



# HHS Public Access

Author manuscript

*J Behav Educ.* Author manuscript; available in PMC 2016 March 01.

Published in final edited form as:

*J Behav Educ.* 2015 March 1; 24(1): 152–166. doi:10.1007/s10864-014-9216-6.

## How to Study the Influence of Intensity of Treatment on Generalized Skill and Knowledge Acquisition in Students with Disabilities

**Paul J. Yoder** and

Special Education Department, Vanderbilt University

**Tiffany Woynaroski**

Department of Hearing and Speech Sciences, Vanderbilt University

### Abstract

Seven empirical studies from this special issue and an overview chapter are reviewed to illustrate several points about studying the possible effects of treatment intensity manipulations on generalized skill or knowledge acquisition in students with disabilities. First, we make a case in favor of studying intensity as separate from complexity and expense of treatment. Second, we encourage researchers to define dependent variables in a way that allows us to determine whether treatment intensity effects on child skills and knowledge are highly generalized versus potentially context bound. Third, we acknowledge that effects of treatment intensity on generalized knowledge and skills likely vary according to student characteristics. Finally, we discuss important research design and measurement issues that are relevant to isolating the likely conditional effects of treatment intensity on generalized outcomes.

---

As highlighted by this special issue, there is increasing interest in investigating aspects of treatment that are being called "treatment intensity." Although educators and service providers regularly have to make decisions about what treatment intensity will most likely result in optimal outcomes for a given child, there has been surprisingly little quality research on this topic to date. Therefore, it has been encouraging to see that, in recent years, many more researchers are asking whether alterations in "intensity" of treatment affect student acquisition of skills and knowledge, and if so, for which students.

This commentary will discuss four issues relevant to answering these questions. First, we will review our definition of intensity and contrast our meaning from the definitions of the term that are used in some of the studies in this special issue. Second, we will discuss the types of effects that we hope will be achieved with more intense treatment – effects on generalized skill and knowledge. Third, we recognize that effects on generalized skill and knowledge will likely vary according to specific child characteristics, as well as the particular treatment and even the outcome of interest. We conclude by offering some guidance regarding how to design studies in a way that will allow us to isolate the likely

conditional effects of treatment intensity on generalized skill and knowledge acquisition of students, using the studies in this special issue to highlight some key points regarding research design, measurement of outcomes, and analytic approach. For this commentary to be useful, we feel that certain discriminations regarding the relative value of terms, designs, and measurement approaches are necessary. However, we do recognize that competent investigators will differ in their approach to studying complex phenomena such as intensity. Our perspective is offered to generate discussion and thought.

### **A Lack of Agreement on the Definition of Intensity**

As Coddling and Lane discussed in the Introduction to this issue, there has been a lack of agreement in the field regarding the definition of treatment intensity. In a 2007 commentary, Warren, Fey, and Yoder encouraged researchers to adopt a common terminology in discussing and investigating the effects of treatment intensity. Warren et al. (2007) proposed that intensity could be conceptualized as a collection of terms, including: a) *dose* (the number of teaching episodes delivered per session), b) *dose frequency* (the number of sessions offered per day/week/month), and c) *total treatment duration* (the number of days, weeks, or months over which treatment is offered). According to this framework, cumulative treatment intensity is the product of dose, dose frequency, and total treatment duration. Additionally, Warren, Fey, and Yoder (2007) discussed the degree to which teaching episodes are distributed versus massed within a session as a function of *dose* and *dose frequency*. In regards to treatment *dose*, a teaching episode could also be called a learning opportunity. A subset of this concept is opportunity to respond. It is important to note that Coddling and Lane (in press) used the term *dose* to include all of the aforementioned dimensions of treatment intensity; therefore, their usage is different from the meaning for the term *dose* as proposed in Warren et al. (2007).

The terminology suggested in Warren et al. (2007) was borrowed from the medical literature. Because educational treatment differs greatly from medical treatment, there is understandable disagreement regarding the application of these terms to educational treatments. Several other frameworks for considering treatment intensity in education have been put forth by others (e.g., Barnett, Daly, Jones, & Lentz, 2004; Mellard, McKnight, & Jordan, 2010), and we realize that there is still far from universal agreement regarding how treatment “intensity” is best conceptualized within the field of education. However, as a common definition continues to elude us, we will use the term *intensity*, as well as the terms *dose*, *dose frequency*, and *dose form* to convey the meanings described by Warren et al. (2007) as a foundation for discussion in the current commentary.

