Master Thesis, Learning in Complex Systems

Autumn 2011

Within-Subject Design in Reading Fluency Research

Bruk av N=1 design i leseflyt forskning

Hildur Valdimarsdottir

Department of Behavioral Science

HØGSKOLEN I OSLO
OG AKERSHUS

Approval Sheet

Supervisor recommends that the present master thesis is handed in for evaluation by the examination committee.

New York, NY, November 11, 2011

Supervisor

Professor, PhD

Student

Clarification of individual contribution in Paper 2

Paper 2 of the master thesis was executed in cooperation with Lilja Yr Halldorsdottir. Both of us searched and gathered references and based on our findings from the literature we designed the intervention together. We contributed equally in implementation of all the phases of the study, i.e. the sessions for each participant were divided evenly between us. In addition, we split reliability and procedural integrity measures between us.

Finally, we processed the results, made the graphs, and wrote the whole manuscript of Paper 2 in tight and complete cooperation.

Hildur Valdimarsdottir

Lilja Yr Halldorsdottir

Acknowledgements

I would like to thank my friend Lilja Yr Halldorsdottir for all the good times while we worked on this manuscript and for exceptional cooperation. I thank the three participants that gave us so much time. I also thank Professor Erik Arntzen for comments and suggestions he made to the manuscript. Finally, I thank my husband and children for being her for me all the time.

Contents

Paper 2: Delayed Multiple Probe Design to Evaluate the Effects of a Multicomponent Intervention to Improve Reading Fluency in Adult Students

Overview Over Figures

Paper 1: Within-Subject Design in Reading Fluency Research

No figures

Paper 2: Delayed Multiple Probe Design to Evaluate the Effects of a Multicomponent Intervention to Improve Reading Fluency in Adult Students

Figure 1 Number of words read correct per minute (WRCM) and errors on the test probes during baseline and withdrawal phases, along with the best reading of each session during intervention- and follow-up phases, for all the participants.

Figure 2 Number of words read correct per minute on the five test probes and the application passages during baseline, withdrawal phase 1, and withdrawal phase 2 for all participants.

Abstract

Reading fluency has become a popular topic in the literacy literature in recent years. It has been stated that reading fluency is one of the most important aspects of reading. The reading fluency literature consists of a number of publications from different disciplines. However, many studies have methodological flaws, do not demonstrate experimental control, or make any efforts for generalization. Those defects have made it impossible to determine the most common practice, repeated reading, as evidence based practice. In Paper 1, the defining features of within- subject designs according to some quality standards that have been proposed for special education will be discussed in relation to reading fluency studies. In addition, it will be discussed how it is possible to increase experimental control in research within precision teaching. In Paper 2, the purpose was to implement a high quality study by using a delayed multiple probe design across participants to evaluate the effects of a multicomponent intervention on reading fluency in adult Norwegian students. The procedure used in the study was based on a study by Lokke, Lokke, and Arntzen (2009). The main findings indicate that reading fluency of all the participants improved after receiving the intervention. Results are discussed with regard to social validity of the study, limitations of conducting a multicomponent intervention, and in terms of whether the present study can be considered as a high quality within-subject experiment according to Horner's et al. (2005)[1] quality indicators.

Key words: Multiple probe design, reading fluency, RESAA, quality standards, within-subject design, stimulus control.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165-179. Retrieved from www.cec.sped.org

Lokke, G. E. H., Lokke, J. A., & Arntzen, E. (2009). Bruk av hurtiglesingsteknikker for å øke lesehastighet hos gutt med kognitiv svikt og reaktiv tilknytningsforstyrrelse. *Norsk Tidsskrift for Atferdsanalyse, 36*(4), 231-240. Retrieved from www.nta.atferd.no

Master Thesis, Master in Learning in Complex Systems

Autumn 2011

Within-Subject Design in Reading Fluency Research

Hildur Valdimarsdottir

Department of Behavioral Science

HØGSKOLEN I OSLO
OG AKERSHUS

Abstract

Fluency has become a popular topic in educational research in the last years. The reading fluency literature consists of many publications from both traditional educational research and from the practice of precision teaching. Despite the amount of research that has been published, regarding reading fluency, there have not yet been published enough high quality researches to determine the most common practice to increase reading fluency, repeated reading, as evidence based practice. In this paper, experimental control in reading fluency research is discussed. Special attention is directed to characteristics of within-subject designs and their use in reading fluency research. There have been proposed some quality standards for within-subject designs in special education to identify evidence practice. Those quality standards are discussed in relation to reading fluency research. In addition, it will be discussed how it is possible to increase experimental control in precision teaching research.

Keywords: Evidence based practice, experimental control, quality standards, reading fluency, within-subject design

Within-Subject Design in Reading Fluency Research

In recent years, *fluency* has become a popular topic in educational research, it traditionally refers to the act of speaking or reading, a person that is fluent speaks or reads smoothly and without effort. In the reading research literature fluency is most often defined by speed, accuracy, and proper expression or prosody (e.g., Kuhn & Stahl, 2003; Rasinski, 2004). Reading speed is used to assess the automaticity of *decoding* a text, decoding is when a vocal response is made upon a written stimuli or symbol. Accuracy is measured with percent of correct read words and prosody is taken as an evidence of that the reader understands the text and is measured with qualitative rubrics or rating scales (Rasinski, 2004).

Within behavior analysis, precision teachers are most tenacious in fluency studies. In precision teaching (PT) fluency can be applied to all types of behavior and fluency research has, for example, been done on dressing (Rolling, 2009), ballet dancing (Lokke, Lokke, & Arntzen, 2008), and instructional outcomes in children with autism (Fabrizio & Moors, 2003). Fluency in PT is a combination of accuracy and speed, defined as an outcome of high rate of behavior characterized with retention, endurance, stability, adduction, and application, shortened with the acronym RESAA (Binder 1996). Reading fluency in PT is measured with the number of words read correct per minute (WRCM) and number of errors per minute.

Research on Reading Fluency and Experimental Designs

Many techniques have been developed to increase reading fluency and numbers of research articles have been published. In the National Reading Panel report (2000) was an extensive meta-analysis of research concerning reading fluency. They found 364 researches that had been published on reading fluency from the year 1990, but only 77 studies met the criteria for inclusion in the analysis based on information provided, only 28 studies used *experimental designs;* 16 used group design and 12 used *within-subject design* (WSD). They discovered that most of the research has displayed *correlational* or ambiguous results.

Correlation means that there is a systematic covariation between two events, but does not infer *functional relationship*, that is, that one variable causes changes in the other (Cooper, Heron, & Heward, 2007). Correlational reading studies report results such as; good readers read more than poor readers (Cunningham & Stanovich, 1998), poor readers remain poor readers, and younger students read slower than older students (Juel, 1988) this does not say anything about whether a functional relationship between the two variables is present. Experimental designs are, however, used to demonstrate *experimental control* and separate correlation from functional relations (Kennedy, 2005). Experimental control is when manipulations of the independent variable produce a predictable change in the target behavior (dependent variable), it also refers to: to what degree of certainty the experimenter can rule out other variables, than the independent variable, as a cause for change in the dependent variable (Cooper et al., 2007). Group studies, most often, involve that one group is exposed to an intervention, and the results are compared to a control group that has not received the intervention. Differences between the groups are presented with different statistical measurement and significant levels (Arntzen, 2010). The group studies that were analyzed by the National Reading Panel were often done with: small samples, lack of experimental control, over short time spans, and no generalization efforts. Therrien (2004) did also a meta-analysis of reading fluency studies. He analyzed studies that implemented the most widely used fluency building strategy across different disciplines, repeated reading (RR) techniques (Samuels, 1979), to increase reading fluency. Therrien found 33 studies; nine (many of them within-subject research) were excluded from the analysis because it was not possible to use them to calculate effect size. Therrien discovered that there was a support for the effectiveness of RR interventions; however, the comparison was made between pretests and posttests with no comparison group or control, which makes it plausible that other independent variables might have affected the gains in reading fluency. Furthermore, Kuhn

and Stahl (2003) intended to do a meta-analysis of studies on reading fluency, but because many of the studies were WSD researches it was impossible and they did a review instead. Kuhn and Stahl mentioned that WSD could be good, but reliance on those designs for a whole practice base could be problematic, most likely because they do not use a conventional comparison group design.

Richards, Taylor, Ramasamy, and Richards (1999) state that educational research is most often group/quantitative designs probably because the practitioners in the field have limited knowledge of WSD. This is a possible consequence of one of the misconceptions that exist around WSD. Kazdin (2011) points out that even though WSD have been used for many years there still exist many misconceptions around the designs in textbooks used for undergraduates. The main misconceptions he discusses are that WSD are low in generality, are not true experiments, and can not reveal causal relationships. Dermer and Hoch (1999) discuss many of the misconceptions and compare the research processes of WSD to group designs. They conclude that to increase the predictability at individual and study level it is necessary to stop aggregating data across subjects and experiments, as is done in meta-analysis, and instead maximize the experimental control at single subject level, as is done in WSD. To accomplish this more awareness about the quality of WSD is necessary.

The studies that are done within PT have not been included in the above mentioned meta-analysis and reviews, probably because precision teachers do not use conventional experimental designs. Instead, they use a special semilogarithmic chart, called the standard celeration chart (SCC), to monitor the student's performance. The SCC provides PT with measurements that can be called superior to other measurements that are often practiced within applied behavior analysis (Vargas, 2003). Boyce (2003) suggests that based on the detailed measurement that the SCC brings to PT "precision measurement" is a better and more attractive name for other scientists than PT. In PT many reading fluency studies have

been published that use quality measures and documentation, with the use of the SCC, but lack control, have poorly described procedures, and lack effort to measure generalization to untrained material (e.g., Malanga, 2003; Sulgrove & McLaughlin, 2004; Teigen, Malanga, & Sweeney, 2001). Boyce (2003) said that "our [PT] global measurement system, research design, and methodology lag behind much of what our peers in other areas of behavior analysis are doing" (p. 62). Precision teachers in general do not publish many researches, but they do share their work among themselves and as such they have discovered many behavioral principles (Lindsley, 1990). Binder and Watkins (1990) explain the lack of published research in PT with the fact that precision teachers are mainly practitioners, not academics, and lack the interest and incentives to publish their work. Lindsley (1990) proposed that there should be made more effort to make it attractive for precision teachers to publish their work in journals that are not only attainable for people in the PT field.

There is a need for reading fluency researches in PT and other disciplines that are more methodological rigorous, such that it will be possible to verify whether functional relationship between the dependent and independent variables is present. The remaining of this paper will discuss the logic behind WSDs, how they can be executed to make more methodological rigorous reading fluency studies, and how experimental control in PT studies can be increased.

Within Subject Research Designs

The current use of within-subject designs (WSD), also called N=1 –, single-subject – or single-case designs, can be traced to the work of Skinner on operant conditioning (Kazdin, 2011). WSD are the experimental deigns that are most often used in applied behavior analysis (Cooper et al., 2007) and have been used in many educational experiments and provided improvements in educational practice (Kennedy, 2005). In behavior analysis, the WSD have been used rather than group designs because it is the individual that is of interested not

statistical measures of average behavior and the primary goal of research is to demonstrate experimental control of behavior rather than statistical control (Arntzen, 2010). In WSD repeated measures of the independent variable, within- and between subjects, are taken; prior, during, and after implementation and compared with each other to demonstrate functional relations rather than pure correlations (Arntzen, 2010). They are also used to demonstrate experimental control for a single participant (Kennedy, 2005) were each participant serves as his own control (Johnston & Pennypacker, 2009). WSD are intended to demonstrate experimental control by eliminating the effects of confounding variables with three important elements: prediction, verification and replication (Cooper et al., 2007). Prediction is the idea that the data path of the dependent variable will remain unchanged if the independent variable has no effect. Verification is when a predicted change in the dependent variable after changes in the independent variable has been confirmed. Replication is when predictions and verifications are repeated within the same study; replication is what separates WSD from typical educational settings (Richards et al., 1999).

There are four main categories of WSD; *withdrawal design*, *multiple treatment design*, *changing criterion design*, and *multiple baseline design*. Each design includes different ways to provide prediction, verification, and replication.

Withdrawal Design

Withdrawal design is the most basic design to show experimental control. The independent variable is alternately presented and withdrawn which makes it possible to see how operant behavior can be turned on and off if a functional relationship is present (Kazdin, 2011). Withdrawal design has also been called reversal design; the term withdrawal is more applicable based on the fact that the independent variable is withdrawn rather than reversed in the second baseline phase (Arntzen, 2010). The behavior during the baseline phase provides a prediction of how behavior patterns will continue if nothing is changed, and it is possible to

predict that if the independent variable affects the behavior the pattern will change. Verification is evident when changes are observed in the behavior after the onset of the independent variable, and again if the behavior returns to the same pattern when the independent variable is withdrawn. When the same change is repeated when the independent variable is presented again replication occurs (Richards et al., 1999).

There are two essential premises for using a withdrawal design; the behavior must be reversible and returning the behavior to baseline level has no undesirable effects for the individual (Arntzen, 2010). Reading is not a reversible behavior because it comes under the control of natural reinforcers (comprehension) when it has been established, and withdrawal design is, therefore, not suitable for demonstration of experimental control in reading fluency research. Some aspects of reading can, however, be reversible, for example, is reading speed reversible to some degree and withdrawal design has been used in reading research in combination with, for example, multiple baseline design (Tan, Moore, Dixon, & Nicholson, 1994) and alternating treatment design (Daly, Martens, Hamler, Dool, & Eckert, 1999).

Changing Criterion Design

Changing criterion design is an alternative to the withdrawal design (Kratochwill et al., 2010) it is characterized by that when implanting the independent variable the criterion rate of a single behavior is gradually changed; experimental control is achieved when a change in the behavior rate along with the changes in the criterion is replicated. Each phase in the design functions as a baseline for the next phase (Kazdin, 2011). Each phase does thus provide new prediction of how behavior will be if nothing is changed, verification takes place either when the length of the treatment phases is varied or when criterion levels is reversed and behavior returns to previously set criterion. Replication is demonstrated each time a predictable change in the behavior occurs (Richards et al., 1999).

Even though the design does not demand reversal of the behavior to show experimental control, a reversal of some degree strengthens the demonstration of experimental control. It is possible to use a changing criterion design in reading researches, but to the best of my knowledge it has not been done, probably because it can be especially difficult to implement properly (Cooper et al., 2007).

