A Longitudinal Study of the Practice Fidelity of a Site-based School Reform

Alan Bain
Charles Sturt University

Abstract

In this study, four years of formal observations of classroom teaching practice were employed to ascertain the practice fidelity of a site-based school reform in a secondary school setting. Those observations were then used as a criterion variable in an examination of differences in the perspectives of administrators, teachers and teaching peers about the reform's implementation. The results showed sustained levels of practice fidelity and statistically significant differences in the ratings of administrators, teachers and peers although those differences reduced overall as the reform progressed. The perspective of administrators was the best predictor of classroom practice in the first three years of the reform although less so in the latter year when teacher and peer responses became better predictors. The implications of the findings are discussed as they relate to the practice fidelity and evaluation of site-based school reform.

International efforts to improve and reform schools have generated an extensive literature on the history, process and efficacy of school effectiveness and change (Berends, Bodilly, & Nataraj Kirby, 2002; Desimone, 2002; Dimmock, 2000; Elmore, 1996; Huberman & Miles, 1984; Sarason, 1982, 1996; Reynolds, Creemers, Stringfield, Teddlie, & Schaffer, 2002; Tirozzi & Uro, 1997; Tyack & Cuban, 1995). Despite the many accounts of school reform efforts, little objective evidence exists about the fidelity with which those reforms are implemented and their classroom impact. A recent review of the fidelity of implementation of K-12 intervention by O'Donnell (2008) identified only 23 studies that provided evidence of the fidelity of implementation. All of these studies were conducted at the primary level, none in secondary schools. Seven pertained to whole school reform efforts and only five included statistical analysis of implementation fidelity findings beyond a descriptive level.

This limited scrutiny of implementation fidelity is common to descriptions of small and large-scale efforts to evaluate school reform initiatives, including national, state and provincial evaluations of reform efforts in the UK, US, Canada and Australia. They include the Playing for Success and Excellence in Cities Programs (UK), Comprehensive

School Reform (US), the Manitoba School Improvement Program (Canada), the Getting it Right Literacy and Numeracy Strategy in Western Australia and The Middle Years Reform Program in Victoria, Australia (Aladjem & Borman, 2006; Berends, Nataraj Kirby, Naftel, & McKelvey, 2001; Doremus, 1981; Earl, Torrance, & Sutherland, 2003; Eastabrook, Fullan, & Bliss, 1977; Elsworth, Kleinhenz, & Beavis, 2004; Fink, 2000; Ridley & Kendall, 2005; Sharp, Eames, Sanders, & Tomlinson, 2005; Sharp, Schagen, & Scott, 2004). As indicated by Gertler, Patrinos and Rubio-Codina (2007) in their guide for the evaluation of international site-based reforms, a need exists to gather "detailed micro-level data over an appropriate time frame that measures the response of individual agents (students, teachers, schools) to the proposed program" (p. 2).

Practice Fidelity

None of the evaluative accounts of site-based reform cited in this study, nor the studies described in the O'Donnell (2008), review established the micro-level "practice fidelity" of classroom implementation through year-over-year structured observation. Practice fidelity is defined here as the integrity with which the pedagogical approaches associated with a reform are implemented across classrooms over time. If, for example, a reform calls for the use of cooperative learning (CL), as many do, a determination of its practice fidelity is made by observing the extent to which the research-based component characteristics of CL including individual accountability, mutual interdependence and task structure (Johnson, Johnson, & Holubec, 1998; Slavin, Farnish, Livingston, Sauer, & Colton, 1994) were implemented routinely in the classroom over time.

Practice fidelity is distinguished from the contemporary focus on implementation fidelity or integrity that addresses whether an overall reform model or approach adhered to the intentions of its developers (Kurki, Boyle, & Aladjem, 2006; Mihalic, 2001; O'Donnell, 2008). This includes whether the professional development of teachers was perceived to be adequate, whether the promised level of consultant support was provided as well as the observation of macro factors associated with classroom practice. The latter may include establishing whether classroom assessment was authentic, or whether lessons were intellectually rigorous or engaging for students. These factors have been the focus of extensive prior observational study including, in an Australian context, the "Queensland School Reform Longitudinal Study" (Lingard, Ladwig, Mills, Bahr, & Chant, 2001). While macro factors are of importance in determining the fidelity with which a program of reform is implemented or in determining the existence of certain classroom approaches, they do not adequately address the veracity with which demonstrable change in practice has occurred in classroom practice over time.