### **Rationale for Excluding Some Concepts from the Definition of Intensity**

We do still feel that the field would benefit from adopting a common definition of treatment intensity. This common definition may ultimately include those concepts put forth in Warren et al. (2007) in combination with concepts put forth by others. Several concepts listed in Table 1 of Coddling and Lane’s overview do not fit into the Warren et al. (2007) framework for intensity and its dimensions. Some of these concepts may benefit our consideration and

investigation of the effects of treatment intensity. However, we feel that it is important to exclude some of these concepts from the definition of treatment intensity.

Warren et al.'s (2007) definition of intensity and its components is intended to capture *how many opportunities a child has to learn in treatment* in a single session, in a certain number of sessions per week or month, over a certain number of weeks or months in total. We would suggest that this concept of treatment intensity should be distinguished from *how many components are involved in treatment* (treatment complexity) and *how much treatment costs in terms of time, money, and expertise* (treatment expense). These attributes of treatment are certainly important. However, they do not always positively covary with intensity of treatment as Warren et al. (2007) defined it. The rationale for making the distinction between intensity and potentially related, but importantly different, concepts, like complexity and expense, is to improve communication about the under-studied topic of intensity. Additionally, lumping complexity and expense considerations under the same umbrella concept with dimensions of intensity that relate to how many opportunities a child has to learn may have an undesirable consequence for practitioners, researchers, and policy makers. For example, it may lead them to overgeneralize logic and findings related to considerations about altering complexity or expense to considerations related to altering intensity as we define it, and vice versa.

### **Dimensions of Intensity Addressed by the Studies in this Special Issue**

Given the lack of consensus about what constitutes treatment intensity, it is not surprising that the studies in this special issue do not all investigate a manipulation of treatment that we would consider treatment intensity. For example, Duhon and colleagues (in press) studied whether adding immediate accuracy feedback to an explicit timing strategy treatment affected class-wide computation fluency in second graders. We would consider this a study of treatment complexity, but not a study of treatment intensity as we have conceptualized it.

Polanin and Espelage (in press) investigated the effects of many Second Step treatment variations, such as financial cost and preparation time incurred by teachers, on student outcomes. We would consider these intervention variations to tap treatment expense versus treatment intensity. However, these investigators did study one dimension of treatment that falls under our definition of intensity - session duration (Polanin & Espelage, in press). This aspect of treatment importantly taps how much treatment a student needs to achieve optimal outcomes, versus how much it would cost us in time or money to provide that amount of treatment.

Other studies in this issue involved one or more independent variables that we consider dimensions of treatment intensity. For example, Ennis and colleagues (in press) studied dose frequency in the sense that they descriptively compare the effect sizes of change for academic engagement and writing skill in their single group of secondary students who received Self-Regulated Strategy Development for fewer sessions per week than is typical to those effect sizes that have been observed in the literature for the same treatment delivered more frequently. Neil and Jones (in press) studied the effects of the number of response opportunities per session, the length of the intervention session, and the distribution of

response opportunities across the session on the acquisition of various skills in young children with Down syndrome. We would consider these manipulations to be related to treatment dose. Ross and Begeny (in press) studied the effects of session duration and group versus individual treatment on the oral reading fluency of second graders who were struggling with reading. We would categorize these independent variables as variations of dose and dose form, respectively. Regardless of whether the studies included in this special issue would meet our own criteria as being studies of treatment intensity or not, they provide us with the opportunity to consider what conditions are necessary to conclude that treatment variations that are being considered “intensity” manipulations affect student skill and knowledge acquisition.

### **Goal of More Intense Treatment is to Optimize Effects on Student Learning**

We presume that the motivation for studying “intensity” of treatment is to identify the amount of treatment that is necessary to achieve optimal effects on student skill and knowledge acquisition. By “optimal” effects, we specifically mean those effects on student skill and knowledge that are demonstrated across settings, activities, persons, materials, and interaction contexts. We have previously referred to these effects as *highly generalized* (Yoder, Bottema-Beutel, Woynaroski, and Sandbank, 2014). Educators universally hope to achieve these types of effects in their students, versus effects that do not generalize outside of intervention. We have previously referred to these effects, which we suspect may be dependent upon the features of the treatment context based on the way that they were measured in the study, as *possibly context-bound* (Yoder, et al., 2014).

