Measuring fidelity of mathematics intervention programme implementations in primary school settings

Fiona Jackson
University of Cambridge

This paper reports on selected findings of a doctoral study exploring varying implementations of a mathematics intervention programme. Most importantly, the research develops methods for measuring fidelity of implementation and identifying instances of positive infidelity.

Keywords: underachievement, process analysis, fidelity,

Introduction

There are varied mathematics interventions currently implemented in UK schools including programmes specifically associated with the National Numeracy Strategy (NNS) and its successors, and other independent programmes. However, there is very little knowledge on how these intervention programmes are implemented or how processes within intervention programmes gain success. As schools begin to adopt these mathematics programmes, implementation rarely transpires as planned from the perspectives of both school personnel and programme developers. Schools may implement a programme as planned by developers but may not meet desired outcomes so a redesign may be required. On the other hand, schools may not implement an intervention consistent with a programme's framework. Fidelity of implementation is part of Roger's (1962) diffusion of innovation theory, originally used to evaluate demonstration projects to determine their appropriateness before disseminating to end users (Dusenbury et al. 2003).

Some studies have researched fidelity of implementation in educational contexts but meta-analyses tend to focus on prevention programmes for health, social and or emotional type issues (Wilson and Lipsey 2007). A few recent studies focussed on fidelity in the area of mathematics education and intervention programmes (Munter et al. 2010; Torgerson et al. 2011) so my research considered these studies and built on their concepts.

Two views on implementation of interventions

Fidelity of implementation is most often described as "The extent to which core components of interventions are delivered as intended" (Dane and Schneider 1998, 23) and is typically measured quantitatively. This view of implementation accepts programme design as correct and non-negotiable by implementers and links high or low fidelity to student outcomes. This then, is an evaluation of deviations by implementers from implementation specifications on the assumption that these diminish programme effectiveness.

The other end of the spectrum on fidelity of implementation takes a more interpretivist stance. This perspective considers some deviations from critical implementation elements may in fact improve the delivery of a programme: this is referred to as 'proadaptation' (Blakely et al. 1987) or 'positive infidelity' (Cordray and Hulleman 2009). Implementers' adaptations may be bidirectional, adapting curriculum to suit their teaching practice and or adapting their instruction to align with curriculum (Remillard 2005). Indeed, these adaptations may be related to craft knowledge or pedagogical content knowledge, viewed as an important aspect of mathematics instruction, where implementers make pertinent changes to processes or instruction to meet needs of individual students. Such changes made by implementers, without compromising an intervention's overall theoretical objectives, it is argued should be considered as valuable contributions to the implementation process.

Fidelity of implementation has become the focus of two recent studies directly related to specific individualised mathematics interventions, Mathematics Recovery (Munter et al. 2010) and Numbers Counts (Torgerson et al. 2011). To a lesser degree, both studies took into consideration the possibility of "positive infidelity". The study of Numbers Count included a process evaluation where views and perceptions of implementers were reported.

Likewise my research elicited views from implementers regarding a mathematics intervention programme, Catch Up® Numeracy. However, my study seeks to identify and understand considerations behind common adaptations made to interventions, thus, acknowledging implementers' expertise.

Measuring fidelity and identifying positive infidelity

Whilst there is extensive literature on measuring fidelity of implementation, there is very little on measuring positive infidelity or positive adaptations. However, there are many studies that acknowledge teachers as interpreters of educative curricula and the ways these resources influence teachers and teaching (Remillard 2005). Teachers must be able to bridge between instructional strategies detailed in educative materials whilst anticipating students' responses in meaningful and manageable ways. This relies on moving away from dependence on materials and a move towards teacher agency (Mc Clain et al. 2009). In these studies, fidelity to curriculum is not measured, preferring instead to understand and record variations in enactments that meet learning needs of students. This model moves away from fidelity of implementation and emphasis on outcomes closer to process analysis; to observe and understand reasons teachers make adaptations to certain elements of an intervention, especially modifications that may increase effectiveness.