Multiple Treatment Design

Multiple treatment design is often used to determine which of two or more treatments is most effective in changing behavior of an individual or group of individuals. It is implemented by alternating between interventions while measuring their effect on the target behavior, different ways of presenting the interventions have been given different names; multi-element design, alternating treatment design, simultaneous design, concurrent schedule design, and multiple schedule design (Arntzen, 2010). In general does each data point function as a prediction of behavior within the same treatment condition and each successive data point functions as verification for previous prediction, all successive data points provide replication of the effects of different treatments (Richards et al., 1999). Multiple treatment design suites well in reading fluency research because; it does not demand reversal of the behavior, and can be used to evaluate which technique is most likely to work in fluency training. Ardoin, McCall, and Klubnik (2007) used, for example, alternating treatment design to compare the effects of drill versus practice opportunities on students reading fluency and Daly et al. (1999) did a brief functional analysis on reading fluency of four students with learning problems with the use of multi-element design.

Multiple Baseline Design

In a multiple baseline design, experimental control is evaluated by introducing intervention on different baselines, there are three main types; across behaviors, participants, and settings (Kazdin, 2011). Each baseline provides a prediction of how the behavior will

continue if nothing is changed, and verification of the effectiveness of the independent variable is provided if the data path changes only at the onset of the independent variable. Replication is provided with repeated verification of predictions across different baselines (Richards et al., 1999).

In a multiple baseline design, the intervention is first implemented on the first baseline while baseline measures for other baselines are continued. Experimental control is demonstrated if changes in independent variable occur only when the intervention is implemented on a certain baseline and other baselines remain stable. In this design, the intervention is not withdrawn, the behavior does, therefore, not have to be reversible, and the rate of the behavior increases or decreases as a pure effect of changes in the independent variable (Cooper et al., 2007). This is the most popular design to use in educational researches, Swanson and Sache-Lee (2000) did a meta- analysis of 85 within- subject researches used in special education and found out that 62% of the studies used multiple baseline design. In reading fluency research, it is most relevant to use multiple baseline design across subjects. Taking prolonged and continues baselines can be problematic because it provides reading practice, which is almost always a part of the intervention. Therefore, it would be more suitable to use a variation of the design called *multiple probe design* (R. D. Horner & Baer, 1978). *Probes* are intermittent assessments of the independent variable when the intervention is not in effect, and provides a prediction of how the behavior would be on daily performance (Kazdin, 2011). When using probes instead of continuous baselines the risk of reactivity is prevented or controlled for.

The Use of Within-Subject Designs in Fluency Research

Many reading fluency studies were WSDs are used have been published (e.g., Ardoin et al., 2007; Chafoules, Martens, Dobson, Weinstein, & Gardner, 2004; Daly, Bonfiglio, Mattson, Persampierei, & Foreman-Yates, 2006; Daly & Martens, 1994; Strong, Wehby,

Falk, & Lane, 2004), however, some of the published work does not include all the information needed to qualify as a highly qualified within-subject research according to criteria that have been proposed to document *evidence based practice* (Chard, Ketterlin-Geller, Baker, Doabler, & Apichatabutra, 2009). Evidence based practice/interventions are interventions that have empirical studies that attest to their effectiveness, to be able to establish evidence based practice high quality and well controlled research are required (Kazdin, 2011). To increase the number of researches that are rigorous enough and well enough controlled to use as a base for evidence based practice some experts in WSD have proposed quality standards for the use of WSD in research (R. H. Horner et al., 2005; Kratochwill et al., 2010). Those standards will now be discussed in relation to reading fluency studies also the researches that are done within PT.

Quality Standards for Evidence Based Practice

R. H. Horner et al. (2005) documented the defining features of WSD, and proposed quality indicators for how to use the designs to document evidence based practice in special education. Based on those quality indicators Chard et al. (2009) developed a checklist to evaluate repeated reading (RR) studies, and determine whether RR could be evaluated as evidence based practice. Even though there is a wide research support for the effectiveness of RR (e.g., Chard, Vaughn, & Tyler, 2002; Therrien, 2004), it was not possible to determine RR procedures as evidence based practice based on the lack of quality in existing research.

Most of the items on Chard et al.s' (2009) checklist concern details of execution that are provided in a study. The more details that are provided the easier it is to replicate the experiment. When an experiment is replicated the generality of the results is demonstrated (Sidman, 1960). A panel of experts representing What Works Clearinghouse (a department created by the U.S. Department of Education's Institute of Education Sciences (IES), in 2002 to be a source of scientific evidence for educational practice), have also created standards to

be used to evaluate whether certain WSD research can be used as a foundation for evidence based practice (Kratochwill et al., 2010).

The quality indicators by R. H. Horner and colleagues are divided into seven categories: participants and settings, dependent variable, independent variables, baseline measures, experimental control and internal validity, external validity, and finally social validity.

Participants and Settings

The first category of R. H. Horner's et al. (2005) quality standards concern the description of participants and settings. Accurate description of the participants and the selection process is necessary such that other researchers have a possibility to replicate the research with similar participants and global descriptions would, thus be insufficient (R. H. Horner et al., 2005). According to Chard et al. (2009) most of the researches they examined, provided information about the participants but not sufficient information for the selection process or description of the settings. In reading fluency research, most of the participants have already been diagnosed with a learning disability and relevant information about the diagnoses should be relatively easy to provide. Adequate description of the settings, whether the intervention was implemented in a controlled setting or a classroom, should also be easily provided to meet the criterion.

Dependent Variable

The second category concerns the dependent variable; description, measurement procedures, measurement validity and description, measurement frequency and data collection on reliability (Chard et al., 2009). The dependent variables should be *operationally defined* to make valid and consistent measures possible, and to allow repetition of the measurement process. Operationally defined means; that the variable must be described using concrete terms such that there is a certainty about occurrence or non occurrence by anyone that reads the definition. This concerns the measurability of the behavior and consistency of

measurements (Kennedy, 2005). In reading fluency studies the dependent variable is most often "number of words read correctly per minute", operational definition would provide information about when a word is correctly read and when it is incorrectly read. Repeated measures of the dependent variable are necessary to be able to compare the individual's performance with his/her prior performance and evaluate the behavior change between phases (R. H. Horner et al., 2005). In Kratochwill et al. (2010) it is demanded that each dependent variable is systematically measured over time by multiple observers and *interobserver agreement* must be collected in each phase, if those demands are not fulfilled the study cannot be used as a base for evidence based practice. According to the analysis by Chard et al. (2009) the factors concerning the dependent variable are often well documented in RR studies, but the descriptions of the measurement validity and reliability discussions were sometimes missing.

Interobserver agreement is an outcome of a procedure used to measure the reliability of behavior recordings of human observers (Kennedy, 2005). This procedure is used to measure the consistency of the dependent variable recordings of two independent observers that should be able to observe the same person and agree whether the target behavior occurred, for how long, to what extent, and so on (Richards et al., 1999). Interobserver agreement does not say anything about the accuracy of the recordings just that two observers agree about the results (Kennedy, 2005). Many formulas have been developed to calculate interobserver agreement (Berk, 1979; Kazdin, 2011; Kennedy, 2005), among those methods are *total agreement*, *interval agreement* and *occurrence/nonoccurrence agreement* (Johnston & Pennypacker, 2009). In most reading fluency research, in which measures of interobserver agreement are taken, interval agreement has been used to calculate reliability. When interval agreement is used agreement is scored if both observers record that behavior has taken place during a specific interval, and disagreement if only one observer records behavior in a given

interval. Then total number of agreements is divided by the total number of agreements plus disagreements and multiplied with 100(Johnston & Pennypacker, 2009). There is a problem concerning the use of interval agreement in reading fluency studies; that concerns whether the behavior, reading word correctly, is frequent, which is often the case, then observers are likely to have high proportions of agreements and that gives a wrong idea of reliability of measurements (Kazdin, 2011). For example, if a student reads 163 words, the primary observer records 4 errors, the student has therefore read 159 correct words, observer two agrees on 157 words then the reliability on agreements is 98,7 %. However, there are 4 errors and the two observers only agree on two of them; that gives 50% reliability on errors. If we only present the reliability for occurrence (correct read words) then the results would indicate that there is a good reliability, but if we also present the results for nonoccurrence (errors) then it is only 50% reliability. In reading fluency research, it would be more suited to use occurrence/nonoccurrence agreement to prevent the inflation of interobserver agreement value. When using this type of agreement two outcomes are presented: for agreement of occurrence and agreement of nonoccurrence (Kennedy, 2005). In PT research reliability measures are rarely provided, but in recent years, some researchers have included it in their studies (e.g., Syrek, Hixson, Jacob, & Morgan, 2007). In educational settings it can be convenient to make video or audio tape recordings of sessions, and have an independent observer record, for example, once a week to measure interobserver agreement.

Independent Variable

The third category of the quality standard from R. H. Horner et al. (2005), concerns the independent variable; description, measurement and fidelity to implementation. The independent variable must be operationally defined and a precise description of materials and procedures must be provided to make replication possible. The independent variable should also be actively implemented by the researcher rather than passively, as it would be in natural

settings, to document experimental control. The independent variable in reading fluency studies is often repeated reading (Samuels, 1979), performance feedbacks (Alber-Morgan, Ramp, Anderson, & Martin, 2007), and error correction (Ardoin et al., 2007). Precise descriptions of each component of an intervention must be provided along with a timeline of when different parts are implemented.

R.H. Horner et al. also recommend that *procedural integrity* is recorded to ensure that the experimenter is implementing the procedure correctly. Procedural integrity refers to correct implementation of the independent variable according to the procedure that is described. Procedural integrity is as important as dependent variable reliability and is necessary to be able to demonstrate a functional relationship (Peterson, Homer, & Wonderlich, 1982), if the independent variable is not implemented with integrity it compromises both internal and external validity of the results (Gresham, Gansle, & Noell, 1993). To measure procedural integrity, there must be; an operational definition of the independent variable, a recording system to gather information about the independent variable, a recording standard, and interobserver agreement must be obtained on trained observers in use of the recording system (Kennedy, 2005). A typical measure of integrity involves calculating percentages of correctly implemented steps according to a procedural checklist (Vollmer, Sloman, & Pipkin, 2008). According to the analysis by Chard et al. (2009) there was not provided adequate descriptions of implantation of the independent variable nor measure procedural integrity in most of the RR researches. In PT decisions about changes in intervention are made based on the performance of each individual and thus some might assume that it is not relevant to make integrity protocols. However, if integrity protocols include predetermined criteria of how and when to change intervention based on the performance of individuals it might be possible to approach procedural integrity in PT studies as well as in other reading fluency studies. As with interobserver agreement measures

would video tape of instructional sessions make it easier to gather information about procedural integrity.

Baseline Measures

The fourth category concerns baseline measures; measurement of dependent variable during baseline and the baseline conditions. Baseline measures should be precisely described and taken until a consistent pattern makes prediction of future responding. Baselines are used to approach patterns of behavior prior to intervention and is comprised of several factors; procedure that is the structure and functions of the environment; who interacts with the participant; what is the curriculum; instructional procedure; and finally material (Kennedy, 2005). Half of the studies in the analys by Chard et al. (2009) did not provide sufficient description of the baseline condition. Even though frequent measures generally represent the pattern of behavior better, it would be best to use probe measures instead of continuous baselines in reading fluency studies to prevent reactivity, whereas repetition is most often a part of the fluency building procedure. When probes are used it is necessary to provide information about when and how often probe measures are taken during the baseline phase. In PT baseline measures sometimes only consist of one measure that is taken the first day of intervention or the day before. That does not provide an accurate picture of how the behavior was whether it was stable or trended in a certain direction prior to the intervention. In the standards by Kratochwill et al. (2010) it is demanded that each phase consists of at least three data points to meet the *Standards with reservations* and at least five data points to meet the Standards without reservations. Probe measures could be a good solution for all reading fluency studies and strengthen the experimental control of the research.

Experimental Control and Internal Validity

The fifth category of the quality standards concerns experimental control and internal validity. It concerns whether the design demonstrates experimental effects, how the designs

control for threats to internal validity and the pattern of results. Evaluation of experimental control is done in several ways; number of replications, changes in data from baseline to intervention, when the effect is observed, the size of the effect in respect to baseline, the precision of the description of the procedure, the reliability of measurements, and consistence with existing data (Arntzen, 2010). The number and degree of replications depends on many factors, for example; the experimental context, status of literature, and nature of the independent variable. There is no fixed number of replications that are sufficient to establish functional relationship; the rule is that replication is used to rule out the affects of confounding variables and the findings must convince others of an existing functional relation (Kennedy, 2005). In the standards from the What Works Clearinghouse panel it is demanded that each experimental effect should be replicated at least three times (Kratochwill et al., 2010). Confounding variables can affect the participant's behavior in several ways and are as such called threats to internal validity, examples of threats are; maturation (natural development of individuals), history (the passage of time and all events that might arise), testing, and instrumentation (Shadish, Cook, & Campbell, 2002). In Chard's et al. (2009) analysis the main flaw concerning experimental control was that limited information was provided regarding what the experimenters did to control for threats to internal validity. A good indicator of threats to internal validity is variability in the data that is not counted for by known variables. The data pattern displayed in educational research is more unclear than those displayed in a laboratory setting because there are many confounding variables present that are difficult to control for, but by using, for example, multiple baseline designs many of those threats to internal validity can be controlled for (Kennedy, 2005).

In a reading intervention several things would need to be accomplished to diminish threats to internal validity and demonstrate experimental control. The participants reading rate would have to change when the intervention is implemented and not before, the sooner

the change occurs the more control is demonstrated. In a multiple-probe design, it is crucial that the baseline for other participants does not change when the independent variable has only been implemented for the first participant. In PT there is not much done to control for threats to internal validity apart from close monitoring of behavior, but if the SCC that is used in PT is combined with, for example, multiple baseline design experimental control could be increased. It would also provide more precision to the within-subject charting methods and thus strengthening both methods (Vargas, 2003).

External Validity

The sixth category concerns external validity or replication of effects; that is that effects must generalize across settings and behaviors. When a study is replicated with the same experimental effects, external validity is strengthened. Johnston and Pennypacker (2009) noted that the same results cannot be precisely replicated, but similar results could be obtained by precisely replicating the procedure. There are two different types of replications for establishing external validity: direct and systematic. *Direct replication* refers to using the exact same procedure in the same setting either with the same participants or different participants (Sidman, 1960). Kennedy (2005) stated that without direct replication, there would be no experiment. *Systematic replication* refers to extending initial findings by that the experimenter changes some aspects from previous research, and when similar effects are gained it demonstrates reliability of the findings and adds to the external validity. Aspects that can be altered in systematic replication include among other: subject types, settings, and target behavior (Cooper et al., 2007). When other researchers replicate a research it is most often considered a systematic replication because there are always some small variations in the experimental preparation. Systematic replication is particularly important to demonstrate the generality and boundaries of the functional relation (Kennedy, 2005). Sometimes experimental effects are not replicated when it has been attempted to either directly or

systematically replicate a research. Those results are not often published in peer-reviewed journals, and that could be a result of that those failures raise more questions about the replication attempt rather than the results of the initial research. Repeated failure to replicate a research raises questions about the integrity of the initial research findings (Kennedy, 2005). Chards' et al. (2009) analysis revealed that none of the studies made any attempt to document potential generalization effects to other learning disabled students. In PT researches there is not an emphasis on external validity, probably based on the fact that their purpose is not to design general interventions that work for everyone but to find out what works for each individual and with comparison of SCC from different individuals general principles of learning might be identified (Binder & Watkins, 1990).

