehoj, Giffin, & Spoth, 2004; Smithson & Pritz, 1994; Wubbels, Brekelmons, & Hoomayers, 1992), study findings demonstrated a statistically significant relationship between self-reported and observed implementation ratings. These findings are significant in light of oppositional reports on the limitations of self-report (Ross et al., 2003).

One factor to consider is the timeframe given to the teacher to report the implementation. On the day of the observation, the teacher completed the Teacher Log, later used to compute the self-reported adherence fidelity score, immediately following the observation and prior to the interview. For the study, all teachers were to record their implementation as they implemented each lesson and return the log at the end of the unit. As such, some teachers may have felt rushed to fill out the log while others may have filled out the log more effectively because the researcher was present.

Additionally, comparing the self-reported and observed adherence scores by fidelity group, the likelihood of the teacher reporting the same thing as did the observer fell on a continuum of "very likely" to "not very likely." The higher the observed AFS, the more likely the teacher was to report implementation in a manner consistent with the observer; the lower the fidelity, the less likely. Several plausible explanations could account for varying accuracy levels in self-reporting.

First, teachers with high fidelity were able to make modifications that were in accord with the *CCM*, thus indicating that they clearly recognized the components and the purpose behind them and inasmuch, were able to recognize where they made changes, whereas teachers with low fidelity were not. Second, teachers with low fidelity did not. This is consistent with the findings of Ross et al. (2003) who reported that teachers and developers might not have the same understanding. Furthermore, as teachers did not receive explicit training in implementation or the theory behind the *CCM*, they may have been unaware of the fact they were changing the curriculum (Hansen & McNeal, 1999) or inhibited by their lack of knowledge or skill, school constraints, or beliefs about students, which may influence teacher efficacy and thus, the self-report (Ross et al., 2003). As such, Greenberg et al. (2005) recommend in-depth training in the program model theory during the pre-implementation phase. Finally, self-report may not have been accurate at lower fidelity levels because of distortion due to guilt, denial, ego enhancement, or social desirability (Resnicow et al., 2998; Ross et al., 2003).

Regardless of their observed fidelity level, over half of the teachers (16 of the 30) were able to report their scores within a five percent or less differential from the observed fidelity score. Future research is needed to understand the continuum of self-reporting.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Blakely, C., Mayer, J., Gottschalk, R., Schmitt, N., Davidson, W., Roitman, D., et al. (1987). The fidelity-adaptation debate: Implications for the implementation of public sector social programs. American Journal of Community Psychology, 15(3), 253-268. doi:10.1007/BF00922697

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research.* Thousand Oaks, CA: Sage.

Durlak, J., & DuPre, E. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*(3-4), 327-350. doi:10.1007/s10464-008-9165-0

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*(2), 237-256. doi:10.1093/her/18.2.237

Fletcher, S., Zimmerman, S., Preisser, J., Mitchell, M., Gould, E., Beeber, A., Reed, P. (2010). Implementation fidelity of a standardized dementia care training program. *Alzheimer's Care Today, 11*(1), 51-60. doi:10.1097/ACQ.0b013e3181cd19fc

Greenberg, M. T., Domitrovich, C. E., Graczyk, P. A., & Zins, J. E. (2005). *The study of implementation in school-based preventative interventions: Theory, research, and practice. (Volume 3).* Rockville, MD: Center for Mental Health Services, Substance Abuse and Mental Health Services Administration.

Hansen, W., & McNeal, R. (1999). Drug education practice: Results of an observational study. *Health Education Research, 14*(1), 85-97. doi:10.1093/her/14.1.85

Kaplan, S. (2005). Layering differential curricula for gifted and talented. In F. Karnes & S. Bean (Eds.), *Methods and materials for teaching gifted students* (pp. 107-132). Waco, TX: Prufrock Press.

Lillehoj, C., Griffin, K., & Spoth, R. (2004). Program provider and observer ratings of school-based preventive intervention implementation: Agreement and relation to youth outcomes. *Health Education and Behavior, 31*(2), 242-257. doi:10.1177/1090198103260514

Renzulli, J. S., & Reis, S. (1985). *The schoolwide enrichment model: A comprehensive plan for educational excellence.* Mansfield Center, CT: Creative Learning Press.

Renzulli, J. S., & Reis, S. (2000). The schoolwide enrichment model. In K. Heller, F. Mönchs, R. Sternberg, and R. Subotnik (Eds.), *The International handbook of giftedness and talent* (2nd ed.) (pp. 367-382). Oxford, England: Elsevier Science, Ltd.

Resnicow, K., Davis, M., Smith, M., Lazarus-Yaroch, A., Baranowski, T., & Baranowski, J. (1998). How best to measure implementation of school health curricula: A comparison of three measures. *Health Education Research, 13*(2), 239-250. doi:10.1093/her/13.2.239

Ross, J., McDougall, D., Hogaboam-Gray, A., & LeSage, A. (2003). A survey measuring elementary teachers' implementation of standards-based mathematics teaching. *Journal for Research in Mathematics, 34*(4), 344-363. Retrieved from http://www.nctm.org/publications/jrme.aspx

Sanchez, V., Steckler, A., Nitirat, P., Hallfors, H., Cho, H., & Brodish, P. (2007). Fidelity of implementation in a treatment effectiveness trial of Reconnecting Youth. *Health Education Research, 22*(1), 95-107. doi:10.1093/her/cy1052

Smithson, J., & Pritz, A. (1994). Measuring classroom practice: Lessons learned from the efforts to describe the enacted curriculum. *The Reform Upclose Study.* Madison, WI: Consortium for Policy Research in Education. Retrieved from http://www.eric.ed.gov

Tashakkori, A., & Teddlie, C. (2003). *Handbook of mixed methods in social and behavioral research.* Thousand Oaks, CA: Sage.

Tomlinson, C. A. (1995). *How to differentiate instruction in mixed-ability classrooms*. Alexandria, VA: Association for Supervision and Curriculum Development.

Tomlinson, C. A. (1999). *The differentiated classroom: Responding to the needs of all learners*. Alexandria, VA: Association for Supervision and Curriculum Development.

Wubblels, T., Brekelmons, M., & Hoomayers, H. (1992). Do teacher ideals distort the self-report of their interpersonal behaviors? Teaching and Teacher Education, 8, 47-58. doi:10.1016/j.ijer.2006.03.003

Appendix B. Tables and Figures

Not included in page count.

Table 1

Summary of Observed AFS

Observed FOI Score			With Outlier		Without Outlier	
		N	Mean	Std. Deviation	Mean	Std. Deviation
Overall	Overall	30	83.27	20.39	85.28	17.47
Teacher Status	Returning Teachers	12	81.58	23.92	86.73	16.74
	New Teachers	18	84.39	18.32	84.39	18.32
Program Type	Self-contained Program	9	92.11	11.70	92.11	11.70
	Pull-out Program	21	79.48	22.31	82.20	18.97
Endorsement	Endorsed	10	77.5	23.91	83.13	16.13
	Not endorsed	20	86.15	18.37	86.15	18.37
Certification	Certified	10	83.00	20.00	83.00	20.00
	Not certified	20	83.00	21.00	86.32	16.70
Masters Degree	Degree	15	86.00	22.00	90.86	14.48
	No Degree	15	80.00	19.00	80.00	19.00
Gifted Masters	Gifted Masters	5	98.00	4.47	98.00	4.47
	No Gifted Masters	25	80.32	21.09	82.63	18.04
Fidelity Rating	High	18	97	6	97	6
	Moderate	5	78	4	78	4
	Low	7	52	15	57	10

Table 2

Summary of Self-reported AFS

	Self-reported FOI Score	N	Mean	Std. Deviation
Overall	Overall	30	90.17	13.46
Teacher Status	Returning Teachers	12	89.67	16.04
	New Teachers	18	90.5	11.94
Program Type	Self-contained Program	9	97.89	3.26
	Pull-out Program	21	86.86	14.84
Endorsement	Endorsed	10	88.1	16.27
	Not endorsed	20	91.20	12.16
Certification	Certified	10	88.5	14.86
	Not certified	20	91.00	13.03
Masters Degree	Degree	15	90.07	14.84
	No Degree	15	90.27	12.46
Gifted Masters	Gifted Masters	5	98.00	4.47
	No Gifted Masters	25	88.60	14.15
Fidelity Rating	High	18	97.85	4.66
	Moderate	5	86.4	11.29
	Low	7	73	13.96

Table 3

Comparison of Observed and Self-reported AFS

Observed Fidelity Level	Mean Point Difference Between Observer and Teacher Scoring	Standard Deviation	Teacher Self-reported AFS Compared to Observer AFS		
			Higher	Lower	Same
High (*N*=18)	2.5	4.51	3	2	13
Moderate (*N*=5)	12.8	10.38	4	1	0
Low (*N*=7)	20.6	14.09	6	0	1