The rationale for incorporating a practice fidelity approach is four fold:

First, extensive longitudinal research has shown that evidence-based teaching practice exerts a profound influence on achievement (Hattie, 2009; Hattie, 2003; Marzano, 1998). Those achievement effects are driven differentially by the presence of the components of those practices (Fraser, Walberg, Welch, & Hattie, 1987; Hattie, 2003; Johnson et al., 1998; Johnson, Johnson, & Stanne, 2000; Slavin, 1990; Prince, 2004). Slavin (1990) established the influence of goal setting and individual accountability in cooperative learning while Hunter (2004) affirmed the importance of guided and independent practice in mastery teaching. From these perspectives, the benefits of school reforms are most likely to accrue when the research-based pedagogical approaches, included by developers to drive achievement effects, are implemented with a high degree of integrity.

Second, while successful reforms involve many aspects of school operation including professional development, administrative support, and school organization, changes in these areas are ultimately designed to influence the way teachers teach and students learn. It is reasonable to conclude that one important way to determine the effects of school reorganization, professional development, and administrative efforts to improve classroom practice, is in the rigor with which those core teaching practices of a reform are implemented day-to-day in classrooms over time.

Third, and as noted previously, the literature on site-based school reform shows highly variable implementation fidelity, and modest effects on student learning (Berends et al., 2001; Borman, Hewes, Overman, & Brown, 2003; Zhang, Shkolnik, & Fashola, 2005). Gertler et al. (2007) state that "after more than a quarter century of site-based management reforms around the world, there is still little conclusive evidence on the effects of these interventions" (p. 35). Borman et al. (2003) indicated that the average effect size for models included in the large scale US Comprehensive School Reform (CSR) program was a modest 0.15. This level of effect falls below the minimum threshold for the educational significance of an innovation (McCartney & Rosenthal, 2000). Modest effects are indicated even in instances where the reform has reported high levels of implementation fidelity (Borman et al., 2005). These data suggest that the focus of efforts to measure implementation may not be addressing the most critical factors that influence student learning. Further, if stronger effects do emerge from efficacy research on site-based reforms, a rigorous case must be made for the attribution of those effects to the design characteristics of the reform models that produce them. This involves determining the authenticity of both the teaching and learning experiences that occur in a reform and the way it is constructed to produce authentic learner outcomes (Cherednichenko, Hooley, Kruger, & Moore 2001). This kind of attribution is essential if reforms models are to sustain financial support for the length of time required to generate positive learning outcomes. Fullan, (2007) notes that the time required to successfully implement

reforms frequently extends beyond the duration of the funding cycles that support them. At present, indirect measures from ratings or surveys and macro-observations constitute the predominant sources of information about the fidelity of implementation of CSR (Zhang et al., 2005). A practice fidelity approach may add important criterion validity to a school reform by focusing the benchmark standard for implementation on features that are likely to realise the stated goals of all reforms to improve student learning and achievement.

Fourth, and most important, a detailed examination of the key features of classroom practice creates the possibility of equally detailed feedback and problem solving about the implementation of a reform. Issues related to the quality of feedback have been a recurring problem in accounts of site-based reforms. Berends et al. (2001) found that none of the designs included in the New American Schools site-based CSR possessed adequate mechanisms for feedback and evaluation. When the key components of pedagogical knowledge are the focus of feedback, the analysis of needs, strengths and weaknesses becomes possible at a level of detail that is more likely to influence teaching and student achievement.

The purpose of this study is to report the practice fidelity of a site-based school reform in a secondary school setting through the structured classroom observation of three pedagogical approaches central to an inclusive education reform initiative. Inclusive education is defined here as an approach that increases the responsiveness of classroom teaching to the needs of all learners. The practice fidelity data were then used as a criterion variable to examine the changing perspectives of school administrators and teachers about the implementation of the reform over a four-year period. Those perspectives were measured using self, peer and administrator questionnaires of classroom implementation of the reform by teachers. Given the lack of data on practice fidelity, the study sought to provide foundational information about the implementation of a reform at the secondary level, and the perspectives of those involved.

Specifically, the study addressed the following five research questions:
Were the teaching practices associated with the reform implemented with fidelity over the four years of study?

1. Were there differences in the perspectives of teachers, their peers and administrators about the implementation of the reform over the four years?
2. Were there differences in the relationship between the ratings of teachers, peers and administrators and the classroom observation over the course of the reform?
3. If there were differences in the perspectives of the three stakeholder groups (teachers, teaching peers and administrators) which of the groups

provided the most accurate predictions of practice fidelity over the four year study period.

4. How did the predictions of the stakeholders change over the four-year period of implementation?

Method

Participants

A total of 78 teachers, 34 females and 44 males participated in the study across conducted over a four-year period. Their teaching areas were as follows: math (11), science (10), English (16), history (11), ESL (6), fine and performing arts (6), languages (5), and instructional support (13). Of the teachers, 37 had 0-3 years of overall teaching experience, 24 had 4-10 years of experience and 17 had more than 10 years of experience. Table 1 describes the composition of the participating faculty including years of participation in the reform program for each of the four years of the study.

