

How satisfaction and self-efficacy for inclusive education matter for Swedish special educators' assessment practices of students with intellectual disability

Monica REICHENBERG¹, Kent LOFGREN²

¹Department of Education and special education,
The University of Gothenburg, Sweden

²Department of Education, Umea University, Sweden

Original Article

Received: 28-April-2019

Revised: 11-June-2019

Accepted: 24-June-2019

Online first 26-June-2019

Abstract

Introduction: Assessment of learning outcomes is integral to both mainstream and special needs comprehensive schools for students with intellectual disabilities (ID). However, assessment of students with ID poses a challenge both to special educators and their cooperation with mainstream teachers in cases of fully included students with ID with an individualised curriculum.

Objectives: We describe and predict the type of assessment practices Swedish special educators in special needs comprehensive schools use for assessment of students with ID.

Methods: Swedish special educators ($n = 148$) were recruited using a non-random sample. To analyse our data, we used the item response model. In addition, we analysed special educators' expected satisfaction with assessments using linear regression and logistic regression.

Results: The study suggests that special educators had the greatest difficulty conducting multiple choice and written assessments. Moreover, the study suggests that satisfaction with assessment and self-efficacy for inclusion matters for predicting types of assessment practice. In addition, the study reports an interaction between job satisfaction for moderately experienced special educators that predicts both types of assessment practice and the special educators' satisfaction with assessment.

Conclusion: We demonstrate how assessment satisfaction, self-efficacy, job satisfaction, and experience matter for special educators' assessment of students with ID.

Keywords: *special educators, intellectual disability, assessment, satisfaction, self-efficacy*

Citation: Reichenberg, M., Lofgren, M. How satisfaction and self-efficacy for inclusive education matter for Swedish special educators' assessment practices of students with intellectual disability. *Journal for ReAttach Therapy and Developmental Diversities*. 2019 Sep 08; 2(1):12-25 <https://doi.org/10.26407/2019jrtdd.1.17>

Copyright ©2019 Reichenberg, M., Lofgren, M. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

Corresponding address:

Monica REICHENBERG

Department of Education and special education

Västra Hamngatan 25, 411 17 405 30 Gothenburg, Sweden

E-mail: monica.reichenberg@ped.gu.se

1. Introduction

Assessment of learning outcomes (knowledge, skills, beliefs, and attitudes) is integral to both mainstream and special needs comprehensive schools for students with intellectual disabilities (ID). Assessment matters (a) to clarify if students have met the curriculum goals (b) to support the development of students (c) for teachers' lesson planning and for decisions concerning special needs instruction. Without assessment, special educators cannot evaluate students with ID. However, assessment of students with ID poses a challenge both to special educators and their cooperation with mainstream teachers in cases of fully included students with ID with an individualised curriculum.

For example, a student with ID (a) might have language and communication difficulties or (b) have limited reading and writing abilities. Many standardised tests require reading and writing abilities. Consequently, special educators often feel that they have insufficient training for assessing students with ID. Because of the difficulty of assessing students with ID, special educators often avoid assessing them by using standardised tests (Tumbull, Tumbull, Wehmeyer, & Park, 2003).

This raises the question of whether special educators should continue to use standardised tests that require writing or multiple choice format or switch to or complement with other forms of assessment. Alternative forms of assessment have gained prominence in education, such as assessing verbal participation or oral assessment. Such forms of assessment offer a complement or even an alternative to the traditional forms of assessment.

The difficulty of assessing students with ID also poses a problem of accountability (Looney, Cumming, van Der Kleij, & Harris, 2018). Without assessment, teachers and policy makers do not know if students with ID meet the standards of the curriculum. Thus, assessment for holding schools responsible for the learning outcomes of students with ID becomes difficult.

The ability to assess students is an ability that many teachers learn through practice at work. Therefore, several studies have been conducted on how teachers in mainstream schools assess students. However little is known about how special educators in special needs comprehensive schools assess students with ID. In the present study, we

sought to gain knowledge about the assessment practices of Swedish special educators in special needs comprehensive schools. Swedish special needs comprehensive schools represent a case of schools criticised for not meeting the curriculum standards and for implementing poor assessment practices (Swedish Schools Inspectorate, 2010).

Our purpose was to describe and predict the type of assessment practices Swedish special educators in special needs comprehensive schools use for students with ID. We posed the following research questions:

- What type of assessment of students with ID do special educators find the most difficult to conduct?
- What predicts special educators' assessment practice score of students with ID?
- What predicts the types of assessment special educators use for students with ID?
- What predicts special educators' satisfaction with assessment of students with ID?

1.1 Teachers' assessment practices

We define educational assessment as the process that establishes what students know and are capable of doing. Assessment can be divided into summative and formative assessment. Summative assessment refers to evaluating the sum of students' learning outcomes through tests (e.g. written or multiple choice) for public reporting, certification, selection, and system accountability. Examples of formative assessment include constant feedback, peer assessment, self-assessment, teacher strategic questioning, and teacher feedback (Black & William, 2005). Summative assessment is typically administrated at the end of a curriculum unit. By contrast, formative assessments are designed for teaching and learning in the classroom and provide diagnostic feedback and progress evaluation (Hattie & Timperley, 2007).

However, we agree with Looney, Cumming, van Der Kleij, and Harris (2018) that the distinction between the two types of assessment can be misleading. For example, on occasion, summative assessment and point-in-time judgement of student achievement can be designed for improving

teaching and learning. On other occasions, formative assessment can be used to provide feedback to students and teachers regarding the next steps for teaching and learning, as assessment for reporting or certification.

Assessing students with ID is pedagogically difficult and requires competence. The difficulty lies in establishing validity and reliability in assessments. Thus, various types of assessments have their pros and cons in terms of validity and reliability. Assessment types with a low degree of validity and reliability impose problems in terms of accountability and equity for students with ID.

Assessments requiring written questions and written answers are inaccessible to many people with ID. However, there are alternative approaches to written tests (Black & William, 2005; Davies, Stock, King, Wehmayer & Shogren, 2017). For example, Bolt, Decker, Lloyd, and Morlock (2011) and Thurlow, Lazarus, and Hodgson (2012) used read-aloud adjustments where another person read aloud test instructions, questions, and answer sets and then recorded the students' responses. Nevertheless, alternative approaches to assessment tend to be underused (Davies et al., 2017; Fujiura et al., 2012; Tanis et al., 2012; Wehmayer & Abery, 2013).

