International Journal of Educational Research xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

International Journal of Educational Research

journal homepage: www.elsevier.com/locate/ijedures

What if compulsory schooling was a 21st century invention? Weak signals from a systematic review of the literature

Jason McGrath*, John Fischetti

School of Education, The University of Newcastle, Hunter Building, University Drive, Callaghan, NSW, 2308, Australia

ARTICLE INFO

Keywords:
Teacher-centred instruction
Student-centred instruction
Standardised assessment
Formative assessment
Compulsory schooling
Future focused learning
21st Century learning
Systematic review
Foresight

ABSTRACT

Counterfactual thinking and foresight strategy are utilised to identify "weak signals" for future practice. We conducted systematic reviews of empirical studies comparing traditional and alternative approaches to instruction or assessment to examine the potential for compulsory schooling to be redesigned. This thorough scan of the scholarly horizon provides an opportunity to take the pulse from the literature published over the first one-eighth of the 21st century. Sixteen "weak signals" were identified from the literature that are discussed in the paper. This review might inform further research into current and future school designs and models, particularly in light of trends to personalise learning that are taking hold around the world.

1. Introduction

Galileo is credited with perhaps the most famous example of counterfactual thinking in history (Palmieri, 2005) with his thought experiment that challenged the dominant thinking of Aristotle about the principles of motion. A contemporary example of counterfactual thinking is a TED talk by Müllenbeck (2017) asking "what if we paid doctors to keep us healthy?"

Many scholars (Abbott, 2010; Griffin, Care, & McGaw, 2011; Hargreaves, 1999; Robinson & Aronica, 2015; Trilling & Fadel, 2009) have argued that the design of compulsory schooling in the western world is a construct of the industrial revolution that needs to be significantly redesigned, transformed or revolutionised for the 21st century.

Two key aspects debated as part of these discussions relate to the role of teacher and learner as well as types of assessment. Robinson and Aronica (2015) recast a theatre analogy by Peter Brook to describe the essence of schooling as the relationship between a learner and a teacher, highlighting the role of the learner and teacher as the core of education (Brown, 2010; Cribiore, 2005; Robinson & Aronica, 2015). Assessment cannot only be seen as a driver of learning but the policies and philosophies feature in debates about the merits of current and future ideas about schooling.

In this paper we use counterfactual thinking to examine the research question, "what if compulsory schooling was a 21st century invention"? A systematic review of the literature has been undertaken in order to identify possible weak signals as a part of utilising a foresight approach to examine (a) the role of the teacher and the learner and (b) assessment practices. This thorough scan of the scholarly horizon provides an opportunity to take the pulse from the literature published over the first one-eighth of the 21st century. This review might inform further research into current and future school designs and models.

Both meta-analysis and systematic reviews have been growing in popularity since the end of the last century as a way of synthesising research around an inquiry question (Ahn, Ames, & Myers, 2012; Littell, Corcoran, & Pillai, 2008). The use of a common

* Corresponding author.
E-mail addresses: Jason.mcgrath@uon.edu.au (J. McGrath), john.fischetti@newcastle.edu.au (J. Fischetti).

https://doi.org/10.1016/j.ijer.2019.02.006
Received 17 July 2018; Received in revised form 8 November 2018; Accepted 7 February 2019

effect size score (Littell et al., 2008) for the former has added to its value at a time when evidence based practice in education is promoted. However, we have chosen to use a systematic review because of the exploratory nature of our inquiry, a preference to unpack aspects such as impact on various equity groups, and assess the overall quality of the available research. To enhance our approach we have drawn on the work of Gough (2007) in relation to systematic reviews. Rather than ascribing a particular value to the type of method used in a piece of research, Gough (2007) argues that we need to analyse the quality of the research method as it is applied within a piece of research and even to consider the relevance of any study in terms of the research questions we set for inquiry.

1.1. Teacher centred versus student centred approaches to teaching and learning

Teacher-centred approaches are rooted in the pre-professional age of teaching, characterised by traditional approaches with few resources and a focus on classroom control as a measure of success (Hargreaves, 1999). The teacher is generally at the front of the classroom engaged in "chalk and talk" instruction involving didactic, traditional teacher centred approaches with a teaching focus on "transmitting information" (Yeung, Craven, & Kaur, 2014, p. 308) that treats skills and knowledge as isolated parts that can be taught separately. Students generally take a passive role relative to that of the teacher (Yeung et al., 2014).

Student-centred approaches are more possible in an era, commencing in many countries in the mid-1980s, of "collegial professionalism marked by the existence of greater and greater student diversity and a concomitant effort to match this with an expanding repertoire of varied teaching strategies" (Hargreaves, 1999, p. 349). The availability of new technologies (Trilling & Fadel, 2009) offers further opportunities to support different approaches to classroom practice.

1.2. Priority for summative versus formative assessment approaches

Guskey (2003) challenged the field to move away from assessment of learning to assessment for learning. Johnson and Burdett (2010, p. 123) highlight some of the tensions, including between:

> summative and formative functions… questioning whether the focus of education should be on the system or the individual, whether learning should be teacher- or student-focused, and whether assessment should primarily summarise learning outcomes or inform future learning (p. 123).

Should assessment practices reflect our changing pedagogical emphases to meet contemporary needs (Griffin et al., 2011)? How do we raise concerns about the current emphasis on high stakes testing which can impact on the former (Trilling & Fadel, 2009)? How do we decouple our assessment and testing practices from the assumptions of intelligence and learning capacity since they are often seen as a driver for teaching and learning (Trilling & Fadel, 2009) and achievement (Black & Wiliam, 2004)? How do we take back the tendency for schools to be identified by testing results as opposed to learning results?

2. Theoretical framework and research questions

The goal of this systematic review of the literature is to examine benefits reflected in current research about different approaches towards compulsory schooling.

Two specific research questions contribute to the overarching counterfactual research question:

1 How might the role of teacher and learner be conceptualised if compulsory education was a 21st century invention?
2 How might assessment practices be conceptualised if compulsory education was a 21st century invention?

Questions to guide the review include (a) what examples of alternative approaches currently exist? (b) to what extent does the evidence of impact show alternative approaches are more or less effective than current approaches? and (c) do alternative approaches offer any benefits for equity groups?

This research contributes to an evaluation of contemporary approaches to teaching, learning and assessment practices by systematically reviewing empirical research available at the beginning of the new millennium.

2.1. Counterfactual thinking

Counterfactual thinking is a way of responding to a rapidly changing world that is highly reflexive (Giddens, 1991; Jenssen, 2010) by expanding simplistic prediction models such as trend forecasting (Garrouste, 2011) to consider "What is the causal relationship of interest?" (Angrist & Pischke, 2009, p. 3). Counterfactual approaches provides a modelling process to unpack the impact of a treatment on various groups in order to consider equity issues, including which groups might benefit (Holm, Jæger, Karlson, & Reimer, 2013; Selwyn & Facer, 2013; Williamson, 2013) from different approaches.

Although there are differences in methods, a quality counterfactual and factual study are similar in that "the plausibility of factual and counterfactual arguments alike rests on the appeal of their assumptions, the tightness of the logic connecting cause to effect and the richness of the evidence that supports" (Lebow, 2009, p. 67). Counterfactual thinking provides a creative but rigorous way of thinking about problems and construct possible futures.

2.2. Foresight strategy

Slaughter and Riedy (2009) defines strategic foresight as the "fusion of futures methods with those of strategic management… the ability to create and maintain a continuous high-quality, coherent and functional forward view, and to use the insights arising in useful organizational ways…" (p. 11). Foresight is linked to development of a "knowledge economy" (Horton, 1999) and "innovation" (Horton, 1999; Miles, 2012) which are being used increasingly by governments and inter-government organisations (Dreyer & Stang, 2013; Jenssen, 2010; Schmidt, 2015) to respond to a rapidly changing world (Hartmann & Stillings, 2015; Jenssen, 2010; Schmidt, 2015). This future focussed work adds the concept of foresight to traditional approaches of organisational planning using hindsight and insight.

Weak signals are "the early signs of possible but not confirmed changes that may later become more significant indicators of critical forces for development, threats, business and technical innovation… [and] …represent the first signs of paradigm shifts, or future trends, drivers or discontinuities" (Saritas & Smith, 2011, p. 297). Identification and critical analysis of weak signals benefits strategic decision-making, planning and action as part of utilising foresight strategies.

3. Methodology

3.1. Systematic review

Reviews of research play an important role in understanding future developments for practice and policy (Hallinger, 2014). The use of a systematic review approach provides a way to focus on the quality of research papers which can enhance the overall review of the literature (Wells & Littell, 2009). This is particularly important given the large amount of research available electronically in today's world (Arthur, Waring, Coe, & Hedges, 2012). Systematic review methods reduce bias (Arthur et al., 2012) and outline criteria and approaches as a priori, including search terms and databases used to source articles.

This review of the literature examines empirical evidence about approaches to schooling, discussed in the literature within the first 18 years of the 21 st century, in order to assess the range of ideas being explored and quality of evaluation undertaken. The review examines the potential impact of approaches to improve outcomes for different equity groups of students (Welch et al., 2016).

3.2. Inclusion and exclusion criteria

The review included sources from multiple aspects of the literature including peer reviewed journal articles, book chapters, theses and conference papers. It was limited to empirical studies written in English and indexed between January 2000 and December 2017. Databases searched were EBSCO Education Research Complete, Education Resources Information Center (ERIC) and Informit.