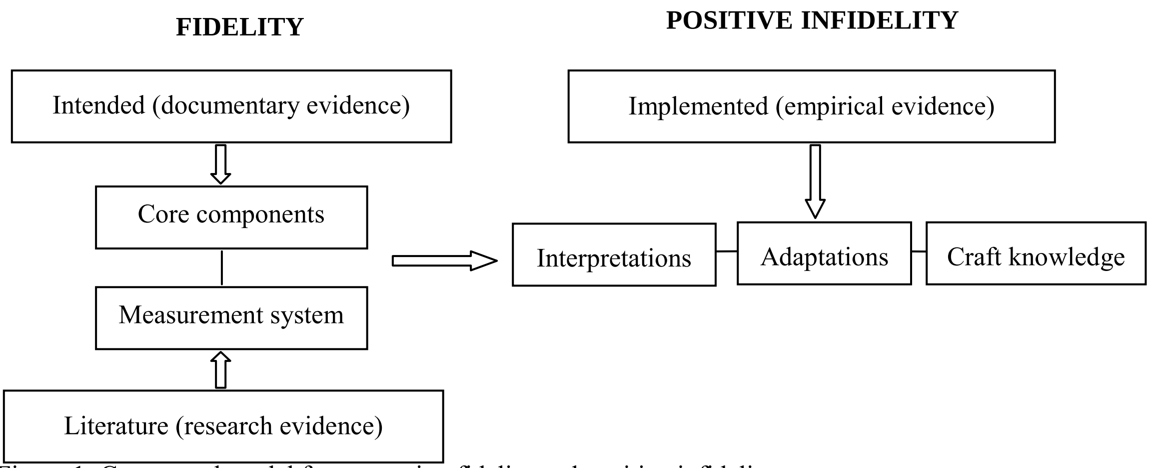

Figure 1: Conceptual model for measuring fidelity and positive infidelity

The conceptual model (Figure 1) framed the research in my study but can be applied to any intervention programme where positive infidelity is being considered.

The first step relates to fidelity and involves reviewing documentary evidence available on a chosen intervention to uncover core components of a programme, more precisely the core theoretical components. However, most programme designers do not state explicitly their core components, that is, the critical components that are responsible for achieving positive outcomes (Ruiz-Primo 2006). It is only through systematic empirical research that a component could be eliminated or modified in some manner to understand its effectiveness on outcomes.

A measurement system is required to ascertain compliance or non-compliance of core programme components. In its most basic form this would be Yes or No, compliant or non-compliant. Next, a thorough reading of research literature on this type of intervention is required to unveil instructional strategies and techniques required to deliver the chosen programme; in this case an individualised mathematics intervention programme, Catch Up Numeracy.

Finally, to identify instances of positive infidelity empirical evidence is required to gain a deeper understanding about where, when and why adaptations are made by implementers within local settings. For example, teaching in a one to one environment relies on many pedagogical skills as teachers and students interact. Practical experience or craft knowledge guides teachers' actions knowing when to give feedback, when to pause to allow students thinking time, how to pose questions relevant to individual student's mathematical understanding, and so forth.

Applying the conceptual model to Catch Up Numeracy

Catch Up Numeracy is an individualised intervention programme aimed at Year 2 to 10 students who are underachieving in school mathematics although this study focused on Year 3 to Year 6 students only. The programme is delivered to students twice each week for fifteen minutes. Distinguishing itself from other current large scale numeracy interventions, Catch Up Numeracy offers training to adults without qualified teacher status or previous mathematics qualifications; normally teaching assistants or SENCOs, both referred to as 'tutors' from this point on in the paper.

Measuring fidelity

Referring to the conceptual model in Figure 1, the first stage of measuring fidelity involved consultation with Catch Up to establish core components related to the programme's success. Nineteen core components were identified and considered in my research. Simultaneously, research literature on individualised mathematics invention programmes at the primary school level was reviewed.

Next a rating system was designed to measure adherence to or deviations from these core components. Designing a rating system was not straight forward as it was possible for tutors to deviate from Catch Up specifications in a number of ways. The way in which this was handled was to classify such deviations and order these classes by seriousness. Thus a fully compliant performance would attract a rating of 3; a performance in which only the least serious deviation was observed would attract a rating of 2; and so on.

Three statements of compliance were developed for each core component which a tutor could meet (Y) or not meet (N); see Table 1 and example below. A fully compliant performance would satisfy all three criteria (YYY profile); a minimally deviant performance would display the least serious type of deviance but comply with the other two criteria (NYY) and so on. This created eight possible combinations or outcomes representing full, partial or no compliance (YYY, NYY, YNY, NNY, YYN, YNN, NYN or NNN) to Catch Up's specifications, which were developed into four possible profiles of agreement or disagreement; a), b), c), or d). An N rating on Statement 1 represented the least deviant action from specifications and statement 3 the most deviant. For example, when a tutor displayed actions a) this meant that they complied with statements 1), 2) and 3) i.e. YYY or rating of 3. As statement 1 was the least serious deviance, non-compliance of statement 1 together with compliance of statements 2 and 3 corresponded to tutors' action b), i.e. NYY or rating of 2. Tutors' actions c) related to two possible combinations YNY and NNY whereby tutors were non-compliant with statement 2 (second most deviant action) but still compliant with statement 3 (most deviant action). Tutors' actions d) represented any remaining profiles not shown in a), b) or c) but relating directly to non-compliance of statement 3. Finally, tutors' action e) was described as 'null' meaning that due to no fault of tutors they were unable to demonstrate adherence to a specific core component. For example, tutors may *intend* to complete a particular action but may be unable to do so due to a student's action thus deeming a rating unfeasible.

Table 1: Fidelity profiles of agreement and disagreement

Tutors' Actions	Statement 1 (Least deviant)	Statement 2	Statement 3 (Most deviant)	Rating
a)	Y	Y	Y	3
b)	N	Y	Y	2
c)	Y N	N N	Y Y	1
d)	Y Y N N	Y N Y N	N N N N	0
e)	-	-	-	null

Identifying instances of positive infidelity

Again, referring to Figure 1 it was important to understand how tutors were interpreting core components, what adaptations were made and what part craft knowledge played towards these adaptations. It is acknowledged that tutors who displayed actions affording a 2 rating (least serious deviance) were often demonstrating acts of positive infidelity. Indeed, it may be argued that positive infidelity actions were sometimes more logical or thoughtful than Catch Up's initial guidelines and may in fact lead to improvements in the programme. An example is now given to demonstrate this phenomenon.

Before teaching sessions commence, students are assessed based on Catch Up's twenty-two subcomponents of numeracy and any identified gaps in understanding then taught in fifteen minute teaching sessions. During assessments, tutors should stop an assessment as soon as students utter an incorrect response and then offer an explanation for a correct answer. However, before fidelity statements were finalised, observations of this practice were completed to identify any deviations from Catch Up's specifications.

Empirical evidence revealed some tutors did not stop assessments immediately but instead checked for student understanding. This was in the form of a tutor question clarifying understanding followed by a student self-correction or a self-initiated correction. Moreover, some tutors stopped an assessment prematurely i.e. before an incorrect student response was uttered due to tutors' misinterpretation of an assessment. Equally, tutors stopped an assessment late i.e. waited until two or more consecutive incorrect answers were uttered before stopping a particular assessment. At whatever point an assessment was stopped, some tutors failed to offer an explanation for a correct answer. Thus deviations from compliance were observed as follows:

- Fail to stop assessment immediately when an incorrect student response was given, but took account of further information about student understanding:
  - From a student correction following tutor check and/or
  - From a student self-initiated correction
- Stop an assessment either prematurely or late
- Fail to offer an explanation for a correct answer when assessment stopped

Thus, corresponding classes of compliance were adjusted to take into consideration observed tutors' deviations as follows:

- Statement 1 - No checking for student understanding after an incorrect response given (Y/N)
- Statement 2 - No stopping an assessment prematurely or late (Y/N)
- Statement 3 - Offer an explanation after an assessment was stopped (Y/N)

In this scenario, a tutor would receive a 2 rating (NYY) if they checked for student understanding after an incorrect response was given (Statement 1) but still complied with Statement 2 and 3. Whilst these actions perhaps demonstrate enhanced pedagogical skills, they are not compliant in terms of fidelity relating to Catch Up's stated specifications. On occasion, further analysis of tutors' actions rated at 1 or 0 was also classified as positive infidelity.

In determining whether these tutors' actions were indeed deviations or acts of positive infidelity, tutors were asked to explain why they deviated from specifications. If tutors offered a rational explanation for adaptations made which may aide improvements to or development of Catch Up Numeracy, then these reasoned ideas were considered examples of positive infidelity. Whilst there is no direct evidence to support the effectiveness of positive infidelity in my study, I argue that it is possible and so requires further investigation.

Summary

Programme designers have to define parameters to ensure tutors adhere to core specifications. Yet designers are also aware of non-essential elements that can be adapted or omitted without consequence. The difficulty lies in communicating this to implementers. If too many parameters are outlined then a programme becomes confusing and too few may compromise its effectiveness. How do designers revise materials to accommodate reasoned adaptations? How do designers distinguish between effects caused by written materials and tutors' interactions with these materials? Accordingly, Berman and McLaughlin (1976) suggest "mutual adaptation" as the most effective strategy for implementation whereby programme designers and implementers accommodate one another in the change process. In this way, implementers offer feedback as to when adaptations are made and in turn designers analyse whether these changes affect programme integrity.

Acknowledgements

I would like to thank Catch Up for their support of this research.

References

Berman, P., and M. W. McLaughlin. 1976. Implementation of educational innovation. *The Educational Forum* 40 (3):345-370.

Blakely, C. H., J. P. Mayer, G. Gottschalk, N. Schmitt, W. S. Davidson, D. B. Roitman, and J. G. Emshoff. 1987. The fidelity-adaptation debate: Implications for the implementation of public sector social programs. *American Journal of Community Psychology* 15 (3):253-268.

Cordray, D. S., and C. S. Hulleman. 2009. Assessing intervention fidelity: Models, methods and modes of analysis. Paper read at Fourth Annual IES Research Conference, at Washington, USA.

Dane, A. V., and B. H. Schneider. 1998. Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review* 18 (1):23-45.

Dusenbury, L., R. Brannigan, M. Falxo, and W. B. Hansen. 2003. A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, Theory & Practice* 18 (2):237-256.

Mc Clain, K., Q. Xhao, I. Visnovska, and E. Bowen. 2009. Understanding the role of intsitutional context in the relationship between teachers and text. In *Mathematics teachers at work: Connecting curriculum materials and classroom instruction*, edited by J. T. Remillard, B. A. Herbel-Eisenmann and G. M. Lolyd. Abingdon, UK: Routledge.

Munter, C., A. Garrison, P. Cobb, and D. Cordray. 2010. Evaluating Math Recovery: Measuring fidelity of implementation. In *SREE Conference 2010: Research into practice* Washington, USA: Society for Research on Educational Effectiveness.

Remillard, J. T. 2005. Key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research* 75 (2):211-246.

Rogers, E.t M. 1962. *The diffusion of innovations*. New York: Free Press.

Ruiz-Primo, M. A. 2006. A multi-method and multi-source approach for studying fidelity of implementation. California, USA: National Center for Research on Evaluation, University of California.

Torgerson, C. J., A. Wiggins, D. J. Torgerson, H. Ainsworth, P. Barmby, C. Hewitt, K. Jones, V. Hendry, M. Askew, M. Bland, R. Coe, S. Higgins, J. Hodgen, C. Hulme, and P. Tymms. 2011. Every Child Counts: The independent evaluation, technical report. York, Birminham and Durham: Universities of York, Birmingham and Durham.

Wilson, S. J., and M. W. Lipsey. 2007. School-base interventions for aggressive and disruptive behavior: Update of a meta-analysis. *American Journal of Preventative Medecine* 32 (2):130-143.