1.2 Special educators' self-efficacy beliefs and assessment practices

As part of socio-cognitive theory, Bandura (1997) developed the concept of self-efficacy to explain how people's beliefs influence their course of action. Special educators' self-efficacy refers to their belief in their capacity to execute an action in teaching. Special educators develop efficacy from prior experience (observing others), self-persuasion, and interactions with colleagues or students. Several studies have stressed the importance of self-efficacy as a predictor of educators' attitudes and behaviour. However, the concept is multidimensional. For example special educators' self-efficacy tends to be domain specific. Special educators might vary in their self-efficacy concerning teaching a subject matter, using Internet communication technology, classroom management, or teaching inclusive education (students with ID). Being confident about a topic is not necessarily the same as being confident in teaching students with ID. Therefore,

the predictive accuracy of self-efficacy might vary depending on the dimension of self-efficacy.

Self-efficacy has been cited as one of the most important variables in special education research. It has predicted a number of teacher work outcomes (e.g. job satisfaction and burnout) (Viel-Ruma, Houchins, Jolivet, & Benson, 2010).

The concept predicts attitudes and willingness to implement various types of special needs instruction. Consequently, we expand the concept of self-efficacy for inclusive education to predict special educators' assessment practice of students with ID. We suggest that self-efficacy for inclusive education varies with the assessment practice applied for students with ID. Greater confidence should promote more complex forms of assessment practice.

1.3 Special educators' emotions and assessment practices

Drawing on the psychology and sociology of emotions, Hargreaves (2000) developed a conceptual framework for understanding emotions in teaching. Hargreaves (2000) argued that teaching is emotionally embedded. Emotional embeddedness means that emotions (e.g. satisfaction, pride, shame, guilt) contribute to explaining teachers' actions and interactions at the workplace. Special educators develop emotions in relationships with colleagues, students, parents, and principals. Therefore, emotions at work mix individual and professional expectations. As an example, dissatisfaction with curricula might keep special educators from implementing new educational policies.

Expanding on Hargreaves' (2000) framework, we contend that satisfaction (or happiness) matters for teachers' assessment practices. First, special educators vary in their satisfaction with tests for conducting assessments. Dissatisfaction with tests might reflect either professional criticism or lack of knowledge.

Second special educators depend on satisfaction with working conditions as well as students, parents, principals, and colleagues. Satisfaction encourages action. In addition, satisfaction estimates special educators' capacity for action. Thus, special educators who can get things done at schools have a greater sense of satisfaction. By contrast, special educators with low satisfaction

might exhibit frustration and alienation at work. Such feelings might reduce their feelings of power over the assessment practices. Such emotions develop during the special educator's career. Therefore emotions might interact with age and experience.

Third, feelings of satisfaction might be grounded in a sense of fairness. Special educators might be more or less satisfied with tests depending on how fair they consider them towards students with ID.

2. Method

2.1 Participants

In this study, 148 special educators from northern and western Sweden participated in a non-random sample (response rate =74%). Nevertheless, we cannot generalise to the Swedish population of special educators. However, we can still make inferences about the process that generated the data. In other words, we can still say something about what is going on in the sample, which might yield insights into what influences special educators' assessment practices. All the special educators teach students with ID. According to The Swedish National Agency for Education (2017), roughly 85% of the teachers in special needs comprehensive schools are females. In our sample, almost all are females (94%). About 85% have a teaching degree, whereas only 29% have a special teacher education degree. However, oversampled educators with a special education degree (52%). The teacher student ratio is close to four, i.e. very small classes and the special educators have worked about, on average, 8 years at their present school.

2.2 Variables

For our study, we analysed several outcome variables (see Table 1). The first set of outcome variables included type of assessment of students with ID: assessment of verbal participation, oral assessment, multiple choice, written assessment, other assessment, and no assessment. The participants responded to six questions (yes/no) concerning type of assessment, and multiple answers were allowed. These types of assessment cannot be classified as either summative or formative.

To study satisfaction with assessment of students with ID, we used three ordinal variables. The

variables asked the participants to rate fully agree (= 4), agree (= 3), disagree (= 2), or fully disagree (= 1) on the fairness of tests for reading ability, writing ability, and mathematical ability for students with ID.

As predictors, we measured participants' age and total years of teaching (discretized into three categories). We also measured efficacy for inclusive education using five questions, scored on a 1–4 scale. To measure job satisfaction, we used a set of 13 questions. To validate the measures, we conducted a factor analysis using principal factoring with the psych package in R (Revelle, 2018). Principal factoring provides a robust alternative to maximum likelihood. At the same time, principal factoring works better with small data sets and has theoretical plausibility compared to principal component analysis.

As can be seen in Table 1, the variables with loadings (coefficients) above 0.3 are in bold. The variables load on the principal factors as expected. The factor loadings indicate no problems with cross-loadings, suggesting that the factors can be treated as perpendicular. The Cronbach's alpha of the lower bound of the reliability of the scale is good for job satisfaction ($\hat{\alpha}=0.85$, **CI[0.88, 0.91]**), acceptable for self-efficacy for inclusion ($\hat{\alpha}=0.77$, **CI[0.71, 0.84]**), and good for self-efficacy for assessment ($\hat{\alpha}=0.89$, **CI[0.85, 0.92]**). Having validated the scales, we computed the average z-scores for each principal factor.

2.3 Strategies for data analysis

To analyse type of assessment of students with ID, we conducted a latent variable analysis (i.e. as a disposition) with the ltm-package (Rizopoulos, 2006) using the item response theory (IRT). We treated assessment practices as a continuous latent variable. A likelihood ratio test indicated that a one-parameter IRT model fit better with the data compared to a Rasch model (LR = 19.36, df = 1, $p < 0.001$). A two-parameter IRT model did not improve the fit (LR = 5.78, df = 3, $p = 0.123$).

In practical terms, we estimated four different difficulty parameters for each variable (i.e. how hard the type of assessment is). However, we only estimated one (unconstrained) discrimination parameter (i.e. how good the variable is at distinguishing between high and low assessment disposition).