A separate search was conducted for each of the specific research questions. For research question one search terms were "chalk and talk" OR "didactic teaching" OR "front of classroom" OR "teacher centred/centered" OR student centred/centered". For research question two search terms were "testing" OR "assessment" AND "system accountability" OR "academic achievement". Common terms added to all searches were AND "learning improvements" OR "learning outcomes" to highlight possible empirical studies and alternative approaches. The terms NOT "higher education, NOT "Universities & Colleges" were used to limit searches to a schooling focus between Kindergarten and high school.

In the screening phase (Gough, 2007) abstracts were reviewed, including to check for a K-12 schooling focus and an empirical study. Reasons for exclusion were tallied using an emergent approach to identify relevant categories. Where relevant to the specific research questions, studies related to teacher training programs were included as part of this criteria.

In the mapping phase (Gough, 2007) studies were assessed in two different ways. Firstly, studies were categorised using an emergent approach. Secondly, studies were assessed in terms of quality based on Gough's "weight of evidence framework" (2007, pp. 221–222). Gough argues that assessing the quality of a study is complex and that it should extend beyond just identifying the type of review, such as preferencing a Randomised Controlled Trial, to include an assessment of the quality of execution of the study which may add value to studies using methods that might in the first instance be considered of a lower quality. The value of a study should also be considered in relation to the purpose of the systematic review. We have developed a four-level rubric (see Table 1) for each aspect of Gough's weight of evidence framework to support our analysis.

3.3. Analytical strategy

Each article identified for full review was read to extract key information which was recorded in a spreadsheet (see Appendix 1). Headings for data extraction were citation and database(s) article was found; research methods, sampling processes, period of time for initiative and any follow up period; studied population including country, subject, cohort; the level of innovation being proposed; outcomes reported, including any benefits for equity and other groups.

Each article was also scored for each level of the rubric developed (Table 1) and a combined score. The scores were used to inform the synthesis of data and development of 'weak signals', including when deciding which studies to emphasise and any limitations to acknowledge. Studies with higher scores on the rubric feature more strongly in the synthesis provide within the findings section.

Educational research was synthesised using weak signals. The significance of the weak signals is to identify "the early signs of possible but not confirmed changes that may later become more significant indicators of critical forces for development, threats, business and technical innovation" (Saritas & Smith, 2011, p. 297). Whilst weak signals do not attempt to predict the future (Horton,

Table 1
Weight of evidence framework for overall assessment of each study.

	1	2	3	4
Methodological quality	Research design is limited.	Research design is logical.	Research design is clearly developed with a focus on evaluation.	Highly developed research design with a strong evaluative focus that is well executed.
	There is some empirical evidence.	The quality of evidence, particularly in terms of validity/ reliability/ generalisability, is limited beyond an evaluation of the sample studied.	The quality of evidence collected is coherent which addresses the research question.	The quality of evidence collected is coherent with high levels of integrity to address the research question.
Methodological relevance	Research Design and/ or execution provide for limited input to research questions.	Research Design provides for some implicit comparison between an alternative approach to a traditional approach.	Research Design provides for some comparison between an alternative approach to a traditional practice that are relevant to the research questions. This might include one off or short time period limited sample size a pilot study.	Research Design provides a sustained exploration between an alternative approach to a traditional practice linked closely to the research questions. This might include extended time period scalable sample size.
Topic relevance	Findings have limited relevance to research questions.	Findings provide an incidental contribution to the research questions.	Findings provide some insights to compare an alternative approach to a traditional practice: Findings provide insights into benefits for equity groups.	Findings compare an alternative approach to traditional practice: Findings provide insights into benefits for equity groups.

Table 2
Summary of samples, length of time for initiatives and Gough weighting scores.

Studies	Part A: (% of sample size for Part A)	Part B: (% of sample size for Part B)	Total (% of total sample)
Total articles	N = 79	N = 77	N = 156
Collected student level data	N = 34 (43%)	N = 56 (73%)	N = 90 (58%)
Including sample sizes of greater than 100 students	N = 19 (24%)	N = 40 (52%)	N = 59 (38%)
Collected data from current teachers	N = 15 (19%)	N = 26 (34%)	N = 41 (26%)
Including sample sizes of greater than 100 teachers	N = 6 (8%)	N = 4 (5%)	N = 10 (6%)
Collected data from pre-service teachers	N = 9 (11%)	N = 2 (3%)	N = 11 (7%)
Including sample sizes of greater than 100 pre-service teachers	N = 3 (4%)	N = 0 (0%)	N = 3 (2%)
Evaluated initiatives that had been in place for at least one year	N = 10 (13%)	N = 34 (44%)	N = 44 (28%)
Weight of Evidence Overall Rating (Gough, 2007)	Median = 9 Mode = 9	Median = 9 Mode = 10	

1999) they can highlight aspects within the research literature that could be explored further in order to influence future directions.

3.4. Limitations

This systematic review is limited to three databases. Given the exploratory purpose (Hallinger, 2014) of this review the search strategy has not been exhaustive (Gough, 2007), such as including hand searches (Gough, 2007). One reviewer conducted the screening and mapping phases.

The search has also been limited to the 21st century as a deliberate strategy to consider current and possible future schooling options. This may have precluded key studies from the 20th century and classic philosophic literature written prior to formal research in education was introduced.

4. Results and discussion

A systematic review was conducted in two parts, one for each of the specific research questions. Table 2 summarises some features of articles that were examined in the mapping phase.

4.1. Systematic review part A: how might the role of teacher and learner be conceptualised if compulsory education was a 21st century invention?

The systematic review for Part A identified 591 possible articles. This was subsequently reduced to 171 studies after the initial scan and 79 articles during the mapping phase. Table 3a provides an overview of the search and features of the articles included.

4.2. Systematic review part B: how might assessment practices be conceptualised if compulsory education was a 21st century invention?

The systematic review for Part B identified 684 possible articles. This was subsequently reduced to 416 studies after the initial scan and 77 articles after the mapping phase. One reference was to a book with various authors and led to two chapters being included at the mapping phase. Table 3b provides an overview of the search and features of the articles included.

4.3. Findings and weak signals

Sixteen weak signals have been distilled from a synthesis of the research following mapping phase from Part A and Part B. The weak signals have been categorised into four overall findings that summarise the educational research into (a) areas which are

Table 3a
Overview of the number of articles searched and some of the features of the articles.

Scanning phase- Potential articles	591
Scanning phase- Included/Mapping phase- Potential articles	171
Mapping phase- Excluded	not K-12 studies (N = 14) not empirical studies (N = 31) not relevant/ peer reviewed (N = 65)
Mapping phase- Included	60
Features of articles included	24 countries and 5 continents Articles related to primary level (N = 12), secondary level (N = 41) Most common subject of studies - Science (N = 31)

Table 3b
Overview of the number of articles searched and some of the features of the articles.

Scanning phase- Potential articles	684
Scanning phase- Included/Mapping phase- Potential articles	416
Mapping phase- Excluded	not K-12 studies (N = 15) not empirical studies (N = 124) not relevant (N = 197) other reasons (N = 4)
Mapping phase- Included	77
Features of articles included	16 countries and 5 continents Articles related to primary level (N = 9), secondary level (N = 21), primary and secondary (N = 45) Most common subject of studies – Mathematics (N = 33), Reading (N = 30), Science (N = 12), History or Social Science (N = 4)

burgeoning, (b) areas that are underrepresented despite their importance, (c) emergent hypotheses for further inquiry, and (d) claims that are contested.

These are summarised in Table 4.

5. Finding one: weak signals point to areas of inquiry that are burgeoning

5.1. Weak signal: extensive inquiry into the impact of standardised assessment

Fifty-five of the 77 total studies included in the mapping phase for Systematic Review Part B related to the impact of standardised assessment. Sixty studies related specifically to or included USA, particularly National Assessment of Educational Progress (NAEP) data in the context of No Child Left Behind policy (NCLB) or more generally in relation to international data collection such as Programme for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS).

Other weak signals developed from these studies are reported below.

5.2. Weak signal: assessment for learning approaches requires support to professionally develop staff and ensure consistency but can lead to learning that sticks

Assessment for learning approaches was extensively considered within Part B Systematic Review.

Table 4
List of the findings and weak signals that emerged from a synthesis of the data.

Finding One: Weak signals point to areas of inquiry that are burgeoning	5.1 *Weak Signal: Extensive inquiry into the impact of standardised assessment*
	5.2 *Weak Signal: Assessment for learning approaches requires support to professionally develop staff and ensure consistency but can lead to learning that sticks*
	5.3 *Weak Signal: Changing teacher practices*
	5.4 *Weak Signal: Some assessment approaches seek to create a personalised approach or "Intelligent accountability"* (Colbert et al., 2012)
	5.5 *Weak Signal: Technology-focussed approaches*
Finding Two: Weak signals point to areas that are underrepresented despite their importance	6.1 *Weak Signal: Lack of large scale, longitudinal comparative studies*
	6.2 *Weak Signal: Student Centred or Assessment for learning practices might offer some potential benefits for indigenous education*
	6.3 *Weak Signal: The impact of poverty beyond the school*
Finding Three: Weak signals point to emergent hypotheses for further inquiry	7.1 *Weak Signal: Significant positive differences for experimental groups focussed on student centred approaches or formative assessment practices, compared with control groups, were more likely when the intervention lasted for more than four weeks*
	7.2 *Weak Signal: Standardised assessment may better gauge top end performance but can lead to criteria compliance that limits the development of advanced learners*
	7.3 *Weak Signal: No evidence of a reduction in gender gaps through standardised assessment*
	7.4 *Weak Signal: Standardised assessment may push for improvement but this could be a proxy measure for school improvement practices*
Finding Four: Weak signals point to claims that are contested	8.1 *Weak signal: some possible advantages of teacher-centred approaches to support lower level students (remedial work) whilst students in the middle and upper bands are more likely to be advantaged from student centred approaches*
	8.2 *Weak Signal: Standardised assessment might be leading to poorer pedagogy for equity groups*
	8.3 *Weak Signal: Standardised assessment is accepted by parents and students but is more likely to promote shallow, test focused pedagogy*
	8.4 *Weak Signal: Monitoring of system level improvements through standardised assessment can lead to distortion of results through the use of exemptions for equity groups*

Challenges of utilising formative assessment approaches includes varying expertise of teacher capacity to use assessment practices focussed on assessment for and as learning (Faria, Greenberg, Meakin, Bichay, & Heppen, 2014; Trumbull, Sexton, Nelson-Barber, & Johnson, 2015) and an over reliance on observation with little value placed on professional learning by teachers (Burke & Wang, 2010). There was a conflict of purpose when these type of assessment processes are used for accountability purposes (Brown, 2011; Charteris, 2016; Frederiksen & White, 2004; Johnson & Burdett, 2010) or items are focussed on preparing for standardised assessment (Burke & Wang, 2010), limiting their effectiveness to provide feedback to students and support improvements in learning.