### **Effects of Treatment Intensity Will Likely Vary According to Child Characteristics**

As Coddling and Lane (in press) indicate, increasing the intensity level of treatment will likely only produce more optimal, highly generalized, effects on skill and knowledge for some students. In other words, we expect that whether or not boosts in treatment intensity will translate to better outcomes will often depend on child profiles (Warren et al., 2007). We have, in fact, observed conditional or “moderated” effects of treatment intensity for some highly generalized outcomes in a recent randomized controlled trial of an early communication intervention in young children with intellectual disabilities (Fey, Yoder, Warren, & Bredin-Oja, 2013; Yoder, Woynaroski, Fey, & Warren, 2014). Therefore, finding that more intensive treatment is better for a specific subgroup of students cannot be taken as evidence that more is better for all students.

### **Intensity Effects are Likely to be Specific to Particular Treatments**

Results of intensity manipulations are also very likely to be specific to the treatment package being tested. Duhon et al. (in press) highlight this point when they indicate that differences between conditions defined by adding an “ingredient” to a treatment package versus the same treatment package without the ingredient can only be attributed to the added ingredient *in combination with* other aspects of the treatment. That is, if increasing the intensity of a particular treatment package enhances child generalized skill development (relative to lower

intensity levels), this does not necessarily mean that increasing the intensity of a similar treatment package or a single component of the treatment package will yield similar results. And this certainly cannot be taken as evidence that more is better for all treatments.

## **Research Design, Measurement, and Analysis Issues Related to Studying Likely Conditional Intensity Effects on Generalized Skill and Knowledge**

Thus, moving forward, there is a great need for studies that examine the likely conditional effects of intensity of specific treatments on generalized skill and knowledge acquisition of students. In an attempt to provide some guidance to researchers who will take on this task, we now discuss research design, measurement, and analysis issues that are relevant to: a) isolating effects of treatment intensity for specific treatments, b) determining whether effects achieved are highly generalized; and c) identifying which specific subgroups of students benefit from more intense treatment. We draw on the studies in this special issue, as well as a recent review chapter, to illustrate key considerations.

### **Experimental Designs Are Best Suited to Investigating Effects of Intensity**

First, how can we design studies in a way that will allow us to evaluate whether increasing the intensity of a particular intervention has an effect on student skill and knowledge? We argue that experimental research designs will be necessary to confidently attribute change in student knowledge and skills to the intensity of treatment, and that some experimental designs will be more useful than others for isolating effects of treatment intensity on student knowledge and skills.

Non-experimental designs won't permit conclusions about the causal effects of treatment intensity on student outcomes. For example, correlational designs allow us to identify associations between treatment intensity and outcome measures of interest, but do not control for alternative explanations for observed associations. Additionally, the results from correlational analyses of intensity effects are ambiguous. For example, in this issue Polanin and Espelage, contrary to expectations, found that the longer teachers taught Second Step lessons, the smaller the treatment effect on their aggregated outcome measure. The most obvious interpretation here would be that a higher dose of Second Step is related to poorer outcomes for students (i.e., lesser effects on student victimization, aggression, etc), perhaps because the message becomes tiresome and/or because students begin to tune out. However, as the authors indicate, this result could also be explained by the theory that students who do not learn from the treatment method or who simply seem to have more problems with aggression, violence, or substance abuse might influence highly motivated teachers to teach longer lessons. Thus, this result, although interesting, does not allow us to draw clear conclusions about the effects of Second Step session duration on generalized skill or knowledge in these middle school students.