Table 1

Outcome and Predictor Variables. Factor Analysis (PAF). A factor analysis with principal axis factoring (PAF), above .3 as blank

Questions	Job Satisfaction	Self-efficacy	Assessment satisfaction	Mean	SD	Min	Max
I think I am good at teaching in general		0.45		3.422	0.548	1	4
I think I am good at teaching students with disabilities		0.75		3.422	0.535	2	4
I think I am good at teaching students with intellectual disabilities		0.79		3.333	0.623	2	4
I think I am good at teaching students with language impairment		0.63		2.889	0.638	1	4
I think I am good at teaching students with autism		0.52		3.124	0.686	1	4
I have clear, reasonable, and meaningful goals	0.51			3.109	0.653	1	4
I have the materials and equipment I need to do a good job	0.38			2.871	0.752	1	4
I know what is expected of me at work	0.66			3.372	0.663	2	4
I can participate in decision-making	0.72			3.34	0.591	2	4
I have a good relationship with my principal	0.62			3.405	0.688	1	4
I have a good relationship with my co-workers	0.69			3.608	0.579	2	4
I have a good relation with parents	0.43			3.628	0.538	2	4
I have had a chance to participate in training and education in the last year	0.32			3.169	0.936	1	4
There is good morale and community at my workplace	0.77			3.238	0.762	1	4
I am satisfied with my workplace	0.75			3.497	0.645	2	4
My opinions are taken seriously	0.68			3.385	0.695	2	4
My co-workers want to do a good job	0.55			3.446	0.587	2	4
I have meaningful and stimulating tasks	0.69			3.622	0.552	2	4
I have the necessary reading tests to be able to assess my students			0.83	2.194	0.992	1	4
I have the necessary writing tests to be able to assess my students			0.90	2.008	0.94	1	4
I have the necessary mathematics tests to be able to assess my students			0.80	2.101	0.959	1	4
Total years teaching				49.707	9.356	23	65
Age in years				20.776	10.771	1	42

Consequently, we considered type of assessment of students with ID as a measure of the special educators' assessment practice disposition. To improve the fit of our model, we used four out of six variables for the one-parameter model and

removed "no assessment" and "other assessment" using a bootstrapped Pearson's chi-squared test ($p = 0.205$, $B = 1000$). Removing these variables seemed logical because they differed substantively from the others.

We checked that the measure was unidimensional (i.e. one and not several). The measure was also highly correlated with the total score (ranging from 0.58 to 0.63) with a Cronbach's alpha of 0.63.

To analyse the binary variables, we conducted binary and ordinal logistic regression (Agresti, 2015). We focused on reporting the probability and marginal change in probability (i.e. the derivative).

We conducted diagnostics of the residuals of the linear and logistic regressions. For the ordinal regression, we ran a proportional odds test (Agresti, 2015). We found no issues.

3. Results

In Figure 1, we summarise the descriptive results. Multiple choice and written assessments were the

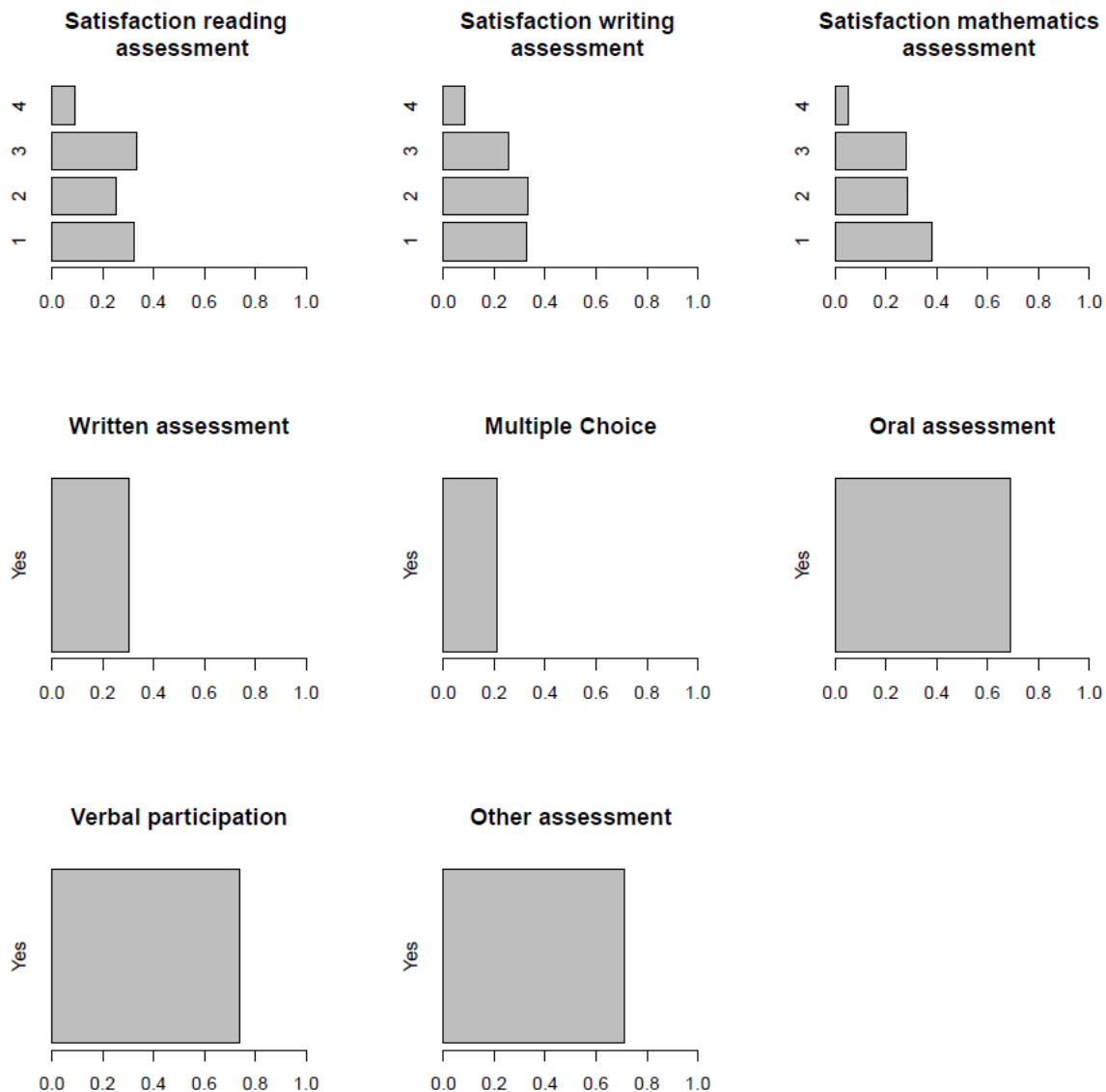


Figure 1. Proportions for assessment variables

In addition, we used the latent variable score as a predictor in linear regression. Finally, we generated tables and graphs with the aid of the following packages: stargazer (Hlavac, 2015), effects (Fox, 2003), and margin (Leeper, 2018). All statistical analyses were conducted using R (R Core Team, 2018).

least popular type of assessment, whereas oral assessment, verbal assessment, and other types of assessment were the most popular among the special educators. Clearly, special educators find speech-based assessment to be most useful. However, we do not know why more than half use other forms of assessment. To get some indication,

we considered the responses to assessment satisfaction.