However, multiple studies investigating assessment approaches that were not reliant on standardised assessments found positives for changing the quality of assessment practices within classrooms. This includes highly structured approaches such as Response to Intervention (VanDerHeyden, Witt, & Barnett, 2005), assessment for learning (Black & Wiliam, 2004; Charteris, 2016; Kloser, 2014) and assessment as learning (Blank & Hewson, 2000; Bourke, 2016; Charteris, 2016). Benefits include assessment that is flexible and adaptive (Kloser, 2014); builds learner capacity in self-assessment to deepen the learning process (Black & Wiliam, 2004; Bourke, 2016; Charteris, 2016; Frederiksen & White, 2004) or "elicit student thinking" (Kloser, 2014, p. 1210); and provides feedback to support learners (Charteris, 2016; Kloser, 2014). A number of the studies found student learning outcomes improved when teachers (Black & Wiliam, 2004; Blank & Hewson, 2000; Faria et al., 2014; Frederiksen & White, 2004) and leaders (Faria et al., 2014) analysed learning data as part of teaching and learning; there was a correlation between the quality of student work with the quality of assignments (Koh & Luke, 2009); and learning gains could be found in post-test data analysis up to six months after learning took place (Blank & Hewson, 2000).

5.3. *Weak signal: changing teacher practices*

The potential to change practice of current teachers (Al-Amoush, Usak, Erdogan, Markic, & Eilks, 2013; Martin & Hand, 2009; Ramnarain & Hobden, 2015; Wang, Hsu, Reeves, & Coster, 2014; Yeung et al., 2014) or pre-service teachers (Al-Amoush et al., 2013; Buldur, 2017; Dickson & Kadbey, 2014; Hudson, 2007; Laronde & MacLeod, 2012; Ucar, 2012) was examined by multiple researchers.

Challenges that were identified included entrenched teacher centred practices (Al-Amoush et al., 2013; Ampadu, 2012), which in some cases might be showing some indication of changing towards more student centred approaches (Aydogdu & Selanik-Ay, 2016), as well as other challenges from student perspectives and involvement (Ampadu, 2012; Kontkanen et al., 2017). A number of studies found variation between teacher perceptions about their practice compared with observation of their practice (Kaya, Kablan, Akaydin, & Demir, 2015; Onurkan Aliusta & Özer, 2017). Mameli and Molinari (2017) found teacher centred approaches were more likely to predict teacher burnout compared with student centred approaches.

The process of "unlearning" traditional practices and moving to more inquiry based approaches requires a large amount of time even with ongoing professional learning (Martin & Hand, 2009; Miranda & Damico, 2015; Wang et al., 2014). Building of teacher agency was found to lead to increased learner agency, such as through the use of formative assessment (Charteris, 2016), or could lead to further benefits if professional support focused on a progression to explicitly develop learner autonomy (Ramnarain & Hobden, 2015). Yeung et al. (2014) found a connection between teacher self-concept and the valuing of learning as influencing the type of teaching approaches adopted. Lebak and Tinsley (2010) was able to show change in teaching approaches towards more student centred approaches as a result of a 14 week period project that utilised action research comprising video observation of teachers' own lessons and peer feedback. Wang et al. (2014) mapped gradual change over two years as part of a collaborative ICT professional learning project.

Ucar (2012) and VanWeelden and Whipple (2007) found it was possible to shape pre-service teacher views over the length of their course although it was a challenging process (Funkhouser & Mouza, 2013). VanWeelden and Whipple (2007) found in a study of pre-service teachers that field training with experienced practitioners supported the capacity to be able to apply accommodations for disabilities and measure improvements in both students with behaviour and cognitive disabilities as a result of using formative assessment approaches.

Researchers across several studies found pre-service teachers had the capacity to value student centred approaches over more teacher centred approaches (Buldur, 2017; Dickson & Kadbey, 2014; Hudson, 2007; Laronde & MacLeod, 2012), including the potential to override previous exposure to didactic teaching methods that they had been exposed to as students (Dickson & Kadbey, 2014). Pre-service teachers were generally more likely to feel they would like to use more experiential style lessons (Laronde & MacLeod, 2012) and more transformative approaches using technology (Tarling & Ng'ambi, 2016). Buldur (2017) used a four year longitudinal study to show change in pre service teacher approaches that occurred in their last two years of training. None of the studies followed up to analyse the influence of the culture of the school where they begin their career.

5.4. *Weak signal: some assessment approaches seek to create a personalised approach or "Intelligent accountability" (Colbert, Wyatt-Smith, & Klenowski, 2012)*

Alternative assessment practices examined student progress involving using either normative assessment (Frederiksen & White, 2004; Seo, Taherbhai, & McGrane, 2015), or criterion and standards based assessment (Caust, 2010; Colbert et al., 2012; Petscher, Kershaw, Koon, & Foorman, 2014) or learning progressions (Mohan, Chen, & Anderson, 2009) to map the learning trajectories of students, make assessment judgements that can be personalised for students, and to track individual progress whether they are in low, middle or high performance bands.

Assessment systems can support teacher capacity (Charteris, 2016; Frederiksen & White, 2004) and are comparable (Caust, 2010;

Colbert et al., 2012) or superior (Mohan et al., 2009) to standardised testing approaches. Caust (2010) and Colbert et al. (2012) studies conclude that while there may be a role for standardised assessment, standard based approaches offer an "intelligent accountability" (Colbert et al., 2012, p. 399) which is rigorous and sustainable, built on skilling up of teachers in assessment and is more able to "manage sensitively the learning development of children" (Caust, 2010, p. 267).

Frederiksen and White (2004) build on some of their earlier research showing that students in "reflective assessment classes" (p. 81), particularly low band students scored higher than control classes, to investigate the use of an inquiry scorer tool to analyse scientific inquiry. Whilst a limitation was time needed to complete an assessment, the focus on developing a rich and valid assessment process included providing quality information that is built around internal and external validity and was shown to deepen teacher understanding about the concept they were teaching. The statistical analysis of scorer responses by teachers supported the development of a normative scoring model and provided a moderation process for checking the validity of scoring.

Two studies highlighted some potential negative impacts of standards approaches to assessment on learning. This includes risk of "criteria compliance" (Bourke, 2016, p. 98) where the assessment parameters narrow student responses or "single shot teaching" (Koh & Luke, 2009, p. 298), where teachers focussed on presenting shallow tasks with more of a summative focus rather than more extended and complex tasks with a formative purpose.

5.5. Weak signal: technology-focussed approaches

Thirteen studies focused on technology that supported student centred approaches.

Four studies (Barlow, 2012; Bayirtepe & Tuzun, 2007; Gupta & Fisher, 2012; Koul, Fisher, & Shaw, 2011) considered student preference with all studies showing most students preferred technology assisted learning environments, including increased student engagement (Bressler & Bodzin, 2016), even when learning outcomes were not significantly greater (Bayirtepe & Tuzun, 2007). Some studies referred to the novelty technology processes as one possible reason for increased engagement (Clarke, 2009) which raises the question about the impact of approaches if they were common in schools every day.

Three studies (Mainali & Heck, 2017; Naidoo & Naidoo, 2011; Yian Yian Oh et al., 2012) found benefits to teaching and learning outcomes using technology focused approaches, particularly when the use of technology moved from a teaching tool to a student learning tool (Funkhouser & Mouza, 2013; Wang et al., 2014). Yian Yian Oh et al. (2012) investigated the use of applets to help students in a Singaporean classroom visualise concepts in physics across a three week unit of work, compared with a traditional approach, and found that "students generally preferred the use of the animations applets as these were found to help them better understand pressure concepts, made physics lessons more interesting and were not considered a waste of time" (p. 38) and there was also a significant improvement in post-test means for treatment group compared with control group.

Studies investigating the potential benefits of technology examined support for assessment practices that had benefits for student assessment as well as supporting teacher capacity (Adie, 2012). This includes for standardised assessment with the potential of Multidimensional Adaptive Assessment to achieve test efficiency through a reduction of items without a loss of test precision (Frey & Seitz, 2011) although another study (Logan, 2015) found that students performed better in some areas with the difference possibly linked to the cognitive load required by students to process a question or because students with high visuospatial ability may perform better. This raises questions about trying to identify what a test is measuring.