Similarly, studying student change within a single group that receives treatment at a different intensity level than is typically represented in the literature is not a persuasive method for studying issues related to intensity. Doing so does not even allow us to confidently conclude that there is a relation between treatment intensity and our outcomes of

interest. For example, in this issue Ennis et al. estimated the effect size of change from pre- to post-treatment for a single group of students who received Self-Regulated Strategy Development. The authors descriptively compared the effect size for pre- to post-treatment change for the students in their study with those effect sizes reported in the larger literature for the same treatment implemented at higher intensity levels. This approach is inadequate for inferring that intensity matters (or not) because it does not control for many threats to internal validity. For example, we cannot be sure that treatment (let alone the intensity of the treatment), rather than factors such as history or maturation, caused changes in student engagement or writing skill because the study unfortunately did not include a control group. Additionally, we cannot be confident in comparing effect sizes for change over the course of treatment in this study with those effect sizes observed for the same treatment delivered at different levels of intensity in previous work because the students in the Ennis et al. (in press) study may be different from students enrolled in other studies on important characteristics.

Usually, most of the change in a nonreversible dependent variable within a single treated group is due to factors other than our treatment. There is a misguided tendency to interpret change within a single group or for a single student (i.e., with an AB design) as a method of identifying whether a student is responding to a treatment alteration, such as intensity. To understand why we consider this misguided, it is useful to review how between-group experiments identify the proportion of the variance in change in the dependent variable that is due to a treatment. Clearly, single-case experimental designs have other ways to control for threats to internal validity. However, noting how group experimental logic does so illustrates a point about change in student knowledge and skill that is not as easily illustrated in single-case experimental designs. Between-group experiments afford partitioning the variance in change in the dependent variable to treatment, or a specific treatment variation, versus other influences by contrasting groups that differ only on the presence of the treatment or the treatment variation. Randomized controlled trials (RCTs) are among the best ways to do this because random assignment to conditions makes it more likely that groups are equivalent at the beginning of treatment.

A strong argument can be made that most of the change students make on generalized skills and knowledge are due to factors other than our treatments. A recent meta-analysis of between-group experiments and quasi-experiments (most of which were RCTs) investigating the size of special education treatment effects on generalized skill and knowledge acquisition in high school students with disabilities found an average effect size of about a Cohen's  $d$  of 1.0 (Scruggs, Mastropieri, Berkeley, & Graetz, 2010). Thus, it is likely that most of our treatment effect sizes on socially important outcomes are less than a Cohen's  $d$  of 2.0. This is relevant because, when the dependent variable is a change score, a Cohen's  $d$  of 2.0 is equal to 50% of the variance in change (Cohen, 1988). Therefore, even when there are significant and large experimental-versus-control group differences on the change in the dependent variable in a well-controlled study, the variability in change on the dependent variable is mostly due to other factors than the treatment. It is extremely likely that the same is true of intensity effects.

## Some Experimental Designs are Better Suited to Studying Intensity than Others

We further propose that, within experimental studies of treatment intensity, those that compare groups or conditions that vary only on one aspect of intensity will be the most useful. Several of the studies in this special issue do compare conditions that vary by only one element. Duhon et al. (in press) provided a particularly useful example of this design. They randomly assigned the second grade students who participated in their study to either a business-as-usual control group or one of two treatment packages, which differed on only one element of treatment complexity (i.e., explicit timing strategy treatment with or without immediate accuracy feedback). This research design allowed the reader to confidently conclude that the treatment package that included feedback had added value relative to the treatment package without feedback *and* that the latter package was still facilitative of fluency over business as usual.

The Marsicano et al. (in press) study provides a counter-example of the principle that compared groups or conditions should vary only on intensity if we want to infer intensity effects on learning. In this combined non-concurrent, multiple baseline across teachers and changing-criterion design study, following the baseline phase, teachers received a “high intensity” or high support consultation package that involved both graphical and verbal performance feedback on their use of milieu math teaching strategies immediately following each observation session. After each teacher demonstrated performance at or above the established criterion for strategy use for three consecutive sessions, he or she received a “low intensity” or low support consultation package, wherein he or she received the same verbal and graphical feedback less frequently, following every three sessions. Because there were multiple differences between baseline and the following “high intensity” (i.e., high support) consultation package that teachers received, the replicated staggered AB shifts in teacher performance in use of milieu math teaching strategies cannot be confidently attributed to one specific component of the consultation package (i.e., graphical versus verbal performance feedback).

Additionally, in the Mariscano et al. (in press) study the contrast between the high and following low intensity condition does not provide evidence that “low intensity treatment is as good as high intensity treatment.” The authors correctly conclude that it only shows that when low frequency feedback follows high frequency feedback, teacher use of teaching methods remained above baseline levels. Importantly, they included the invariant sequencing of intensity levels in their interpretation.