Year of Program	Years of Experience in Reform Program *N* (%)			
	1	2	3	4
1	48 (100)			
2	20 (37)	34 (63)		
3	8 (16)	16 (32)	26 (52)	
4	8 (17)	4 (8.5)	17 (36)	18 (38)

Table 1: Teacher Participation by Experience in the Program

Five administrators also participated, 3 male and 2 female. Each administrator had in excess of 10 years teaching experience. None of the participating teachers and administrators had specific practical knowledge of the teaching approaches described here prior to the start of the reform project undertaken by the school. All participants were provided with a six-week training program prior to teaching at the school. The program included training in the pedagogies that were the subject of the observations and in the use of the observation protocols employed in this research.

The study was conducted in a co-educational independent secondary school in the United States (grades nine through thirteen) with an enrolment of three hundred and fifty students. Two-thirds of the students board at the school and enroll from twenty-eight states and sixteen countries. The school accepts students across the ability

spectrum, approximately 25% of whom meet generally accepted classification criteria for learning disability (Mastropieri & Scruggs, 2004). The overall performance profile for students entering the school approximated that of the average US secondary school on standardised tests of achievement (Bain & Ross, 2000).

Beginning in 1992, the school engaged in a school reform based upon a comprehensive approach known as the "Self-Organizing School" (Bain, 2007). The Self-Organizing School design integrates the development of a school-wide plan, research based strategies and methods, professional development, external technical support, measurable student outcomes and a comprehensive plan for evaluation (e.g., Comprehensive School Reform Demonstration Guidelines in Desimone (2002)), embedded within a broader theoretical framework derived from a study of complex adaptive systems. The reform included the development of over 2000 hours of differentiated curriculum based on the pedagogies that are examined in this study and a suite of software tools for delivering all aspects of the approach (Bain & Parkes, 2006), a feedback and evaluation system based on student, teacher, teaching peer and administrative feedback, a human resource model based on teaching and administrative teams and a professional development model and program conducted annually for 16 years. A complete description of the design, its theoretical underpinnings, elements and research exceeds the scope of the present study and can be found in (Bain, 2007; Bain, Fallon, & Smith, 1999; Brosnan, 1996; Brown, 2000; Dimmock, 2000; McCord, 1999). This includes a comparative external evaluation of the approach that compared the performance of the Self-Organizing School design with three other site-based reforms (Weston & Brooks, 2008).

Criterion variable

The criterion variable used to determine the practice fidelity of this reform was the classroom implementation of three pedagogical approaches as determined by direct classroom observation, namely, peer assisted learning, cooperative learning and explicit teaching, these were the classroom centerpieces of the site-based reform under investigation and are widely acknowledged as the cornerstones of inclusive educational practice (Ashman & Elkins, 2004; Mastropieri & Scruggs, 2004).

Observations

Observers (department heads, and school's administrators) employed one or more of three electronic observation protocols to observe 50-55 minute class sessions in order to determine the practice fidelity of the inclusive pedagogies. The observations were undertaken as part of the school's ongoing cycle of feedback for professional growth and career progression and not for the specific purpose of this study. They represent a sample of the ongoing operation of the school's program. An event recording observation approach, described previously (Bain & Parkes, 2006) was employed in the study. All observers participated in a two day workshop on each of the pedagogies

included in the reform During the workshop the observers were trained to look for the features of the pedagogies under observation throughout the 55 minute lesson observation and record those features as present, absent or not observed.

Each of the observation protocols included items that described the essential characteristics of the practice under observation. For example, an explicit teaching protocol (18 items) provided the observers with an opportunity to determine whether the lesson purpose was stated and whether an anticipatory set, modeling, guided and independent practice were present. The cooperative learning protocol (19 items) required observers to determine the presence of appropriate groupings, task structure, and interdependence, while for peer tutoring (21 items) the observers looked for appropriate tutor direction, clear guidelines and appropriate evaluation. These items were derived from the research-based characteristics of the approaches.

The protocols reported overall percent implementation integrity for each observation based upon the items observed present as a percentage of the total observed present and observed missing. Each observation protocol included space for a narrative reflection completed by the observer. The narrative placed the objective classroom data within the context of the classroom, the curriculum and the teacher's professional growth plan.