Social Validity

The last category concerns *social validity* of the results; the importance of the dependent variable, the magnitude of the behavior change, the practicality and the cost effectiveness of the implementation of the independent variable, and the typical nature of implementation of the independent variable (Chard et al., 2009). Social validity is particularly important in applied research because researchers are working directly with people with various problems. All educational research is applied and takes often place in social settings, the concept of social validity was developed to understand this social context (Kennedy, 2005). Social validity is a subjective measure of the desirability and adequacy of interventions. Wolf (1978) was one of the first to introduce the concept of social validity into the field of applied behavior analysis. He proposed that social validity should be estimated concerning goals, procedures, and outcomes of interventions. These domains have since then been the base for systematic study of subjective data. However, there have not yet been established standardized methods for evaluation of social validity, but as Baer and Schwartz (1991) have pinpointed it is difficult to determine its importance without such standardized methods.

Subjective evaluation (Kazdin, 2011) and *sustainability* (Kennedy, 2005) are among methods that can be used to evaluate social validity in reading fluency studies. Sustainability is the most recent method and refers to the degree of maintenance of effect over some time. It derives from the idea that "if an intervention is sustained over time, it must have some qualities that are consistent with what is meant by social validity" (Kennedy, 2005, p. 229). Sustainability has good face validity; if participants maintain the intervention by them self then there must be something reinforcing about the intervention. It is, however, limited because it demands extended periods of time before data can be gathered and other factors might maintain the gains of the intervention (Kennedy, 2005). If there is not a time for sustainability measures then subject evaluation method is an option, it involves that the participant, social significant persons, and/or specialists regarding the target behavior are involved when goals, procedures and outcomes of the intervention are evaluated. Information is gathered with questioners, multiple choice questions, and structured- or open end interviews (Kazdin, 2011). The data are analyzed either with descriptive statistics or qualitative analysis; it is most important that the process and presentation represent the characteristics of the data in the best possible way. The limitation with the use of this method is that questions are often designed to get favorable answers and people's perception of behavior change might not reflect the real change of behavior, but it can be a helpful in adding qualitative information to data based on experimental analysis of behavior (Kennedy, 2005). According to Chard et al. (2009) analysis social validity was often reported and seems, therefore, not to be a problem in reading fluency studies. Most studies within PT are inherently high in social validity because the target behavior is chosen based on its need for improvement and changes in intervention are based on the students performance and the goal is often set as doubling the rate of behavior in one week (Binder & Watkins, 1990).

Summary

There are a number of reading fluency studies that have been published, but there is still a need for research that is of high quality. Many of the researches that have been published are correlation or have methodological flaws, which weaken the research base of reading fluency interventions. The lack of high quality research in the reading fluency literature might be a result of that most educational researchers, including precision teachers, have another purpose for conducting a research than, for example, behavior analysts. There are many different reasons for practitioners to conduct an experiment; Sidman (1960) mentioned as an example five reasons; to evaluate hypothesis, indulge the investigators curiosity, to try out new methods or techniques, to establish the existence of behavioral phenomenon, and to explore the conditions under which behavior occurs. Even though researchers have different reasons for conducting a research, sharing a successful research such that others can replicate it should always be the goal (Kennedy, 2005). Most of the criterion for high quality within-subject research has to do with explicit descriptions of what is done to make replication possible. It is not sufficient to have a strong experimental design if the standards are not fulfilled; and without possibility of replication of results the research does not benefit others. If the high quality criteria that are outlined by R. H. Horner and colleagues (2005) and by the What Works Clearinghouse panel (2010) are considered when executing a WSD more methodological rigorous reading fluency studies are made and the research base would grow stronger. With a strong research base a possibility for evidence based intervention increases. Experimental control in PT studies could also be strengthened if WSD would be included in the procedure that is used. Future literacy research should focus on improving the existing research base with more high quality research with more emphasis on experimental control and external validity.

References

Alber-Morgan, S. R., Ramp, E. M., Anderson, L. L., & Martin, C. M. (2007). Effects of repeated readings, error correction, and performance feedback on the fluency and comprehension of middle school students with behavior problems. *The Journal of Special Education, 41*, 17-30. doi: 10.1177/00224669074100102O1

Ardoin, S. P., McCall, M., & Klubnik, C. (2007). Promoting generalization of oral reading fluency: Providing drill versus practice opportunities. *Journal of Behavioral Education, 16*, 55-70. doi: 10.1007/s10864-006-9020-z

Arntzen, E. (2010). Experimentelle design, med spesial vekt på ulike typer av N=1 design. In E. Arntzen & J. Tolsby (Eds.), *Studenten som forsker i utdanning og yrke*. Kjeller: Hogskolen i Akershus.

Baer, D. M., & Schwartz, I. S. (1991). If reliance on epidemiology were to become epidemic, we would need to assess its social validity. *Journal of Applied Behavior Analysis, 24*, 231-234. doi: 10.1901/jaba.1991.24-231

Berk, R. A. (1979). Generalizability of behavioral observations: A clarification of interobserver agreement and interobserver reliability. *American Journal of Mental Deficiency, 83*, 460-472. Retrieved from www.psycnet.apa.org/psycinfo/1979-25026-001

Binder, C., & Watkins, C. L. (1990). Precison teaching and direct instruction: Measurably superior instructional technology in schools. *Performance Improvement Quarterly, 3*, 74-96. doi: 10.1111/j.1937-8327.1990.tb00478.x

Boyce, T. E. (2003). Moving from precision teaching to precision measurement: We need to just DO IT. *European Journal of Behavior Analysis, 4*, 59-86. Retrieved from www.ejoba.org

Chafoules, S. M., Martens, B. K., Dobson, R. L., Weinstein, K. S., & Gardner, K. B. (2004). Fluent reading as the improvement of stimulus control: Additive effects of performance-based interventions to repeated reading on students' reading and error rates. *Journal of Behavioral Education, 13*, 67-81. doi: 10.1023/JOBE.0000023656.45233.6f

Chard, D. J., Ketterlin-Geller, L. R., Baker, S. K., Doabler, C., & Apichatabutra, C. (2009). Repeated reading interventions for students with learning disabilites: Status of Evidence. *Exceptional Children, 75* (3), 263-281. Retrieved from www.cec.sped.org

Chard, D. J., Vaughn, S., & Tyler, B-J. (2002). A synthesis of research on effective internvetions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities 35*, 386-406. doi: 10.1177/00222194020350050101

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2 ed.). Upper Saddle River, NJ: Merill Prentice Hall.

Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator, 22,* 8-15. Retrieved from http://www.aft.org/newspubs/periodicals/ae/index.cfm

Daly, E. J., Bonfiglio, C. M., Mattson, T., Persampierei, M., & Foreman-Yates, K. (2006). Refining the experimental analysis of academic skills deficits: Part II. Use of brief experimental analysis to evaluate reading fluency treatments. *Journal of Applied Behavior Analysis, 39,* 323-331. doi: 10.1901/jaba.2006.13-05

Daly, E. J., & Martens, B. K. (1994). A comparison of three interventions for increasing oral reading performance: Application of the instructional hierarcy. *Journal of Applied Behavior Analysis, 27,* 459-469. doi: 10.1901/jaba.1994.27-459

Daly, E. J., Martens, B. K., Hamler, K. R., Dool, E. J., & Eckert, T. L. (1999). A brief experimental analysis for identifying instructional components needed to improve oral reading fluency. *Journal of Applied Behavior Analysis, 32,* 83-94. doi: 10.1901/jaba.1999.32-83

Dermer, M. L., & Hoch, T. A. (1999). Improving descriptions of single-subject experiments in research texts written for undergratuates. *The Psychological Record, 49,* 49-66. Retrieved from http://thepsychologicalrecord.siuc.edu/

Fabrizio, M. A., & Moors, A. L. (2003). Evaluating Mastery: Measuring instructional outcomes for children with autism. *European Journal of Behavior Analysis, 4,* 23-36. Retrieved from www.ejoba.org

Gresham, F. M., Gansle, K. A., & Noell, G. H. (1993). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis, 26,* 257-263. doi: 10.1901/jaba.1993.26-257

Horner, R. D., & Baer, D. M. (1978). Multiple-probe technique: A variation of the multiple baseline. *Journal of Applied Behavior Analysis, 11,* 189-196. doi: 10.1901/jaba.1978.11-189

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71,* 165-179. Retrieved from www.cec.sped.org

Johnston, J. M., & Pennypacker, H. S. (2009). *Strategies and tactics of behavioral research* (3rd ed.). New York: Routledge.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80,* 437-447. doi: 10.1037//0022-0663.80.4.437

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2 ed.). New York, NY: Oxford University Press.

Kennedy, C. H. (2005). *Single-case designs for educational research*. Boston: Allyn & Bacon.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from What Works Clearinghous website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology, 95*, 3-21. doi: 10.1037/0022-0663.95.1.3

Lindsley, R. O. (1990). Our aims, discoveries, failures, and problem. *Journal of Precision Teaching, 7*, 7-17. Retrieved from www.celeration.org

Lokke, G. E. H., Lokke, J. A., & Arntzen, E. (2008). Precision teaching, frequency -building and ballet dancing. *Journal of Precision Teaching and Celeration, 24*, 21-27. Retrieved from www.celeration.org

Malanga, P. (2003). Using repeated readings and error correction to build reading fluency with at risk elementary students. *Journal of Precision Teaching and Celeration, 19*, 19-27. Retrieved from www.celeration.org

National Institute of Child Health and Human Development. (2000). Report of the national Reading Panel. *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* Report of the subgroups (NIH Publication No. 004769). Retrieved from http://www.nationalreadingpanel.org/Publications/subgroups.htm

Peterson, L., Homer, A. L., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis, 15*, 477-492. doi: 10.1901/jaba.1982.15-447

Rasinski, T. V. (2004). Assessing reading fluency. *Pacific Resources for Education and Learning*. Retrieved from http://www.prel.org/media/137837/10_assessing-fluency.pdf

Richards, S. B., Taylor, R. L., Ramasamy, R., & Richards, R. Y. (1999). *Single subject research: Applications in educational and clinical settings*. Belmont: Wadsworth.

Rolling, K. M. (2009). Presisjonsopplæring (PO) i opplæring i påkledning for jente med ADHD. *Norsk Tidskrift for Atferdsanalyse, 36*, 39-50. Retrieved from www.nta.atferd.no

Samuels, S. J. (1979). The method of repeated readings. *The Reading Teacher, 32*, 403-408. Retrieved from http://www.jstor.org/stable/2019790

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin Company.

Sidman, M. (1960). *Tactics of scientific research: evaluating experimental data in psychology*. New York: Basic Books.

Strong, A. C., Wehby, J. H., Falk, K. F., & Lane, K. L. (2004). The impact of a structured reading curriculum and repeated reading on the performance of junior high students with emotional and behavioral disorders. *School Psychology Review, 33,* 561-581. Retrieved from www.nasponline.org

Sulgrove, M. K., & McLaughlin, T. F. (2004). The effect of an additional timed reading on reading rate. *Journal of Precision Teaching and Celeration, 20,* 9-16. Retrieved from www.celeration.org

Swanson, H. L., & Sachse-Lee, C. (2000). A meta-analysis of single-subject-design intervention research for students with LD. *Journal of Learning Disabilites, 33*, 114-134. doi: 10.1177/002221940003300201

Syrek, A. L., Hixson, M. D., Jacob, S., & Morgan, S. (2007). A comparison of error correction procedures on word reading. *Journal of Precision Teaching, 23,* 2-13. Retrieved from www.celeration.org

Tan, A., Moore, D. W., Dixon, R. S., & Nicholson, T. (1994). Effects of training in rapid decoding on the reading comprehension of adult ESL learners. *Journal of Behavioral Education, 4*(2), 177-189. doi: 10.1007/BF01544111

Teigen, T., Malanga, P. R., & Sweeney, W. J. (2001). Combining repeated readings and error correction to improve reading fluency. *Journal of Precision Teaching and Celeration, 17*, 58-67. Retrieved from www.celeration.org

Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education, 25,* 252-261. doi: 10.1177/07419325040250040801

Vargas, J. S. (2003). Precision teaching and Skinner's legacy. *European Journal of Behavior Analysis, 4*, 80-86. Retrieved from www.ejoba.org

Vollmer, T. R., Sloman, K. N., & Pipkin, P. (2008). Practical Implications of data reliability and treatment integrity monitoring. *Behavior Analysis in Practice, 1,* 4-11. Retrieved from http://www.abainternational.org

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*, 203-214. doi: 10.1901/jaba.1978.11-203

Master thesis, Learning in Complex Systems

Autumn 2011

Delayed Multiple Probe Design to Evaluate the Effects of a Multicomponent Intervention to Improve Reading Fluency in Adult Students

Hildur Valdimarsdottir

Lilja Yr Halldorsdottir

Department of Behavioral Science

HØGSKOLEN I OSLO
OG AKERSHUS

Abstract

It is socially important for people who read slowly to improve their reading fluency since reading is an essential part of almost every aspect of daily life. The literature for building reading fluency consists of many studies that are often non-experimental or not of high quality. The purpose of the present study was to implement a high quality study, based on quality indicators proposed by Horner et al. (2005), by using a delayed multiple probe design across participants to evaluate the effects of a multicomponent intervention on reading fluency in adult Norwegian students. Prior to the study all the participants read below average reading speed and one had the diagnosis dyslexia. The procedure was based on a study by Lokke, Lokke, and Arntzen (2009) and consisted of a reading support, performance criterion, response prompts, repeated reading, performance feedback, and error correction procedure. Training consisted of 1-min timings implemented every weekday and each passage was read four times in each session or until a predetermined performance criterion was reached, whichever occurred first. The performance criterion for each participant was set as 1,5x acceleration or 50% increase, from average reading speed on baseline. When the performance criterion was reached a new passage was introduced in the following session. After the intervention, baseline passages were represented to evaluate generalization to untrained passages. In addition, other aspects of reading were tested to evaluate learning outcomes of fluent performance. The main findings indicate that reading fluency of all the participants improved after receiving the intervention. Results are discussed with regard to social validity of the study, limitations of conducting a multicomponent intervention, and in terms of whether the present study can be considered as a high quality within-subject experiment according the quality indicators by Horner et al. (2005).