This highlights the limitations of multiple baseline designs that use sequential phases of treatment conditions as it relates to the study of treatment intensity. First, the lack of replicated differences between high then low intensity conditions cannot be confidently interpreted as meaning that the less intense condition is “as good as” the more intense condition. Multiple explanations exist for no-difference findings. Second, the presence of replicated differences between high then low intensity conditions cannot be confidently interpreted as meaning one intensity level is better than another. The invariant sequencing does not exclude alternative explanations for the findings. If results had shown that performance was consistently higher in the high intensity condition than in the following

lower condition, lower performance in the latter condition could be explained by boredom (a type of maturation threat). If results had shown that performance was consistently higher in the lower intensity condition than in the preceding higher intensity condition, the results could be explained by familiarity (a type of maturation threat). Thus, this experimental design is not particularly well suited to contrasting treatment intensity conditions.

## Two Research Designs That Allow Studying Intensity Effects on Non-reversible Dependent Variables

Part of the problem with contrasting treatment intensity conditions using sequential phases is that many dependent variables are non-reversible. Nonreversible dependent variables maintain well-above-baseline levels many months after the offset of treatment. This is the type of dependent variable that is most relevant to informing us about intensity effects on generalized skill and knowledge acquisition. Two research designs that will be particularly useful for studying these types of dependent variables are multi-element or adapted alternating treatments single case designs and randomized between-groups experiments.

**Multi-element/adapted alternating treatments design**—Although distinguishable, these designs are discussed as one class here. Three studies in this special issue used a multi-element/adapted alternating treatments design (Haegle & Burns, in press; Neil & Jones, in press; Ross & Begeny, in press). These designs are well-suited to asking questions about effects of treatment intensity on non-reversible dependent variables, but they are accompanied by two related requirements. First, when different sets of targets are assigned to the conditions, it is required that the targets sets are equivalent in difficulty among conditions. Second, when the same targets are addressed among conditions, the conditions of assessment must be equivalent among conditions. The difficulty in convincing skeptical readers that the set of targets or assessment tasks are of equivalent difficulty across conditions is the most limiting issue for this design. When the argument for equivalence is weak or absent, we have no way of knowing whether differences between conditions truly reflect more optimal student skill and knowledge acquisition or are simply a matter of easier material in the superior intensity condition.

When the number of targets to be assigned to conditions is small, randomization alone is insufficient to convince a skeptic that the assumption of equivalence has been met. For example, Neil and Jones (in press) randomly assigned five responses to five conditions (one response per condition). Sampling theory tells us that randomly assigning only one unit to each condition is much less likely to produce equivalent target sets across conditions than randomly assigning a much larger number of units per condition (Glennerster & Kudzai, 2013).

When there is no information about among-condition comparability of the difficulty of the skills assessed or the testing contexts, the principle of parsimony requires doubting equivalence. Ross and Begeny (in press) provided an example of this problem. They used different sets of reading passages for each condition. Unfortunately, no information was provided regarding the relative difficulty of the passages between conditions within a child.

General readability was restricted between 2.85 and 3.99 grade level, but this potential variability is too wide to satisfy skeptics.

**Between-group randomized experiments**—One advantage of between-group randomized experiments is that the same targets can be used for both groups. Additionally, between-group experiments provide powerful control over the threat of maturation. Of course, other threats to internal validity also need to be controlled. The most important of the threats to internal validity in a between-group experiment is selection bias, and thus investigators must demonstrate that groups are equivalent before treatment on the dependent variable or the “pretest” and, ideally, all variables that covary with the dependent variable.

The research design used by Duhon et al. (in press) illustrates this approach. These investigators stratified students on initial computation fluency and randomly assigned the second grade students who enrolled in their study to groups. Randomization at the student level increased the probability of pretreatment equivalence between groups. Duhon and colleagues (in press) then demonstrated that randomization successfully resulted in pretest means that were very closely matched among groups. Showing non-significant differences and small effect sizes on other pretreatment variables that covary with gain in fluency would have further strengthened their implicit argument for between-group pretreatment equivalence. Another strength of this study was the researchers’ assessment of students in a way that ensured unbiased assessment among groups. Thus, this study provides a good example of the use of a randomized between-groups experimental design for a treatment variation that the investigators considered to be intensity.