Technology approaches can also support formative assessment though the use of web based formats (Wang, Wang, Wang, & Huang, 2006) which can incorporate additional features to assist student learning achievement compared with a pen and paper format, finding that students reported positive student perceptions (on a 5-point scale) for features including 'query scores' (4.20), 'all pass and then reward' (4.18), 'provide with no answer' (4.07), 'repeat the test' (4.03), 'monitor answering history' (3.96), and 'ask questions' (3.88)".

Technology can also assist building of teacher capacity and validity of results. This includes through online moderation of student work samples across schools to support a standards-based assessment approach (Adie, 2012) and through the design of an assessment tool that is valid and rigorous in its substance to support both assessment for learning processes and can also provide accountability data (Frederiksen & White, 2004). Whilst identifying benefits for assessment and learning, both studies identify issues such as an increase in time required by teachers to engage with these processes.

The research exploring technology-focussed approaches overlaps both the role of the teacher and learner as well as assessment practices.

6. Finding two: weak signals point to areas that are underrepresented despite their importance

6.1. Weak signal: lack of large scale, longitudinal comparative studies

Both of the systematic reviews found a lack of large scale, long term studies that sought to either compare didactic and alternative approaches or formative assessment strategies for teachers across a system. Despite much debate in the literature generally arguing the benefits of student centred versus traditional approaches (Gao, 2014) there was no major study found that sought to extensively address this question through a large scale, longitudinal comparative study. We found no study that sought to evaluate formative assessment approaches across at least one educational system over time.

Whilst ten studies were based on an initiative that was in place for at least one year, most studies involved a short intervention with no further follow up to consider long term benefits or challenges. The systematic review process highlights opportunity for further research to focus on evaluating initiatives over a longer period of time as well as seeking to follow up the impact of initiatives

post intervention.

6.2. Weak signal: student centred or assessment for learning practices might offer some potential benefits for indigenous education

Two studies considered approaches that strengthened the cultural or community context as the basis for learning. This included combining indigenous ways of knowing and traditional school curriculum through student centred inquiry approaches that utilise expertise of indigenous elders and families (Dublin, Sigman, Anderson, Barnhardt, & Topkok, 2014).

Similarly, in a three-year study across ten schools with high and low density populations of American Indian and Alaskan native students, Trumbull et al. (2015) found formative assessment strategies can support indigenous ways of knowing although teacher capacity is generally low in this area.

These studies highlight an underrepresentation within the research about future schooling in terms of consideration of impacts on outcomes for indigenous students. They also highlight an opportunity to reconsider how indigenous ways of knowing can inform the design of schooling to support outcomes for all students.

6.3. Weak signal: the impact of poverty beyond the school

Researchers investigating developing countries and standardised assessment sought to analyse factors both within and beyond the school that have an impact on outcomes such as poverty (Marshall et al., 2012; Osses, 2014; Timmermans, Bosker, Wolf, Doolaard, & Werf, 2014) that are often not reported on when seeking to quantify school effectiveness over other policy levers outside of the educational domain in developed systems. These authors considered school effectiveness within a broader socio-economic context that extends beyond the domain of schooling to consider broader socio-political progress of a community.

The lack of any similar report within developed countries, examining the impact of changing poverty levels on standardised assessment results, highlights a possible gap in the literature that could be further explored.

7. Finding three: weak signals point to emergent hypotheses for further inquiry

7.1. Weak signal: significant positive differences for experimental groups focussed on student centred approaches or formative assessment practices, compared with control groups, were more likely when the intervention lasted for more than four weeks

Overall, 21 studies utilised a student control group as part of their research design in order to compare learning results between experimental and control groups, including eight studies that analysed academic outcomes. Eight studies, utilising general student centred approaches (Ajiboye & Ajitoni, 2008; Huang, Chien, Cheng, & Guo, 2012; Kibirige & Lehong, 2016; Odom, Stoddard, & LaNasa, 2007; Polly, Margerison, & Piel, 2014) or student centred technology approaches (Mainali & Heck, 2017; Naidoo & Naidoo, 2011; Yian Yian Oh et al., 2012) found significant positive difference for the student centred experimental groups.

The results of multiple studies found no significant difference in results for initiatives lasting one to two weeks (Bayirtepe & Tuzun, 2007; Chang & Tsai, 2005; Tieng & Eu, 2014) or less than a four week period (Buabeng-Andoh, 2012; Chen, 2016; Kiliç & Topsakal, 2011; Wu & Huang, 2007). Some studies which found no significant differences in outcomes between teacher-centred and student-centred approaches reported benefits on motivation and attendance (Smit, de Brabander, & Martens, 2014) or other aspects of engagement (Wu & Huang, 2007).

One study by Chang (2002) utilised multimedia approaches over one week and found small to medium benefits in utilising a teacher centred approach compared with a student centred approach. Kalender and Berberoglu (2009) analysed more than 29 000 student Science results from the Turkish Student Assessment Program and found a positive relationship to outcomes using teacher centred approaches and a negative relationship for student centred approaches. Sturm and Bogner (2008) found a significant positive difference from a teacher centred approach during an introductory lesson compared with a student centred approach, however, the retention test six weeks later showed no significant difference between the two approaches.

Two studies (Black & Wiliam, 2004; Blank & Hewson, 2000) utilised a student control group as part of the research design examining assessment practices. Both studies found significant positive differences in academic outcomes as a result of formative assessment approaches. Black and Wiliam (2004) described an 18 months study involving six schools across two school districts in England and found 0.27 median effect size on learning (mean 0.32) compared to the control group. Blank and Hewson (2000) found that learning gains using formative approaches to assessment are more likely to occur in the longer term with permanent change to thinking reflected in post-test data up to six months later.

The emergence of significant positive differences for student centred approaches or formative assessment warrants further inquiry. In particular, it seems important to promote research using comparative data for students with similar demographics, measuring longer-term interventions, and include post-intervention evaluation.

7.2. Weak signal: standardised assessment may better gauge top end performance but can lead to criteria compliance that limits the development of advanced learners

Comparisons between teacher judgement with results from standardised testing found similarities between teacher based judgements and testing approaches (Caust, 2010; Colbert et al., 2012; Gibbons & Chevalier, 2008; Hay, 2008). However, standardised assessment may better gauge top-end performance with high band students more likely to be provided lower scores with teacher

judgement than through testing (Dickinson & Adelson, 2016; Gibbons & Chevalier, 2008) and there are also possible variations from outlier teacher judgements (Caust, 2010; Colbert et al., 2012).

At the same time, standardised assessment may limit achievement in the top bands compared to other approaches (Mohan et al., 2009). A standards-based approach can develop learning progressions that are ambitious and set high expectations for teaching and assessment. As part of a three-year study investigating the design of learning progressions in science, Mohan et al. (2009) extrapolate from their four assessment levels that current teaching approaches and high stakes testing generally focus on achievement up to the first three levels but do not adequately focus learning on level four where learning concentrates on evaluation of information and capacity to impact on the learner's engagement as a citizen to inform future decision making.

A hypothesis for further inquiry relates specifically to the impact of teaching and assessment for top-end academic students. In particular, (a) are top-end students being limited because of a ceiling effect caused by standardised assessment, and (b) why might teachers be less able to identify or describe the performance of top end students. Both questions provide an opportunity to consider how top-end performance might be unleashed.

7.3. Weak signal: no evidence of a reduction in gender gaps through standardised assessment

In relation to gender equity (Matějů & Smith, 2015; McGraw, Lubienski, & Strutchens, 2006; Reilly, Neumann, & Andrews, 2015), McGraw et al. (2006) found that the gender gap is not diminishing in NAEP data over time despite improvements overall. They also found, when surveyed, that girls were less likely to say they are good at or like Maths but will say they understand it. Matějů and Smith (2015) found that boys were more likely to perform better on high stakes tests while Reilly et al. (2015) found small sex based means favour males in science and maths but male results are also more variable than female.

By further exploring the continuation of gender differences, there is an opportunity to consider how any redesign of schooling should take account of the need to respond to this challenge.

7.4. Weak signal: standardised assessment may push for improvement but this could be a proxy measure for school improvement practices

Benefits of standardised assessment include providing a mechanism for state level tracking of improvements (Somers, Zhu, & Wong, 2011), "instructional variables" (Fitchett & Heafner, 2013, p. 306; Lubienski, 2006) or equity gaps over time for specific groups (Anderson & Chang, 2011; Fitchett & Heafner, 2013; Lubienski, 2006) that push or promote improvements. The data can also be used to support mapping of curriculum in a logical and appropriate sequence (Schmidt & Houang, 2012), identify professional learning needs (Polesel, Rice, & Dulfer, 2014) and to compare teacher judgements with test scoring (Heafner & Fitchett, 2015).

Nichols, Glass, and Berliner, (2012) found the impact of "increased testing pressure" (p. 26) had some impact on both reading and mathematics achievement, possibly in different ways but largely because "under pressure, teachers grow more efficient at training students for the test" (p. 27). Dee and Jacob (2011) analysed NAEP data sets between 1990 and 2002 to investigate the impact of the No Child Left Behind (NCLB) policy and found significant improvements in some areas, smaller improvements in other areas, and limited impacts on equity gaps.

There is also some evidence that success in achieving positive changes in data appears to be linked with effective school accountability processes (Dee & Jacob, 2011; Lee, 2006; Simmonds & Webb, 2013) and that improvement from standardised testing in the US was most linked to states that previously did not have robust school accountability systems (Dee & Jacob, 2011; Lee, 2006; Simmonds & Webb, 2013); when successful leaders were able to convert external pressure to internal commitment to improve the quality of approaches within a school (Knapp & Feldman, 2012); or where there was strong resource inputs (Lee & Lee, 2006). The narrow, test focussed approach of standardised testing becomes a proxy measure for school improvement processes which are more likely to consider a wider range of factors to support evaluation and development, be leadership led and resourced.

