Student and Teacher Perceptions of Goal Attainment during Intervention with the Self-Determined Learning Model of Instruction

Karrie A. Shogren

Tyler A. Hicks

Sheida K. Raley

Jesse R. Pace

Graham G. Rifenbark

Kathleen Lynne Lane

University of Kansas

Journal of Special Education

2020

Author Note

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A170008 to University of Kansas. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education

Abstract

A major instructional focus of interventions designed to promote self-determination, such as the Self-Determined Learning Model of Instruction (SDLMI), is to engage students in learning to set their goals, identify action plans, and evaluate their performance. However, little is known about how students define their goal attainment outcomes, or the degree to which students and teachers agree the attainment of goal set using the SDLMI in inclusive general education classes. This study examined the relation between student and teacher ratings of goal attainment during the first semester of a longitudinal, cluster randomized controlled trial of the SDLMI, as well as the impact of student disability status and teacher supports for implementing the SDLMI (i.e., online resources versus online resources + in-person coaching) on goal attainment. Findings suggested the feasibility of engaging students with and without disabilities in rating their goal attainment process during SDLMI instruction in secondary schools, with Kappa analysis indicating that, when credit is given for at least partial agreement between students and teachers, there is a fair amount of inter-rater agreement using conventional interpretation criteria. Importantly, however, conclusions drawn about the impact of student (i.e., disability status) and teacher factors (i.e., teacher implementation supports) on goal attainment outcomes are impacted by whether student or teaching ratings of goal attainment are utilized as the outcome measure. Implications for future research and practice are described.

Keywords: self-determination, goal attainment, the Self-Determined Learning Model of Instruction, Goal Attainment Scaling

There is growing recognition in the education field that supporting adolescents to set and go after goals is essential to promoting self-determination (Shogren et al., 2015). Learning to self-regulate the process of setting goals and evaluating progress toward goal attainment facilitates key executive abilities linked to success in multiple domains, particularly for students with disabilities (Shogren, Burke, et al., 2019; Shogren et al., 2012). Although enhancing self-determination has been identified as a critical outcome for secondary students (National Technical Assistance Center on Transition, 2016), promoting self-direction of the goal setting process and self-evaluation of goal attainment outcomes requires specific instructional strategies and individualized outcome measures. Goal Attainment Scaling (GAS; Kiresuk & Sherman, 1968) has been used as an individualized outcome measure across many disciplines, including education, special education, disability, and rehabilitation, as it provides a systematic framework through which to evaluate the attainment of goals (Krasny-Pacini et al., 2013; Roach & Elliott, 2005; Ruble et al., 2012; Schlosser, 2004). The general framework for GAS has not changed substantially since its introduction by Kiresuk and Sherman (1968). The first step is to identify an individualized goal, followed by the development of an individualized five-point rating scale or scoring "rubric" operationalizing expected outcomes ranging from -2 (*much less than expected*) to +2 (*much more than expected*) with 0 being the expected level of attainment. The final step is rating goal attainment based on the personalized rating scale. While the GAS process can seem relatively straightforward, there are complexities in the application of GAS, particularly when used as an outcome measure within secondary education research. Research teams have suggested criteria for increasing reliability and validity of GAS scores (Krasny-Pacini et al., 2016; Shankar et al., 2020); yet, consensus has not been established.

Previous applications of GAS typically involved researchers or clinicians supporting the development of measurable goals leading the creation of GAS rubrics and associated ratings. However, when using GAS in secondary schools to evaluate the outcomes of self-determination interventions, the conditions under which goals are set and ways in which GAS rubrics are both created and rated can be complex. For example, the *Self-Determined Learning Model of Instruction* (SDLMI; Shogren et al., 2018; Wehmeyer et al., 2000) is an evidence-based intervention designed to be implemented by trained facilitators (e.g., general and special education teachers, related service personnel) that can be overlaid on any area of instruction (e.g., academic content, transition planning). Trained SDLMI facilitators deliver targeted instruction to teach students to use a series of questions to guide themselves through the process of setting goals, building action plans, and evaluating their progress toward goal attainment. The SDLMI is delivered over an academic semester, with instruction organized around 12 Student Questions divided into three phases (Set a Goal, Take Action, Adjust Your Goal or Plan). Instruction guided by the 12 Student Questions continues across semesters to support students to build abilities and skills associated with self-determination as they refine their goals and action plans to achieve desired outcomes (Raley et al., 2018 for additional details on implementation).

Increasingly, the SDLMI has been applied in inclusive, secondary core content classrooms to promote enhanced academic goal attainment, self-determination, and academic achievement for students with and without disabilities. The premise is that *all students* can benefit from instruction in goal setting and attainment, building key self-regulatory and executive abilities that contribute to success in school and beyond (Shogren et al., 2016). In pilot work in inclusive, mathematics classrooms, benefits for both general education teachers as well as students with and without disabilities were found, including enhancements in the attainment of goals that facilitate success in mathematics (Raley et al., 2018; Raley et al., in press). Yet, in using SDLMI in such settings, unique issues related to the assessment of goal attainment emerge.

A major instructional focus of self-determination interventions is directly engaging students in learning to set their goals, identify action plans, and evaluate their performance. However, little is known about having students lead the GAS process. While researchers have stressed the importance of directly involving research participants in goal setting when using GAS (e.g., Krasny-Pacini et al., 2016; Shankar et al., 2019; Shogren, Dean, et al., in press), there is not agreement on how to involve participants in developing the GAS rubric, making ratings, and determining the implications for data analysis and interpretation of outcomes. Such issues are particularly important to consider when scaling up evidence-based interventions that target goal attainment in secondary settings, such as the SDLMI, as requiring teachers or research team members to establish individualized GAS rubrics for each student in an inclusive, general education class would be time intensive. Further, students may benefit from the instructional focus on learning not only to set goals but to identify targeted outcomes and evaluate their own goal attainment. Additionally, for some student-selected goals, external raters (e.g., teachers, research team members) may not be the best source of information regarding the student's current level of performance. For example, researchers have found that when the SDLMI is utilized in inclusive, secondary classes that target core content areas (e.g., English Language Arts [ELA], Science), students typically set goals that facilitate academic success, including enhancing study skills, or increasing engagement or attending to instruction more effectively (Raley, Shogren, Brunson, et al., 2020). Even though these goals are related to the core content curriculum, teachers may not know what reasonable expectations are for each student or have enough information to rate performance across all environments. Thus, examining the impact of students leading the GAS process and level of correspondence between teacher and student ratings of student-directed GAS is needed to inform decision making about the use of GAS.