### How to Determine that Effects of Treatment Intensity are Highly Generalized

Thus, there are several research designs that will allow us to isolate causal effects of treatment intensity on student outcomes. The second question is, how can we be confident that the treatment intensity effects on student knowledge and skill that we observe are highly generalized? To ensure that the effects we achieve will generalize outside of treatment, our measure of skill or knowledge acquisition must differ from our treatment context on many dimensions. In contrast, when there is not functional independence between conditions or when dependent variables readily reverse when the treatment is withdrawn, the change is potentially context-bound. More information about the important distinction between context-bound behavior change and highly generalized skill acquisition can be found in Yoder, Bottema-Beutel, Woynaroski, and Sandbank (2014).

Certainly, context-bound behavior changes can represent incremental steps towards generalized skill or knowledge. However, it is our job as researchers to show the link between potentially context-bound behavior changes and generalized skill and knowledge acquisition. It is also our job to clearly situate our dependent variables along the continuum between demonstrated context-bound behavior change and highly generalized skill or knowledge use.

It is likely that many investigators are not aware of a key issue relevant to using a dependent variable that clearly taps generalized knowledge and skill. Single subject designs are widely used in special education research (Hammond & Gast, 2010). Many of these designs require

a relatively immediate change in the dependent variable after the onset of the treatment phase in order for one to be confident that change is due to the independent variable (Lieberman, Yoder, Reichow, & Wolery, 2010). This requirement leads many of us to use potentially context-bound behavior change as our dependent variable.

For example, Neil and Jones (in press) used within-treatment session data as their dependent variable, and their results were used to make statements about intensity effects on learning. These investigators did observe large shifts in level immediately following treatment onset. However, because the dependent variable was measured in the intervention sessions and because the defining difference between conditions was a difference in the number of opportunities that children had to respond, the difference in performance does not necessarily reflect generalized skill or knowledge acquisition for the two children with Down syndrome who participated in the study. Even when we are looking at spontaneous responses, sessions in which more opportunities for a response are given set up an expectation for more frequent use of the trained response than sessions in which fewer opportunities for a response are given.

Similarly, the study by Ross and Begeny (in press) involved dependent variables that are potentially context bound, rather than clearly generalized. We say this because their outcome measures of reading fluency were reading passages that were highly similar to (in the case of the word overlap gain scores, which were derived using passages that included about 80% of the words from the target passage used at the beginning of the intervention session), or identical to (in the case of retention gain scores), the passages that were used in their intervention. These measures were administered under conditions that were extremely similar to the treatment sessions (and were, in fact, administered in the treatment sessions). We cannot be confident that similar gains in reading fluency would generalize to unfamiliar reading materials presented in different contexts. The potentially context bound nature of these dependent variables limits our ability to draw strong conclusions about the importance of the findings for “teaching reading.” As researchers, we can assist our audience in drawing appropriate conclusions about the implications of our studies by discussing our results in light of whether our dependent variables suggest potentially context bound versus highly generalized effects.

### **Identifying the Characteristics of Students who Respond to Intensity Manipulations**

Thus far we have discussed how we may best design studies to isolate effects of treatment intensity and how we can be confident in concluding effects observed are highly generalized. However, we mentioned earlier that such highly generalized effects are likely to be observed for only some students. So the next question is, how do we determine which students will benefit from increased treatment intensity?

**A widespread, but ineffective way to identify the characteristics of subgroups who respond to intensity manipulations**—Earlier, we indicated that most of the change in a single treatment group is probably due to nontreatment factors (assuming we are examining generalized skill and knowledge acquisition). Thus, correlates of such change within a single group do not tell us about the student characteristics that might define the

subgroups for which intensity alterations matter. They only tell us about correlates of change, most of which is unrelated to the treatment. Ennis et al. (in press) attempted to use this approach to argue that age and behavior problems were characteristics that predicted response to their treatment. The problem is that one has to partition out the variance of the change in the dependent variable that is due to the treatment to test whether that part of the variance is correlated with student characteristics of interest.