Keywords: Multicomponent, multiple probe design, quality indicators, reading fluency, repeated reading, RESAA, stimulus control

Delayed Multiple Probe Design to Evaluate the Effects of a Multicomponent Intervention to Improve Reading Fluency in Adult Students

In a modern society which requires its members to have good reading skills in almost every aspect of daily life it is socially important for people who read slowly to improve their *reading fluency* to prevent them from falling behind. Reading fluency includes reading fast and accurately (Binder, 1988) and is generally considered one of the most important components of reading. Many studies concerning improvement of reading fluency have been published across different disciplines (e.g., Alber-Morgan, Ramp, Anderson, & Martin, 2007; S.P. Ardoin, Eckert, & Cole, 2008; Carroll, McCormick, & Cooper, 1991; Chafouleas, Martens, Dobson, Weinstein, & Gardner, 2004; Conte & Hintze, 2000; Hughes, Beverley, & Whitehead, 2007; Lokke, Lokke, & Arntzen, 2009; Polk & Miller, 1994; Samuels, 1979; Teigen, Malanga, & Sweeney, 2001).

Precision teachers have been tenacious in fluency studies, but *precision teaching* (PT) is a behavioral analytical method rooted in Skinner's free operant conditioning laboratories. Before fluency building in PT starts *fluency aims* are set but the aims predict the response rate at which the student will achieve specific learning outcomes that have been identified with various acronyms, for example *RESAA* (Johnson & Layng, 1996). RESAA stands for; *retention* (maintenance of performance between two points in time without practice), *endurance* (the ability to perform for prolonged periods of time), *stability* (the performance is not easily distracted), *application* (a composite behavior is more easily acquired when its subcomponents are fluent), and *adduction* (a fluent performance facilitates the occurrence of a non-trained skill), which are the characteristics of fluent responding (Johnson & Layng, 1996). Fluency building in PT typically consists of short repeated daily timings (sprints) of the target behavior until fluency aims have been reached. The target behavior is a *free operant* measured with *rate* (i.e., frequency or count per minute) of responses rather than

percent correct based on the notion that rate is more sensitive to changes in the environment than percents (Lindsley, 1992). The student records his/her own response rate on a *standard celeration chart* (SCC) and changes in rate over time are denoted as *celeration* (derived from acceleration and deceleration). Instructional decisions are based on systematic evaluation of the student's performance that is displayed on the chart. A typical sprint is 1-min timing, but timings can last for longer or shorter intervals. It has been documented (Binder, Haughton, & Eyk, 1990) that sprints even shorter than 1-min can help building fluency if longer timings have not been effective. After achieving fluency at a shorter interval the timing period can be increased systematically to build endurance (Binder et al., 1990).

The methods typically used in reading fluency building in PT resemble in many aspects *oral repeated reading* (RR) *interventions*, which are the most referred fluency building strategies used across distinct disciplines (e.g., Kuhn & Stahl, 2003; Therrien, 2004). One of the oldest oral RR intervention is Samuels' *method of repeated readings* (Samuels, 1979), involving that the same passage is read repeatedly until a certain *performance criterion* is reached, then a new passage is introduced and the same procedure implemented again. Ardoin et al. (2008) explained the improvement in reading fluency after RRs such that the behavior is brought under greater stimulus control of the text, where repeated exposure to a text establishes each word as a *discriminative stimulus* (S^D) for the next word. Many modified versions of Samuels' method have been developed, for example in which a passage is read repeatedly for fixed number of times without performance criterion (e.g., S.P. Ardoin, Williams, Klubnik, & McCall, 2009). However, studies have shown that including performance criterion provides better results when building reading fluency (e.g., Chard, Vaughn, & Tyler, 2002; Therrien, 2004). Other versions include different multicomponent intervention packages that involve RR. Most often those packages include *response prompts* (e.g., Alber-Morgan et al., 2007), some kind of *performance feedback* (e.g., Alber-Morgan et

al., 2007), and some type of *error correction procedure* (e.g., S. P. Ardoin, McCall, & Klubnik, 2007).

Response prompts are additional antecedent stimuli that operate directly on the target response to aid correct responding in the presence of a specific S^D (Cooper, Heron, & Heward, 2007). In RR studies response prompts are verbal instructions that are delivered to the reader prior to each reading (e.g., Alber-Morgan et al., 2007; S. P. Ardoin et al., 2007). In some publications the instructions have not been specified (e.g., Chafouleas et al., 2004; O'Shea, Munson, & O'Shea, 1984; Teigen et al., 2001) but it has been documented that the content of the instructions can be important with regard to the purpose of the given study (e.g., Binder & Watkins, 1990; Samuels, 1979; Therrien, 2004). For example, the practice of PT has revealed that emphasizing accuracy can prevent the student from progressing and emphasizing speed has been shown to result in increased learning (Binder & Watkins, 1990).

Performance feedback in RR interventions involves providing information about the reader's performance immediately after reading a passage and it has been documented to positively affect oral reading fluency (Conte & Hintze, 2000; Eckert, Ardoin, Daly, & Martens, 2002). Performance feedback is most often provided orally but can also be delivered graphically. In PT, oral- and graphic performance feedbacks are combined such that the student gets oral information about his/her fluency score and then charts it on a SCC (Binder, 1996).

Studies have shown that RR with some kind of error correction procedure is superior to RR without error correction to improve reading fluency (Chard et al., 2002; Hardardottir, 2007; Therrien, 2004). W*ord drill* and *phrase drill* are examples of error correction procedures that have been shown to be effective (O'Shea et al., 1984). When using phrase drill the reader is provided with contextual cues (i.e., the words surrounding the error word) to recognize the target word in context and thereby increases stimulus control of the text

above when a single word is drilled. Errors can either be drilled right from the text or by using physical- or electronic flashcards (Levy, Abello, & Lysynchuk, 1997; O'Shea et al., 1984). With persistent errors, using *stimulus prompts* can possibly increase stimulus control. To the best of our knowledge, stimulus prompts have not been used as a part of error-correction procedure in fluency studies but it would be interesting to incorporate such prompts when the reader repeatedly makes the same errors.

Another strategy that has been developed to improve reading fluency is *speed reading* but that is a collection of techniques that are generally accepted in popular science. One of the techniques is called *meta-guidance* and it consists of using a reading support, such as a finger or a pen, to control the eye movements of the reader (Stangeland & Forsth, 2001). According to Stangeland and Forsth (2001), low reading rate is often related to high number of eye-fixations and they claim that meta-guidance can be helpful in controlling the reader's eye movements. Even though it is often claimed that speed reading techniques are effective few studies have been published to manifest it. However, speed reading studies that have been published have documented remarkable effects (e.g., Calef, Pieper, & Coffey, 1999; Macalister, 2010; Stangeland, 1998). Unfortunately, the description of the procedure in published studies is often deficient which raises problems for demonstration of external validity. In addition, internal validity is often threatened because those studies are often *non-experimental* (e.g., Calef et al., 1999; Stangeland, 1998) and the same applies to other types of reading fluency studies (e.g., Legault, Maloney, & Giroux, 2001; Malanga, 2003; Teigen et al., 2001). However, in some reading fluency studies *within-subject designs* (e.g., Bonfiglio, Daly, Martens, Lin, & Corsaut, 2004; Chafouleas et al., 2004; Daly & Martens, 1994; Valleley & Shriver, 2003) or group designs (e.g., Berends & Reitsma, 2006; Tan & Nicholson, 1997) have been used but despite of that many of the studies still have some methodological flaws (Chard, Ketterlin-Geller, Baker, Doabler, & Apichatabutra, 2009; Lyon

& Moats, 1997) which results in difficulties with replication and evaluation of *experimental control*. Horner et al. (2005) proposed some quality indicators for within-subject designs to identify *evidence-based practice* in special education. Those indicators include description of participants, settings and baseline, description and measurement of the dependent variable, description and manipulation of the independent variables, along with evaluation of internal-, external-, and social validity (see details in Horner et al., 2005). Chard et al. (2009) created a 4-point Likert scale (see pp. 269-270) to evaluate to which degree experimenters fulfill the quality indicators of Horner et al. They applied the Likert scale to examine the quality of a sample of within-subject design studies that used RR with participants with, or at risk for, learning disabilities (LD). Even though it has been claimed that RR is an evidence-based strategy (e.g., Therrien, 2004), Chard et al. found out that no study that they examined met the criterion for being of a high quality and thus concluded that RR could not be considered as an evidence-based practice for students with LD.

In the present study the rating scale of Chard et al. will be used to evaluate to which degree the experiment fulfills the quality indicators of Horner et al. The main purpose of the study is to improve reading fluency in adult students with the use of a multicomponent intervention consisting of a reading support, performance criterion, response prompts, repeated reading, performance feedback and error correction procedure. The procedure is based on a study by Lokke et al. (2009). They used PT methods to evaluate the effects of an intervention package combined of speed reading techniques (meta-guidance) and repeated reading with an error correction procedure, to improve reading fluency of a 14-year-old boy with reactive attachment disorder, with promising results. The procedure of the present study differs in several ways from Lokke et al.'s procedure. First, another type of error correction procedure will be used. Even though Barbetta et al. (1994) found out that immediate error correction (such as Lokke et al. used) gave better results in reducing errors than delayed error

correction, we choose to use delayed error correction because as Binder and Watkins (1990) highlighted emphasizing accuracy above speed can prevent learning. In addition, the average reading rate of the participants in the present study is faster than in the study by Lokke et al. and therefore it is considered more difficult to correct errors during reading. Stimulus prompts will also be used with persistent errors. Second, Lokke et al. used non-experimental design and SCC to evaluate the progress, but even though we use a procedure that is similar to methods often used in PT we choose to use *delayed multiple probe design* (Cooper et al., 2007) across participants, both for evaluation of the participants' progress and to base instructional decisions on, in an attempt to increase experimental control. The SCC will only be used as a part of the performance feedback, but not as an evaluation tool to base instructional decisions on. This will limit the possibility of making changes in the procedure on individualized level but in an attempt to deal with this limitation some predetermined criteria will be set to decide when to implement different parts of the intervention allowing changes based on each individual performance. In other respects, the procedure of the present study is similar to the procedure of Lokke et al. Additional reading measurements will also be taken to evaluate learning outcomes of fluent performance or what precision teachers talk about as RESAA.

Method

Participants

Three Norwegian speaking bachelor students at the university participated in the study. The participants did not know each other and had no contact prior or during the experiment. All the participants volunteered to participate after having seen an announcement at the university website where students with reading problems were requested. Before the volunteers were accepted as participants each of them was interviewed about relevant background information concerning his/her reading problems and asked about preferences

for reading material to be used in the experiment. After the interview each volunteer read a passage based on his/her reading preferences and the passage was used as a reading test to determine reading speed. The criterion for participation was reading slower than 150 words per minute on the reading test but fluency is often achieved when the reading rate is between 150–250 words per minute with 0–2 errors (Kubina & Starlin, 2003). If the volunteer met this criterion baseline probe measures started for him/her immediately. The first volunteer that met the criterion was assigned as Participant 1 and was first to receive intervention.

Participant 1. Josh, a 35-year-old male, was diagnosed with dyslexia, according to the *Aston Index test,* in high school by the *Educational Psychological Service* in Norway (*Pedogogisk-psykologisk tjeneste*/PPT). In the interview, Josh reported that his slow reading both affected his study and his free time reading, that is, he never succeeded to read through the entire reading list in school and almost never read in his free time due to his slow reading. When taking exams in school, Josh was permitted to use computer with a correction program and was allowed extra time. For Josh, the reading test in the experiment was a passage from a crime novel and he read 144 words per minute.

Participant 2. Rose, 22-year-old female. In the interview Rose reported that she had gone through a diagnostic process for dyslexia via the student foundation at the university (*Studentsamskipnaden i Oslo og Akershus*/SIO) because of reading problems and suspicion of dyslexia, but according to *LOGOS* diagnostic test she did not reach the criteria. Rose informed that her reading problem consisted of slow reading, “jumping” back and forth in the text, and lack of reading comprehension and she meant that those problems had affected her grades in school. In college, she had been allowed extra time when taking exams and at the university some of the exams were read aloud to her. In the experiment, the reading test for Rose was a passage from her current reading list and she read 138 words per minute.

Participant 3. Annie, 42-year-old female. In the interview Annie informed that when she was in college she had gone through dyslexia diagnostic process performed by the PPT, but did not reach the criteria for diagnosis of dyslexia according to the Aston Index test. Annie reported that due to her slow reading she never had time to read through the entire reading list in school and thus she wanted to increase her reading speed. The reading test in the experiment was a passage from her current reading list and she read 142 words per minute.

None of the participants had tried any methods to increase their reading speed.

Before the experiment started the participants received information about the study. The information included; the purpose of the study (i.e., increasing reading speed and accuracy), methods of data collection and registration, that the study was a part of the experimenters' master theses at the university, voluntary participation and anonymity, that the experimenters were bound to secrecy, and that the study was reported to the *Norwegian Social Science Data Services* (NSD). All of the participants signed an informed consent (Appendix A).

Apparatus and Settings

All the reading passages were texts on white A4 sheets with the font Verdana, size 14 pt. and 1.5 spacing. Each passage was around 300 words to ensure reading to be a free operant.

A digital timer was used to time 1-min or 30 s reading sprints and all sessions were recorded on a digital dictaphone. A computerized PowerPoint 2007 slideshow was used to practice errors. The slideshow was programmed such that each slide was present until pressed Enter and it looped continuously until pressed Esc.

Errors were recorded with a pencil on the experimenter's own exemplar of the target passage. Microsoft Word 2007 documents of each passage were used to find out number of words read each time. In addition, a data recording sheet (Appendix B) was used to record all important data from each session; the participants' initials, date, the name of the passage,

words read per minute (WRM), number of words read correctly per minute (WRCM), number of errors, what errors were made, type of error correction procedure, type of reading support, and the initials of the experimenter. Daily per minute SCCs were used to record the best reading of each session for each participant.

In some instances the reading sessions were executed through the Internet via Skype ™ communication program and a hardwired web cam was used to monitor the participant's use of the reading support.

All the phases of the experiment took place in a quiet room that included at least one table and two chairs, so that the experimenter and the participant had a proper working situation, and a laptop to record data and train errors. An exception from this setting was when sessions were implemented through Skype, but then the experimenter and the participant were in separate quiet rooms similar to the other setting.