The present study is situated in a three-year cluster randomized controlled trial (C-RCT) examining the impacts of differing SDLMI implementation supports for teachers (online resources versus online resources + in-person coaching) in inclusive, secondary core content classes. In the trial, students with and without disabilities receive instruction from their general and special education teachers and directly engage in setting goals, developing GAS rubrics, and rating their goal attainment. In addition to students' ratings of goal attainment, teachers also provided ratings of their perceptions of students' goal attainment, using student-created GAS rubrics. The purpose of this analysis is to address the following research questions:

1. How much agreement is there between student and teacher ratings of student goal attainment outcomes?
2. Does the impact of teacher implementation support condition (online only versus online + coaching) on goal attainment outcomes vary across student and teacher ratings?
3. Does the impact of student disability status (disability versus no disability) on goal attainment outcomes vary across student and teacher ratings?

Method

Overall Study Design

Data used to explore the relations between student and teacher ratings of goal attainment came from the first semester (Fall 2018) of a three-year C-RCT described previously. The overall purpose of this ongoing, longitudinal trial is to examine the impact of differing intensities of SDLMI implementation supports for teachers (online resources versus online resources + in-person coaching) on teacher and student outcomes, including student goal attainment outcomes. When implemented in inclusive classrooms, SDLMI implementers (e.g., general and special education teachers) from each school collaborate as a team to plan instruction; therefore, we randomized schools into implementation support groups at the school-level to control for spillover effects. In the first year of the C-RCT, seven schools from the Mid-Atlantic area of the United States were randomized (four schools to online only; three schools to online + coaching). However, one school in the online + coaching group could not participate in the first year of the trial as majority of the implementers at the school were ill during the summer training. In the overall C-RCT, 1,002 students across the six participating high schools engaged in the SDLMI in inclusive classrooms and contributed data at some point during the first year of the project.

Participants and Setting

For the current analysis, data from a subset of students included in the overall C-RCT were utilized; specifically, 647 students (528 [81.6%] in the online only group from four schools; 119 [18.4%] in the online + coaching group from two schools) who had goal attainment scaling contributed data from the first semester of project implementation (Fall 2018). Table 1 presents demographic data on the 647 students in the present analyses from administrative records with a small amount of missing data (<2%) backfilled from a student self-report demographic survey. Most students from the overall study sample who did not have GAS data either joined the project during the second semester or were enrolled but did not contribute any outcome data (n = 271; 27.0% of total sample). A smaller percentage of students (8.4%; n = 84) were excluded because while they set goals and created a GAS rubric they indicated that they did not complete their goal and could not make a GAS rating during the fall semester. Implementing teachers included 12 general education and five special education teachers who collectively taught 20 ninth grade English Language Arts (ELA) or 16 ninth grade Science classes. Most teachers identified as female (n = 15, 88.2%) and two (11.8%) identified as male. Teachers identified as

White/European American (*n* = 15, 88.2%), African American/Black (*n* = 1, 5.9%), and Hispanic/Latinx (*n* = 1, 5.9%). All general education teacher participants were certified in the subject areas they taught (i.e., ELA or Science) and special education teacher participants were certified to provide special education supports. With regard to collaboration, two general education teachers (11.8%) reported that they did not collaborate at all with other teachers. Other teacher participants reported collaborating in diverse ways across general and special education, including co-assessing student performance and progress (*n* = 11, 58.8%), co-planning lessons (*n* = 9, 52.9%), co-teaching some class sessions (*n* = 9, 52.9%), and co-teaching all classes (*n* = 6, 35.3%). Across the six high schools, class sizes ranged from 13 to 29 students.

Procedures

Each school identified general and special educators to be co-trained to implement the SDLMI as a part of the overall C-RCT. All SDLMI general and special education implementers attended a standardized, two-day SDLMI in-person training provided by the research team in summer 2018. Implementing teachers followed specific protocols for SDLMI whole-class implementation (Raley et al., 2018; Shogren, Raley, et al., 2019). General and special education teachers were trained to provide two weekly SDLMI mini-lessons (i.e., 15-minute instructional sessions) at the beginning of their class instruction. Implementers were provided with SDLMI mini-lessons (e.g., Student Question guides) to support their students in cycling through the three phases of the SDLMI twice an academic year, once per semester. Teachers were empowered to modify the SDLMI mini-lesson materials provided at the in-person training to align with their students' learning and engagement needs; however, they were required to meet the overall Teacher Objectives for each mini-lesson to enable students to answer the Student Question targeted in the lesson. After completing Phase 1 (Set a Goal) during each academic semester, students set a goal related to academic learning (e.g., "I want to improve upon doing my homework more and turning it in on time") with support from teachers.

SDLMI Online Resources and In-Person Coaching

Implementing general and special education teachers received one of two types of implementation support based on random assignment of their school: (a) online modules disseminated every two weeks via email (online only) or (b) online modules and in-person coaching provided monthly by trained SDLMI coaches (online + coaching). The SDLMI online resources were web-based modules that provided implementers with additional instructional strategies, video examples, and materials to supplement their SDLMI implementation that aligned with the three phases of the SDLMI and were disseminated throughout the academic year. The online modules were not designed to be interactive. Teacher participants assigned to the online + coaching group received the online modules plus monthly, in-person coaching from trained SDLMI coaches. Coaches had previous experience as teachers, administrators, and/or coaches and completed a standardized two-day training during Summer 2018 to learn how to implement the guiding principles of the SDLMI Coaching Model (Hagiwara et al., 2020). During the two-day training, coaches learned to conduct a SDLMI mini-lesson observation and then an observation of the teacher leading core content instruction using the *SDLMI Fidelity Measure: Inclusive, General Education Version* (Shogren, Raley, et al., in press). Coaches conducted six individual coaching sessions with each implementing teacher over the academic year (one coaching session per SDLMI phase each semester). Coaching sessions included involved a 30-minute observation using the procedures learned in the training, and a 30-minute conversation in which coaches prompted teachers to reflect on their implementation up to that point in the academic year and provided feedback and resources to address teachers' implementation needs. Every coaching session ended with coaches and teachers collaboratively setting a goal and action plan for teacher implementation prior to the next coaching session.