Using correlates of change in a single group is a much larger problem when the dependent variable is not readily reversible. In research designs focusing on readily reversible dependent variables (e.g., withdrawal designs), a functional relation between the treatment and a dependent variable can be said to mean that the treatment (or added intensity of the treatment) is the *primary* cause of the change in the dependent variable.

**Effective methods for identifying the characteristics of subgroups whose generalized skill or knowledge acquisition changes as a function of intensity manipulations**—We discuss here two effective methods for identifying characteristics of responders to intensity manipulations. One is a very novel method used by Haegele and Burns in this special issue. The other is a tried and true method that has been around for a long time. Ironically, no study in this issue used it. So we will use one of our own studies to illustrate this method.

Haegele and Burns (in press) used an interesting design logic to link the effect of an intensity alteration to a child characteristic. They select a student characteristic that is intrinsically linked to the intensity manipulation and compare matched versus mismatched conditions. The aspect of intensity they investigated was the number of unknown items taught within a set with a known item. We consider this manipulation a type of dose form manipulation. The child characteristic they assessed was called “acquisition rate” (AR). AR was quantified as the number of taught words that a student could accurately rehearse and recall at least one day later within contexts that systematically varied the ratio of unknown to known words. This interesting measure is based on a theory that posits that teaching “too much” or “too little” new information relative to old information affects efficiency and success of learning and recall. Additionally, theory and empirical evidence suggest that this AR is quantifiable on the individual level using one intervention session and one recall session. These investigators used a multi-element design to compare the gain in the number of words retained among: a) individually-identified AR-matched, b) generally-defined ‘too easy’ and c) generally-defined ‘too hard’ conditions.

Assuming that design assumptions were met and that their dependent variable links to the socially important outcome of “learning to read,” this design allowed them to conclude that matching individually-defined AR optimizes a component skill to learning to read. Importantly, they confirmed their very specific predictions. Such confirmation signals probable replicability. However, this design and treatment-to-child matching method will probably have limited application in studying intensity manipulations. The obstacles to doing so are the lack of sufficiently specific theory and the limited set of dependent variables that are independent and yet equally learnable within a child.

A more common method of identifying the subgroup of students whose generalized skill or knowledge will change as a function of treatment, or treatment intensity specifically, is to find a statistical interaction between a pretreatment child characteristic and group assignment in a randomized between-group experiment. In a recent RCT, we found that the children who played with nine or more different objects in a 15 minute structured play assessment prior to treatment learned more spoken words when receiving five, 1 hour sessions of a treatment per week than when receiving only one, 1 hour session of the same treatment (Fey, Warren, Yoder, & Bredin-Oja, 2013). Children who scored under this level of functional play or “object interest” did not respond differentially to the different dose frequencies of the same treatment. The very specific level at which intensity manipulations are detected can be computed without point-by-point (i.e., multiple significance) testing. The value of this design is that it uses the ability of RCTs to partition out the variance due to the intensity manipulation, which can then be tested for association with many types of child characteristics, including malleable and constraint variables (Yoder & Compton, 2004).

### Analytic Approaches Best Suited to Detecting Effects of Treatment Intensity

We have suggested that effects of treatment intensity will be specific not only to student subgroups, but also specific to the treatment and even the outcome of interest. Because intensity effects are likely to be treatment-specific and even outcome-specific, we suspect that meta-analytic logic may be less likely to yield useful information than analyses carried out for a dataset from a single experimental study that carefully manipulates intensity within a particular treatment. In this issue, Polanin and Espelage took a meta-analytic approach to examining the effects of the Second Step: Student Success Through Prevention program, a preventative program that seeks to proactively target aggression, violence, and substance abuse, on seven outcomes (e.g., peer victimization, physical aggression) in middle school students. These investigators found that the average effect sizes for the associations of several aspects of treatment expense and treatment effect, when analyzed across all seven outcomes used in their study, were small and non-significant (see their Table 4 notes).

Although Polanin and Espelage’s (in press) findings are about expense-to-treatment-effect associations *across outcome measures* (i.e., not effects of treatment intensity as we conceptualize it across different treatments), we predict that similar results will be observed for future meta-analyses examining the association of intensity with treatment effects across different treatment types. If intensity effects are treatment-specific, then the effect size of varying intensity levels across different types of interventions should be small. However, this does not necessarily mean that effects of treatment intensity within a particular intervention for particular outcomes will be similarly small.