The pattern could explain why special educators are unlikely to use written or multiple choice assessments. Most special educators did not strongly agree that they have the adequate tests to assess students in mathematics and writing. By contrast, a large share disagreed or strongly disagreed. Overall, special educators were fairly satisfied with reading tests. Next, we considered how difficult special educators find the various types of assessment.

Multiple choice is the most difficult type of assessment, closely followed by written assessment (Figure 2). Verbal participation and oral assessment are the easiest types of assessment. Consequently, special educators with a low assessment practice score can still conduct these

types of assessments. We plotted the coefficients for the Rasch model with their confidence intervals. None of the coefficients overlapped with zero. The difficulty parameters with greater values suggest greater difficulty. We observed that the discrimination coefficient was about 2, which suggests that the model has good discrimination between variables.

Next, we attempted to predict the assessment practice using the latent variable score (see Table 2). The variable score can be interpreted similar to logit scale.

The analysis suggests that there is a curved linear relationship (inverse U-shape) between satisfaction with assessment and the assessment practice score ($\frac{\partial y}{\partial x} \approx -0.45$ in logits).

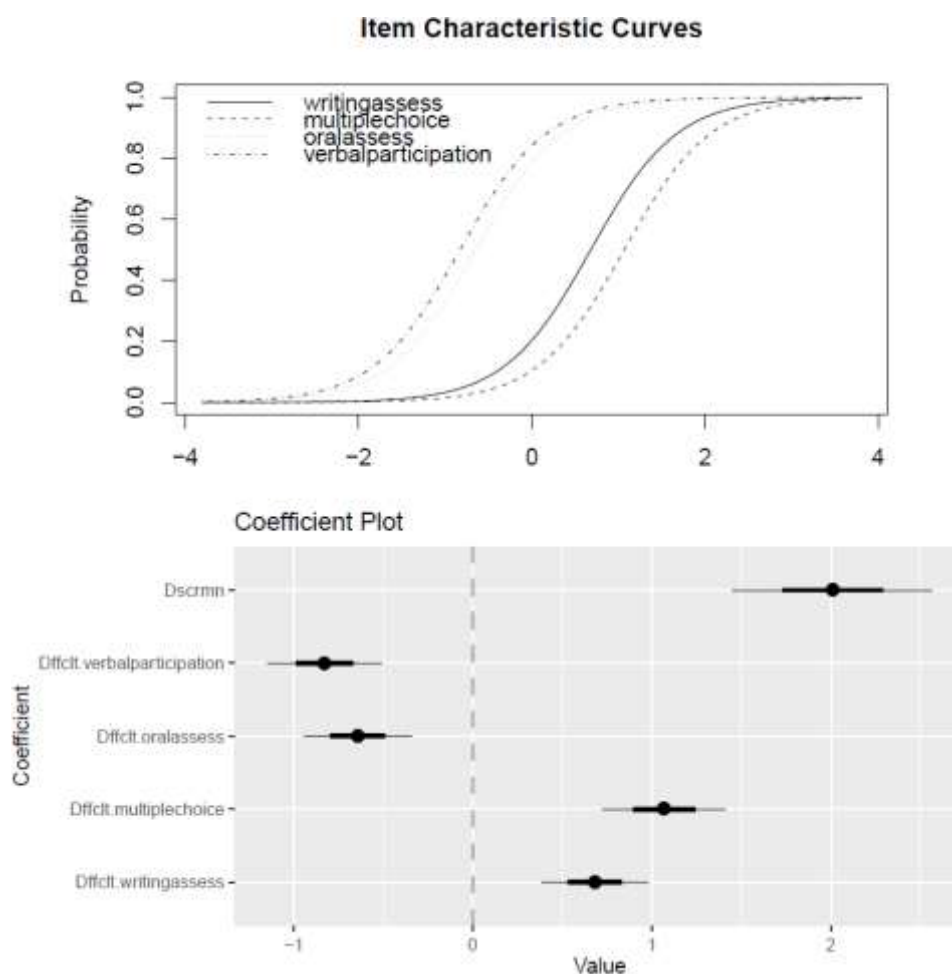


Figure 2. One-parameter item response theory model.

In addition, we found an interaction between total years of teaching and job satisfaction for special

educators with intermediate experience (compared to inexperienced special educators). We can

interpret the interaction by taking the derivative with respect to job satisfaction. One additional standard deviation in job satisfaction for special educators with moderate teaching experience is associated with a 0.122 increase in the latent score,

on average, adjusting for other predictors. Overall, the model explains 22% of the variation in the outcome. Moreover, the average deviation of the error is 0.68.