Dependent Variable

The dependent variable was reading fluency, that is, speed and accuracy, measured with WRCM and errors. A word was scored correct if there was a point-to-point correspondence between the textual stimuli and the vocal response. Self-corrections were also scored as correct. A vocal response without point-to-point correspondence to the textual stimuli was scored as an error, including word omissions and word insertions. WRCM were calculated by subtracting the errors from the total number of WRM.

Design

A delayed multiple probe design across participants was used to evaluate the effects of a multicomponent intervention package that consisted of a reading support, performance criterion, response prompts in form of vocal instructions (faded in a predetermined way during the intervention, see procedure), repeated reading, performance feedback, and an error correction procedure composed of a drill and stimulus prompt, on reading fluency. All the

parts of the intervention package were implemented concurrently when the intervention started for each participant. The design was delayed such that baseline measures did not start concurrently for the participants but as soon as each of them was accepted as a participant in the experiment. The intervention started for Participant 1 when three acceptable participants had been approved for the project. For Participant 2 the intervention started two weeks later and four weeks later for Participant 3. After the intervention phase the intervention package was withdrawn twice, with three months apart, where the procedure used in the baseline phase was repeated. The last phase of the design was follow-up.

Procedure

The type of reading material that was used in baseline-, intervention-, and withdrawal phases was decided in cooperation with each participant. The experimenters selected reading passages by convenience based on each participant's reading material preferences. All the passages used in the experiment were in the participants' native language, Norwegian. The passages that were used in baseline- and withdrawal phases were given the names Test probes 1–5 and the passages that were used in intervention- and follow-up phases were named Passages 1–10. All the passages that each participant read during baseline-, intervention-, and withdrawal phases were from the same book/article and had therefore greater overlap in content words than if they had been from different books/articles. The test probes were also used in the withdrawal phases to evaluate generalization to untrained passages. To make sure that all the passages which each participant read were of similar difficulty level the experimenters estimated the difficulty of the passages based on average word length of each passage (Evensen & Vagle, 2003). The passages selected for Josh were all from a crime novel, and were customized to have the average word length between 4.2–4.8 letters per word (LW). The passages for Rose and Annie were from separate articles on

their reading lists at the university and were customized to have average word length between 5.0–5.7 LW for Rose and between 5.2–5.6 LW for Annie.

Baseline probes. In each baseline session the experimenter set the timer to one minute, gave the participant a test probe and the following instructions:"Read out loud as fast as you can until the timer rings. Read all the words and try not to do errors. You can start when you are ready." When the participant started reading, the experimenter started the timer and followed the reading on his own exemplar of the text and marked errors if occurred. Then the experimenter registered WRCM and number of errors without any comments to the participant.

The duration of the baseline phase was four months for Josh, three months for Rose, and one and a half month for Annie. Five probes, with varied interval, were taken during the baseline phase for each participant. Because both the instructions used on baseline and 1-min timings were also part of the intervention package the number of probes was limited to five to prevent *practice effect* (Cooper et al., 2007), provisionally that the baseline was stable or trended downwards. The probes were used as pretests for the withdrawal phases.

Intervention. On the first intervention day each participant was informed about his/her personal performance criterion which was predetermined as 1.5x acceleration of the average reading rate (which is a 50% increase in rate) during baseline rounded to the next whole number with three or less errors for each passage. The experimenter gave the participant a choice between different types of reading supports that is; a pen, a Mikado pin, a knitting needle, or own index finger. All the participants choose to use their index finger as a reading support in all the sessions, except for Josh who used a pen in the first session but his finger in all the other sessions. The experimenters roughly explained the main characteristics of the SCC.

In the beginning of each session eye movements were trained with the reading support. The participant received one page with a text that was irrelevant for the fluency building and was instructed how to use the reading support. A correct use was defined as holding the reading support under the first line of the passage, moving it from left to right, having the eyes focused on the text, and letting them follow the support but not vice versa. At the end of each line the support was supposed to be moved fast and gently, with the eyes following it, to the beginning of the next line and then continuing with the same procedure through the whole text. The participant did the reading support training three times before the first reading in the first three sessions with instructions from the experimenter to increase the speed each time. To maintain training of correct eye movements reading support training was also done at least once in the beginning of all other sessions.

After the reading support training the experimenter set the timer to one minute and gave the participant relevant passage and instructions. Before the first reading of the first three sessions the instructions were:

> Read out loud until the timer rings. Read as fast as you can, read all the words and try not to do errors. Hold the reading support under the line you are reading, move it a little bit faster than you read such that your eyes follow the support but not vice versa. You can start reading when you are ready.

Before the first reading of other sessions the instructions emphasized speed, accuracy, and correct use of the reading support, but prior to all other readings it was sufficient to only emphasize speed. As soon as the participant started reading the experimenter started the timer and followed the reading on his own exemplar of the text, marked errors if occurred without the participant seeing the markings, and marked how far the participant read.

After the 1-min timing, the experimenter gave the participant performance feedback about total WRM, accuracy, and speed contingent on his reading and recorded it on a

registration sheet. If the participant had made any errors in the reading sprint the experimenter informed him/her immediately about the errors and modeled them correctly. Before rereading the passage, errors from previous reading sprint were drilled, that is, the participant repeatedly read the errors. Phrase drill was used if the error was an incorrectly read word from another word class than noun, a word insertion, or a word omission. In phrase drill a phrase of two to four words from the text, containing the error, was drilled. However, if the error was a noun that was read incorrectly then word drill was considered sufficient. If the participant made three or less errors, each word/phrase was drilled right from the text until read correctly five times successively. However, if the number of errors exceeded three, the words/phrases were drilled on PowerPoint slides in a computer, until all the words/phrases were correctly read five times successively.

After the drill or performance feedback, whichever was relevant (dependent on whether errors were made in previous reading or not), the same passage was read again (RR) and was followed by the same procedure; performance feedback, pointed at errors, modeling, and error correction, whichever was relevant each time. If the participant made any of the same errors in two sequential readings a stimulus prompt was provided by underlining the particular error with a pencil after the drill. Then the passage was read again with the prompt present. The prompt was erased from the text when the target word/phrase had been read correctly.

When a passage was read for the first time the session consisted of only one reading (Polk & Miller, 1994), performance feedback, and modeling of errors. Other sessions also included error correction procedure and RR, where the whole procedure was repeated four times or until the predetermined criterion was reached, whichever occurred first. When the criterion for a passage had been reached, the number of times the participant had read the passage was recorded on the registration sheet and a new passage was introduced in the

following session. The whole procedure was replicated with ten different passages for each participant. If the same passage had been read 15 times without reaching the performance criterion the same procedure was used with 30 s timings, instead of 1-min timings, until the criterion was reached. When training of all the ten passages was finished the passages that had been trained with 30 s timings were introduced again for 1-min timings until the criterion was reached again.

In the end of each session the participant recorded the best reading of the session on a SCC. Intervention sessions took place once a day on working days for all the participants and were 31 sessions for Josh, 24 for Rose, and 38 for Annie.

Withdrawal 1. The first withdrawal phase was implemented right after the intervention ended. A posttest was implemented, that is the test probes from baseline were represented with the same procedure as in baseline to evaluate generalization effects of the intervention to untrained passages.

Withdrawal 2. A second posttest was implemented approximately three months after the intervention phase ended, that is, all the test probes were tested once again with the same procedure as used in baseline and withdrawal phase 1.

Additional reading measurements. Different aspects of reading were also tested to evaluate the characteristics of fluent reading; RESAA.

Follow-up/retention. Follow up measures were used to evaluate retention. For each participant, three random passages from the intervention phase were tested again, once each in a randomized order, with the same procedure as in the intervention phase except without error correction and repeated reading. Follow-up was carried out about three months after the intervention ended.

Application. Application was tested with the same procedure as used during baseline. Each participant read the same passage three times; once before the intervention started

(pretest), once right after the intervention ended (posttest 1), and once approximately three months after the intervention (posttest 2) (exact timing decided by convenient). The passages that were used were more difficult than the passages used in baseline- and intervention phases with the average word lengths of 5.1 LW for Josh, 6.0 LW for Rose, and 6.2 LW for Annie. Those passages were also from another book/article than the other passages and had therefore less overlap in content words.

Endurance and stability. For Annie, one 2 min timing with a random passage of approximately 600 words composed of two sequential intervention passages was used to test endurance. Annie was also exposed to a stability test where she read a random passage that had been used in the intervention for 1-min with the radio playing in the background. Except for the duration of the timing of the endurance test and the noise in the background during the stability test, the procedure for those tests was the same as used in baseline.

Reading comprehension/adduction. Adduction was evaluated with reading comprehension tests. The participants were exposed to two types of comprehension tests; recall measure- and question answering tests. The recall measure test included a 1-min reading of a novel text (that was not related to other passages used in the experiment) and a 1-min recall of facts from the text, measured by the number of correct recalled facts (e.g., Beneke, 1991; McDowell, McIntyre, Owen, & Keenan, 1998; Polk & Miller, 1994). A correct fact was defined as a word, phrase, or a sentence that involved correct information from the text such as names, dates, subject-verb-object relation (e.g., “John ate fish”), and adjective-noun relation (e.g., “yellow car”). Repetitions of correct facts and words or sentences that had no informative value (e.g., “that had been”) and incorrect information were not scored. Before implementation of the recall measure tests the experimenter informed the participant that after reading the text he/she would have 1 min to write down any recalled facts. The experimenter gave the same instructions as before readings in baseline

and after the test he recorded number of WRM and number of correct facts. The recall measure test was implemented immediately after a 1-min timing of a test probe, once during the baseline phase, and another similar recall measure test was carried out once during withdrawal phase 1. The exact timing of the execution of these tests was decided by convenience.

The question answering tests included 20 multiple choice questions from the Norwegian novels *Mayday Mayday* and *Sitt livs chanse* (Stangeland & Forsth, 2001). The experimenter instructed the participant to read the novel silently and informed him/her that after the reading he/she was supposed to answer multiple choice questions from the text. The experimenter timed the reading and recorded number of WRM afterwards. After the reading the experimenter gave the participant 20 written multiple choice questions about the novel, which he/she answered independently by marking with a pencil. The experimenter then recorded the number of correct answers. The question answering tests were carried out immediately after a 1-min timing of a test probe; the test from *Mayday Mayday* was implemented during the baseline phase and the one from *Sitt livs chanse* during the withdrawal phase. The exact timing of the tests was decided by convenience.

Reliability and procedural integrity. Reliability was evaluated by taking interobserver agreement (IOA) scores; a second observer independently recorded number of correct words and errors while listening to an audio tape recording of sessions. Agreement was defined as both experimenter and observer agreeing on whether a word was correctly or incorrectly read. IOA was calculated by dividing the total number of agreements by total number of agreements plus disagreements and multiplied by 100 %. IOA data was collected for each participant by using a randomized sample of 33% of all the sessions in each phase. The average IOA for all the participants across phases was 99% with the range of 97% to 100%.

A procedural integrity protocol (Appendix C) was developed and a procedural integrity checklist (Appendix D) was made based on the protocol. To assess the experimenters' adherence of the procedure, procedural integrity (PI) was observed regarding correct implementation of; instructions, reading support, error correction, recording of time, and performance feedback. A second observer independently scored whether the experimenter correctly implemented the procedure according to the checklist or not by listening to an audiotape of the session. PI was assessed from a randomized sample of 33 % of all the sessions in each phase for all the participants. PI was calculated by dividing the number of steps implemented correctly (by the experimenter) by total number of opportunities and multiplying it by 100 %. The average PI was 97%, with range from 89 % to 100 %. The PI score was only once lower than 90% and that low value occurred when the experimenter failed to start the timer on time before all the readings in a single intervention session. Thus, it is possible that WRM in this particular session was incorrectly recorded.

Social validity. After the last session of the follow-up phase social validity was evaluated based on the three social validity criteria suggested in a milestone article by Wolf (1978). The three criteria concern whether the target behavior is of social importance, whether the procedure is socially accepted, and whether the results are of social importance. The evaluation was carried out such that the experimenters interviewed the participants about their acceptability of the intervention and satisfaction of the results. The Interview contained eight open-ended questions (Appendix E) concerning the participants' likeability of the different components of the procedure, whether they noticed effects of the intervention in general reading (i.e., generalized effects), and whether they planned to continue using some of the intervention components when reading in general.

Results

The average reading rate of the test probes during baseline was 143 WRCM and 2 errors (range 0–5 errors) for Josh, 131 WRCM and 2 errors (range 1–3 errors) for Rose, and 140 WRCM and 1 error. This resulted in performance criterion of 214, 197, and 210 WRCM, for each participant respectively.

In the first intervention session Josh read 120 words correct per minute (WRCM) and 5 errors while using a pen as a reading support, but in the first reading of session 2 he read 172 WRCM and 3 errors while using his finger as a reading support for the first time.

During the intervention phase the average reading rate and number of errors of the first readings of each passage for Josh was 166 WRCM (range 120–179) and 3 errors (range 0–6). However, in the first reading of the intervention Josh read only 120 WRCM which is much lower than all the other scores and without this outlier the average reading rate was 171 WRCM (range 159–179). For Rose the average reading rate and number of errors of the first readings was 168 WRCM (range 149–204) and 2 errors (range 0–3) and for Annie it was 175 WRCM (range 166–192) and 2 errors (range 0–8). Rose and Annie read faster in all the first readings in the intervention phase than during all the test probes on baseline, but Josh read slower in the first reading of Passage 1 and Passage 7 than during baseline but faster in the first reading of all the other passages used in the intervention.

When rereading a passage, the number of WRCM was always higher than in the previous session except in one session for Josh (Intervention session 23) and two sessions for Annie (Intervention sessions 3 and 24). To reach the performance criterion for the passages used in the intervention Josh read each passage at the average of 8.7 times, with eight to nine readings of the four first passages, 19 readings (15x 1-min timings + one 30 s timing + three 1-min timings) of the fifth passage, but only four to five readings of the last three passages. Rose read the passages at the average of 5.9 times to reach the criterion with the range of five

to nine readings for the first eight passages, except for the fifth passage where she reached the criterion in the first reading. Rose reached the criterion for the last two passages in four readings. The average number of readings for Annie was 10.8 times. She needed seven to 15 readings to reach the criterion for Passages 2 to 5, 19 readings (15x 1-min timings + one 30 s timing + three-min timings in both cases) for Passages 1 and 6, and four to five readings for the rest of the passages except for the last one which she read 12 times.

Figure 1 displays number of WRCM and errors during baseline-, withdrawal-, and follow-up phases, and the best reading of each session plus errors during the intervention phase for all the participants.

Insert Figure 1

During the intervention phase a stimulus prompt was used four times for Josh and Rose, and two times for Annie. None of the participants needed stimulus prompt for the same error in two sequential readings.