Goal Attainment Outcome Measurement

As mentioned previously, Goal Attainment Scaling (GAS; Kiresuk & Sherman, 1968) has been frequently used to document the attainment of individualized goals in the disability field. GAS requires that a range of personalized and differentiated levels of goal attainment be specified (i.e., a GAS rubric) using standardized procedures. GAS rubrics include five levels of possible attainment: much less than expected (-2), somewhat less than expected (-1), expected outcome (0), somewhat more than expected (1), and much more than expected range of outcomes (2). These levels of goal attainment are to be directly linked to the goal and reasonable expectations of attainment within a specific timeframe. Researchers have developed specific procedures to promote reliability and validity of GAS ratings (Krasny-Pacini et al., 2016), including different protocols for identifying goals for GAS and for developing and rating attainment using the GAS rubric (Shogren, Dean, et al., in press). In this C-RCT, students identified their own goals and developed their own GAS rubrics, after receiving instruction delivered by teachers as part of SDLMI implementation. Students entered their goal in a customized online platform after completing Phase 1 and created their GAS rubric at the same time. SDLMI instruction continued and after completing Phase 3 of the SDLMI focused on self-evaluating goal attainment, students and teachers separately logged their independent ratings of goal attainment using the student-created GAS rubric. Students and teachers made GAS ratings approximately eight weeks after students identified their goal and created GAS rubrics.

Analysis Plan

To address Research Question 1, we examined inter-rater agreement on goal attainment outcomes (i.e., GAS ratings) across students and teachers using standard weighted κ (Kappa) analysis which uses a precise formula to assign partial credit for near, but not exact, inter-rater agreement on an ordinal scale (Landis & Koch, 1977). Conventionally, κ coefficients between ranges of .0-.10 indicate an agreement level equivalent to chance; .11-.20, subtle agreement; 0.21-.40, fair agreement; .41-.60, moderate agreement; .61-.80, substantial agreement; and .91-1, equivalent to perfect agreement. To address Research Questions 2 and 3, we used multivariate analysis to jointly regress student and teacher ratings of goal attainment outcome on teacher implementation supports (0 = Online Only; 1 = Online + Coaching) or student disability status (0 = No disability; 1 = Disability). Because the trial design focused on sets of nested units (students nested in schools), all modeling was done in a multilevel framework. Multilevel analysis decomposes variance in outcomes across units of analysis (i.e., students, schools) so that inferences account for data dependency (Baldwin et al., 2014). To accommodate the number of schools (n = 6), we initiated Bayesian modeling using the Markov Chain Monte Carlo (MCMC) procedure in SAS 9.4 (PROC MCMC; *SAS/STAT® 14.3 User's Guide*, 2017). Because of their high stability, MCMC algorithms recover complex models otherwise inaccessible with reduced sample sizes (Bolstad, 2010). Our diffuse priors meant that Bayes estimates coincided with ones obtainable by familiar maximum likelihood methods (Kruschke, 2011). We had a small amount of missing data on predictors (<1% of students had missing disability status data) and a more sizable amount of missing data on teacher goal attainment outcomes (we had all student ratings, but 61% of corresponding teacher ratings were missing). Although all participating teachers agreed to rating student-developed GAS rubrics when they joined the multi-year project, high demands on teachers' time during the academic year (e.g., benchmark grading, developing Individualized Education Program [IEP] plans) limited their abilities to contribute these data.

However, we found that the pattern of missingness was unrelated to the value of the goal attainment outcome (e.g., student rating was not a predictor of a missing teacher rating), and we treated all missing data as model parameters and, accordingly, estimated them (Chen, 2013). Details about computation (e.g., priors) are available in supplemental materials.

Multivariate, multilevel analysis allowed us to jointly regress separate goal attainment outcomes ratings (students and teachers) on the same predictors. We also examined whether our predictors – teacher implementation supports (school-level), disability status (student-level) and their cross-level interaction – might differ between raters. As an example, it could be that disability mattered in ratings of goal attainment for teachers but not students. As such, there were five possible scenarios for each type of rater: (a) no effects; (b) only teacher implementation support effect; (c) only disability effect; (d) only main implementation support and student disability effects; or (e) implementation support, student disability, and interaction effects. Given the multivariate analysis jointly modeled data from teachers and students, the number of scenarios to consider increased from 5 to 25 ($=5^2$). As an example, the no effects scenario for teacher ratings could be paired with any of the five scenarios for student ratings.

To determine the scenario best aligned with the data and manage multiple models in one analysis we used a machine learning technique, Bayesian model averaging (BMA; Hoeting et al., 1999). Overall, BMA analysis provides a principled procedure guided by model probabilities rather than *p*-values (Morey & Rouder, 2011). BMA allows for weighing of models by probabilities to evaluate effect hypotheses (Howson & Urbach, 2006). Following standard Bayesian procedures, we confirmed all models were viable via a posterior predictive check (Lynch & Western, 2004). Second, a Bayesian fit statistic appropriate for Gaussian outcome data, the Deviance Information Criteria (DIC), was used to assign each model a relative probability (an estimated probability that, out of all models, it will best predict new data). Then, we pooled effects across the full set of models to derive inferences. We retained all models regardless of the magnitude of their probability, so as not to use arbitrary significance thresholds. Next, we (a) calculated the conditional probability that an effect contributed explanatory power in the individual models of student- and teacher-rater data, (b) pooled estimates across models including the effect to gauge the direction and size of the effect using data from each type of rater, and then, bringing together our individual models for each type of rater, (c) derived the weighted difference in the effect size estimates between type of rater. Adapting guidelines set forth by Viallefont et al. (1998), we considered effect probabilities of .15 or less to indicate strong evidence for no effect, effect probabilities close to .5 as weak (or uncertain) evidence, and effect probabilities .85 or above as strong evidence.

Results

Research Question 1: Degree of Agreement Between Student and Teacher Raters

Table 2 presents a descriptive crosstabulation of the student and teacher goal attainment ratings. As shown in Table 2, 125 (49.8%) of goal attainment outcomes are in exact agreement between the student and teacher rater. Spearman correlation (r_s) analysis also indicates the two score distributions are moderately related, $r_s(251)$ = .447, 95% CI [.341, .541]. More generally, both appear to approach normalcy, with most goals being attained at expected levels. There are higher levels of agreement across raters at the center of the outcome distribution (expected levels of attainment) with most disagreement at the tail-ends of the outcome distribution (i.e., much less than expected [-2] or much more than expected [2]). When student and teacher raters are given partial credit for near agreement, Kappa (κ) analysis indicated that, overall, there was a fair amount of agreement, κ=.370, 95% CI (.282, .458). That is, student and teacher perceptions of students' goal attainment outcomes provide divergent but complementary information, each likely worth considering on its own in analysis and instructional planning.