Table 2
Linear Regression and Binary Logistic Regression

	Outcome variable:				
	Assessment Practice <i>OLS</i> (1)	Multiple Choice <i>Logistic</i> (2)	Written assessment <i>Logistic</i> (3)	Verbal participation <i>Logistic</i> (4)	Oral assessment <i>Logistic</i> (5)
age45,55	-0.279 (0.183)	-1.165 (0.737)	0.056 (0.620)	-1.712* (0.819)	-0.487 (0.638)
age55,65	-0.061 (0.192)	-0.537 (0.727)	0.362 (0.666)	-1.043 (0.821)	0.195 (0.691)
totalteachyears15,25	0.075 (0.169)	0.322 (0.637)	-0.607 (0.575)	1.153 (0.722)	-0.131 (0.607)
totalteachyears25,42	0.003 (0.205)	0.001 (0.853)	-0.741 (0.746)	0.388 (0.761)	0.248 (0.735)
z.satisfaction	-0.451* (0.195)	-1.065 (0.744)	-0.365 (0.631)	-2.300* (0.974)	-1.285 (0.833)
z.efficacy	0.055 (0.100)	0.974* (0.447)	-0.261 (0.344)	0.071 (0.404)	-0.111 (0.372)
z.satisfactionassess	0.155 (0.083)	0.158 (0.317)	0.782* (0.331)	0.117 (0.299)	0.342 (0.282)
z.satisfactionassess (squared)	-0.302** (0.098)	0.269 (0.367)	-0.949* (0.418)	-1.480** (0.450)	-0.854* (0.343)
totalteachyears15,25:z.satisfaction	0.573* (0.253)	1.491 (0.994)	0.296 (0.838)	2.452* (1.226)	1.837 (0.992)
totalteachyears25,42:z.satisfaction	0.269 (0.260)	0.745 (1.104)	-0.507 (0.932)	1.485 (1.195)	1.464 (1.075)
Constant	0.329* (0.159)	-1.435* (0.601)	0.097 (0.540)	2.768*** (0.767)	1.871** (0.613)
Observations	103	103	103	103	103
R ²	0.218				
Residual Std. Error	0.684				

Note. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

We graphed the results in Figure 3: the latent score on the y-axis, and the satisfaction with assessment on the x-axis. The line indicates the predicted score with a 95% shaded confidence interval. The x-axis also includes a “rug” indicating sample observations. Those with moderate satisfaction with

assessment had the highest expected assessment practice score after adjusting for other predictors. In other words, those special educators with low/high satisfaction with assessment had a low assessment practice score when compared to those with moderate assessment satisfaction.

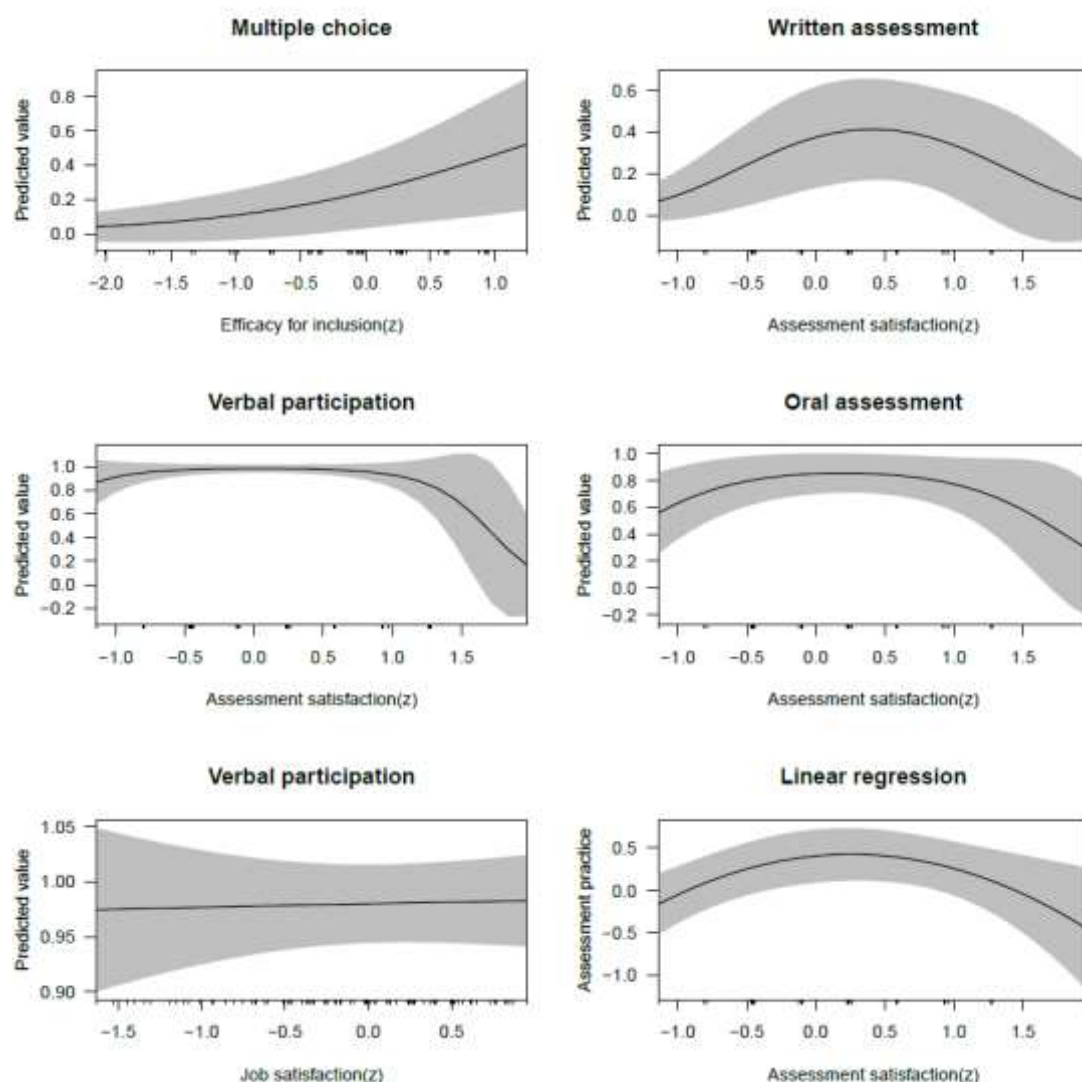


Figure 3. Predicted values for types of assessment and assessment practices.

In Table 2, we also report the individual types of assessment. We estimated four logistic regressions. Three of four indicate the same pattern, namely a curved linear relationship between assessment satisfaction and the propensity of assessment. Written assessment, verbal participation, and oral assessment all convey a curved linear relationship. By contrast, conducting multiple choice is unrelated to assessment satisfaction. Instead, self-efficacy for inclusion seems to be positively related to conducting multiple choice assessments. An additional standard deviation increase in self-efficacy for inclusion is associated with more than twice the odds of conducting multiple choice assessment ($\exp(\beta) = 2.7$), on average, after adjusting for other predictors.

All other predictors are not statistically significant, with the exception of the interaction between job

satisfaction and total years of teaching concerning verbal participation and written assessment. Note that when interpreting interactions, we focused on the interaction term.

As an indication of the model fit to the data, we computed the correlation (Agresti, 2015) between the fitted values and the outcome. We used a jack-knife approach (and averaged). The model of verbal participation provides the best fit ($r_{jack} = 0.53$), or 29% reduction in the misclassification. The multiple choice, written assessment, and oral assessment provide comparably low fit to the data ($r_{jack} = [0.36, 0.40, 0.38,]$).

To understand the patterns, we plotted the predicted probabilities for the outcomes and the self-efficacy, assessment satisfaction, and job satisfaction in Figure 3. The fitted probabilities are on the y-axis, and the predictor is on the x-axis. Again, the graphs

include a shaded 95% confidence interval and a “rug” indicating sample observations. The panels show how the propensity for each type of assessment follows a U-curve, with the exception of self-efficacy for inclusion. However, the curve for the two predictors seems nonlinear.

We also included a plot of the marginal change for each predictor in Figure 4. The marginal change is simply the derivative. For the linear regression, this means that we get a linear approximation for the score, which is negative. We suggest that these plots provide a better understanding of the pattern than can be deduced from the estimates in Table 2 alone.

Finally, we turn to the issue of what predicts assessment satisfaction. In Table 3, we present an ordinal logistic regression model of each individual type of assessment satisfaction: reading, writing, and mathematics. The results of writing and mathematics were similar. We also included an interaction with job satisfaction and total teaching years. We found that job satisfaction for special educators with intermediate experience was associated with a higher likelihood of being satisfied with mathematics tests for assessments, compared to special educators with low teaching experience, on average, after adjusting for other predictors.

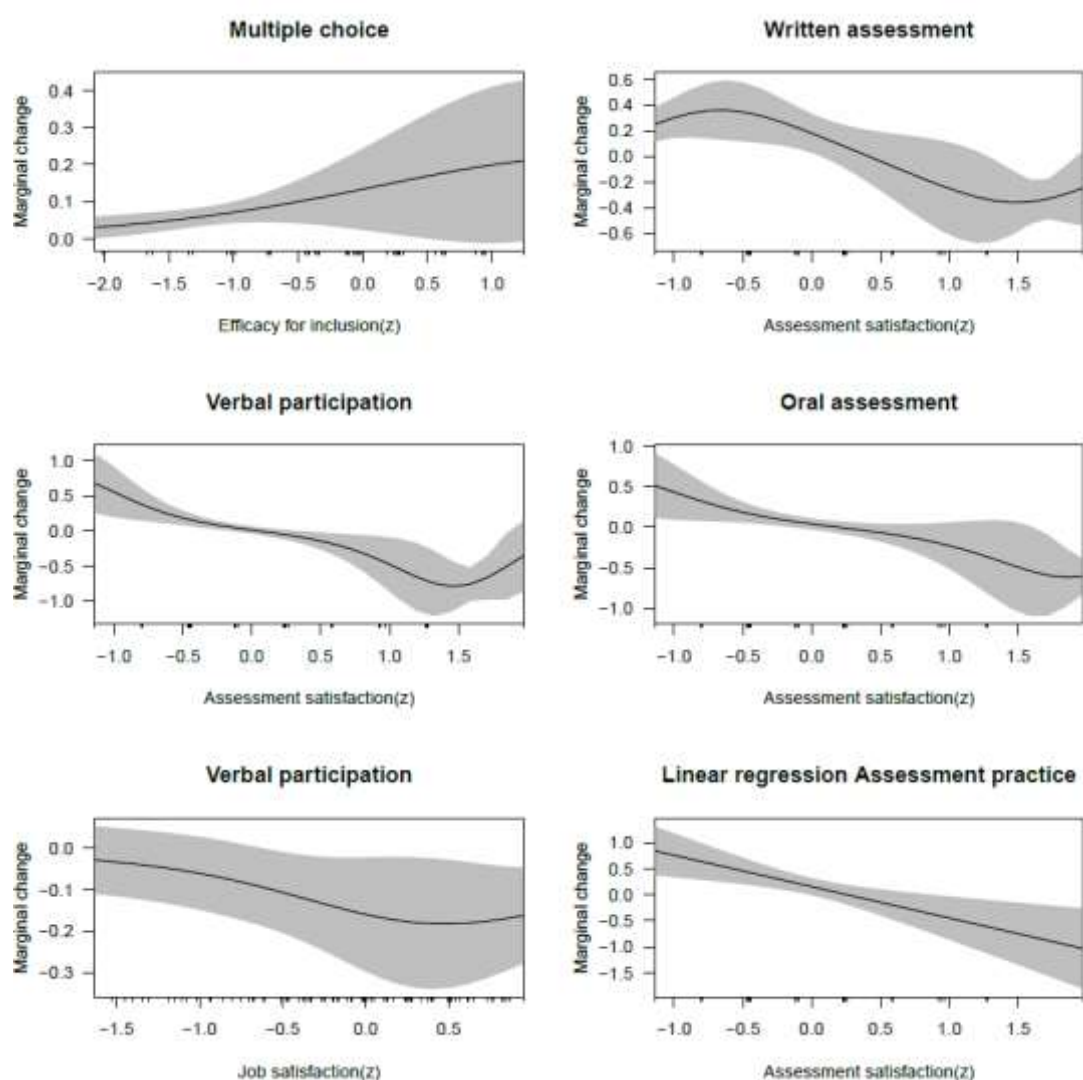


Figure 4. Marginal change in predicted values for types of assessment and assessment practices.

The same pattern holds for writing tests but not for reading tests. To assess the model fit, we computed

the correlation between the fitted probabilities and the outcome using jack-knife estimation. For the

mathematics model, we estimated a low to moderate correlation ($\hat{r}_{jack} = [0.38, 0.39, 0.10, -0.37]$). The correlation was slightly lower for satisfaction with written assessment ($\hat{r}_{jack} = [0.33, 0.32, 0.16, -0.31]$).

Although the lower order terms in the written assessment model are not statistically significant, this does not invalidate the interaction. Because interpreting coefficients in logistic regression models is difficult (Agresti, 2015), and interactions in particular, we plotted the fitted probabilities for mathematics assessment satisfaction in Figure 5.