On posttest 1 during the first withdrawal phase the average reading rate for Josh was 172 WRCM and 3 errors (range 2–4 errors), which equals 1.21x acceleration or increase of 29 words from the average reading rate on the pretests on baseline (i.e., the test probes). For Rose the average reading rate was 178 WRCM and 1 error (range 0–2 errors), that is 1.36x acceleration or increase of 47 words, and for Annie it was 177 WRCM and 1 error (range 1–2 errors), which equals 1.28x acceleration or increase of 37 words between the pre- and posttests.

On the second posttest during the later withdrawal phase Josh had the average reading rate of 173 WRCM and 3 errors (range 2–5 errors) which equals acceleration of 1.21x or increase of 30 words from the pretest on baseline. The average reading rate for Rose was 184 WRCM and 1 error (range 0–3 errors) or 1.40x acceleration which equals increase of 53 words, and

171 WRCM for Annie and 2 errors (range 1–3 errors) or acceleration of 1.22x or increase of 31 words.

In the follow-up phase, Josh read Passages 2, 5, and 8 at the average rate of 190 WRCM and 2 errors. Rose read Passages 6, 7, and 8 at the average rate of 185 WRCM and 1 error, and Annie read Passages 4, 6, and 10 at the average rate of 189 WRCM and 1 error.

For Josh, the reading rate of the application passage was 136 WRCM (1 error) on the pretest and 156 WRCM (0 errors) on the first posttest which equals 1.15x acceleration or 20 words increase. On the second posttest, the reading rate for the same passage was 153 WRCM (1 error), or 1.13x acceleration which equals an increase of 17 words from the pretest. Rose read her application passage for the first time (pretest) at the rate of 120 WRCM (1 error). The second time (posttest 1) the rate was 152 WRCM (0 errors) which is a 1.27x acceleration or an increase of 32 words, and the last time (posttest 2) she read 136 WRCM (2 errors) which is a 1.13x acceleration or 16 words increase from the first reading. The reading rate of the application passage for Annie was 128 WRCM (1 error) on the pretest. In the first posttest, she read the passage at 154 WRCM (1 error) which equals 1.20x acceleration or an increase of 26 words from the pretest. In the last reading Annie read 158 WRCM (4 errors), or 1.23x acceleration which is a 30 word increase from the pretest.

A comparison, of the pretest and the two posttests of all the untrained passages (i.e., the test probes and the application passages), for each participant is displayed in Figure 2.

Insert Figure 2

Annie read 188 WRCM and 1 error in the endurance test and 195 WRCM with no error in the stability test.

On the pre-recall measure test Josh read 138 words/min and recalled 5 correct facts, but on the post-recall measure test he read 154 words/min and recalled 4 correct facts. Josh

answered 18 questions correct on both of the question answering tests, but had the reading speed of 141 WRM and 187 WRM on the pre- and posttests, respectively. Rose read 152 words/min on the pre-recall measure test and 194 words/min on the post-recall measure test, and recalled 6 correct facts on both of the tests. On the pre-question answering test she read 159 words/min and answered 16 questions correct, but on the posttest she read 267 words/min and answered 17 questions correct. On the pre-recall measure test, Annie read 162 words/min and recalled 6 correct facts, but on the post-recall measure test she read 190 words/min and recalled 8 correct facts. Annie answered 18 and 17 correct questions, respectively on the question answering tests and had the reading speed of 165 words/min and 216 words/min, respectively.

The results of the interview, that was used to evaluate social validity, were that the participants overall liked the intervention procedure and would recommend it to others. They all stated that they had continued to use the reading support when reading difficult texts and planned to continue using it. Concerning the social importance of the results, Rose and Annie reported noticeable increase in reading fluency when reading in other settings than in the present study. In addition, both of them mentioned being more confident when reading out loud for others and both had received positive feedbacks about their reading presentation from their classmates. Rose also felt like she comprehended more of the content of what she read outside of the present study. Josh reported faster reading after the intervention when using a reading support but did not notice any remarkable difference in neither accuracy nor comprehension. The only drawback the participants mentioned was the demand of daily attendance.

Discussion

The purpose of the present study was to improve reading fluency, that is, speed and accuracy, in three adult students with the use of a multicomponent intervention. The intervention

consisted of a reading support, performance criterion, response prompts, repeated reading (RR), performance feedback, and error correction procedure. The main findings indicate that all the participants' reading fluency improved following the intervention and they were overall pleased with the results.

Regarding accuracy, the number of errors was low and relatively stable for all the participants through all the phases, including the baseline phase, indicating floor effect. As a result of low number of errors during baseline an error correction procedure might have been unnecessary. However, the participants rarely made the same errors in two sequential readings during the intervention, and when stimulus prompts were used they never made the same error in the following reading. This implies that the error correction procedure was effective for specific errors even though overall number of errors did not decrease.

During baseline the reading rate of the test probes was stable for Rose and Annie. The baseline was not as stable for Josh whereas the score of Test probe 4 was considerably higher than the scores of the other test probes, but because Test probe 5 trended sharply in the opposite direction of the performance criterion it was considered acceptable to start the intervention. It should be mentioned that Test probe 4 had the fewest letters per word (LW) of all the test probes and based on its high score it can be inferred that number of LW was a good estimator of the difficulty level of the passages.

Before reading Passage 1 in the intervention for the first time the procedure was distinct from baseline by the use of a reading support and corresponding changes in instructions. Rose and Annie read faster in their first reading of the intervention than they did on all the test probes during baseline which indicates that the reading support and the changes in the instructions might have been contributing factors to increase their reading speed. However, Josh read slower in the first reading of the intervention than during all the test probes on baseline and all the subsequent intervention sessions. A plausible explanation is that it was

difficult to move the pen, which he used as a reading support in the first session, in a right way and at right pace such that the pen might have disrupted him rather than helped him. According to a self-report, Josh informed that he had more control when using his finger than a pen as a reading support. Together, those factors indicate that using a finger as a reading support contributed to increased reading speed for all the participants. However, an individualized performance criterion of a specific rate was also set prior to the first reading of the first intervention session. As studies have shown, setting such a performance criterion/specific goal is superior to not setting a performance criterion or only a general goal, such as "read as fast as you can", regarding behavior improvement (Chard et al., 2002; Locke & Latham, 2002; Therrien, 2004). In the present study, setting a specific performance criterion might have functioned as an *evocative altering establishing operation* (Michael, 2000). In future research it might be interesting to examine further the mere effects of a reading support and the mere effects of performance criterion on reading speed, for example by using *alternating treatment design* (Kazdin, 2011).

The reading rate of the first readings of new passages displayed an upward trend toward the individualized performance criterion which confirms the findings of Lokke et al. (2009). The present results also displayed that RR of the same passage most often resulted in increased reading rate from the previous session for all the participants which is in accordance to the results of many other RR interventions studies (e.g., Polk & Miller, 1994; Samuels, 1979; Teigen et al., 2001). This indicates that when presenting the same sequences of words repeatedly for the participants a stronger stimulus control of the text was developed (S.P. Ardoin et al., 2008).

In the end of the intervention, fewer readings of each passage were required to reach the individualized performance criterion (an exception is the last passage for Annie) which is in concordance to the findings of Samuels (1979). Ardoin et al. (2007) pointed out that when

people get multiple opportunities to read the same words, but in different passages, it contributes to stronger stimulus control of those words, and promotes the development of generalization of those words when they appear in new passages. We consider this a reasonable explanation for the results of the present study.

During the intervention, Josh and Annie did not reach their performance criteria after 15 readings for one and two passages, respectively. When they read those passages with 30 s timing they both reached the criteria after only one reading. When those passages were introduced again with 1-min timing (after training of all the passages was finished) they both reached their criteria after three readings. For Josh, this particular passage had the highest difficulty level (4.7 LW) of the passages that he read during the intervention which is likely to have affected his performance. The same accounts for one of the passages (5.6 LW) for Annie which, in addition, was the first passage used in the intervention. The other passage that Annie had trouble with was at a moderate difficulty level (5.4 LW) compared to other passages used in the intervention. However, with closer examination the experimenters found out that the first quarter of the passage was much more difficult (5.8 LW) than the rest. This indicates that the difficulty level of a passage seems to be an important factor in fluency building which is in accordance to the findings of other researchers on reading fluency (S.P. Ardoin et al., 2009; Christ & Ardoin, 2009). Even though we made an effort to control for the difficulty level of the passages an even smaller difference in difficulty level might have been necessary. Many procedures have been developed to estimate the difficulty level of texts (Evensen & Vagle, 2003; McShane, 2005), but it is very difficult to ensure that all the passages used in a single study are of the exact same difficulty level and no procedures currently exist to do so (Christ & Ardoin, 2009). Annie needed 12 readings for Passage 10 even though that passage had rather low difficulty level (5.3 LW). The reason for this decreased performance in the end of the intervention is puzzling, especially because she only

needed four to five readings for the previous three passages. It is most likely that some confounding variables affected Annie's performance; she mentioned for example being eager to finish the intervention because of upcoming exams in her study program at the university. To increase experimental control it would have been more suited to train passages until reaching stable performance for each participant rather than terminating the intervention after predetermined number of passages.

When the test probes were introduced again during the first withdrawal phase all the participants increased the reading rate of each passage compared to baseline measures. Besides, the average reading rate was also higher in the first withdrawal phase than the average rate of the first readings of the intervention phase. This further indicates stimulus control of single words when the participants read untrained passages with great word overlap. Moreover this strongly indicates generalization effect, that is, that the intervention was effective in increasing the reading rate of untrained passages (i.e., the test probes). For all the participants, the increase in reading rate of untrained passages was also maintained when the test probes were introduced again in the second withdrawal phase, which further supports generalization effects. Interestingly, on few test probes (twice for Josh and three times for Rose) the reading rate on the second withdrawal phase was slightly higher (23 and 7 words for Josh and 16, 3, and 6 words for Rose) than in the first withdrawal phase. This could be an example of practice effects since it was the third time the participants were exposed to the same passages, but because a significant time gap was between each exposure we consider it unlikely. However, it should be mentioned that the difference in reading rate between the two withdrawal phases was only 1 and 6 words on average for Josh and Rose, respectively, and thus the average difference is insignificant.

For each participant, the reading rate of the passages tested during follow-up (retention) was not entirely maintained. However, even though the rate during the follow-up phase was

not as high as the performance criterion it was both significantly higher than the baseline measures and the first readings of the passages used in the intervention. As such, compared to baseline all the participants showed increased reading rate of the passages that were directly trained during the intervention after a period of time had elapsed from the end of the intervention. This is in accordance with the findings of Hughes et al. (2007), that tested for RESAA after an implementation of PT intervention to improve reading fluency. Annie's reading rate on both the endurance and the stability tests was higher than the average reading rate of the first readings of all the passages during intervention although it was not as high as the performance criteria. For all the participants, the results of the application test following the intervention showed an increase in reading rate of untrained and more difficult passages with less overlap in content words than the passages used in the intervention. This supports the findings of Therrien (2004) which reviewed RR studies and found out that RR can be effective in improving reading fluency of untrained passages. In addition, the present results were maintained three months later for Josh and Annie. The results of the pre- and post comprehension (adduction) tests (i.e., the recall- and comprehension tests) indicate that there was not a remarkable change in general reading comprehension for any of the participants. Those findings do not confirm the results of Therrien's (2004) meta-analysis which indicated the potential of repeated reading interventions to increase comprehension in regard to new passages. Even though the participants of the present study were all slow readers, compared to average reading rate for adults, they had longer learning history and read faster than typical participants of RR intervention studies. Studies have shown that people that begin intervention at higher reading level gain less improvement in comprehension than those that begin at lower level (Kuhn & Stahl, 2003) and that can explain the present results. The present results are also in concordance with the results of Vallely and Shriver (2003) which studied generalized effects of RR intervention on comprehension in secondary students, and

did not obtain remarkable change. Besides, it is also worth mentioning that even though there was not an improvement in reading comprehension after the intervention, the comprehension level was maintained (i.e., the number of correct answers/recalled facts were similar) despite the increased reading speed of all the participants. That is, increased reading speed was not at the expense of reading comprehension. It is possible that ceiling effects can explain why the number of correct answers/recalled facts did not increase; regarding the question answering tests the participants answered majority of the questions correct in the pretest and therefore had little opportunity to improve on the posttest, and regarding the recall measure tests studies have shown (Gleitman, Fridlund, & Reisberg, 1999) that on average people are only able to recall 7 ± 2 items, which applies to the present results. In summary, the results of all those additional reading measurements (i.e., the RESAA tests) indicate that the performance criteria used in the present study were suitable to predict the specific learning outcomes of fluency or RESAA.

Even though all the participants improved their reading fluency on trained- (follow-up) and untrained (withdrawal and application) passages after the intervention the effects size was rather small; or from 1.13x to1.40x acceleration (while it was 1.5x for each passage during the intervention phase). It is possible that an emphasis on error correction restrained increases in speed, but as mentioned before an emphasis on accuracy can prevent learning (Binder & Watkins, 1990). This further supports our conclusion that the error correction procedure might have been too extensive or even unnecessary.

A limitation of the present study is that the intervention was multicomponent making it difficult to determine which components were essential in producing the observed effects. However, the components used in the intervention were decided after detailed research of the reading fluency literature (e.g., Alber-Morgan et al., 2007; S. P. Ardoin et al., 2007; Chard et al., 2002; Eckert et al., 2002; Hardardottir, 2007; Lokke et al., 2009; O'Shea et al., 1984; Polk

& Miller, 1994; Samuels, 1979; Stangeland & Forsth, 2001; Therrien, 2004). Despite, it would be interesting for future research to either do a functional analysis, for example by using an alternating treatment design , or to use a *decreasing intensity design* (Barnett, Daly, Jones, & Lentz, 2004), to find out what factors are essential for the target behavior change and thereby to increase experimental control. In this case the use of decreasing intensity design would include that the whole multicomponent package would be implemented in the beginning to gain control over reading fluency relatively quickly and then single components of the package would systematically be withdrawn to reveal what components are essential for the behavior change.

Additional goal of this study was to implement a high quality within-subject experiment according to the quality indicators proposed by Horner et al. (2005). We applied the rating scale created by Chard et al. (2009) to evaluate the quality of the present study. Chard et al. required an average score of 3 on each indicator to evaluate a study to be of a high quality. According to our ratings the present study had the average of 3–4 points on each quality indicator and thus met the standards for high quality research according to Horner et al. However, it is important to mention that no objective external raters rated the present study but only the experimenters that conducted it and thus the ratings might be subjective. On the social validity component of the rating scale the present study had the average score of 3, but only two points on two of the subcomponents. Despite, the participants were overall satisfied with the intervention procedure. Rose and Annie made more positive remarks concerning the generalized effects of the intervention than Josh, but it is possible that those remarks were provided to please the experimenters. However, we consider it more likely that more intense intervention would have been required for Josh since he had more extensive reading problems from the beginning. It might be interesting to further investigate the difference between adult slow readers, with and without confirmed diagnosis of dyslexia, when using

the procedure in the present study but studies have for example shown that people diagnosed with dyslexia have more eye fixations while reading than normal readers (DeLuca, Borrelli, Judica, Spinelli, & Zoccolotti, 2002). Whereas the target behavior was considered socially important and the participants' reports provide support for both social acceptance of the procedure and social importance of the results, the present study can be determined as overall socially valid according to the three evaluation criteria mentioned in Wolf (1978).