This weak signal highlights the need to consider the impact of various policy levers, and combinations, for change. Further inquiry should seek to establish causal rather than correlational attributes about how various policy settings. If it emerges that standardised assessment has become a proxy measure for school improvement processes there is potential to promote change in policy settings that would be important and achievable.

8. Finding four: weak signals point to claims that are contested

8.1. Weak signal: some possible advantages of teacher-centred approaches to support lower level students (remedial work) whilst students in the middle and upper bands are more likely to be advantaged from student centred approaches

Some researchers found benefits of teacher centred approaches for struggling students (Maeng & Kim, 2011; Mays, 2012; Stipek, 2004; Wu & Huang, 2007) including from a student perspective (Garrett & Shortall, 2002). This includes improving remedial reading performance of students with low literacy levels (Mays, 2012). Maeng and Kim (2011) documented one teacher moving between more student centred focus for successful learners and more teacher centred for unsuccessful learners with changes in discourse in the classroom occurring in response to teacher interaction with student needs at different points of the lesson. This choice of approach is contested by others (see 8.2) who emphasise links to socio-economic or other equity factors playing a greater part in determining the approach used.

Most studies in Part A of the systematic review focused on exploring benefits and challenges of student centred approaches compared with the traditional teacher centred approaches. Benefits of student centred approaches included increased affective

outcomes (Bayirtepe & Tuzun, 2007; Dublin et al., 2014; Garrett & Shortall, 2002; Naidoo & Naidoo, 2011; Oludipe & Awokoya, 2010; Rass, 2010; Smit et al., 2014) such as a reduction in anxiety about learning (Bayirtepe & Tuzun, 2007; Garrett & Shortall, 2002; Oludipe & Awokoya, 2010); motivation, fun, confidence or self-efficacy as a learner (Dublin et al., 2014; Garrett & Shortall, 2002; Naidoo & Naidoo, 2011; Smit et al., 2014); and strengthened teacher-student relationships (Rass, 2010; Smit et al., 2014).

There was also evidence of positive change in learner practices as a result of being involved with student centred approaches (Martin & Hand, 2009; Seng, 2014) including more ownership of learning and voice in lessons and amongst peers (Martin & Hand, 2009); and students taking a deeper approach to learning when they perceive their teachers adopting more student centred approaches or a surface approach when the approach is more teacher centred (Beausaert, Segers, & Wiltink, 2013).

At the same time, some challenges with student centred approaches pointed to risks around less capacity to provide teacher feedback or correct errors (Ajiboye & Ajitoni, 2008; Naidoo & Naidoo, 2011; Wood, 2017).

Ajiboye and Ajitoni (2008) found most benefits for learners were achieved when a blended approach was used to enhance student centred approaches, including significantly positive results being achieved for low, middle and higher groupings of students. Although teacher activities were the same for the student centred and blended groups, in the latter, students were required to work independently on work that is marked by the teacher prior to group collaboration. This highlights an important role for the teacher to check understanding and correct misconceptions early in the learning process to fully realise benefits available from student centred approaches.

8.2. Weak signal: standardised assessment might be leading to poorer pedagogy for equity groups

There was some evidence that an overemphasis on the results of standardised assessment might be leading to promotion of poorer quality pedagogy for some equity groups.

One study (Watanabe, 2008) compared pedagogical approaches between regular and academically gifted classes taught by the same teachers and found that, whilst classes were taught similar content, instruction differed in approach with the academically gifted classes having less of a focus on explicit test preparation which provided more time:

> to engage in other curricula that teachers deemed important… received more opportunities to practice a wider range of reading and writing skills, engaging in more challenging instruction and assignments and received more written and immediate feedback on essay assignment (Watanabe, 2008, p. 500).

As a result, the quality of instruction offered to regular students was less rich.

Similarly, in a mathematics study, Lubienski (2006) found that black and Hispanic students were more likely to be more regularly tested using multiple choice questioning than white students. These studies highlight some of the tensions identified earlier in this paper by Johnson and Burdett (2010) about the purpose of assessment and how best to reduce equity gaps.

The extent to which teaching approaches might be linked to equity groups rather than ability levels were also examined. Although Stipek (2004) found that the percentage of students below grade level was a greater predictor of teachers using more teacher centred approaches compared with the proportion of students from low income families or ethnic composition of the class she posits that the increased use of teacher centred approach might also be based on a view “that children must master ‘the basics’ before they can benefit from problem-solving, critical thinking, or more inquiry-oriented, constructivist practices” (p. 522), influenced by parental or student preference, level of staff training and leadership, and classroom management approaches seeking to maintain control on more challenging classes.

Gao (2014) analysed TIMSS 2011 data across low, middle and upper achievement groupings across the United States, Chinese-Tapei and Singapore and found that low achievement student groups did not consistently favour particular didactic or inquiry based strategies and that specific inquiry based strategies were not an indicator of success for medium or high achievement groupings across countries, concluding consideration should relate to the cultural needs of students.

8.3. Weak signal: standardised assessment is accepted by parents and students but is more likely to promote shallow, test focused pedagogy

Whilst teachers may not favour testing, parents and students seem more comfortable with the idea (Polesel et al., 2014; Riek, 2014) and even spending most of the school day focused on literacy and numeracy (Riek, 2014).

However, multiple studies identified challenges about the limited and negative impacts of standardised assessments, particularly within a high stakes testing environment. This includes narrowing what is taught based on what is tested (Anderson, 2012; Dee & Jacob, 2011; Polesel et al., 2014) and being used as the main data to make judgements about a school’s effectiveness (Kruger, Mihye, & Treagust, 2013; Polesel et al., 2014). Some studies report negative impacts on pedagogy (Anderson, 2012; Polesel et al., 2014) such as preferencing testing outcomes over opportunities to develop practices that would support effective learning (Auerbach & Collier, 2012); being in opposition to other aims such as middle schooling principles (Pendergast & Swain, 2013); deskilling of teachers (Burke & Wang, 2010; Caust, 2010; Colbert et al., 2012; Pendergast & Swain, 2013; Polesel et al., 2014); and that it can lead to poorer outcomes for equity groups (DeCuir, 2014; Watanabe, 2008).

Researchers also challenged the quality of policy positions based on standardised assessment in a high stakes context. This includes achievement targets being unrealistic (Dee & Jacob, 2011; Hambleton, Sireci, & Smith, 2009; Konstantopoulos & Hedges, 2008; Lee, 2010). Hambleton et al. (2009) found the NCLB targets of 100% proficiency set in the USA by 2014 were not even achieved by Singapore, a high performing country which came closest. Dee and Jacob (2011) found that, although there had been significant improvement in Year 4 Mathematics results over time, the target had still not been achieved by more than 60%.

These findings are particularly important given that one of the reasons often cited for standardised tests is a lack of confidence in schooling systems. It is also important because of some policy responses to achieving or not achieving targets such as performance pay (Thompson & Price, 2012), closing schools or moving executive teams even when improvement is found within the school but results are short of immediate targets (MacMahon, 2011). Konstantopoulos and Hedges (2008) analyses NAEP data in reading, mathematics and science between 1971 and 1996 to examine the question that titles the study, “How Large an Effect Can We Expect from School Reforms?” (p. 1611) and concludes that “the danger is that real reform that improves the quality of education must not be judged by standards that preordain its evaluation as a failure" (p. 1636). These policy levers can be deleterious to actual practice, improvement processes and cultures within some schools.

8.4. Weak signal: monitoring of system level improvements through standardised assessment can lead to distortion of results through the use of exemptions for equity groups

Exemption rates used in standardised testing and alternative assessments in disabilities (Braun, Jinming, & Vezzu, 2010; Saven, Anderson, Nese, Farley, & Tindal, 2016) as well as English as a Second Language (ESL) (Torres-Guzm´n, Abbate, Brisk, & Minaya-Rowe, 2002) were examined to consider the impact of trying to accurately understand overall state progress in NAEP over time versus increasing the number of exemptions to suggest improvement. Some states in the USA have supposedly recorded significant improvement in their own state-wide testing but this is not reflected in NAEP data (Lee, 2010). DeCuir (2014) investigated historical documents in Louisiana, a very strong high stakes environment, and found results had not improve significantly but dropout rates increased for black students and social promotion was not reduced. Torres-Guzm´n et al. (2002) found limited teacher and leader capacity to unpack ESL (assessment) data as a result of overemphasis on standardised testing- improved skills would have provided them with a more positive indication of student achievement.

9. Conclusion

This systematic review of the literature provides some insights about the role of teacher and learner, as well as assessment practices, to improve classroom practice that can support greater alignment between the work of practitioners and policy makers. We point to some weak signals about the confidence many scholars and practitioners have in the current dominant system.

Examining what schools might look like if they were invented today can help us model possible futures, including consideration of their impact across equity groups. There are multiple alternatives emerging within schools and in new models of schooling, but this still leaves the core “old school” structures intact that most students attend—transmission teaching with teacher-centred constructs for assessment. Having educators help drive the conversation from a learner and learning focused perspective could be beneficial to overcoming the tendency to continue to reproduce what has been shown above to be mostly obsolete.

Declarations of interest

Author 1: Jason is a doctoral student at The University of Newcastle, supported through an Australian Government Research Training Program Scholarship.