Teachers usually participated in observations once per semester although teachers experiencing difficulty or those who requested additional support or feedback were observed more frequently, up to seven times annually. Each observation event was scheduled as part of an email exchange between observer and teacher. The selection of the specific lessons to be observed frequently resulted from an expression of interest by teachers for feedback on a given methodology, and/or as part of the broader ongoing process of curriculum development and refinement. Teacher and observer would meet briefly prior to the observation to discuss points of interest and foci. The objective feedback on the teaching practice and the narrative provided teachers with information that showed their facility with the methodology. A meeting for critical reflection and exchange followed each observation between teacher and observer.

In order to complete an observation, each observer selected an unobtrusive location in the classroom. The observer then logged a laptop computer onto the school network selecting the required protocols from the feedback tools described previously. When available, the observer also opened the curriculum software (also located on the network) and selected the actual lesson being taught. In reviewing the lesson within the context of the curriculum, the observer could situate the pedagogy under observation within the broader curriculum goals and methods of which it was a part. The latter was especially important for completing the narrative component of the observation protocol.

While observer presence has long been recognised as a source of influence on the data gathered in classroom observation studies (Blease, 1983; Samph, 1976), students and teachers in this study school quickly became acclimatised to the presence of others in the classroom. The overall frequency of classroom observations, frequent visitors to the school, and the on-going exchange of feedback resulted in a normalizing of observer presence as part of the classroom environment and overall culture of the school.

Ratings of implementation

Ratings and interview data provided by the teachers involved in CSR are the most common form of implementation measurement. In many instances, those views represent the sole integrity measure (Berends et al., 2001; Faddis et al., 2000). In the Self-Organizing School project, self, peer and supervisor questionnaires were completed each year using rating protocols that were part of the suite of evaluation tools deployed at the school. They were then compared with the ongoing observations described above. Each questionnaire included 30 items in four categories directly associated with the reform. They were "student learning", "implementation of the design", "teamwork", and "professional growth". Raters were asked to judge whether the behavior of interest was present: 0 (*never*), 1 (*rarely*), 2 (*sometimes*), 3 (*mostly*) or 4 (*always*).

For example, in the student learning category, ratings were made on the "use of the teaching practices observed in classes", the extent to which the "classroom was differentiated" and whether "technology was used effectively". In the implementation section, ratings were made of the "implementation of team plans", knowledge of the school's processes and the roles taken in "collaborative decision-making". Items under the teamwork heading included "making expectations clear", "the effectiveness of communication" and "the quality of problem solving". In the professional development section, items focused on "translating professional development into practice" and "seeking support and resources". All items reflected a key feature of the school reform. Teachers nominated a peer to complete the questionnaire. Peers were usually selected from members of the same teaching team. The team leader and administrator responsible for the team on which the teacher served undertook supervisor feedback. The questionnaires, like the observations, also included an opportunity for sharing a narrative reflection.

Results

The results of the study are described in response to each of the research questions:

Question 1: Were the teaching practices implemented with fidelity over the four years of the study?

Four hundred and eighty one observations were undertaken over a four-year period. The average number of annual observations per teacher in each year were 2.45 (Year 1), 2.31 (Year 2) 2.45 (Years 3) and 2.65 (Year 4). Table 2 describes the number of observations along with mean and standard deviation scores by year over the four years of the study.

Year	Observations		
	Total	Mean	*SD*
Year 1	118	87.14	10.58
Year 2	117	90.90	8.26
Year 3	121	90.96	7.52
Year 4	125	91.80	8.34

Table 2: Mean and Standard Deviation Scores for Observations of Teaching Practice

The average percent implementation for all groups in all years ranged from 87% in Year 1 to in excess of 91% in Year 4 indicating that the inclusive pedagogies were implemented in classrooms with high levels of practice fidelity in each year. The greatest gain occurred between year one and two after which the results stabilised around the 90% level for the remaining years. The findings indicate a consistent base of practice fidelity evidence that was sustained over time and could be employed as a criterion for an analysis of teacher (self), teaching peer (peer) and supervisor perspectives. A univariate analysis of variance indicated no statistically significant differences in observation over the four years ($F(3, 174)=2.25$, $p=.08$).

Question 2: Were there differences in the perspectives of teachers, their peers and administrators about the implementation of the reform over the four years?

Questionnaires were completed twice per year for all teachers over a four-year period. Table 3 describes the mean and standard deviation scores for each of the stakeholder groups for the four years of the study.

Each of the groups rated the implementation of the reform at a level between three and four, indicating that they assessed engagement with the essential features *mostly* to *always*. These ratings were consistent with the levels of practice fidelity (87-91%) reported in Table 2.

Question 3: Did the ratings of the implementation of the reform alter over time and did those differences diminish or increase as the reform progressed?

	Mean and SD Scores by Rating Type					
	Self		Peer		Supervisor	
Year	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	3.15	.43	3.43	.37	3.04	.45
2	3.35	.32	3.60	.34	3.30	.35
3	3.53	.31	3.71	.29	3.46	.34
4	3.46	.29	3.66	.30	3.42	.35

Table 3: Mean and Standard Deviation scores of Implementation by Surveys

Figure 1 provides a graphic representation of the self, peer and supervisor questionnaires over the four years.

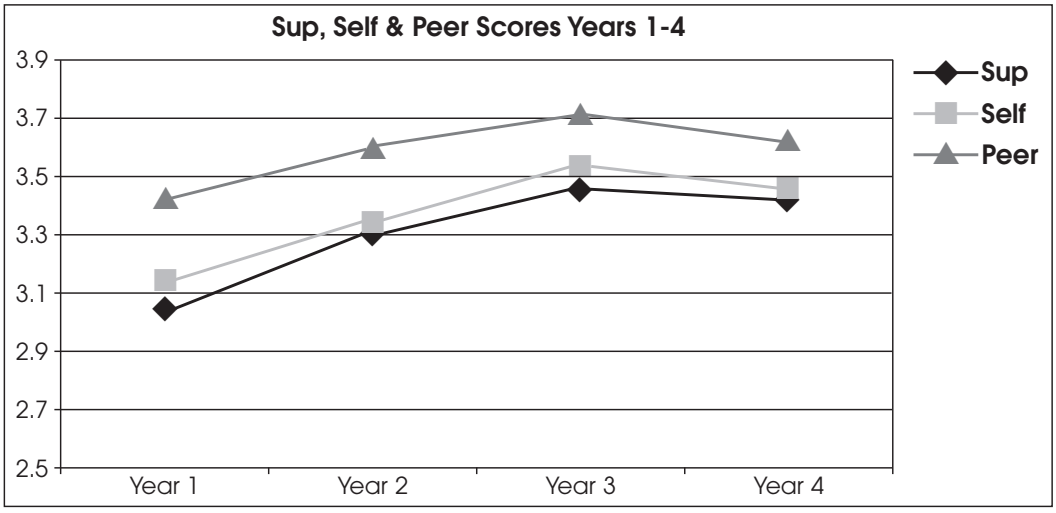

Figure 1: Ratings of Implementation by Teachers, Peers and Supervisors

The ratings of each group followed the increase in observed practice fidelity from year one to two and continued to increase through year three falling slightly in year four. Overall, this pattern of rating was consistent with the data derived from observations, with the exception of the small decrease in ratings between years 3 and 4 when observations showed a slight increase in practice fidelity. Peers rated their colleagues consistently higher than self and supervisor ratings in all years. Supervisors provided the lowest ratings in all four years, although their ratings were highly similar to teachers' self-ratings in the fourth year.

An Analysis of Variance revealed statistically significant differences in overall ratings across the four years ($F(3, 473)=18.14$, $p=.000$), while a second analysis of variance

revealed statistically significant differences across the stakeholder groups ($F(2, 474)=23.60$, $p=.000$). Table 4 describes multiple comparisons indicating the differences across the three rater groups.

Group	Comparison	Mean Difference	Std. Error	Sig.	Confidence Interval (95%)	
					Lower	Upper
Sup						
	Self	-.05	.04	.643	-.15	.05
	Peer	-.27	.04	.000	-.37	-.17
Self						
	Peer	-.21	.04	.000	-.31	-.18

Table 4: Multiple Comparisons for Rater Type

The table shows statistically significant differences between the ratings of supervisors and peer ratings, and self and peer ratings regarding perceptions about the implementation of the reform by individual teachers.

Question 4: Were there differences in the relationship between the ratings of teachers, peers and supervisors and the classroom observations of practice fidelity over the course of the reform?

Group	Year			
	1	2	3	4
Sup	.77	.63	.56	.32
Teacher	.26	.57	.33	.46
Peer	.00	.25	.23	.56

Table 5: Correlation between Ratings and Observations for Supervisors, Teachers and Peers

Table 5 describes the correlations between observations and ratings for supervisors, teachers and peers over the four years. Overall, the strongest correlations were recorded by supervisors although the strength of the relationship between their ratings and observations weakened over time. Conversely, the strength of the correlations for teachers increased from a relatively low level in the first year to modest levels overall. The strength of the correlations for teaching peers strengthened from virtually no relationship in the first year of data collection to a correlation of $r=.56$ in the fourth year.

Question 5: Which of the groups provided the most accurate predictions of practice fidelity over the four-year study period?