The present study extends previous studies on reading fluency by that the participants were normally developed adult students since not many reading fluency studies have been published with that target population (Kruidenier, 2002; Strong, Wehby, Falk, & Lane, 2004). Another extension is that it was tested for all the aspects of RESAA, but according to Doughty et al. (2004) few published fluency studies have done that. In addition, the study improves the reading fluency literature by being evaluated as a high quality within-subject research according to the rating scale developed by Chard et al. which is based on the quality indicators to identify evidence-based practice for special education by Horner et al. (2005). It would be interesting to have objective raters rate the present study with the rating scale to see whether their ratings would result in similar scores. We recommend future researchers in reading fluency studies to make an effort to conduct high quality within-subject studies, for example by using the quality indicators proposed by Horner et al., in an attempt to increase experimental control and strengthen the research base.

References

Alber-Morgan, S. R., Ramp, E. M., Anderson, L. L., & Martin, C. M. (2007). Effects of repeated readings, error correction, and performance feedback on the fluency and comprehension of middle school students with behavior problems. *The Journal of Special Education, 41*, 17-30. doi: 10.1177/00224669070410010201

Ardoin, S. P., Eckert, T. L., & Cole, C. A. S. (2008). Promoting generalization of reading: A comparison of two fluency-based interventions for improving general education student's oral reading rate. *Journal of Behavioral Education, 17*, 237-252. doi: 10.1007/s10864-008-9066-1

Ardoin, S. P., McCall, M., & Klubnik, C. (2007). Promoting generalization of oral reading fluency: Providing drill versus practice opportunities. *Journal of Behavioral Education, 16*, 55-70. doi: 10.1007/s10864-006-9020-z

Ardoin, S. P., Williams, J. C., Klubnik, C., & McCall, M. (2009). Three versus six rereadings of practice passages. *Journal of Applied Behavior Analysis, 42*, 375-380. doi: 10.1901/jaba.2009.42-375

Barbetta, P. M., Heward, W. L., Bradley, D. M., & Miller, A. D. (1994). Effects of immediate and delayed error correction on the acquisition and maintenance of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis, 27*, 177-178. doi: 10.1901/jaba.1994.27-177

Barnett, D. W., Daly, E. J., Jones, K. M., & Lentz, F. E. (2004). Response to intervention: Empirically based special service decisions from single-case designs of increasing and decreasing intensity. *The Journal of Special Education, 38*, 66-79. doi: 10.1177/00224669040380020101

Beneke, W. M. (1991). Precision teaching to enhance reading skills of introductory psychology students. *Journal of Precision Teaching, 8*(2), 37-43. Retrieved from www.celeration.org

Berends, I. E., & Reitsma, P. (2006). Remediation of fluency: Word specific or generalised training effects? *Reading and Writing, 19*, 221-234. doi: 10.1007/s11145-005-5259-3

Binder, C. (1988). Precision Teaching: Measuring and attaining exemplary academic achievement. *Youth Policy, 10*(7), 12-15. Retrieved from http://www.binder-riha.com/

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19*(2), 163-197. Retrieved from http://www.binder-riha.com/

Binder, C., Haughton, E., & Eyk, D. V. (1990). Increasing endurance by building fluency: Precision teaching attention span. *Teaching Exceptional Children, 22*(3), 24-27. Retrieved from http://www.binder-riha.com/

Binder, C., & Watkins, C. L. (1990). Precision teaching and direct instruction: Measurably superior instructional technology in schools. *Performance Improvement Quarterly, 3*, 74-96. doi: 10.1111/j.1937-8327.1990.tb00478.x

Bonfiglio, C. M., Daly, E. J., Martens, B. K., Lin, L-H. R., & Corsaut, S. (2004). An experimental analysis of reading interventions: Generalization across instructional strategies, time, and passages. *Journal of Applied Behavior Analysis, 37,* 111-114. doi: 10.1901/jaba.2004.37-111

Calef, T., Pieper, M., & Coffey, B. (1999). Comparisons of eye movements before and after a speed-reading course. *Journal of the American Optometric Association, 70*(3), 171-181. Retrieved from http://optometryjaoa.com/

Carroll, C. L., McCormick, S., & Cooper, J. O. (1991). Effects of a modified repeated reading procedure on reading fluency of severely disabled readers. *Journal of Precision Teaching, 8*(1), 16-26. Retrieved from www.celeration.org

Chafouleas, S. M., Martens, B. K., Dobson, R. L., Weinstein, K. S., & Gardner, K. B. (2004). Fluent reading as the improvement of stimulus control: Additive effects of performance-based interventions to repeated reading on students' reading and error rates. *Journal of Behavioral Education, 13,* 67-81. doi: 10.1023/B:JOBE.0000023656.45233.6f

Chard, D. J., Ketterlin-Geller, L. R., Baker, S. K., Doabler, C., & Apichatabutra, C. (2009). Repeated reading interventions for students with learning disabilites: Status of Evidence. *Exceptional Children, 75*(3), 263-281. Retrieved from http://www.cec.sped.org

Chard, D. J., Vaughn, S., & Tyler, B-J. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities, 35,* 386-406. doi: 10.1177/00222194020350050101

Christ, T. J., & Ardoin, S. P. (2009). Curriculum-based measurement of oral reading: Passage equivalence and probe-set development. *Journal of School Psychology, 47,* 55-75. doi: 10.1016/j.jsp.2008.09.004

Conte, K. L., & Hintze, J. M. (2000). The effects of performance feedback and goal setting on oral reading fluency within curruiculum-based measurement. *Assessment for effective intervention, 25,* 85-98. doi: 10.1177/073724770002500201

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2 ed.). Upper Saddle River, NJ: Merill Prentice Hall.

Daly, E. J., & Martens, B. K. (1994). A comparison of three interventions for increasing oral reading performance: Application of the instructional hierarchy. *Journal of Applied Behavior Analysis, 27,* 459-469. doi: 10.1901/jaba.1994.27-459

De Luca, M., Borrelli, M., Judica, A., Spinelli, D., & Zoccolotti, P. (2002). Reading words and pseudowords: An eye movement study of developmental dyslexia. *Brain and Lanugage, 80,* 617-626. doi: 10.1006/brln.2001.2637

Doughty, S. S., Chase, P. N., & O'Shields, E. M. (2004). Effects of rate building on fluent performance: A review and commentary. *The Behavior Analyst, 27*(1), 7-23. Retrieved from www.abainternational.org

Eckert, T. L., Ardoin, S. P., Daly, E. J., & Martens, B. K. (2002). Improving oral reading fluency: A brief experimental analysis of combining an antecedent intervention with consequences. *Journal of Applied Behavior Analysis, 35*, 271-281. doi: 10.1901/jaba.2002.35-271

Evensen, L. S., & Vagle, W. (2003). *Kvalitetssikring av læringsutbyttet i norsk skriftlig (KAL-prosjektet)*. Retrieved from Institutt for språk- og kommunikasjonsstudier, NTNU website: http://prosjekt.hihm.no/r97-kal

Gleitman, H., Fridlund, A. J., & Reisberg, D. (1999). *Psychology* (5th ed.). New York, NY: Norton and Company.

Hardardottir, B. E. (2007). En undersøkelse av effekten av antall korreksjoner på feillesing av ord hos elever med psykisk utviklingshemning. *Norsk Tidsskrift for Atferdsanalyse, 34*(1), 1-12. Retrieved from www.nta.atferd.no

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Council for Exceptional Children, 71*(2), 165-179. Retrieved from http://www.cec.sped.org

Hughes, J. C., Beverley, M., & Whitehead, J. (2007). Using precision teaching to increase the fluency of word reading with problem readers. *European Journal of Behavior Analysis, 8*(2), 221-238. Retrieved from http://www.ejoba.org/

Johnson, K. R., & Layng, T. V. J. (1996). On terms and procedures: Fluency. *The Behavior Analyst, 19*(2), 281-288. Retrieved from www.abainternational.org

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2 ed.). New York, NY: Oxford University Press.

Kruidenier, J. (2002). *Research-based principles for adult basic education: Reading instruction* [Adobe Digital Editions version]. Retrieved from http://lincs.ed.gov/publications/pdf/adult_ed_02.pdf

Kubina, R. M., & Starlin, C. M. (2003). Reading with precision. *European Journal of Behavior Analysis, 4*(1&2), 13-21. Retrieved from http://www.ejoba.org/

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology, 95*, 3-21. doi: 10.1037/0022-0663.95.1.3

Legault, A., Maloney, M., & Giroux, N. (2001). Learning rates with direct instruction, precision teaching and the corrective reading series. *Journal of Precision Teaching and Celeration, 17*(2), 89-91. Retrieved from www.celeration.org

Levy, B. A., Abello, B., & Lysynchuk, L. (1997). Transfer from word training to reading in context: Gains in reading fluency and comprehension. *Learning Disability Quarterly, 20*, 173-188. doi: 10.2307/1511307

Lindsley, O. R. (1992). Precision teaching: Discoveries and effects. *Journal of Applied Behavior Analysis, 25*, 51-57. doi: 10.1901/jaba.1992.25-51

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*, 705-717. doi: 10.1037//0003-066X.57.9.705

Lokke, G. E. H., Lokke, J. A., & Arntzen, E. (2009). Bruk av hurtiglesingsteknikker for å øke lesehastighet hos gutt med kognitiv svikt og reaktiv tilknytningsforstyrrelse. *Norsk Tidsskrift for Atferdsanalyse, 36*(4), 231-240. Retrieved from www.nta.atferd.no

Lyon, G. R., & Moats, L. C. (1997). Critical conceptual and methodological considerations in reading intervention research. *Journal of Learning Disabilities, 30*, 578-588. doi: 10.1177/002221949703000601

Macalister, J. (2010). Speed reading courses and their effect on reading authentic texts: A preliminary investigation. *Reading in a Foreign Language, 22*(1), 104-116. Retrieved from www.nflrc.hawaii.edu/rfl

Malanga, P. (2003). Using repeated readings and error correction to build reading fluency with at risk elementary students. *Journal of Precision Teaching and Celeration, 19*(2), 19-27. Retrieved from www.celeration.org

McDowell, C., McIntyre, C., Owen, B., & Keenan, M. (1998). Even more challenging reading. *Journal of Precision Teaching and Celeration, 15*(2), 6-12. Retrieved from www.celeration.org

McShane, S. (2005). *Applying research in reading instruction for adults: First steps for teachers.* [Adobe Digital Editions version]. Retrieved from http://lincs.ed.gov/publications/pdf/applyingresearch.pdf

Michael, J. (2000). Implications and refinements of the establishing operation concept. *Journal of Applied Behavior Analysis, 33*, 401-410. doi: 10.1901/jaba.2000.33-401

O'Shea, L. J., Munson, S. M., & O'Shea, D. J. (1984). Error correction in oral reading: Evaluating the effectiveness of three procedures. *Education and Treatment of Children, 7*(3), 203-214.

Polk, A. L., & Miller, A. D. (1994). Repeated readings and precision teaching: Increasing reading fluency and comprehension in sixth through twelfth grade boys with emotional disabilities. *Journal of Precision Teaching, 12*(1), 46-66. Retrieved from www.celeration.org

Samuels, S. J. (1979). The method of repeated readings. *The Reading Teacher, 32*, 403-408. Retrieved from http://www.jstor.org/stable/20194790

Stangeland, T. K. (1998). Langtidseffekter av kurs i hurtiglesing: En undersøkelse av leseferdighetene til elever ved en videregående skole (SPUN-rapport 1998-2). Oslo: Stiftelsen for Pedagogisk Utvikling i Norge.

Stangeland, T. K., & Forsth, L-R. (2001). *Hurtiglesing, superlesing, fotolesing*. Aurskog: Aquarius Forlag.

Strong, A. C., Wehby, J. H., Falk, K. B., & Lane, K. L. (2004). The impact of a structured reading curriculum and repeated reading on the performance of junior high students

with emotional and behavioral disorders. *School Psychology Review, 33*(4), 561-581. Retrieved from www.nasponline.org

Tan, A., & Nicholson, T. (1997). Flashcards revisited: Training poor readers to read words faster improves their comprehension of text. *Journal of Educational Psychology, 89,* 276-288. doi: 10.1037//0022-0663.89.2.276

Teigen, T., Malanga, P. R., & Sweeney, W. J. (2001). Combining Repeated readings and error correction to improve reading fluency. *Journal of Precision Teaching and Celeration, 17*(2), 58-67. Retrieved from www.celeration.org

Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education, 25,* 252-261. doi: 10.1177/07419325040250040801

Valleley, R. J., & Shriver, M. D. (2003). An examination of the effects of repeated readings with secondary students. *Journal of Behavioral Education, 12*, 55-76. doi: 10.1023/A:1022322422324

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*, 203-214. doi: 10.1901/jaba.1978.11-203