Research Question 2: Impact of Rater on Teacher Implementation Supports Effect

Table 3 summarizes results of BMA analysis, including side-by-side comparisons of effect estimates from the most probable models. In this BMA analysis, 25 models in total were retained, however, the models shown in Table 3 were the most probable given available data and, consequently, had greater weight in BMA analysis. The overall conditional probability given available data for an effect of teacher implementation support (online only versus online + coaching) was .578 (or 57.8%) using student-rated goal attainment outcome data and .565 (or 56.5%) using teacher data, which constitute weak evidence. We also examined if the effect of implementation support was moderated by student disability status. The low conditional probability for such an effect (≈20%), irrespective of type of rater, provides reasonably strong evidence that disability status was not a moderator of the relationship between teacher implementation supports and goal attainment outcomes.

However, when comparing conclusions drawn when using student versus teacher ratings of goal attainment as the outcome, the BMA analysis indicated that using student-rated goal attainment data leads to different estimates of the effect size of the relationship between teacher implementation supports and goal attainment outcomes than using teacher ratings. Using data from student raters, the overall effect estimate—obtained by optimally averaging across models with this effect—was .192 (in standard deviation units), an effect size considered to be small. Only negligible deviance from this overall effect estimate was found in the individual estimates of models using only student rater data (*SD* = .020), which means this estimate was stable across individual models. However, using goal attainment ratings from teachers, the overall estimate for the same effect was only -0.01, which is, practically, null. This effect was also stable across individual models (*SD* = 0.058). Such divergence in effect size, nearly .200 in standard deviation units, shows a divergence across student and teacher raters that could impact conclusions drawn about the outcomes of teacher implementation supports. That is, if student ratings of their goal attainment are used, then the effect size estimate of differing implementation supports for their teachers on goal attainment outcomes is small; however, if teacher ratings are used, then the estimated effect size is very close to non-existent.

Research Question 3: Impact of Rater on Student Disability Status Effect

The third research question focused on whether the impact of student disability status (no disability versus identified disability) on goal attainment outcomes differed if student or teacher ratings of goal attainment were utilized. The overall conditional probability for the student disability effect was .599 (or 59.9%) using student ratings of goal attainment but increased to .807 (80.7%) using teacher ratings of goal attainment. As previously mentioned, these probabilities, especially in small sample contexts, indicate that when students rate their own goal attainment there is moderate evidence for an effect of disability status, and when teachers rate student goal attainment, the data in turn suggest strong evidence. Moreover, BMA analysis indicated that the estimated effect sizes of disability status is slightly different across raters. When using student ratings, the overall effect is -.177 in standard deviation units (a subtle effect), whereas, when using teacher ratings, the effect size is -.263 in standard deviation units (a small effect). For all individual models, there was only negligible deviance from this overall effect size estimates, which shows estimates were stable. Interestingly, although a difference of .100 appears quantitatively negligible, qualitatively it ended up being enough to push the estimate into the small effect size range using standard criteria with teachers indicating more of an effect of student disability status on student goal attainment outcomes. Thus, determination of the practical importance of disability status, at least when using conventional criteria, is sensitive to whether student ratings or teacher ratings of goal attainment are used.

Discussion

Key Findings and Implications Future Research and Practice

Agreement Between Students and Teachers on Goal Attainment Outcomes

Our findings suggest the majority of students with and without disabilities were able to set goals using the SDLMI, and more importantly to the present analyses, establish GAS rubrics and make GAS ratings. Further, a majority of teachers were able to make ratings of student goal attainment using GAS rubrics created by students, suggesting the face validity of the rubrics that were established (e.g., teachers did not report that they were unable to make meaningful ratings using the rubrics). Only 8.4% of the overall student sample indicated that they were unable to rate their goal attainment during the fall semester as they decided not to complete their goal. Of this subset, 13% had an identified disability which is generally proportional to the students with disabilities included in the sample. However, what is not known is the degree to which disability or unmet disability-related needs impacted this group of students. Ongoing research, from a tiered support framework (Shogren et al., 2016), is needed to explore how to support all students, including students with disabilities, to successfully engage in goal setting and attainment as well as how to identify students who may be struggling so that more intensive interventions, indicated by goal attainment data (or lack thereof), can be implemented.

When looking at the level of inter-rater agreement, partial and exact, between students and teachers, key considerations emerge. Kappa analysis found only a fair amount of agreement on the level of student goal attainment outcomes. Agreement was easier to achieve at the expected levels of performance (a rating of 0 on the GAS rubric) than at the extremes of the GAS rubric (much less [-2] or much more than expected [2]). The analyses suggested different conclusions about goal attainment outcomes could be drawn based on the rater, particularly when ratings suggest much greater or much less than expected attainment. Ongoing research is critically needed to determine the factors that influence divergence in ratings by students and teachers and the significance of such divergence. Although low inter-rater agreement is typically treated as only an unfortunate nuisance (e.g., a function of measurement error), in this case, divergence might be of more substantive interest. Could it be that students and teachers are providing different perspectives, likely influenced by unique contextual factors, of goal attainment outcomes? Further, how do student and teacher ratings correspond to actual student skills and use of these skills in general education classrooms? Are students or teacher ratings more aligned with actual performance? Currently, given the limited knowledge of the reasons for the divergences, it does not appear valid to weigh teacher or student ratings over the other. Instead, such outcomes must be examined jointly in analyses, with implications for conclusions drawn clearly indicated. Ongoing research is needed to examine factors that predict ratings, including teacher and student factors, as well as to explore the alignment of student and teacher ratings with observational data on actual behavior.

Additionally, ongoing research is needed to see how student and teacher ratings of goal attainment will diverge and overlap over time, particularly with ongoing student instruction and teacher training. For example, in research on changes in student self-determination outcomes as a function of SDLMI intervention, researchers have begun to find a pattern when data is collected more frequently. Specifically, students initially rate their self-determination relatively high, then show an initial drop in ratings after one semester of instruction (hypothesized to result from learning and adjustment in understanding of one's self-determination), and then increase at the end of the year (hypothesized to result from more informed ratings of self-determination abilities and growth with instruction; Raley, Shogren, Rifenbark, et al., 2020). Thus, the degree to which similar patterns are seen over time in student and teacher ratings of goal attainment over multiple semesters are needed, particularly to determine if teacher and student ratings become more aligned or more divergent over time and what contextual factors influence these patterns. Such data could inform the interpretation of inferences drawn about goal attainment outcomes and may suggest that triangulating data from students and teachers with observational data is more important at different stages of intervention. Through subsequent academic semesters of the multi-year C-RCT, observing if these patterns maintain or change as students and teachers gain more experience in using the SDLMI will be an area of focus.

