

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

# Computers & Education

journal homepage: [www.elsevier.com/locate/compedu](http://www.elsevier.com/locate/compedu)

## Implementation fidelity in computerised assessment of book reading



Keith Topping\*

University of Dundee, Scotland, United Kingdom

### ARTICLE INFO

#### Article history:

Received 16 March 2017

Received in revised form 24 August 2017

Accepted 21 September 2017

Available online 23 September 2017

#### Keywords:

Evaluation methodologies

Gender studies

Improving classroom teaching

Pedagogical issues

Teaching and learning strategies

### ABSTRACT

Measuring the implementation fidelity (IF) or integrity of interventions is extremely important, since without it a positive or negative outcome cannot be interpreted. However, IF is actually measured relatively rarely. Direct and indirect methods of measurement have been used in the past, but tend to over-emphasize teacher behaviour. This paper focuses on student behaviour collated through computers - an interesting alternative. It deals with the reading of real books and reading achievement, for which variables a very large amount of computerised data was available – on 852,295 students in 3243 schools. Reading achievement was measured pre-post with STAR Reading, a computerised item-banked adaptive norm-referenced test of reading comprehension. IF came from the Accelerated Reader (AR), which measures understanding of independent reading of real books the student has chosen by a quiz. Results showed higher IF was related to higher achievement. Neither IF nor reading achievement related to socio-economic status. Primary (elementary) schools had higher IF and achievement than secondary (high) schools. Females had higher IF and achievement than males. Students of higher reading ability implemented AR at a higher level, but did not gain in reading at a higher level. However, this computerised method of measuring IF with book reading showed limited reliability, no greater than methods emphasising teacher behaviour. In future, IF measures emphasising student response and those emphasising teacher behaviour need to be blended, although the latter will never generate the sample size of the former. This may be true of implementation fidelity in areas other than book reading.

© 2017 Elsevier Ltd. All rights reserved.

In these evidence-based times, there is much emphasis on randomised controlled trials (RCTs) as the “gold standard” of good research. However, as [Stockard \(2010\)](#) among others points out, there are concerns about the external validity of such findings. There are a number of collections of such research intended to impact on practice, but only one (The What Works Clearinghouse - WWC) is sponsored by the US government. Extraordinarily, the WWC largely disregards the issue of IF, assuming that on average it “washes out” in the reviews they promote.

Another issue with RCTs is that by definition they allocate the intervention randomly. In education, this means to teachers who may or may not have the slightest interest in implementing the intervention. [Wehby, Maggin, Johnson, and Symons \(2010\)](#) studied the effect that teacher choice of intervention had on their level and quality of implementation. A total of 69 teachers (88% female; 68% general education, 32% special education) working with K-6 students participated. Implementing a

\* School of Education, University of Dundee, Dundee DD1 4HN, United Kingdom.

E-mail address: [k.j.topping@dundee.ac.uk](mailto:k.j.topping@dundee.ac.uk).

preferred intervention was related to higher degrees of initial and sustained IF as well as greater numbers of actual implementers.

Clearly, there are issues here about implementation fidelity. But how might it be defined?

## 1. Definition of implementation fidelity

Implementation fidelity (or integrity) was initially defined as the degree to which an intervention or treatment was implemented as planned, intended, or originally designed. However, this only specified the behaviour of the interventionist, not that of the recipients of the intervention. By contrast, [Schulte, Easton, and Parker \(2009\)](#) included features related to the delivery of the intervention, how the intervention was received by the participants, and how the participants were able to use the learned skills in a natural environment. Of course, the question then arises of which of these many indices are most related to outcome ([Durlak & DuPre, 2008](#)).

Despite the importance of treatment fidelity, historically it has been frequently overlooked in research and practice. Since the emphasis has moved towards “evidence-based” interventions, measuring the quality of intervention has become an increasing preoccupation. Clearly, there is little point attempting to implement an evidence-based intervention and measure the outcomes if there is no parallel attempt to see whether the method has actually been implemented. As [Carroll et al. \(2007\)](#) express it, IF acts as a potential moderator of the relationship between interventions and their intended outcomes. Unless IF is assessed, in a circumstance of poor outcome we cannot know whether the program did not work or merely was not implemented properly, or both. Indeed, even in a circumstance of good outcome, we also cannot know whether the program actually worked and was responsible for the positive outcome.

[Dane and Schneider \(1998\)](#) and [Schulte et al. \(2009\)](#) among others espoused five elements in IF often found in the previous literature: adherence to an intervention, exposure or dose, quality of delivery, participant responsiveness and program differentiation (the extent to which key factors in effectiveness are identified). Measuring IF is not easy – researchers quickly found that it was both complex and expensive. Not all interventions clearly specified what the teacher had to do and in what order. Indeed, some of them had optional teacher behaviours, assuming that no two teachers would implement alike. Indirect attempts which simply asked teachers whether they had implemented well were often found not to correlate with outcomes. Direct attempts which used observational methods (to avoid teacher subjectivity) were expensive (and consequently usable only on a small scale) and could still suffer from observer effects – what the teacher did when observed might not have been typical of what they did when not observed. Another issue was whether any professional development prior to the intervention was one-off, or whether it was several sessions with time in-between for reflection and discussion with colleagues, or included ongoing coaching to shape teacher behaviour as the program was implemented. Teacher behaviour is the focus of much of the literature. [Schulte et al.'s \(2009\)](#) inclusion of participant responsiveness has been largely overlooked. There is also an issue about how often IF should be assessed, since many of the reports in the literature are of short-term interventions.

## 2. The current paper

This paper is set in the context of the effectiveness of book reading and emphasises student response rather than teacher behaviour. The focus on book reading is because this area generates the largest amount of data of any computerised assessment. A parallel assessment of mathematics is available, but the number of users and amount of data is considerably smaller. The paper compares and contrasts two different kinds of participant indicators of IF with growth in reading achievement. Of course other measures may be relevant, but this study deploys measures of student response to counter-balance the existing over-emphasis on teacher behaviour. In this paper both outcome and IF measures are completed locally but scored online centrally, and the results fed back locally, all by computer. This central scoring enables the collection of large samples of data. Both IF and outcome measures generate a number of variables directly from student responses.

## 3. Previous research on IF in reading

The present paper interrogates the literature on IF in book reading by exploring research on indirect measures (self-reports completed subjectively by teachers and head teachers) and direct measures (completed by observation, although still far from “objective” given possible observer effects). Curiously, studies appear to have done one or the other – there are very few studies which have directly compared the two. After this, a further section explores the literature (such as it is) on IF using the novel measures deployed here – computer-assisted item-banked adaptive reading outcome assessment and assessment of comprehension of real books after they have been read.

### 3.1. Methodology of the literature review

The Social Sciences Citation Index (SSCI) and the Educational Research Information Centre (ERIC) were searched from 1995 to date (the terms implementation/treatment fidelity/integrity had little currency prior to this date). Search terms were “book reading” AND “implementation fidelity” OR “implementation fidelity” OR “treatment integrity” OR “treatment fidelity”. The inclusion criterion specified relevance to the research questions (see below). Only 33 hits resulted from the first search of titles and abstracts. A further criterion of incorporating substantive data was implemented. On reading the full text, a number

Download English Version:

<https://daneshyari.com/en/article/4936779>

Download Persian Version:

<https://daneshyari.com/article/4936779>

[Daneshyari.com](https://daneshyari.com)