Author 2: None.

Appendix 1: Supplementary materials

A spreadsheet is available that provides the data for each study included in the review of literature https://tinyurl.com/whatifschools.

Acknowledgements

We thank the anonymous reviewers for their feedback, comments and suggestions.

References

Abbott, J. (2010). *Overschooled but undereducated: How the crisis in education is jeopardizing our adolescents.* A&C Black.

Adie, L. (2012). Learning as identity and practice through involvement in online moderation. *Educational Assessment, Evaluation and Accountability, 24*(1), 43–56.

Ahn, S., Ames, A. J., & Myers, N. D. (2012). A review of meta-analyses in education: Methodological strengths and weaknesses. *Review of Educational Research, 82*(4), 436–476. https://doi.org/10.3102/0034654312458162.

Ajiboye, J. O., & Ajitoni, S. (2008). Effects of full and quasi-participatory learning strategies on Nigerian senior secondary students’ environmental knowledge: Implications for classroom practice. *International Journal of Environmental and Science Education, 3*(2), 58–66.

Al-Amoush, S., Usak, M., Erdogan, M., Markic, S., & Eilks, I. (2013). Pre-service and in-service teachers’ beliefs about teaching and learning chemistry in Turkey. *European Journal of Teacher Education, 36*(4), 464–479.

Ampadu, E. (2012). *Investigation into the teaching and learning of mathematics in junior secondary schools: The case of Ghana.* Anglia Ruskin University.

Anderson, K. J. (2012). Science education and test-based accountability: Reviewing their relationship and exploring implications for future policy. *Science Education, 96*(1), 104–129.

Anderson, R., & Chang, B. (2011). Mathematics course-taking in rural high schools. *Journal of Research in Rural Education, 26*, 1–10.

Angrist, J., & Pischke, J. (2009). *Mostly harmless econometrics: An empiricists’ companion*.

Arthur, J., Waring, M., Coe, R., & Hedges, L. V. (2012). *Research methods and methodologies in education.* SAGE.

Auerbach, S., & Collier, S. (2012). Bringing high stakes from the classroom to the parent center: Lessons from an intervention program for immigrant families. *Teachers College Record, 114*(3), n3.
Aydogdu, B., & Selanik-Ay, T. (2016). Determination of teacher characteristics that support constructivist learning environments. *Eurasian Journal of Educational Research, 63*, 293–310.
Barlow, T. (2012). The end of' chalk and talk'. *Teaching Science, 58*(1), 56–59.
Bayirtepe, E., & Tuzun, H. (2007). The effects of game-based learning environments on students' achievement and self-efficacy in a computer course. *Hacettepe Universitesi Egitim Fakultesi Dergisi- Hacettepe University Journal of Education, 33*, 41–54.
Beausaert, S. A., Segers, M., & Wiltink, D. P. (2013). The influence of teachers' teaching approaches on students' learning approaches: The student perspective. *Educational Research, 55*(1), 1–15.
Black, P., & Wiliam, D. (2004). The formative purpose: Assessment must first promote learning. *Yearbook of the National Society for the Study of Education, 103*(2), 20–50.
Blank, L. M., & Hewson, P. W. (2000). A metacognitive learning cycle: A better warranty for student understanding? *Science Education, 84*(4), 486.
Bourke, R. (2016). Liberating the learner through self-assessment. *Cambridge Journal of Education, 46*(1), 97–111. https://doi.org/10.1080/0305764X.2015.1015963.
Braun, H., Jinming, Z., & Vezzu, S. (2010). An investigation of Bias in reports of the national assessment of educational progress. *Educational Evaluation & Policy Analysis, 32*(1), 24–43. https://doi.org/10.3102/0162373709351137.
Bressler, D. M., & Bodzin, A. M. (2016). Investigating flow experience and scientific practices during a mobile serious educational game. *Journal of Science Education and Technology, 25*(5), 795–805.
Brown, L. (2010). Nurturing relationships within a space created by "Indigenous ways of knowing": A case study. *The Australian Journal of Indigenous Education, 39*(S1), 15–22.
Brown, G. T. (2011). Teachers' conceptions of assessment: Comparing primary and secondary teachers in New Zealand. *Assessment Matters, 3*, 45–70.
Buabeng-Andoh, C. (2012). An exploration of teachers' skills, perceptions and practices of ICT in teaching and learning in the Ghanaian second-cycle schools. *Contemporary Educational Technology, 3*(1), 36–49.
Buldur, S. (2017). A longitudinal investigation of the preservice science teachers' beliefs about science teaching during a science teacher training programme. *International Journal of Science Education, 39*(1), 1–19.
Burke, G., Jr., & Wang, Y. (2010). Methods and uses of classroom assessments employed in teaching grades three through five in five school districts in the Mississippi Delta. *Education, 130*(4), 657–665.
Caust, M. K. (2010). *Measuring student progress in school: A role for teacher judgement.* Townsville Qld: James Cook University.
Chang, C.-Y. (2002). The impact of different forms of multimedia CAI on students' science achievement. *Innovations in Education and Teaching International, 39*(4), 280–288.
Chang, C. Y., & Tsai, C. C. (2005). The interplay between different forms of CAI and students' preferences of learning environment in the secondary science class. *Science Education, 89*(5), 707–724.
Charteris, J. (2016). Dialogic feedback as divergent assessment for learning: An ecological approach to teacher professional development. *Critical Studies in Education, 57*(3), 277–295.
Chen, L.-L. (2016). Impacts of flipped classroom in high school health education. *Journal of Educational Technology Systems, 44*(4), 411–420.
Clarke, P. J. (2009). A caribbean pre-service mathematics teacher's impetus to integrate computer technology in his practice. *International Journal for Technology in Mathematics Education, 16*(4).
Colbert, P., Wyatt-Smith, C. M., & Klenowski, V. (2012). A systems-level approach to building sustainable assessment cultures: Moderation, quality task design and dependability of judgement. *Policy Futures in Education, 10*(4), 386–401.
Cribiore, R. (2005). *Gymnastics of the mind: Greek education in hellenistic and roman.* Egypt: Princeton University Press.
DeCuir, E. (2014). The Louisiana Educational Assessment Program: Implications for students and schools. *Curriculum & Teaching Dialogue, 16*(1/2), 31–46.
Dee, T. S., & Jacob, B. (2011). The impact of no child left behind on student achievement. *Journal of Policy Analysis & Management, 30*(3), 418–446. https://doi.org/10.1002/pam.20586.
Dickinson, E. R., & Adelson, J. L. (2016). Choosing among multiple achievement measures. *Journal of Advanced Academics, 27*(1), 4–22. https://doi.org/10.1177/1932202X15621905.
Dickson, M., & Kadbey, H. (2014). "That's not the way i was taught science at school!" How preservice primary teachers in Abu Dhabi, United Arab Emirates are affected by their own schooling experiences. *Science Education International, 25*(3), 332–350.
Dreyer, I., & Stang, G. (2013). Foresight in governments–practices and trends around the world. *Yearbook of European Security*, 7–32.
Dublin, R., Sigman, M., Anderson, A., Barnhardt, R., & Topkok, S. A. (2014). COSEE-AK ocean science fairs: A science fair model that grounds student projects in both western science and traditional native knowledge. *Journal of Geoscience Education, 62*(2), 166–176.
Faria, A.-M., Greenberg, A., Meakin, J., Bichay, K., & Heppen, J. (2014). Replicating the relationship between teachers' data use and student achievement: The urban data study and the data dashboard usage study. *Society for Research on Educational Effectiveness*, 1–7.
Fitchett, P. G., & Heafner, T. L. (2013). Making critical connections between social studies teaching and student achievement using NAEP data explorer. *Teacher Educator, 48*(4), 296–310. https://doi.org/10.1080/08878730.2013.827921.
Frederiksen, J. R., & White, B. Y. (2004). Designing assessments for instruction and accountability: An application of validity theory to assessing scientific inquiry. *Yearbook of the National Society for the Study of Education, 103*(2), 74–104.
Frey, A., & Seitz, N.-N. (2011). Hypothetical use of multidimensional adaptive testing for the assessment of student achievement in the programme for international student assessment. *Educational & Psychological Measurement, 71*(3), 503–522. https://doi.org/10.1177/0013164410381521.
Funkhouser, B. J., & Mouza, C. (2013). Drawing on technology: An investigation of preservice teacher beliefs in the context of an introductory educational technology course. *Computers & Education, 62*, 271–285.
Gao, S. (2014). Relationship between science teaching practices and students' achievement in Singapore, Chinese Taipei, and the US: An analysis using TIMSS 2011 data. *Frontiers of Education in China, 9*(4), 519–551.
Garrett, P., & Shortall, T. (2002). Learners' evaluations of teacher-fronted and student-centred classroom activities. *Language Teaching Research, 6*(1), 25–57.
Garrouste, C. (2011). *Towards a benchmark on the contribution of education and training to employability: Methodological note.* European Union Centre in Taiwan.
Gibbons, S., & Chevalier, A. (2008). Assessment and age 16+ education participation. *Research Papers in Education, 23*(2), 113–123. https://doi.org/10.1080/02671520802048638.
Giddens, A. (1991). *Modernity and self-identity: Self and society in the late modern age.* Stanford University Press.
Gough, D. (2007). Weight of evidence: A framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education, 22*(2), 213–228.
Griffin, P. E., Care, E., & McGaw, B. (2011). *Assessment and teaching of 21st century skills.* Dordrecht; London: Springer.
Gupta, A., & Fisher, D. (2012). Technology-supported learning environments in science classrooms in India. *Learning Environments Research, 15*(2), 195–216.
Guskey, T. R. (2003). How classroom assessments improve learning. *Educational Leadership, 60*(5), 6.
Hallinger, P. (2014). Reviewing reviews of research in educational leadership: An empirical assessment. *Educational Administration Quarterly, 50*(4), 539–576.
Hambleton, R. K., Sireci, S. G., & Smith, Z. R. (2009). How do other countries measure up to the mathematics achievement levels on the national assessment of educational progress? *Applied Measurement in Education, 22*(4), 376–393. https://doi.org/10.1080/08957340903221675.
Hargreaves, A. (1999). Schooling in the new millennium: Educational research for the postmodern age. *Discourse: Studies in the Cultural Politics of Education, 20*(3), 333.
Hartmann, D., & Stillings, C. (2015). Using scenarios in multinational companies across geographic distances—A case from the chemical industry. *Foresight, 17*(5), 475–488. https://doi.org/10.1108/FS-11-2014-0076.
Hay, P. J. (2008). *Social construction of abilities and conduct of assessment in Senior PE.* St Lucia Qld: University of Queensland.
Heafner, T. L., & Fitchett, P. G. (2015). An opportunity to learn US history: What NAEP data suggest regarding the opportunity gap. *High School Journal, 98*(3), 226–249. https://doi.org/10.1353/hsj.2015.0006.

Holm, A., Jæger, M. M., Karlson, K. B., & Reimer, D. (2013). Incomplete equalization: The effect of tracking in secondary education on educational inequality. *Social Science Research, 42*(6), 1431–1442.

Horton, A. (1999). A simple guide to successful foresight. *Foresight, 1*(1), 5–9. https://doi.org/10.1108/14636689910802052.

Huang, C. M., Chien, L. Y., Cheng, C. F., & Guo, J. L. (2012). Integrating life skills into a theory-based drug-use prevention program: Effectiveness among junior high students in Taiwan. *The Journal of School Health, 82*(7), 328–335.

Hudson, P. (2007). High-impact teaching for science. *Teaching Science: The Journal of the Australian Science Teachers Association, 53*(4).

Jenssen, S. (2010). Municipal visions: Reflexive futures between paradigm and practice. *Futures, 42*(4), 345–354.

Johnson, M., & Burdett, N. (2010). Intention, interpretation and implementation: Some paradoxes of assessment for learning across educational contexts. *Research in Comparative and International Education, 5*(2), 122–130.

Kalender, I., & Berberoglu, G. (2009). An assessment of factors related to science achievement of Turkish students. *International Journal of Science Education, 31*(10), 1379–1394.

Kaya, S., Kablan, Z., Akaydin, B., & Demir, D. (2015). Teachers' awareness and perceived effectiveness of instructional activities in relation to the allocation of time in the classroom. *Science Education International, 26*(3), 344–357.

Kibirige, I., & Lehong, M. J. (2016). The effect of cooperative learning on grade 12 learners' performance in projectile motions, South Africa. *Eurasia Journal of Mathematics, Science & Technology Education, 12*(9), 2279–2292.

Kiliç, O., & Topsakal, U. U. (2011). *The effectiveness of using student and teacher centered analogies on the development of the students' cognitive and affective skills. Paper presented at the Asia-Pacific forum on science learning and teaching*.

Kloser, M. (2014). Identifying a core set of science teaching practices: A delphi expert panel approach. *Journal of Research in Science Teaching, 51*(9), 1185–1217.

Knapp, M. S., & Feldman, S. B. (2012). Managing the intersection of internal and external accountability: Challenge for urban school leadership in the United States. *Journal of Educational Administration, 50*(5), 666–694. https://doi.org/10.1108/09578231211249862.

Koh, K., & Luke, A. (2009). Authentic and conventional assessment in Singapore schools: An empirical study of teacher assignments and student work. *Assessment in Education: Principles, Policy & Practice, 16*(3), 291–318. https://doi.org/10.1080/09695940903319703.

Konstantopoulos, S., & Hedges, L. V. (2008). How large an effect can we expect from school reforms? *Teachers College Record, 110*(8), 1611–1638.

Kontkanen, S., Dillon, P., Valtonen, T., Eronen, L., Koskela, H., & Väisänen, P. (2017). Students' experiences of learning with iPads in upper secondary school–a base for proto-TPACK. *Education and Information Technologies, 22*(4), 1299–1326.

Koul, R. B., Fisher, D., & Shaw, T. (2011). An application of the TROFLEI in secondary-school science classes in New Zealand. *Research in Science & Technological Education, 29*(2), 147–167.

Kruger, M., Mihye, W., & Treagust, D. F. (2013). Teachers' perceptions on the changes in the curriculum and exit examinations for biology and human biology. *Australian Journal of Teacher Education, 38*(3), 41–58. https://doi.org/10.14221/ajte.2013v38n3.5.

Laronde, G., & MacLeod, K. (2012). Modeling various teaching methods in a faculty of education in science education: Chalk and talk, virtual labs or hovercrafts. *Journal of College Teaching & Learning (Online), 9*(2), 107.

Lebak, K., & Tinsley, R. (2010). Can inquiry and reflection be contagious? Science teachers, students, and action research. *Journal of Science Teacher Education, 21*(8), 953–970.

Lebow, R. N. (2009). Counterfactuals, history and fiction. *Historical Social Research/Historische Sozialforschung,* 57–73.

Lee, J. (2006). Input-guarantee versus performance- guarantee approaches to school accountability: Cross-state comparisons of policies, resources, and outcomes. *Peabody Journal of Education (0161956X), 81*(4), 43–64. https://doi.org/10.1207/s15327930pje8104_3.

Lee, J. (2010). Trick or treat: New ecology of education accountability system in the USA. *Journal of Education Policy, 25*(1), 73–93. https://doi.org/10.1080/02680930903377423.

Lee, C., & Lee, G. G. (2006). Information gain and divergence-based feature selection for machine learning-based text categorization. *Information Processing & Management, 42*(1), 155–165. https://doi.org/10.1016/j.ipm.2004.08.006.

Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis.* Oxford University Press.

Logan, T. (2015). The influence of test mode and visuospatial ability on mathematics assessment performance. *Mathematics Education Research Journal, 27*(4), 423–441. https://doi.org/10.1007/s13394-015-0143-1.

Lubienski, S. T. (2006). Examining instruction, achievement, and equity with NAEP mathematics data. *Education Policy Analysis Archives, 14*(14), 1–30.

MacMahon, B. (2011). The perpetuation of risk: Organizational and institutional policies and practices in a title 1 school. *Journal for Critical Education Policy Studies (JCEPS), 9*(2), 199–215.

Maeng, S., & Kim, C. J. (2011). Variations in science teaching modalities and students' pedagogic subject positioning through the discourse register and language code. *Science Education, 95*(3), 431–457.

Mainali, B. R., & Heck, A. (2017). Comparison of traditional instruction on reflection and rotation in a Nepalese high school with an ICT-rich, student-centered, investigative approach. *International Journal of Science and Mathematics Education, 15*(3), 487–507.

Mameli, C., & Molinari, L. (2017). Teaching interactive practices and burnout: A study on Italian teachers. *European Journal of Psychology of Education, 32*(2), 219–234.

Marshall, J., Chinna, U., Hok, U., Tinon, S., Veasna, M., & Nissay, P. (2012). Student achievement and education system performance in a developing country. *Educational Assessment, Evaluation & Accountability, 24*(2), 113–134. https://doi.org/10.1007/s11092-012-9142-x.

Martin, A. M., & Hand, B. (2009). Factors affecting the implementation of argument in the elementary science classroom. A longitudinal case study. *Research in Science Education, 39*(1), 17–38.

Matějů, P., & Smith, M. L. (2015). Are boys that bad? Gender gaps in measured skills, grades and aspirations in Czech elementary schools. *British Journal of Sociology of Education, 36*(6), 871–895.

Mays, D. (2012). Tackling illiteracy in year 7 of the comprehensive school. *Support for Learning, 27*(3), 123–128.

McGraw, R., Lubienski, S. T., & Strutchens, M. E. (2006). A closer look at gender in NAEP mathematics achievement and affect data: Intersections with achievement, race/ethnicity, and socioeconomic status. *Journal for Research in Mathematics Education, 37*(2), 129–150.

Miles, I. (2012). Dynamic foresight evaluation. *Foresight, 14*(1), 69–81. https://doi.org/10.1108/14636681211210378.

Miranda, R. J., & Damico, J. B. (2015). Changes in teachers' beliefs and classroom practices concerning inquiry-based instruction following a year-long RET-PLC program. *Science Educator, 24*(1), 23.

Mohan, L., Chen, J., & Anderson, C. W. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *Journal of Research in Science Teaching, 46*(6), 675–698.

Müllenbeck, M. (2017). *What if we paid doctors to keep people healthy? [Video file].* Retrieved fromhttps://www.ted.com/talks/matthias_mullenbeck_what_if_we_paid_doctors_to_keep_people_healthy.

Naidoo, N., & Naidoo, R. (2011). Using blended learning to facilitate the mathematical thought processes of primary school learners in a computer laboratory: A case study in calculating simple areas. *Journal of College Teaching & Learning (TLC), 4*(7).

Nichols, S. L., Glass, G. V., & Berliner, D. C. (2012). High-stakes testing and student achievement: Updated analyses with NAEP data. *Provas de consequências severas e desempenho do aluno: análise atualizada de dados NAEP, 20*(20), 1–30.

Odom, A. L., Stoddard, E. R., & LaNasa, S. M. (2007). Teacher practices and middle-school science achievements. *International Journal of Science Education, 29*(11), 1329–1346.

Oludipe, D., & Awokoya, J. O. (2010). Effect of cooperative learning teaching strategy on the reduction of students' anxiety for learning chemistry. *Journal of Turkish Science Education, 7*(1), 30.

Onurkan Aliusta, G., & Özer, B. (2017). Student-centred learning (SCL): Roles changed? *Teachers and Teaching, 23*(4), 422–435.

Osses, A. (2014). *Measuring school performance in high-stakes accountability systems: Evidence from and proposals for the Chilean accountability policy.* Parkville Vic: University of Melbourne.

Palmieri, P. (2005). 'Spuntar lo scoglio più duro': Did Galileo ever think the most beautiful thought experiment in the history of science? *Studies in History and*

Philosophy of Science, 36, 223–240. https://doi.org/10.1016/j.shpsa.2005.03.001.
Pendergast, D., & Swain, K. (2013). Competing interests? NAPLAN and middle schooling assessment practices. *Middle Years of Schooling Association, 13*(1), 4–17.
Petscher, Y., Kershaw, S., Koon, S., & Foorman, B. R. (2014). *Testing the importance of individual growth curves in predicting performance on a high-stakes reading comprehension test in Florida.* REL 2014-006. Regional Educational Laboratory Southeast.
Polesel, J., Rice, S., & Dulfer, N. (2014). The impact of high-stakes testing on curriculum and pedagogy: A teacher perspective from Australia. *Journal of Education Policy, 29*(5), 640–657. https://doi.org/10.1080/02680939.2013.865082.
Polly, D., Margerison, A., & Piel, J. A. (2014). Kindergarten teachers' orientations to teacher-centered and student-centered pedagogies and their influence on their students' understanding of addition. *Journal of Research in Childhood Education, 28*(1), 1–17.
Ramnarain, U., & Hobden, P. (2015). Shifting South African learners towards greater autonomy in scientific investigations. *Journal of Curriculum Studies, 47*(1), 94–121.
Rass, R. A. (2010). Drama in chalk and talk classrooms. *Journal of Language Teaching and Research, 1*(4), 378–381.
Reilly, D., Neumann, D. L., & Andrews, G. (2015). Sex differences in mathematics and science achievement: A meta-analysis of national assessment of educational progress assessments. *Journal of Educational Psychology, 107*(3), 645–662. https://doi.org/10.1037/edu0000012.
Riek, R. M. (2014). *Arts education in a culture of performativity: A case study of what is valued in one Queensland school community.* Nathan Qld: Griffith University.
Robinson, K., & Aronica, L. (2015). *Creative schools: Revolutionizing education from the ground up.* Australia: Penguin.
Saritas, O., & Smith, J. E. (2011). The big picture–trends, drivers, wild cards, discontinuities and weak signals. *Futures, 43*(3), 292–312.
Saven, J. L., Anderson, D., Nese, J. F. T., Farley, D., & Tindal, G. (2016). Patterns of statewide test participation for students with significant cognitive disabilities. *Journal of Special Education, 49*(4), 209–220. https://doi.org/10.1177/0022466915582213.
Schmidt, J. M. (2015). Policy, planning, intelligence and foresight in government organizations. *Foresight, 17*(5), 489–511. https://doi.org/10.1108/FS-12-2014-0081.
Schmidt, W. H., & Houang, R. T. (2012). Curricular coherence and the common core state standards for mathematics. *Educational Researcher, 41*(8), 294–308. https://doi.org/10.3102/0013189X12464517.
Selwyn, N., & Facer, K. (2013). *Introduction: The need for a politics of education and technology. The politics of education and technology*. Springer1–17.
Seng, E. L. K. (2014). Investigating teachers' views of student-centred learning approach. *International Education Studies, 7*(7), 143.
Seo, D., Taherbhai, H., & McGrane, J. (2015). The role of student growth percentiles in monitoring learning and predicting learning outcomes. *Educational Assessment, 20*(2), 151–163. https://doi.org/10.1080/10627197.2015.1028621.
Simmonds, M., & Webb, P. T. (2013). Accountability synopticism: How a think tank and the media developed a quasimarket for school choice in British Columbia. *International Education Journal: Comparative Perspectives, 12*(2), 21–41.
Slaughter, R., & Riedy, C. (2009). The state of play in the futures field: A metascanning overview. *Foresight, 11*(5), 6–20.
Smit, K., de Brabander, C. J., & Martens, R. L. (2014). Student-centred and teacher-centred learning environment in pre-vocational secondary education: Psychological needs, and motivation. *Scandinavian Journal of Educational Research, 58*(6), 695–712.
Somers, M.-A., Zhu, P., & Wong, E. (2011). *Whether and how to use state tests to measure student achievement in a multi-state randomized experiment: An empirical assessment based on four recent evaluations.* NCEE 2012–4015: US Department of Education.
Stipek, D. (2004). Teaching practices in kindergarten and first grade: Different strokes for different folks. *Early Childhood Research Quarterly, 19*(4), 548–568.
Sturm, H., & Bogner, F. X. (2008). Student-oriented versus teacher-centred: The effect of learning at workstations about birds and bird flight on cognitive achievement and motivation. *International Journal of Science Education, 30*(7), 941–959.
Tarling, I., & Ng'ambi, D. (2016). Teachers pedagogical change framework: A diagnostic tool for changing teachers' uses of emerging technologies. *British Journal of Educational Technology, 47*(3), 554–572.
Thompson, G., & Price, A. (2012). Performance pay for Australian teachers: A critical policy historiography. *The Social Educator, 30*(2), 3–12.
Tieng, P. G., & Eu, L. K. (2014). Improving students' Van Hiele level of geometric thinking using geometer's sketchpad. *Malaysian Online Journal of Educational Technology, 2*(3), 20–31.
Timmermans, A. C., Bosker, R. J., Wolf, I. F., Doolaard, S., & Werf, M. P. (2014). Value added based on educational positions in Dutch secondary education. *British Educational Research Journal, 40*(6), 1057–1082.
Torres-Guzm´n, M. E., Abbate, J., Brisk, M. E., & Minaya-Rowe, L. (2002). Defining and documenting success for bilingual learners: A collective case study. *Bilingual Research Journal, 26*(1), 23.
Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times.* San Francisco, California: John Wiley & Sons.
Trumbull, E., Sexton, U., Nelson-Barber, S., & Johnson, Z. (2015). Assessment practices in schools serving American Indian and Alaska Native Students. *Journal of American Indian Education, 54*(3), 5–30.
Ucar, S. (2012). How do pre-service science teachers' views on science, scientists, and science teaching change over time in a science teacher training program? *Journal of Science Education and Technology, 21*(2), 255–266.
VanDerHeyden, A. M., Witt, J. C., & Barnett, D. W. (2005). The emergence and possible futures of response to intervention. *Journal of Psychoeducational Assessment, 23*(4), 339–361.
VanWeelden, K., & Whipple, J. (2007). Preservice music teachers' predictions, perceptions, and assessment of students with special needs: The need for training in student assessment. *Journal of Music Therapy, 44*(1), 74–84.
Wang, K. H., Wang, T. H., Wang, W. L., & Huang, S. C. (2006). Learning styles and formative assessment strategy: Enhancing student achievement in web-based learning. *Journal of Computer Assisted Learning, 22*(3), 207–217.
Wang, S.-K., Hsu, H.-Y., Reeves, T. C., & Coster, D. C. (2014). Professional development to enhance teachers' practices in using information and communication technologies (ICTs) as cognitive tools: Lessons learned from a design-based research study. *Computers & Education, 79*, 101–115.
Watanabe, M. (2008). Tracking in the era of high stakes state accountability reform: Case studies of classroom instruction in North Carolina. *Teachers College Record, 110*(3), 489–534.
Welch, V., Petticrew, M., Petkovic, J., Moher, D., Waters, E., White, H., ... Wells, G. (2016). Extending the PRISMA statement to equity-focused systematic reviews (PRISMA-E 2012): Explanation and elaboration. *Journal of Clinical Epidemiology, 70*, 68–89.
Wells, K., & Littell, J. H. (2009). Study quality assessment in systematic reviews of research on intervention effects. *Research on Social Work Practice, 19*(1), 52–62.
Williamson, B. (2013). Networked cosmopolitanism? Shaping learners by remaking the curriculum of the future. In N. Selwyn, & K. Facer (Eds.). *The politics of education and technology* (pp. 39–59). Springer.
Wood, A. B. (2017). Pre-twentieth century literature in the year 9 classroom: Student responses to different teaching approaches. *English in Education, 51*(3), 308–326.
Wu, H. K., & Huang, Y. L. (2007). Ninth-grade student engagement in teacher-centered and student-centered technology-enhanced learning environments. *Science Education, 91*(5), 727–749.
Yeung, A. S., Craven, R. G., & Kaur, G. (2014). Teachers' self-concept and valuing of learning: Relations with teaching approaches and beliefs about students. *Asia-Pacific Journal of Teacher Education, 42*(3), 305–320. https://doi.org/10.1080/1359866X.2014.905670.
Yian Yian Oh, E., Treagust, D. F., Koh, T. S., Phang, W. L., Ng, S. L., Sim, G., & Chandrasegran, A. (2012). Using visualizations in secondary school physics teaching and learning: Evaluating the efficacy of an instructional program to facilitate understanding of gas and liquid pressure concepts. *Teaching Science: The Journal of the Australian Science Teachers Association, 58*(4).