Figure 5 strengthens our interpretation. The probability of scoring a 1 (*strongly disagree*) was

higher for inexperienced special educators for higher values of job satisfaction, after adjusting for other predictors. By contrast, the probability of scoring a 1 (*strongly disagree*) was lower for intermediate experienced special educators for higher values of job satisfaction, on average, after adjusting for other predictors. The probability of scoring a 3 was lower for inexperienced special educators for higher values of job satisfaction, on average, after adjusting for other predictors. By contrast, the probability of scoring a 3 was higher for intermediate experienced special educators for higher values of job satisfaction, on average, after adjusting for other predictors.

Table 3
Ordinal Logistic Regression

	Outcome variable:		
	readingtests (1)	writingtests (2)	mathtests (3)
z.efficacy	-0.062 (0.267)	0.005 (0.274)	0.209 (0.267)
age45,55	-0.352 (0.488)	-0.064 (0.493)	0.261 (0.489)
age55,65	-0.048 (0.519)	0.068 (0.534)	0.452 (0.527)
totalteachyears15,25	0.223 (0.445)	0.636 (0.460)	0.940* (0.461)
totalteachyears25,42	-0.157 (0.545)	-0.427 (0.583)	-0.397 (0.564)
z.satisfaction	-0.596 (0.524)	-0.773 (0.528)	-1.084* (0.519)
totalteachyears15,25:z.satisfaction	0.782 (0.684)	1.552* (0.732)	1.675* (0.677)
totalteachyears25,42:z.satisfaction	0.284 (0.693)	0.063 (0.726)	0.751 (0.707)
Observations	103	103	103

Note. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

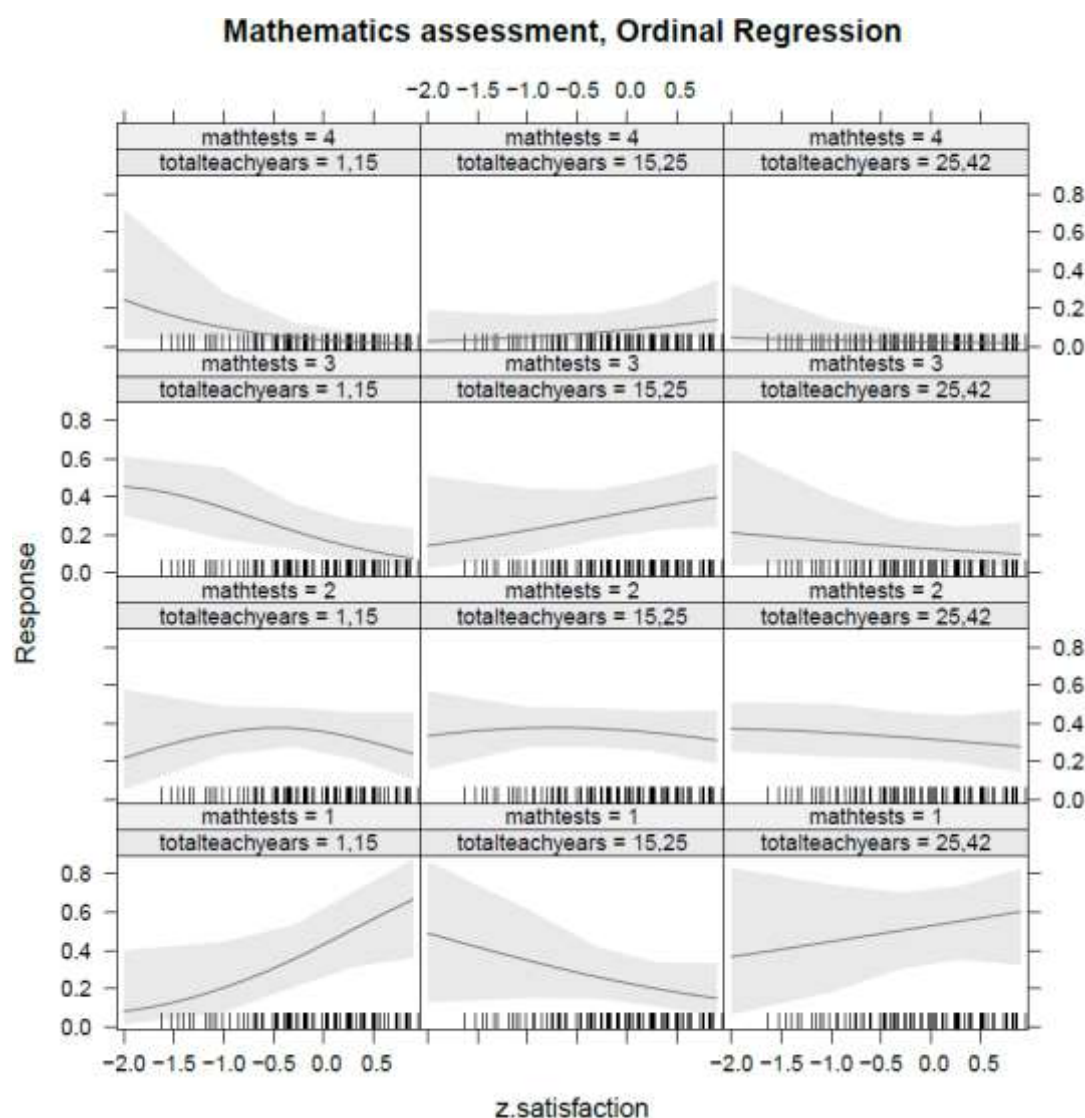


Figure 5. Predicted values for satisfaction with mathematics assessment.

4. Discussion

Assessment of learning outcomes (knowledge, skills, beliefs, and attitudes) is central in both mainstream and special needs comprehensive schools (Black & Wiliam, 2010; Bolt et al., 2011; Davies et al., 2017; Thurlow, Lazarus, & Hodgson, 2012). Assessment matters to clarify if students have met the curriculum goals, to support the development of students, for decisions concerning special needs instruction and for teachers' lesson planning. Although most studies have focused on teachers' assessment in mainstream schools, less is known about special educators' assessment in special needs comprehensive schools. Our purpose was to describe and predict the type of assessment practices Swedish special educators conduct in special needs comprehensive schools for students with ID.