Effects of Teacher Implementation Supports and Disability Status

As noted, another critical reason for exploring the agreement between teacher and student ratings of goal attainment is to inform the source of outcome data utilized to draw inferences about the outcomes of an intervention. Rarely have researchers directly examined the impact of different data sources on conclusions drawn, and the findings of this analysis suggest that the rater of goal attainment outcomes has the potential to influence the conclusions drawn, suggesting the importance of collecting data from multiple sources and perspectives and analyzing data from each source rather than simply collapsing or characterizing differences as measurement error. Although ongoing research is needed, the findings from this study suggest a critical need to attend to the role of differing perceptions of outcomes in the analysis of intervention efficacy. When examining the impact of student versus teacher ratings in estimating the effect of teacher implementation supports, student ratings of goal attainment suggested a larger impact as teachers received more intensive supports for implementation. While, this was still a small effect size, it was nearly .2 standard deviation units larger than the nearly non-existent effect that teachers' ratings of student goal attainment suggested. This suggests that, during the first semester of engaging in the SDLMI with teachers receiving implementation supports, students may be seeing a larger impact of factors that influence their teachers' implementation of the SDLMI on their goal attainment than teachers are seeing. Although future research is needed to confirm these effects, this is an interesting finding that suggests that there could be differences in how students and teachers perceive the delivery of SDLMI instruction. It also suggests that gathering student perspectives of their teacher's implementation of evidence-based interventions could be useful and perhaps inform training, coaching, and the intensification of supports for teachers to effectively implement complex interventions. This is even more important given research that suggests an interactive relationship between teacher perceptions of their implementation and student outcomes, with students influencing teachers and teachers then influencing students during the academic year (Shogren et al., 2020). However, a recent review of coaching interventions in inclusive, secondary contexts identified that only 58.3% of included studies reported on both student and implementer outcomes, and only 42.9% of that subset of articles reported on the interaction between student and implementer outcomes (Raley, Shogren, & Hagiwara, 2020). Therefore, there is a need to evaluate the interaction between student and implementer outcomes, including collecting student perspectives on their teacher's implementation when support like coaching is provided, to assess accurately the impact and sustainability of an adopted intervention on student outcomes (Cook & Odom, 2013).

With regard to the effect of student disability status, we found that student disability status led to lower ratings of goal attainment after SDLMI intervention across student and teachers, consistent with previous research, but that this effect was substantially larger for teacher ratings. Specifically, if we only used student ratings of goal attainment, we would find a very small effect of disability (only detectable with statistics) on goal attainment outcomes, but when using teacher ratings we found a small effect size of disability. Future research is needed to better understand the practical implications of these discrepancies in ratings across student and teacher raters. For example, do students overestimate their strengths while teachers identify areas of additional instructional needs and supports? Or are teachers' expectations of students' capacities shaped by students having an identified disability as has been suggested by previous research (Shogren et al., 2014)? Or are both meaningful yet independent, self-perceptions of outcomes that could predict observed outcomes in distinct ways? Given that most of the teacher raters were general educators, the degree to which experience and expertise in disability and disability-related support needs also needs to be further considered. For example, to what degree are teachers prepared to identify students who based on disability or other factors need more intensive supports, particularly for building self-determination abilities? Additional work is also needed to identify the impact of other student-level factors on expectations and ratings. For example, researchers have found that the divergence between student and teacher ratings of self-determination becomes even greater for students from minoritized backgrounds (e.g., Black/African American). Although we were not able to explore these interactive factors given our sample size, future work is needed examining both student and teacher factors that influence goal attainment ratings, with specific focus on implications for understanding different perspectives as well as creating contexts where diverse perspectives can be discussed and instructional implications determined and planned for in a culturally sustainable way (Shogren, 2011). Practically, using multiple sources of data (self-report, proxy-report, and objective observations of self-determination abilities) appears necessary, particularly in secondary, general education settings. Adding observational data may provide a useful means to enable both student and teachers to reflect on their ratings and the factors that influence their ratings of goal attainment. A growing body of research suggests that each provides unique information for a comprehensive assessment (e.g., Baroody et al., 2016).

Limitations and Future Directions

In this section, we highlight limitations that must be considered in interpreting findings. First, as noted, this analysis utilizes data from the first semester of a three-year C-RCT. In this multi-year study, schools, teachers, and students are being added over time in cohorts and while this provides multiple opportunities to replicate these findings, it also means that such replication will be critical to understand the veracity of these implications. With reduced sample sizes and 61% of the student ratings missing a corresponding teacher rating, BMA analysis results must be interpreted cautiously because of the greater sampling variability in small samples and, albeit explicitly modeled, the extra uncertainty missing data adds. Yet, such replications in the overall study and by other researchers will allow for more systematic consideration of implications of the raters selected for outcome measures in large-scale, classwide implementation of evidence-based practices as well as the relation between these self- or proxy-report measures, observational data and standardized measures (e.g., academic progress and achievement indicators). Second, the logic of BMA analysis, a conservative procedure that acknowledges uncertainty in which effects should be included in linear regression, presumes that one model in the set of models considered is the correct one. This assumption, albeit reasonable given our guiding theory, warns against generalizing results to models not considered. Specifically, we did not address other student-, teacher-, and school-level factors that may influence the relationship between implementation of the SDLMI and outcomes. Further understanding of contextual factors that impact outcomes will both allow for the identification of malleable factors (e.g., teacher training, more intensive instruction for students on rating goal attainment) that could be targeted in intervention, as well as factors that may not be malleable but could inform teacher training and supports. For example, better understanding of how diverse students define meaningful goal attainment may be an area of need in teacher preparation and training around goal setting and attainment interventions, like the SDLMI. Third, this BMA analysis was based on cross-sectional data and only examined goal attainment outcomes after one semester of SDLMI instruction. Ongoing, longitudinal data analysis is needed on students outcomes over time, if the factors that affect student and teacher ratings over time change, and if student persistence in the intervention influences goal attainment outcomes, particularly the agreement between student and teacher ratings. Finally, the rate of missing teacher data suggested that identifying time to assess individual students' goal attainment using student-developed GAS rubrics was challenging for secondary general and special educators. Therefore, future research should explore how data management systems can be structured to create a less time-intensive process as well as identifying how secondary school schedules can be structured to ensure teachers have dedicated and protected time to report on student outcomes to guide research and practice.

Conclusion

The present analysis utilized data from the first semester of a three-year, C-RCT examining the impact of different intensities of supports (online only versus online + coaching) on teacher implementation of the SDLMI in inclusive, core content classes with students with and without disabilities. The focus of the current analysis was to examine goal attainment outcomes during the first semester of the C-RCT, specifically agreement across student and teacher ratings. And, when differences were found, this analysis explored if the different sources (student versus teacher) of outcome data would influence the conclusions drawn about effects of teacher implementation supports or student disability status on the relationship between SDLMI implementation and goal attainment outcomes. Such work is important both to inform the use of GAS as an outcome measure in this and other trials, as well as to guide ongoing, longitudinal work examining goal attainment outcomes. The findings further establish the feasibility of students with and without disabilities in inclusive, general education classes setting goals using the SDLMI and establishing and making ratings using Goal Attainment Scaling. Although there was only fair agreement overall between student and teacher ratings, there was still relatively strong agreement at expected levels. However, educational researchers and practitioners must acknowledge that student- and teacher-reports often diverge and that the source of data utilized can influence the inferences drawn. Ongoing research is needed to further elucidate the reasons for these divergences, the degree to which self- and other-report data aligns with actual behavior in the general education classroom, the role of triangulation of multiple sources of assessment data to drawn inferences about intervention efficacy and to inform instructional decision making, and the degree to which self- and other-perceptions unique predict and inform intervention outcomes.

References

Baldwin, S. A., Imel, Z. E., Braithwaite, S. R., & Atkins, D. C. (2014). Analyzing multiple outcomes in clinicial research using multivariate multilevel models. *Journal of Consulting and Clinicial Pyschology, 82*(5), 920-930. https://doi.org/10.1037/a0035628

Baroody, A. E., Rimm-Kaufman, S. E., Larsen, R. A., & Curby, T. W. (2016). A multi-method approach for describing the contributions of student engagement on fifth grade students' social competence and achievement in mathematics. *Learning and Individual Differences*, *48*, 54-60. https://doi.org/10.1016/j.lindif.2016.02.012

Bolstad, W. M. (2010). *Understanding computational Bayesian statistics*. John Wiley & Sons.

Chen, F. (2013). Missing no more: Using the MCMC procedure to model missing data (Paper 436-2013). SAS Global Forum.

Cook, B. G., & Odom, S. (2013). Evidence-based practices and implementation science in special education. *Exceptional Children*, *79*(2), 135-144. https://doi.org/10.1177/001440291307900201

Hagiwara, M., Shogren, K. A., Lane, K. L., Raley, S. K., & Smith, S. A. (2020). Development of the Self-Determined Learning Model of Instruction coaching model: Implications for research and practice. *Education and Training in Autism and Developmental Disabilities*, *55*(1), 17-27.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, *14*(4), 382-417.

Howson, C., & Urbach, P. (2006). *Scientific reasoning: The Bayesian approach*. Open Court.

Kiresuk, T. J., & Sherman, R. E. (1968). Goal attainment scaling: A general method for evaluating comprehensive community mental health programs. *Community Mental*

Health Journal, 4(6), 443-453. https://doi.org/10.1007/BF01530764

Krasny-Pacini, A., Evans, J., Sohlberg, M. M., & Chevignard, M. (2016). Proposed criteria for appraising goal attainment scales used as outcome measures in rehabilitation research. *Archives of Physical Medicine and Rehabilitation, 97*(1), 157-170. https://doi.org/https://doi.org/10.1016/j.apmr.2015.08.424

Krasny-Pacini, A., Hiebel, J., Pauly, F., Godon, S., & Chevignard, M. (2013). Goal attainment scaling in rehabilitation: A literature-based update. *Annals of Physical and Rehabilitation Medicine, 56*(3), 212-230. https://doi.org/10.1016/j.rehab.2013.02.002

Kruschke, J. K. (2011). *Doing Bayesian data analysis: A tutorial with R and BUGS*. Elsevier.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. https://doi.org/10.2307/2529310

Lynch, S. M., & Western, B. (2004). Bayesian posterior predictive checks for complex models. *Sociological Methods and Research, 32*(3), 301-335. https://doi.org/10.1177/0049124103257303

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing null hypotheses. *Psychological Methods, 16*(4), 406-419. https://doi.org/10.1037/a0024377

National Technical Assistance Center on Transition. (2016). *Evidence-based practices and predictors in secondary transition: What we know and what we still need to know*. Author.

Raley, S. K., Shogren, K. A., Brunson, L. Y., Gragoudas, S., & Pace, J. R. (2020). Examining goals set by students with and without disabilities engaging in the Self-Determined Learning Model of Instruction in inclusive, secondary classes. *Manuscript submitted for publication*.

Raley, S. K., Shogren, K. A., & Hagiwara, M. (2020). A review of existing research on coaching in inclusive, secondary classrooms. *Manuscript submitted for publication*.

Raley, S. K., Shogren, K. A., & McDonald, A. (2018). Whole-class implementation of the Self-Determined Learning Model of Instruction in inclusive high school mathematics classes. *Inclusion, 6*(3), 164-174. https://doi.org/10.1352/2326-6988-6.3.164

Raley, S. K., Shogren, K. A., Rifenbark, G. G., Lane, K. L., & Pace, J. R. (2020). The impact of the Self-Determined Learning Model of Instruction on student self-determination in inclusive, secondary classrooms. *Manuscript under development*.

Raley, S. K., Shogren, K. A., Rifenbark, G. G., Thomas, K., McDonald, A. F., & Burke, K. M. (in press). Enhancing secondary students' goal attainment and self-determination in general education mathematics classes using the Self-Determined Learning Model of Instruction. *Advances in Neurodevelopmental Disorders*.

Roach, A. T., & Elliott, S. N. (2005). Goal attainment scaling: An efficient and effective approach to monitoring student progress. *Teaching Exceptional Children, 37*(4), 8-17. https://doi.org/10.1177/004005990503700401

Ruble, L., McGrew, J. H., & Toland, M. D. (2012). Goal attainment scaling as an outcome measure in randomized controlled trials of psychosocial interventions in autism. *Journal of Autism and Developmental Disorders, 42*(9), 1974-1983. https://doi.org/10.1007/s10803-012-1446-7

SAS/STAT® 14.3 User's Guide. (2017).

Schlosser, R. W. (2004). Goal attainment scaling as a clinical measurement technique in communication disorders: a critical review. *Journal of Communication Disorders, 37*(3), 217-239. https://doi.org/10.1016/j.jcomdis.2003.09.003

Shankar, S., Marshall, S. K., & Zumbo, B. D. (2020). A systematic review of validation practices for the Goal Attainment Scaling measure. *Journal of Psychoeducational Assessment, 38*(2), 236-255. https://doi.org/10.1177/0734282919840948

Shogren, K. A. (2011). Culture and self-determination: A synthesis of the literature and directions for future research and practice. *Career Development for Exceptional Individuals*, *34*, 115-127. https://doi.org/10.1177/0885728811398271

Shogren, K. A., Burke, K. M., Anderson, M. A., Antosh, A. A., LaPlante, T., & Hicks, T. A. (2020). Examining the relationship between teacher perceptions of implementation of the SDLMI and student self-determination outcomes. *Career Development and Transition for Exceptional Individuals, 43*(1), 53-63. https://doi.org/10.1177/2165143419887855

Shogren, K. A., Burke, K. M., Antosh, A. A., Wehmeyer, M. L., LaPlante, T., Shaw, L. A., & Raley, S. K. (2019). Impact of the Self-Determined Learning Model of Instruction on self-determination and goal attainment in adolescents with intellectual disability. *Journal of Disability Policy Studies, 30*(1), 22-34. https://doi.org/10.1177/1044207318792178

Shogren, K. A., Dean, E. E., Burke, K. M., Raley, S. K., & Taylor, J. L. (in press). Goal Attainment Scaling: A framework for research and practice in the intellectual and developmental disabilities field. *Intellectual and Developmental Disabilities*.

Shogren, K. A., Palmer, S. B., Wehmeyer, M. L., Williams-Diehm, K., & Little, T. D. (2012). Effect of intervention with the Self-Determined Learning Model of Instruction on access and goal attainment. *Remedial and Special Education*, *33*(5), 320-330. https://doi.org/10.1177/0741932511410072

Shogren, K. A., Plotner, A. J., Palmer, S. B., Wehmeyer, M. L., & Paek, Y. (2014). Impact of the Self-Determined Learning Model of Instruction on teacher perceptions of student capacity and opportunity for self-determination. *Education and Training in Autism and Developmental Disabilities, 49*(3), 440-448.

Shogren, K. A., Raley, S. K., & Burke, K. M. (2019). *SDLMI Teacher's Guide Supplement: Implementing the SDLMI with the Whole Class*. Kansas University Center on Developmental Disabilities.

Shogren, K. A., Raley, S. K., Burke, K. M., & Wehmeyer, M. L. (2018). *The Self-Determined Learning Model of Instruction: Teacher's Guide*. Kansas University Center on Developmental Disabilities.

Shogren, K. A., Raley, S. K., Rifenbark, G. G., Lane, K. L., Bojanek, E. K., Karpur, A., & Quirk, C. (in press). The Self-Determined Learning Model of Instruction: Promoting implementation fidelity. *Inclusion*.

Shogren, K. A., Wehmeyer, M. L., & Lane, K. L. (2016). Embedding interventions to promote self-determination within multi-tiered systems of supports. *Exceptionality, 24*(4), 213-224. https://doi.org/10.1080/09362835.2015.1064421

Shogren, K. A., Wehmeyer, M. L., Palmer, S. B., Forber-Pratt, A. J., Little, T. J., & Lopez, S. J. (2015). Causal agency theory: Reconceptualizing a functional model of self-determination. *Education and Training in Autism and Developmental Disabilities, 50*(3), 251-263.

Viallefont, V., Raftery, A. E., & Richardson, S. (1998). *Variable Selection and Bayesian Model Averaging in Case-Control Studies*. University of Washington.

Wehmeyer, M. L., Palmer, S. B., Agran, M., Mithaug, D. E., & Martin, J. E. (2000). Promoting causal agency: The Self-Determined Learning Model of Instruction. *Exceptional Children, 66*(4), 439-453. https://doi.org/10.1177/001440290006600401

Table 1

Sample Demographics

Characteristic	Online Only		Online + Coaching		Total	
	n	%	*n*	%	*n*	%
Gender						
Female	255	48.3	50	42.0	305	47.1
Male	273	51.7	68	57.1	341	52.7
Missing	0	0.0	1	0.8	1	0.2
Disability						
No disability	427	80.9	102	85.7	529	81.8
Autism spectrum disorder	5	0.9	3	2.5	8	1.2
Emotional or behavioral disorder	2	0.4	1	0.8	3	0.5
Hearing impairment	1	0.2	0	0.0	1	0.2
Intellectual disability	5	0.9	0	0.0	5	0.8
Learning disabilities	61	11.6	9	7.6	70	10.8
Other health impairment	20	3.8	2	1.7	22	3.4
Physical disability	2	0.4	0	0.0	2	0.3
Speech language impairment	4	0.8	0	0.0	4	0.6
Missing	1	0.2	2	1.7	3	0.5
Race/Ethnicity						
African American/Black	205	38.8	33	27.7	238	36.8
American Indian/Alaska Native	2	0.4	1	0.8	3	0.5
Asian American	16	3.0	6	5.0	22	3.4
Hispanic or Latinx	37	7.0	20	16.8	57	8.8
Hawaiian Native or Pacific Islander	2	0.4	0	0.0	2	0.3
Two or more races	16	3.0	3	2.5	19	2.9
White/European American	250	47.3	53	44.5	303	46.8
Missing	0	0.0	3	2.5	3	0.5
Grade						
9th	515	97.5	116	97.5	631	97.5
10th	9	1.7	0	0.0	9	1.4
11th	0	0.0	2	1.7	2	0.3
Missing	4	0.8	1	0.8	5	0.8
English language learner (ELL) status						
No	512	97.0	108	90.8	620	95.8
Yes	15	2.8	6	5.0	21	3.2
Missing	1	0.2	5	4.2	6	0.9
Free and reduced price lunch status						
No	275	52.1	55	46.2	330	51
Yes	244	46.2	51	42.9	295	45.6
Missing	9	1.7	13	10.9	22	3.4

Note. Total of percentages for each category may not be 100% due to rounding.

Table 2

Crosstabulation of Student and Teacher Ratings of Goal Attainment Outcomes (n = 647)

		Teacher						
		-2	-1	0	1	2	Missing	Total
		n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Student	-2	10 (1.6)	4 (0.6)	5 (0.8)	1 (0.2)	0 (0)	16 (2.4)	36 (5.6)
	-1	11 (1.8)	32 (5)	12 (1.8)	4 (0.6)	3 (0.4)	97 (15)	159 (24.6)
	0	7 (1)	23 (3.6)	63 (9.8)	13 (2)	5 (0.8)	169 (26.2)	280 (43.4)
	-1	1 (0.2)	10 (1.6)	12 (1.8)	14 (2.2)	5 (0.8)	84 (13)	126 (19.6)
	2	0 (0)	3 (0.4)	4 (0.6)	3 (0.4)	6 (1)	30 (4.6)	46 (7)
	Total	29 (4.6)	72 (11.2)	96 (14.8)	35 (5.4)	19 (3)	396 (61.2)	647 (100)

Note. Kappa analysis excludes the 61% of the student ratings (n = 646) with no corresponding teacher ratings. These teacher ratings were assumed missing at random. Kappa analysis results, with partial credit included for near agreement, indicate that, although not substantial, there was a fair amount of at least partial agreement between types of rater, κ=.3701, 95% CI (.282,.4581). That is, students and teachers gave complementary but separate perspectives on goal attainment outcomes.

Table 3

Bayesian Model Averaging Results

Model Parameters	P(Effect≠0)	Weighted Estimate (SD)	Most Probable Models Considered		
			Model 1	Model 2	Model 3
Student Rater Data					
Fixed Effects					
Intercept (β_0^1)	100%	-.069 (0.039)	-.101	-.023	-.010
Implementation Support (β_1^1)	57.8%	.192 (0.020)	.186		
Disability (β_2^1)	59.9%	-.177 (0.018)	-.193	-.174	-.179
Interaction (β_3^1)	17.8%	.137 (0.031)	.135		
Variance Components					
Residuals (τ_{u0}^2)		.926 (.001)	.925	.924	.925
School (τ_{u0}^2)		348. (.035)	.371	.310	.309
Teacher Rater Data					
Fixed Effects					
Intercept (β_0^2)	100%	-.166 (.031)	-.167	-.148	-.176
Implementation Support (β_1^2)	56.5%	.010 (.035)			-.035
Disability (β_2^2)	80.2%	-.263 (.058)	-.313	-.318	-.321
Interaction (β_3^2)	20.7%	.366 (.006)			
Variance Components					
Residuals (τ_{u2}^2)		1.088(.004)	1.086	1.086	1.085
School (τ_{u3}^2)		.468 (.072)	.394	.398	.493
Rater Type Correlations					
Rater-Type Residuals (r_{st})		.431 (.041)			
Rater-Type School Effects (r_{st})		.049 (.015)			
Model Fit					
DIC (Smaller is better)			3922	3923	3923
Model Probabilities			12.2%	8.4%	7.5%

Note. The first column shows parameters considered in BMA analysis. The second column, P(Effect≠0), shows the probability the effect contributes any explanatory power—obtained by summing the probabilities of all models including the effect. The third column, weighted Estimate, shows the optimally weighted effect estimate—obtained by averaging estimates of all models including the effect by model weight. The last three columns show parameter estimates for all models in the set with at least a 10% probability of being correct. In total, we considered 25 different effect configurations in this BMA analysis.

Article Title: Student and Teacher Perceptions of Goal Attainment during Intervention with the Self-Determined Learning Model of Instruction

Supplemental Material

This section provides further detail on the Bayesian analysis. To address research questions, we implemented multivariate multilevel model [M-MLM] analysis. An abridged example model is below.

$$Y_{kts}^{1} = \beta_{000}^{1} + \beta_{001}^{1} Grp_{s} + \beta_{100}^{1} Disability_{kts} + \beta_{101}^{1} Grp_{s} * Disability_{tks} + u_{0s}^{1} + u_{1ks}^{1} + e_{tks}^{1}$$

$$Y_{kts}^{2} = \beta_{000}^{2} + \beta_{001}^{2} Grp_{s} + \beta_{100}^{2} Disability_{kts} + \beta_{101}^{2} Grp_{s} * Disability_{tks} + u_{0s}^{2} + u_{1ks}^{2} + e_{tks}^{2}$$

Y_{kts}^{1} and Y_{kts}^{2} denote a goal attainment outcome rating student *k* of teacher *t* at school *s*, Grp_{s} (coded: 0/1) indicates if school is in the "online + coaching" group, $Disability_{tks}$ (coded: 0/1) indicates if a student has a disability, β_{000}^{1} and β_{000}^{2} are intercepts for outcomes, β_{001}^{1} and β_{001}^{2} are the incremental effect of Grp_{tks} on goal attainment data, β_{100}^{1} and β_{100}^{2} are the incremental effect of $Disability_{tks}$ on goal attainment data, and β_{101}^{1} and β_{101}^{2} are the interaction effect between Grp_{s} and $Disability_{tks}$, and u_{0s}^{1}, u_{0s}^{1}, and u_{1s}^{1}, u_{1s}^{1} are random school schools on intercept and slopes. Random upper level effects (u) for intercepts and residuals (e) were correlated across separate outcomes and multivariate normally distributed, with mean zero and estimated variance.

All parameters in the model were given independent, weakly informed priors to facilitate Bayesian estimation. These priors were:

$$\beta \sim N(0, \sigma_{\beta}^{2} = 10{,}000)$$

$$\mu_{0} \sim N(0, \tau_{\mu_{0}}^{2} = 1)$$

$$\mu_1 \sim N(0, \tau_{\mu_1}^2 = 1)$$

$$\tau^2 \sim IGamma(Shape = 2, scale = 2)$$

These priors, as weakly informed priors, were selected so that Bayesian estimates would closely approximate equivalent ML estimates. To fit this model to data, PROC MCMC used the Gibbs sampler to obtain posterior distributions for each parameter. Based on trial and error, we requested 505,000 MCMC iterations but discarded the first 5,000 and thereafter kept every 100th iteration to reduce the amount of auto-correlation in the final MCMC chain. We also examined the quality of the MCMC output through convergence diagnostic plots obtained via PROC MCMC (available upon request).