Year	Group	*B*	*SE B*	*β*	*t*	*p*
1	Sup	21.14	.79	.75	26.45	.000
	Teacher	.29	.73	.01	.40	.68
	Peer	-5.96	.82	-.20	-7.24	.000
2	Sup	10.72	.76	.46	14.09	.000
	Teacher	7.61	.91	.27	8.33	.000
	Peer	3.62	.72	.14	4.99	.000
3	Sup	10.51	.74	.47	14.12	.000
	Teacher	5.52	.84	.21	6.51	.000
	Peer	-.20	.94	.00	-.22	.826
4	Sup	1.89	.80	.07	2.34	.019
	Teacher	10.94	.94	.36	11.55	.000
	Peer	13.21	.72	.50	18.14	.000

Table 6: Regression table for ratings as practice validity predictor for Supervisors, Teachers and Peers

Table 6 describes the results of a regression analysis that shows the extent to which ratings predicted practice fidelity for supervisors teachers and peers over the four years of the study. In the initial year of the study only supervisors' ratings were strong statistically significant predictors of the practice fidelity of the reform. Peer ratings in that year were negative predictors of classroom practice. In the second year all three stakeholder groups' ratings were predictive of classroom practice at a statistically significant level. Supervisor ratings continued to be the strongest predictors followed by teachers' self-ratings and peers. This pattern continued in the third year although, peer ratings again were negative predictors. In the fourth year of the study the pattern of prediction was inverted. Supervisor ratings were the weakest predictors while peers became the strongest indicators of classroom practice followed by teacher self-ratings.

Discussion

The data derived from the classroom observations undertaken in this study provide initial evidence of the practice fidelity of a reform in a secondary school setting. At present, there are no studies that have generated such data on either implementation or practice fidelity (O'Donnell, 2008). The data show that it is possible to sustain the

implementation of the core teaching practice of a reform with high levels of practice fidelity over time and stand in contrast to the findings of existing school reform studies which show fading implementation and increased variability from teacher-to-teacher within individual schools over time (Berends et al., 2002; Cook et al., 1999; Datnow, 2005; Muncey & McQuillan, 1996).

At a descriptive statistical level of analysis, the ratings of stakeholders about the implementation showed a general concurrence with the classroom observations. Ratings by teachers, peers and supervisors all fell above three on a four point scale indicating that over all four years each group felt that teachers were implementing the reforms most to all of the time.

However, closer scrutiny revealed statistically significant differences in those ratings across individual groups and in the extent to which they predicted the practice fidelity in classrooms. Peers consistently generated higher ratings of the implementation of the reform followed by teachers' self-ratings and supervisors. Ratings for all groups were highest in the third year dropping slightly in the fourth. Correlations between classroom practice and the ratings ranged from low to moderate and were highly variable depending upon year and stakeholder group indicating that the judgments of participants about a reform may vary substantially from that which is occurring in classrooms.

These results indicate that even in circumstances where a reform is achieving high levels of practice fidelity, the perspectives of key stakeholders may vary substantially over time especially in the extent to which they predict what is happening in classrooms. This is an important finding given that few reforms seem to achieve higher levels of implementation year over year (Datnow, 2005) and the evaluations of those reforms tend to rely heavily on the indirect judgments, ratings and perspectives of others in determining their efficacy (Berends et al., 2001; O'Donnell, 2008; Zhang et al., 2005). In the present study, the disparity in the predictive quality of ratings was greatest in the initial years of the reform, indicating that a reliance on indirect measures may be more problematic in the early years of implementation.

This variability in the predictive quality of the ratings may reflect the different foci of stakeholders in a reform process and the ways in which a reform program matures in an organization. Research on teacher perspectives about school reforms by (Schmidt & Datnow, 2005) indicates that in a reform process teachers are much more emotionally focused on the classroom implications of change and its impact on their practice, than instrumental school-wide implementation issues. While the leaders are focused on accuracy in terms of school wide practice fidelity, teachers are more focused on the personal and emotional impact of the change on their professional lives (Bain, 2007). The longer time taken by teachers to build comfort, understanding

and capacity with the specific practice-related competencies in a reform may generate variability in their perspectives about implementation when compared with the more instrumental drivers of leadership.

The poorer predictive validity of administrator ratings in the fourth year may indicate that as the reform develops and becomes more broadly embedded in the professional lives of teachers and a school, an instrumental focus becomes less predictive of the totality of the implementation of a reform. Teachers and their peers, as the agents in a reform may possess a broader conceptualization of practice and a more complete understanding of what is occurring.

Conclusion

In summary, what is clear from the results of this study overall, is that the judgments about the implementation of a reform may vary substantially over time and across stakeholders. This finding should be cautionary for evaluators who are reliant upon ratings by stakeholders as an implementation fidelity or evaluative measure, and especially given the knowledge that ratings by others constitute the predominant measure of choice in evaluations of school reforms (O'Donnell, 2008). The results also suggest that the nature, process and timing of the delivery of feedback to teachers needs to be considered carefully in terms of their readiness to receive such input as part of their overall engagement in the process of adopting a school reform.

References

Aladjem, D., & Borman, K. M. (2006). *Examining comprehensive school reform*. Washington, DC: The Urban Institute Press.

Ashman, A. F., & Elkins, J. (Eds.). (2004). *Educating children with diverse abilities* (2nd ed.). French's Forest, NSW: Pearson Education/Prentice Hall.

Bain, A. (2007). *The self-organizing school: Next generation comprehensive school reforms*. Lanham, MD: Rowman & Littlefield.

Bain, A., Fallon, M., & Smith, D. (1999). Designing the future. In T. Hillman & C. Thorn (Eds.), *Oh What a Web We Weave*. Gilman, NH: Avocus Press.

Bain, A., & Parkes, R. J. (2006). Curriculum authoring tools and inclusive classroom teaching practice: A longitudinal study. *British Journal of Educational Technology*, *37*(2), 177-189.

Bain, A., & Ross, K. (2000). School re-engineering and SAT-1 performance: A case study. *The International Journal of Educational Reform*, *9*(2), 148-154.

Berends, M., Bodilly, S. J., & Nataraj Kirby, S. (2002). *Facing the challenge of whole school reform, new American schools after a decade*. Santa Monica, CA: Rand.

Berends, M., Nataraj Kirby, S. N., Naftel, S., & McKelvey, C. (2001). *Implementation and performance in New American schools: Three years into scale-up*. Santa Monica, CA: Rand.

Blease, D. (1983). Observer effects on teachers and pupils in classroom research. *Educational Review*, *35*(3), 213-217.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). *Comprehensive school reform and student achievement: A meta-analysis*. Baltimore, MD: Center For Research on the Education of Students Placed at Risk.

Borman, G. D., Slavin, R. E., Cheung, A., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2005). Success for All: First-year results from the national randomized field trial. *Educational Evaluation and Policy Analysis*, *27*(1), 1-22.

Brosnan, M. (1996). Make it new: Brewster Academy reinvents itself. *Independent School*, *Spring*, 12-16.

Brown, J. (2000). Putting vision into practice. *Converge*, *3*(2). Retrieved December 3, 2000, from http://www.convergemag.com

Cherednichenko, B., Hooley, N., Kruger, T., & Moore, R. (2001). Longitudinal Study of School Restructuring. Retrieved October 24, 2009, from http://www.aare.edu.au /01pap/kru01380.htm

Cook, T. D., Habib, F.-N., Phillips, M., Settersten, R. A., Shagle, S. C., & Degirmencioglu, S. M. (1999). Comer's school development program in Prince George's County, Maryland: A theory-based evaluation. *American Educational Research Journal*, *36*(3), 543-597.

Datnow, A. (2005). *The sustainability of comprehensive school reform models in changing district and state contexts. Educational Administration Quarterly, 41*(1), 121-153

Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research*, *72*(3), 433-479.

Dimmock, C. A. J. (2000). *Design the learning-centered school: A cross cultural perspective*. London, UK: Falmer Press, Garland Inc.

Doremus, R. R. (1981). What ever happened to... John Adams High School? *Phi Delta Kappan*, *63*(3), 199-202.

Earl, L., Torrance, N., & Sutherland, S. (2003). *Manitoba school improvement program final evaluation report*. Toronto, ON: The Ontario Institute for Studies in Education of the University of Toronto.

Eastabrook, G., Fullan, M., & Bliss, J. (1977). Action research in the school involving students and teachers in classroom change. In R. Carlton, L. Colley & N. Mackinnon (Eds.), *Education and change in society*. Toronto, ON: Gage.

Elsworth, G., Kleinhenz, E., & Beavis, A. (2004). *Evaluation of the Middle Years Reform program*. Melbourne, Vic.: RMIT.

Elmore, R. (1996). Getting to scale with good educational practice. *Harvard Educational Review*, *66,* 60-78.

Faddis, B. J., Beam, M., Hahn, K. J., Willardson, M., Sipe, D., & Ahrens-Grey, P. (2000). *The implementation of the comprehensive school reform demonstration program: The work of 40 schools in seven Midwest states*. Retrieved December 3, 2000, from http://www.eric.ed.gov/

Fink, D. (2000). *Good schools/real schools: Why school reform doesn't last*. New York, NY: Teachers College Press.

Fullan, M. (2007). *The new meaning of educational change* (4th ed.). New York: Teachers College Press.

Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research*, *11*(2), 147-252.

Gertler, P., Patrinos, H. A., & Rubio-Codina, M. (2007). *Impact evaluation for school-based management reform* (No. 10). Washington, DC: Poverty Reduction and Economic Management.

Hattie, J. (2003, 19th-21st October). *Teachers make a difference: What is the research evidence?* Paper presented at the Australian Council for Educational Research Conference on Building Teacher Quality, Melbourne, Victoria.

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Huberman, A. M., & Miles, M. B. (1984). *Innovation up close*. New York: Plenum.

Hunter, R. (2004). *Madeline Hunter's mastery teaching: Increasing instructional effectiveness in elementary and secondary schools*. Thousand Oaks, CA: Corwin Press.

Johnson, D. W., Johnson, R. T., & Holubec, E. J. (1998). *Cooperation in the classroom*. Edina, MI: Interaction Book Company.

Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). *Cooperative learning methods: A meta-analysis*. Minneapolis, MN: University of Minnesota.

Kurki, A., Boyle, A., & Aladjem, D. K. (2006). Implementation: Measuring and explaining the fidelity of CSR implementation. *Journal of Education for Students Placed at Risk*, *11*(3/4), 255-277.

Lingard, B., Ladwig, J., Mills, M., Bahr, N., & Chant, D. (2001). *Queensland School Reform Longitudinal Study*. Brisbane, Australia: Education Queensland

Marzano, R. J. (1998). *A theory-based meta-analysis of research on instruction*. Aurora, Colorado: Mid-continent Regional Educational Laboratory.

Mastropieri, M. A., & Scruggs, T. E. (2004). *The inclusive classroom: Strategies for effective instruction* (2nd ed.). Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.

McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development*, *71*(1), 173-180.

McCord, M. (1999). Exploring the frontier of technology education reform. *New Hampshire Business Review/TechNet*, *12*(February), 33-36.

Mihalic, S. (2001). The importance of implementation fidelity. *Blueprints News*, *2*(1), 1-2.

Muncey, D. E., & McQuillan, P. J. (1996). *Reform and resistance in schools and classrooms: An ethnographic view of the coalition of essential schools*. New Haven: Yale University Press.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K 12 curriculum intervention research. *Review of Educational Research, 78*(1), 33-84.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education, 93*(3), 223-231.

Reynolds, D., Creemers, B., Stringfield, S., Teddlie, C., & Schaffer, G. (2002). *World class schools: International perspectives on school effectiveness*. London, UK: Routledge Palmer.

Ridley, K., & Kendall, L. (2005). *Evaluation of Excellence in Cities primary pilot 2001-2003* (Research Report No. 675). London, UK: National Foundation for Educational Research.

Samph, T. (1976). Observer effects on teacher verbal behavior. *Journal of Educational Psychology, 68*(6), 736-741.

Sarason, S. B. (1982). *The culture of school and the problem of change*. New York, NY: Teachers College Press.

Sarason, S. B. (1996). *Revisiting the culture of school and the problem of change*. New York, NY: Teachers College Press.

Schmidt, M., & Datnow, A. (2005). Teachers' sense-making about comprehensive school reform: The influence of emotions. *Teaching and Teacher Education, 21*(8), 949-965.

Sharp, C., Eames, A., Sanders, D., & Tomlinson, K. (2005). *Postcards from research engaged schools*. Slough: NFER.

Sharp, C., Schagen, I., & Scott, E. (2004). *Playing for success: The longer term impact. A multilevel analysis* (No. RB593). Nottingham, UK: National Foundation for Educational Research.

Slavin, R. E. (1990). *Cooperative learning: Theory, research and practice*. Englewood Cliffs, NJ: Prentice-Hall.

Slavin, R. E., Farnish, A. M., Livingston, M. A., Sauer, D. C., & Colton, B. S. (1994). *The John Hopkins team learning project: Using student team learning*. Baltimore, MD: The John Hopkins University.

Tirozzi, G. N., & Uro, G. (1997). Education reform in the United States: National Policy in support of local efforts for school improvement. *American Psychologist, 52*(3), 241-250.

Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.

Weston, M. E., & Brookes, D. M. (2008). Critical constructs as indicators of a shifting paradigm.

in education: A case study of four technology-rich schools. *Journal of Ethnographic & Qualitative Research, 2*, 281-291.

Zhang, Y., Shkolnik, J., & Fashola, O. (2005). *Evaluating the implementation of Comprehensive School Reform and its impact on growth in student achievement*. Paper presented at the Annual meeting of the American Educational Research Association, Montreal, Canada.