Our findings contribute to the refined understanding of how assessment satisfaction, self-efficacy, and job satisfaction matter for special educators' assessment of students with ID. Because special educators require greater assessment practice of students with ID, we believe these findings contribute to educational theory, special education programmes, and educational policy. Several studies have shown the importance of self-efficacy for special educators (Viel-Ruma, Houchins, Jolivette, & Benson, 2010). However, our study stresses the importance of the specific assessment satisfaction as opposed to self-efficacy for inclusion. Although self-efficacy for inclusion clearly matters, it was not as important as predicting assessment conduct. We choose to interpret the predictive importance of satisfaction

as lending support to Hargreaves (2000) that assessment is indeed emotionally embedded.

In the context of education, our study relates to the educational discussion on assessment to ensure accountability of the curriculum. Although, assessment has been a topic of much criticism, it still serves to safeguard students with ID:s progression. Our study also relates to the educational discussion on educators' emotions and self-efficacy in predicting special educators' practices and attitudes. Here assessment seems to be one specific example out of many to understand why educators tend to implement a policy or not.

4.1 Limitations

Our study has several limitations. First, we used a non-random sample. Although this does not invalidate the use of inferential statistics, we cannot generalise our findings. At best, we can predict patterns in the data that might be interesting to researchers, teachers, and special educators as such. Second, due to the small sample size, our study lacks statistical power. Thus, we can at best discover large differences. Third, our study has measurement issues. Ideally, we should have fitted a measurement model for our predictors. Beyond reliability, our measure of satisfaction with assessment only included questions about testing and lacked other aspects of assessment. We are also aware that self-efficacy for inclusion can be measured in several ways. Fourth, we have not assessed the possibility of a reciprocal association: assessment practices may perhaps also predict greater feelings of job satisfaction.

4.2. Recommendations

We suggest that special teacher educators equip pre-service special educators (and in service) with the necessary competence to include different kind of assessments. Specifically, special educators may need skills to both develop and administrate written tests and multiple choice tests. Including exposing pre-service special educators (and in service) to a variety of test methods. Promoting inclusion is important but not sufficient. From a policy perspective, we suggest that more attention to assessment is given in special education programmes for assessing students with ID. Finally, we encourage test-developers to design more adequate mathematics tests for students with ID.

5. Conclusions

We found that special educators had the greatest difficulty conducting multiple choice and written assessments. The special educators found it easier to conduct oral assessments and assess verbal participation.

We found that special educators' assessment satisfaction had a curvilinear relation to their assessment practice score. In practical terms, special educators with moderate assessment satisfaction had the highest assessment practice score. In addition, job satisfaction interacted with moderate educator experience for assessment.

We found that there is a curvilinear relationship between special educators' assessment satisfaction with verbal participation, oral assessment, and written assessment. By contrast, self-efficacy for inclusive education predicted conducting multiple choice assessments. In addition, job satisfaction interacted with moderate teacher experience for assessment by verbal participation and written assessment.

We found an interaction between total teaching years and job satisfaction. Special educators with higher job satisfaction and moderate teaching experience are more likely to be satisfied with writing and mathematics tests compared to inexperienced special educators.

Conflicts of interests

Authors declare no conflict of interests.

References

- Agresti, A. (2015). *Foundations of linear and generalized linear models*. Hoboken, NJ: John Wiley & Sons.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W H Freeman/Times Books/ Henry Holt & Co.
- Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81–90.
- Bolt, S. E., Decker, D. M., Lloyd, M., & Morlock, L. (2011). Students' perceptions of accommodations in high school and college. *Career Development for Exceptional Individuals*, 34(3), 165–175.
- Davies, D. K., Stock, S. E., King, L., Wehmeyer, M. L., & Shogren, K. A. (2017). An accessible testing, learning and assessment system for people with intellectual disability. *International Journal of Developmental Disabilities*, 63(4), 204–210.

- Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, 8(15), 1–27.
- Fujiura, G. T., & RRTC Expert Panel on Health Measurement. (2012). Self-reported health of people with intellectual disability. *Intellectual and Developmental Disabilities*, 50(4), 352–369.
- Hargreaves, A. (2000). Mixed emotions: Teachers' perceptions of their interactions with students. *Teaching and Teacher Education*, 16(8), 811–826.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Hlavac, M. (2015). *Stargazer: Well-formatted regression and summary statistics tables*. R package (Version 5.2).
- Leeper, T. J. (2018). margins: Marginal Effects for Model Objects. R package (Version 0.3.23).
- Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: Teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, 25(5), 442–467.
- R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Revelle, W. (2018). *psych: Procedures for personality and psychological research*. Evanston, IL: Northwestern University.
- Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17(5), 1–25.
- Swedish Schools Inspectorate. (2010). <https://www.skolinspektionen.se/sv/Beslut-och-rapporter/Publikationer/Granskningsrapport/Kvalitetsgranskning/Undervisningen-i-svenska-i-grundsarskolan/>
https://sir.is.skolverket.se/reports/rwervlet?cmdkey=common&geo=1&report=grsar_personal&p_sub=1&p_flik=G&p_ar=2016&p_lankod=&p_kommunkod=&p_skolkod=&p_hmantyp=&p_hmankod= Accessed 21 June 2019.
- Tanis, E. S., Palmer, S., Wehmeyer, M. L., Davies, D., Stock, S., Lobb, K., & Bishop, B. (2012). Self-report computer-based survey of technology use by people with intellectual and developmental disabilities. *Intellectual and Developmental Disabilities*, 50(1), 53–68.
- The Swedish National Agency for Education. (2017). Grundsärskolan Personalstatistik Läsåret 2016/17.
- Thurlow, M. L., Lazarus, S. S., & Hodgson, J. R. (2012). Leading the way to appropriate selection, implementation, and evaluation of the read-aloud accommodation. *Journal of Special Education Leadership*, 25(2), 72–80.
- Tumbull III, H. R., Tumbull, A. P., Wehmeyer, M. L., & Park, J. (2003). A quality of life framework for special education outcomes. *Remedial and Special Education*, 24(2), 67–74.
- Viel-Ruma, K., Houchins, D., Jolivet, K., & Benson, G. (2010). Efficacy beliefs of special educators: The relationships among collective efficacy, teacher self-efficacy, and job satisfaction. *Teacher Education and Special Education*, 33(3), 225–233.
- Wehmeyer, M. L., & Abery, B. H. (2013). Self-determination and choice. *Intellectual and Developmental Disabilities*, 51(5), 399–411.