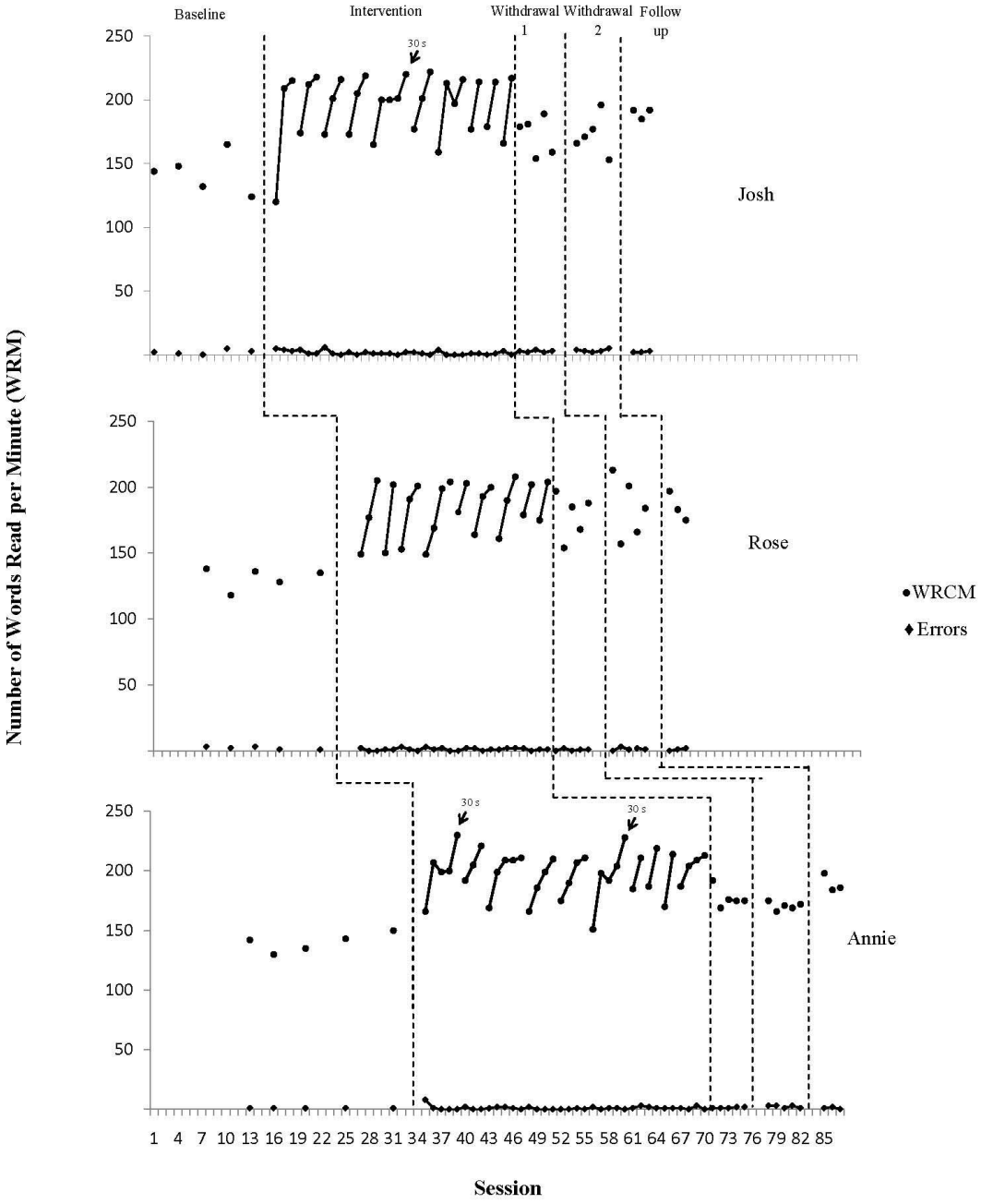

Figure 1. Number of words read correct per minute (WRCM) and errors on the test probes during baseline and withdrawal phases, along with the best reading of each session during intervention- and follow-up phases, for all the participants. The 30 s markings indicate sessions with 30 s timing.

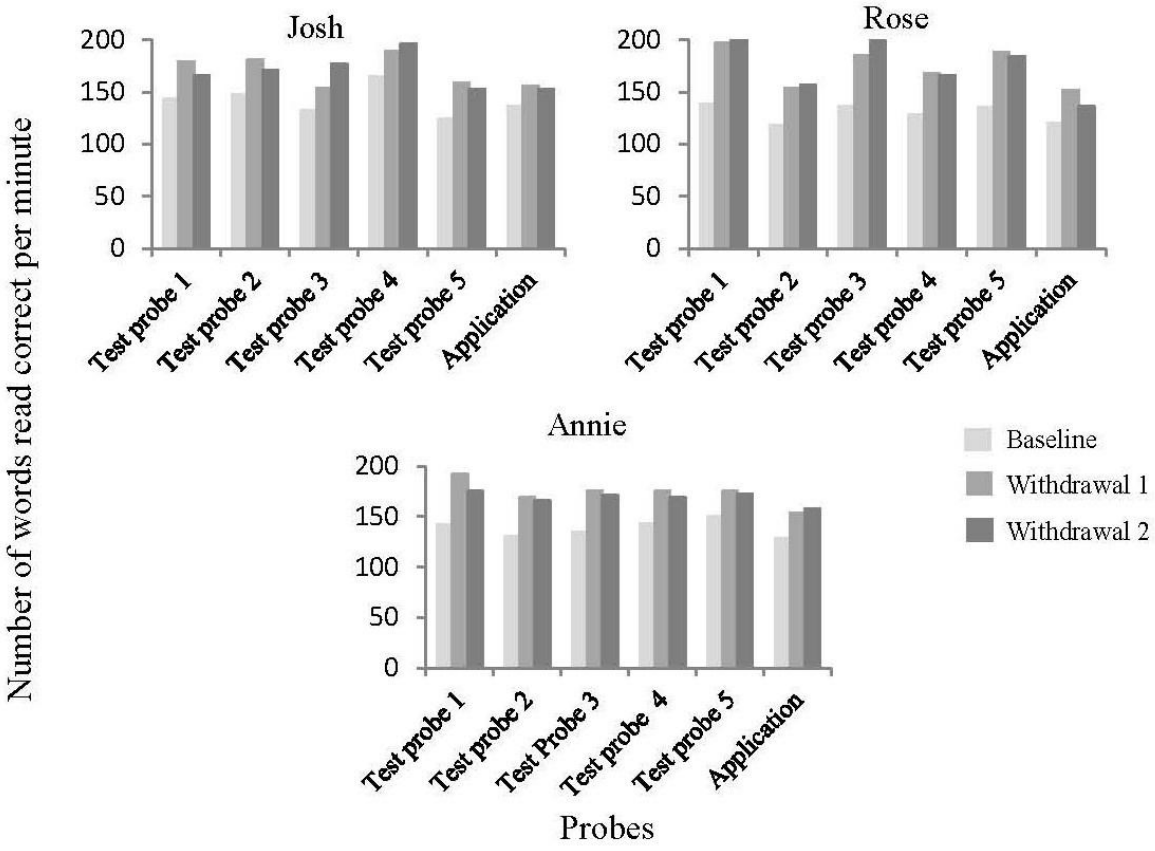

Figure 2. Number of words read correct per minute on the five test probes and the application passages during baseline, withdrawal phase 1, and withdrawal phase 2 for all the participants.

Appendix A
Informed consent

Kjeller, _____2010

Jeg _____ undertegner herved samtykke å delta i et forskningsprosjekt som er en del av mastergradsprosjektet til Hildur Valdimarsdottir og Lilja Yr Halldorsdottir ved studiet Master i læring i komplekse systemer ved Høgskolen i Akershus. Jeg har fått forklaring av målet og metodene som blir brukt i forskningen. Jeg har blitt informert om at jeg kan trekke meg ut av prosjektet på hvilket som helst tidspunkt uten noen konsekvenser for meg.

Jeg gir mitt samtykke til at data som blir samlet inn blir brukt og publisert i mastergradsoppgaven til Hildur Valdimarsdottir og Lilja Yr Halldorsdottir ved Høgskolen i Akershus og i en artikkel i et fagtidsskrift. Jeg er klar over at jeg har lov til å se alle data angående meg som blir samlet inn i forskningen og at dataene blir anonymisert ved prosjektslutt.

_____ _____

Signatur Dato

Initials:

Appendix B

Data Recording Sheet

Date																				
WRM																				
WRCM																				
ERROR																				
What error																				
Correction procedure																				
Reading support																				
Passage																				
Researcher																				
NB																				

Date																				
WRM																				
WRCM																				
ERROR																				
What error																				
Correction procedure																				
Reading support																				
Passage																				
Researcher																				
NB																				

Appendix C

Procedural Integrity Protocol

A comment: Whenever an item is irrelevant color the corresponding box grey

1. **Reading support instructions**
    a. The instructions are provided before the first reading support training in the first three sessions. The instructions must emphasize:
        i. Holding the reading support under the line, if not, do not check this.
        ii. Letting the eyes follow the reading support, if not, do not check this.
2. **Reading support training**
    a. At least three trainings in the beginning of the first three sessions, if fewer trainings, do not check this.
    b. At least one training in the beginning of other sessions, if no training, do not check this.
3. **Performance criterion**
    a. The personalized performance criterion stated in the beginning of relevant session, if not, do not check this.
4. **Error correction procedure**
    a. Points at errors from last reading; if not, do not check this
    b. Models the correct reading of the errors; if not, do not check this
    c. Drill
        i. If there are more than three errors, the errors shall be put on a power point slides but if there are three or less errors the errors are drilled right from the text but not put on slides. If this is done in another way; do not check this.
        ii. Rereading implemented correctly; if not, do not check this. Correct rereading includes:
            1. If **not** on slides each error is read independently at least five times correctly in a row.
            2. If on slides all the words/phrases are read in a sequence and all of them are read correctly five times in a row.
        iii. Words/phrase drill used appropriately; if not, do not check this**.**
            1. If the error is an incorrectly read noun only the noun must be drilled.
            2. If the error is a word omission, insertion of a word, replacement of a word or an incorrectly read word in another word class then a noun, a phrase of two to four words including the error is drilled.
    d. **Prompt** used appropriately; if not, do not check this.
        i. If the same error is made in two consecutive readings the error is highlighted on the participant's copy of the text before next reading.
5. **Instructions correctly given**.
    a. The instructions given in the baseline-, withdrawal- and follow-up phases are:

1. **Norsk: ”**Les høyt til stoppeklokken ringer. Les så fort som du kan, les alle ordene og prøv ikke å gjøre feil. Du kan begynne å lese når du er klar”.
2. **English:** ”Read out loud as fast as you can until the timer rings. Read all the words and try not to do errors. You can start when you are ready.”

ii. The instructions do not have to have point to point correspondence, but all the items on the procedural integrity checklist for those phases must be said.

b. The instructions given before the first reading on the first three intervention sessions are:
1. **Norsk:**”Les høyt til stoppeklokken ringer. Les så fort som du, les alle ordene og prøv ikke å gjøre feil. Hold lesepinnen under linjen som du leser, flytt pinnen litt fortere enn du leser så at øynene følger pinnen men ikke omvent. Du kan begynne å lese når du er klar”.
2. **English:”**Read out loud until the timer rings. Read as fast as you can, read all the words and try not to do errors. Hold the reading support under the line you are reading, move it a little bit faster than you read so your eyes follow the support but not vice verse. You can start reading when you are ready”.

ii. The instructions do not have to have point to point correspondence, but all the items on the procedural integrity checklist for the first three intervention sessions must be said.

c. Before the first reading of other sessions speed, accuracy, and reading support must be emphasized, otherwise, do not check this.
d. Before all other readings the instructions must emphasize speed, otherwise, do not check this.

Other components of the instructions above can be mentioned but are not necessary.

6. **Starts the timer before the second word of the text has been read.**
   a. If the participant has finished reading the second word, do not check this.
7. **Timing**
   a. All timings shall be 1 min except if the participant has read the same passage 15 times then the timing shall be 30 s, otherwise, do not check this.
8. **Marks errors while the participant is reading**
   a. If errors are marked afterwards, e.g. by listening to a recording or discussing errors afterwards with another observer or the participant, do not check this.
9. **Performance feedback**
   a. Gives feedback based on performance, i.e. **provides the number of words read;**
      i. Total read words; if not, do not check this

ii. **Correct read words**; if not, do not check this
iii. **Errors**; if not, do not check this

10. SCC

a. The participant records the best score of the session on a Standard Celeration Chart, if not, do not check this.

Appendix D

Procedural Integrity Checklists

Procedural integrity checklist for baseline-, withdrawal-, and follow-up phases:

Instructions correct	
Read out loud	
Read as fast as you can	
Read all the words	
Try not to do errors	
Timer started on time (see protocol)	
Error marked during reading	

Procedural integrity checklist for the first three intervention sessions:

***Int.** part**., date (session), int.** obs.*	**YES**			
Reading support instructions	If not relevant color the boxes grey			
Reading support under the line				
Eyes follow support				
Reading support training				
At least three trainings before reading				
Performance criterion stated				
Error correction procedure	If not relevant color the boxes grey			
Points out errors from last reading				
Models error correctly				
Appropriate drill procedure (slides or not, see protocol)				
Rereading correctly implemented (see protocol)				
Words or phrases drilled appropriately. (see protocol)				
Prompt used when relevant/prompt not used if irrelevant (see protocol)				
Instructions correct				
Read out loud				
Read as fast as you can				
Read all the words				
Try not to do errors				
Reading support under the line				
Reading support faster than reading				
Eyes follow support				
Timer started on time (see protocol)				
Timing for 1 min/30 s (see protocol)				
Error marked during reading				
Performance feedback / Number of words read				
Total words				
Correct words				

Errors				
About errors	If not relevant color the boxes grey			
Points out errors from last reading				
Models error correctly				
SCC				

Procedural integrity checklist for all the intervention sessions from session four:

Int. *part.***,** *date* ***(session), int.*** *obs.*	**YES**			
Reading support training				
At least one training before reading				
Error correction procedure	If not relevant color the boxes grey.			
Points at errors from last reading				
Models errors from last reading correctly				
Appropriate drill procedure (slides or not, see protocol)				
Rereading correctly implemented (see protocol)				
Words or phrases drilled appropriately (see protocol)				
Prompt used when relevant/prompt not used if irrelevant (see protocol)				
Instructions correct				
Speed emphasized				
Accuracy emphasized				
Remember the reading support				
Timer started on time (see protocol)				
Timing for 1 min/30 s (see protocol)				
Error marked during reading				
Performance feedback/ Number of words read				
Total words				
Correct words				
Errors				
About errors	If not relevant color the boxes grey			
Points at errors from reading				
Models errors from reading correctly				
SCC				

Appendix E

Interview guide

1. Legger du merke til noen endringer i forbindelse med lesing etter du startet i prosjektet?

   a) Leser du fortere?
   b) Leser du nøyaktigere?
   c) Husker du mer av det som du leser?
   d) Forstår du mer av det som du leser?

2. Var deltakelsen i prosjektet nyttig på noen måte?

   a) Hvis ja, på hvilke måter da?
   b) Hva synes du er den største fordelen ved deltakelsen i prosjektet?

3. Var det noen ulemper ved deltakelsen i prosjektet?

   a) Hvis ja, hvilke ulemper?

4. Hvis du fikk tilbud til å fortsette med treningen ville du takke ja til det?

5. Kommer du til å trene videre selvstendig?

6. Ville du anbefale andre å trene leseflyt på denne måten (dvs. lese samme teksten ofte, trene på feil og bruke lesestøtte)?

7. Hva synes du om bruken av lesestøtten?

   a) Tror du at du kommer til med å fortsette å bruke lesestøtte hvis du skal lese en tekst fort?

8. Hva synes du om bruken av Standard endringsskjemaet?

   a) Var det nyttig å registrere selv?