### Summary

The interest in treatment intensity is ever-increasing, and the definition of intensity is clearly still-evolving. We feel that intensity is importantly different from treatment complexity and expense and that there is a need to consider *how much treatment is needed* to achieve optimal student outcomes separate from determining *how complex treatment is* or *how much that treatment costs*. Eventually, we want to know whether more intense treatment facilitates highly generalized skill and knowledge acquisition in our students. Generalized intensity

effects are likely to be treatment specific and to be larger in certain subgroups of children. Fortunately, we as researchers have at our disposal all of the design and analysis tools necessary to determine whether more intensive treatments will help us achieve highly generalized effects on important student outcomes, and if so, for which students.

## References

- Codding RS, Lane KL. Spotlight on treatment intensity: An important and often overlooked component of intervention inquiry. *Journal of Behavioral Education*. (in press).
- Cohen, J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ., editor. L. Erlbaum Associates; 1988.
- Duhon G, House S, Hastings K, Poncy BC, Solomon BG. Adding immediate feedback to explicit timing: An option for enhancing treatment intensity to improve mathematics fluency. *Journal of Behavioral Education*. (in press).
- Ennis RP, Jolivet K, Terry NP, Fredrick LD, Alberto PA. Classwide teacher implementation of self-regulated strategy development for writing with students with E/BD in a residential facility. *Journal of Behavioral Education*. (in press).
- Fey M, Warren S, Yoder P, Bredin-Oja S. Is more better? Milieu Communication Teaching in toddlers with intellectual disabilities. *Journal of Speech, Language and Hearing Research*. 2013; 56:679–693.
- Glennerster, R.; Kudzai, T. *Running randomized evaluations: a practical guide*. Princeton: Princeton University Press; 2013.
- Haegele K, Burns M. Effect of modifying intervention set size with acquisition rate data among students identified with a learning disability. *Journal of Behavioral Education*. (in press).
- Hammond D, Gast DL. Descriptive analysis of single subject research designs: 1983–2007. *Education and Training in Autism and Developmental Disabilities*. 2010; 45:187–202.
- Lieberman R, Yoder P, Reichow B, Wolery M. Expert visual analysis of multiple-baseline across participant data showing delayed changes in the dependent variable. *School Psychology Quarterly*. 2010; 25:28–44.
- Marsicano RT, Morrison JQ, Moomaw SC, Fite NM, Kluesener CM. Increasing math milieu teaching by varying levels of consultation support: An example of analyzing intervention strength. *Journal of Behavioral Education*. (in press).
- Neil NM, Jones EA. Studying treatment intensity: Lessons from two preliminary studies. *Journal of Behavioral Education*. (in press).
- Polanin J, Espelage D. Using a meta-analytic technique to assess the relationship between treatment intensity and program effects in a cluster-randomized trial. *Journal of Behavioral Education*. (in press).
- Ross S, Begeny JC. An examination of treatment intensity with an oral reading fluency intervention: Do intervention duration and student-teacher instructional ratios impact intervention effectiveness? *Journal of Behavioral Education*. (in press).
- Scruggs TE, Mastropieri MA, Berkeley S, Graetz JE. Do special education interventions improve learning of secondary content? A meta-analysis. *Remedial & Special Education*. 2010; 31(6):437–449.
- Warren S, Fey ME, Yoder PJ. Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews*. 2007; 13:70–77. [PubMed: 17326112]
- Yoder PJ, Bottema-Beutel K, Woynaroski T, Sandbank M. Social communication intervention effects vary by dependent variable type in preschoolers with autism spectrum disorders. *Evidence-based Communication Assessment and Intervention*. 2014; 7(4):150–174. [PubMed: 25346776]
- Yoder PJ, Compton D. Identifying predictors of treatment response. *Mental Retardation and Developmental Disabilities Research Reviews*. 2004; 10:162–168. [PubMed: 15611985]
- Yoder, P.J.; Symons, F. *Observational measurement of behavior*. New York, NY: Springer Publishing Company; 2010.

Yoder PJ, Woynaroski T, Fey M, Warren S. Effects of dose frequency of early communication intervention in young children with and without Down syndrome. *American Journal on Intellectual and Developmental Disabilities*. 2014; 119(1):17–32. [PubMed: 24450319]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript