Effect of an inclusive physical education (IPE) training workshop on trainee teachers' self-efficacy

Neville, Ross; Makopoulou, Kyriaki; Hopkins, Will

DOI:
10.1080/02701367.2019.1650877

Effect of an Inclusive Physical Education (IPE) Training Workshop on Trainee Teachers' Self-Efficacy

Ross D. Neville,[1] Kyriaki Makopoulou,[2] & Will G. Hopkins[3]

1. School of Public Health, Physiotherapy and Sports Science, University College Dublin, Ireland
2. School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham Edgbaston, Birmingham, United Kingdom.
3. Institute for Sport and Health, Victoria University, Melbourne, Australia

Corresponding Author: Ross D. Neville, School of Public Health, Physiotherapy and Sports Science, University College Dublin, Ireland, [ross.neville@ucd.ie] +353 (0) 1 716 3419.

Abstract

Purpose: Children with special educational needs and disabilities (SEND) are often excluded from physical education (PE). Opportunities exist to enhance the competence of trainee teachers, thus preparing them for the demands of inclusive PE (IPE). In this study, we examined the effect of IPE training workshops on trainee teachers' self-efficacy for including SEND children within PE. **Method:** Data was collection from 366 trainee teachers across 44 workshops (63% females; 57% primary-school trainees; age, 23 ± 5 y, mean ± standard deviation; age range 18 – 53 y). A pre-post design was used to assess the effect of the workshop and possible moderating factors. Data were analyzed using mixed modelling. **Results:** IPE training had a large beneficial effect on trainees' self-efficacy (1.61, ±0.21; standardized effect size, 99% confidence limits). Mixed modelling revealed substantial differences in the mean outcome between workshops. The standard deviation (SD) representing this variation between workshops was moderate-sized (SD=0.30, ±0.30). We also found that trainees with an above average sense of self-efficacy at baseline (mean +1SD) were likely to benefit more. Workshop size (0.35, ±0.42) and trainees' perception of workshop quality (0.39, ±0.19) were substantial moderators of changes in self-efficacy. Other moderators with possible substantial effects were trainee specialization and completion of a school-based placement. **Conclusion:** This study supports the viability of embedding IPE within teacher training. Attention should be paid to trainees' starting points prior to scheduling training. The size of workshops and quality of delivery should also be considered when developing IPE competencies within trainees.

Keywords: professional learning; continuing professional development; quality of professional learning; special educational needs and disabilities

International research shows that children with special educational needs and disabilities (SEND) are still not accessing or being fully included into educational programs within mainstream schools (Rieser, 2013). This problem of inclusion is particularly acute when it comes to SEND children accessing high-quality physical education (PE), and the research evidence shows the signs of a double-bind here: SEND children are disproportionately affected by social-emotional, behavioral, health conditions related to physical inactivity, and exclusion from PE (Bloemen et al., 2015); PE teachers, however, often lack the knowledge, confidence, or pedagogical training to adapt their lessons in ways that would suitably include them (Rekaa, Hanisch, & Ytterhus, 2018).

There is a longstanding recognition within PE of the challenges associated with including children who vary in physical ability and learning preference (Makopoulou & Thomas, 2016). Policy statements and empirical studies on the attitudes of teachers and on the suitability of teacher education for inclusion have really only emerged over the last ten to fifteen years, however (Rekaa et al., 2018). Among the more pressing challenges – in light of international consensus on the need for curricular reform to promote inclusion (Rieser, 2013) – is the design of interventions which can expand the scope and quality of initial teacher training (ITT). The basis for this commitment, to inclusive pedagogy as a foundation for effective teaching and learning, is found within the Salamanca Statement (United Nations Educational, Scientific, and Cultural Organization [UNESCO], 1994), which states "pre-service training programs should provide to all student teachers, primary and secondary alike, positive orientation toward disability, thereby developing an understanding of what can be achieved in schools with locally available support services" (p. 27).

In spite of this edict, in 2009, the Organization of Economic Cooperation and Development (OECD, 2009) reported that ITT was still not having enough of an impact on teacher preparedness for inclusive education. Similar findings have also been reported at length in the

inclusive PE (IPE) literature since then. Research shows that, while children with SEND enjoy being physically active (Coates & Vickerman, 2010; Rekaa et al., 2018), their progression in PE is still hampered by a broad range of factors, including: school culture (Tripp, Rizzo, & Webbert, 2007); class sizes (Ammah & Hodge, 2005; Hodge et al., 2009); lack of paraprofessional support (An & Goodwin, 2007; Murata & Jansma, 1997); parental concerns about such a lack of support (Qi & Ha, 2012); lack of peer acceptance and activities designed to promote greater social interaction (Coates & Vickerman, 2010; Place & Hodge, 2001); over-emphasis on sporting activities (Qi & Ha, 2012) and, subsequently, on competition and winning and losing (Blinde & McCallister, 1998; Coates & Vickerman, 2010; Goodwin & Watkinson, 2000; Fitzgerald & Stride, 2012; Seymour, Reid, & Bloom, 2009); and, from the perspective of SEND children themselves, a lack of appropriateness in the design of PE classroom activities, which they feel are poorly matched to their individual their abilities and needs (Maher, 2017; Spencer-Cavaliere & Watkinson, 2010).

In light of this research evidence, it is hardly surprising to find that many teachers receive insufficient training, feel inadequately prepared, therefore often lacking the confidence and practical competence to work with SEND children in PE (Hodge, Ammah, Casebolt, LaMaster, & O'Sullivan, 2004; Klavina, Block, & Larins, 2007; Lienert, Sherrill, & Myers, 2001; Vickerman, 2007; Vickerman & Coates, 2009). In fact, research has even shown that, where perceptions of confidence and practical competence in IPE are reported by teachers, ITT is itself not acknowledged as being formative of their attitudes towards inclusion and/or their abilities vis-à-vis inclusive practice (Coates, 2012).

To address shortcomings in teachers' preparedness for IPE, it has been suggested that embedding training within ITT can equip trainees with the tools and practical competence to underpin future pedagogy. This has often been referred to in the PE literature as the 'infusion' approach (DePauw & Karp, 1994; Kowalski, 1995; Tripp & Rizzo, 2006), whereby knowledge of

SEND and pedagogical models for including SEND pupils within PE lessons is distributed across the ITT curriculum. It stands to reason that trainee teachers' understandings about IPE can be developed with an infusion approach using training courses – or, indeed, further expanded, challenged, and even scrutinized depending upon their level of prior experience (Hodge, Davis, Woodward, & Sherrill, 2002). However, much of the infusion research to date has focused on addressing attitudes towards inclusion rather than on developing actual abilities to deliver inclusive pedagogy (Hardin, 2005; Hutzler, 2003; Kowalski & Rizzo, 1996). For example, in their recent systematic review, Hutzler, Meier, Reuker, and Zitomer (2019) concluded that, whilst infusing knowledge across the ITT curriculum remains a promising idea for enhancing PE pedagogy, future studies are needed to better establish its effects – particularly in relation to PE teachers' self-efficacy towards including children with different types of SEND in their classes.

In view of the limited evaluative research evidence on effective IPE training within ITT, the present study sought to: (i) measure the impact of an ITT-based national professional development program on trainee teachers' self-efficacy, and (ii) determine the extent to which program-related and/or trainee-specific characteristics moderated the observed changes in self-efficacy. The focus on self-efficacy aligns well with Hutzler et al.'s (2019) recent call for new research studies and, crucially, it was directly aligned with the stated aim of the IPE program: "[To] [i]mprove [course] participants' confidence and competence in their own skills to provide a high quality, inclusive PE experience for all young people" (Youth Sport Trust, 2019). Self-efficacy has been defined by Bandura (1997) – and others since then in PE (Tschannen-Moran & Woolfolk Hoy, 2001; Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998) – as the belief in one's ability to organize and execute courses of action to produce desired outcomes in a given task (Morris, Usher, & Chen, 2017). It is therefore clear that self-efficacy accurately captures the practical competence relating to inclusive pedagogy that IPE program designers sought to enhance in trainees.

To the best of our knowledge, this study is among only a very small number studies (Block, Taliaferro, Harris, & Krause, 2010; Taliaferro, Hammond, & Wyant, 2015) to focus directly on factors affecting the IPE-related self-efficacy of trainee PE teachers – despite widespread recognition that self-efficacy is an important predictor of future teaching quality, adaptability, and aspirations (Zee & Koomen, 2016). Therefore, in this study, we provide some initial empirical evidence on the extent to which infusion training programs within ITT can be a part of a future within education that aspires towards high-quality IPE for all.

Methods

The IPE Program

The intervention in question was a national professional development program (herein 'the program') funded by [insert funder here] and developed by [insert major partners here] with the main aim of preparing teachers in England for high-quality IPE. The model of inclusion upon which the program was developed was Black and Stevenson's 'Inclusion Spectrum' (Stevenson, 2009). The Inclusion Spectrum recognizes that all pupils can be included and should be challenged to progress in their learning in PE. The burden of responsibility is therefore upon the teacher to design a learning environment which is supportive of these ends using four different types of activities: (i) open (invite activity that does not emphasize individual differences); (ii) modified (provide differentiated instruction to individuals within a whole-group activity); (iii) parallel (divide the class into ability groups); (iv) separate (use temporary interventions with individual pupils to ensure that their activity is (re) aligned with the learning objectives of the lesson) activities. The guiding premise of the program, therefore, was that inclusion is not a 'specialist' topic but a core competency of effective teachers. And using these different types of activities, the program sought to develop trainee teachers' understanding of the ways in which PE can be differentiated to meet pupils' diverse and often complex needs.

The program was evaluated over the period September 2015 to May 2016. Our involvement in the program resulted in data being collected at 44 IPE workshops from 418 primary and secondary school trainees at their respective Higher Education Institute (HEI) in England. The content and structure of the workshop was designed centrally and reviewed annually by the major partners. Each IPE workshop was designed to last approximately six hours and was delivered on a single day as an additional training unit embedded within the trainees' university course. The delivery of each workshops was the responsibility of a national faculty of tutors, who were either PE teachers working in secondary or special schools with tutoring experience, specialist university lecturers, or independent IPE consultants. Detailed workshop material was produced and disseminated to ensure program fidelity. Tutors were also invited to participate in 'tutor development days' twice a year, during which time IPE workshop material was presented, explained and debated, and practical sessions given to illustrate examples of effective delivery.

As a matter of process evaluation, we observed a subsample of (n=7) IPE workshops. The average overall time on task at workshops, the proportions of time devoted to theoretical and practical aspects of inclusion, as well the proportions of time devoted to tutor- and participant-led activities, are summarized in Table 1. The process evaluation concluded that the theory of instruction underpinning the program clearly and reliably reflected contemporary approaches to professional learning. And, whilst there was some very minor variation in the ways in which core principles of inclusion were applied by tutors across workshops (as represented by the standard deviations shown in Table 1), examples of high quality implementation included: time for discussion of both theoretical and practical matters; reflection upon existing knowledge and practice; demonstrations of actual practical activities that reflected best practice use of the Inclusion Spectrum; facilitation of discussion and sharing of experiences; and even opportunities for planning and acting out lessons based on a wide variety of pre-defined and trainee-prompted SEND scenarios. In fact, as shown in Table 1, we observed that approximately 40% of the time

within these IPE workshops was devoted to trainee-centered learning, wherein trainees were tasked with designing and modifying inclusive PE activities and to share outcomes with their peers. In this respect, the program was exemplary of the type of 'hands-on' infusion practice that has been suggested by Hutzler and Bar-Eli (2013), Taliaferro et al (2015), Block, Grenier, and Hutzler (2017), and others outside of PE (Van Laarhoven, Munk, Lynch, Bosma, & Rouse, 2007), whereby trainees are challenged to actively engage with and critically evaluate case studies and examples of inclusion as it is likely to play out in "real life" (Azzarito & Ennis, 2003, p. 179).

[INSERT TABLE 1 HERE]

Research Design and Participants

A pre-post design was used to assess the effect of the workshop and possible moderating factors. Data were collected at two time-points: immediately prior to the workshops to obtain baseline data (Time 1, or T1); and immediately afterwards to ascertain changes that had occurred as a result of participation (T2). T1 and T2 data collection for all variables was administered as a paper-and-pencil questionnaire that was distributed to participants by the IPE tutor on the day. All workshop attendees were invited to participate in the research by default (though they could choose to opt out). Questionnaires were collected by the IPE tutor on the day of the workshop and were sent back to the program administrators at [insert major partners here]. Packs of completed questionnaires were sent to the researchers for data entry and analysis by the program design and implementation team at routine reporting intervals (every ~3 months). Of the 418 trainees involved in data collection, 366 had complete demographic and pre-post data, and were therefore eligible for inclusion in this study. This study was approved by the [insert name of institutional review board]. All participants provided informed written consent to participate. Baseline characteristics for the trainee teachers are shown in Table 2.

[INSERT TABLE 2 HERE]

Dependent Variable

As noted in the introduction, self-efficacy was identified as the primary outcome variable. The rationale for choosing the self-efficacy as the construct against which to evaluate the effect of the program was based on research evidence linking it to teaching quality in general (OECD, 2009; Zee & Koomen, 2016), and to teacher competence in the delivery of differentiated (i.e., inclusive) instruction in particular (Block, Taliaferro, Harris, & Krause, 2010; Chao, Chow, Forlin, & Ho, 2017; Taliaferro, Hammond, & Wyant, 2015; cf. Taliaferro & Harris, 2014). Self-efficacy has been defined broadly in Bandura's (1997) social-cognitive theory as belief in one's ability to organize and execute the courses of action required to produce desired outcomes. Building on Bandura's work in the context of PE, Tschannen-Moran and Woolfolk Hoy (2001) have since defined teacher self-efficacy as the belief about one's ability to "bring about desired outcomes of student engagement and learning, even among those students who may be difficult or unmotivated" (p. 783).

A bespoke eight-item inventory was used to measure self-efficacy in the context of the present study. The precise self-efficacy inventory was developed by the research team in collaboration with the major program partners. The process was iterative, and we drew extensively upon Tschannen-Moran and Woolfolk Hoy's (2001; Tschannen-Moran et al., 1998) 'Teacher Sense of Efficacy Scale' (TSES) (cf. Block, Hutzlter, Barak, & Klavina, 2013; Depaepe & König, 2018; Humphries, Herbert, Daigle, & Martin, 2012). An initial set of items was developed following the research team's observation of two separate IPE courses and reflecting upon the ways in which the content of the IPE courses aligned with items on the TSES. These items were then further refined during a piloting process that took place at two additional IPE courses. The final set of items that were used for this study were agreed in collaboration with the program designers to ensure that the inventory was assessing aspects of trainees' self-efficacy that mapped

onto program content. This was critical, since it meant that trainees self-efficacy could be assessed against the underlying principles upon which the program had been developed.

The final inventory comprised statements assessing trainees' sense of self-efficacy in three main areas which have been shown (in previous research) to reflect effective pedagogical practice: ability to differentiate tasks ("*How confident are you in your ability to…Give different tasks to different groups of learners* (*at the same time*) *to meet their diverse needs*?"), assess pupils ("*…Assess students' learning an use this information to further their learning*?"), and facilitate independent learning ("*…Provide opportunities to all students, including SEND students, to be independent learnings*?"). A 7-point Likert scale ranging from 1 (*not at all confident*) to 7 (*completely confident*) was used. Principal components analysis confirmed a single-component solution for the self-efficacy inventory. This component had an eigenvalue of 5.90 in comparison to the nine other possible components, which had eigenvalues of ≤1.06. Alpha reliability was 0.92 and 0.93 at T1 and T2, respectively.

Moderating Variables

A subject-characteristics inventory was incorporated into the questionnaire at T1 to obtain data on age, sex, teaching specialization (primary or secondary school specialists), and teaching practice (completed or not-yet-started). An adapted inventory for quality of professional learning (Ingvarson, Meiers, & Beavis, 2005) was incorporated into the follow-up questionnaire at T2. Further rationale for our choice of moderators is presented in the following subsections.

Age and Sex. Evidence is mixed about the extent to which teacher characteristics (such as age and sex) influence engagement in professional learning activities and moderate their impact (Gore et al., 2017). Research is equally inconclusive about the extent to which attitudes towards inclusion are moderated by these variables (Avramidis & Norwich, 2002). Research on attitudes towards inclusion in PE reveal similar findings and are therefore not additionally instructive of possible effects here (Rekaa et al., 2018). Despite uncertainty in the extant research, both age and

sex were still incorporated into the analysis as potential moderators. Given the large sample size and diverse cohort of trainees being assessed in this study, we believed that this dataset presented an opportunity to further confirm or disconfirm findings which have largely come from smaller scale studies. Age and sex were analyzed as a linear numeric and binary nominal predictor, respectively.

Teaching specialization. Research shows that there are differences in what primary and secondary school teachers want and need to learn in professional learning settings, in order to practice effectively (Murphy & de Paor, 2017). Research in PE, more particularly, has found that primary and secondary school teachers have very different needs and are prepared in very different ways for the task of facilitating inclusive education (Maher & Macbeth, 2014). An additional factor that requires consideration here is that, in England (where this study was conducted) primary PE is taught by generalist teachers, whilst, in secondary schools, PE is taught by subject specialists. A major consequence of this, according to the Office for Standards in Education in the United Kingdom (Ofsted, 2013), is that primary-school teachers have very limited opportunities to learn how to teach PE in ITT courses. Therefore, we would expect to observe differences in the relative effect of IPE workshop attendance on trainee teachers with different specializations and career trajectories. Whether trainees were specializing as a primary or secondary school teachers was analyzed as a binary nominal predictor.

Teaching practice. Research has shown that access to school-based placements during ITT is an important predictor of teaching self-efficacy (Gurvitch & Metzler, 2009). Whether trainees describe their placements as being either positive or negative, research shows that the experience gained turns out to be a key determinant of self-efficacy and future teaching practices (Zach, Harari, & Harari, 2012). A recent study outside of PE (Schwab, Hellmich, & Görel, 2017) found that exposure to inclusive education whilst on placement is among the most important factors explaining teacher self-efficacy in this area. In light of these outcomes, we would expect to

find the completion of a school-based placement to be a substantial moderator of self-efficacy change. Whether or not trainees had completed their school-based teaching placement was analyzed as a binary nominal predictor.

Workshop quality perceptions. Perceptions about the quality of the design and delivery of pedagogical interventions can be an important factor explaining their effectiveness and impact. Capturing the extent to which IPE program impact was moderated by trainee' quality perceptions was considered an important component of the analysis. Perceptions about workshop quality were obtained immediately following the training. Here, trainees were asked about the extent to which they had opportunities to participate in activities that the CPD literature identified as critical to supporting knowledge growth and skills development. A 13-item workshop quality inventory was developed drawing upon the Quality of Professional Learning Index (Ingvarson et al., 2005). Trainees were asked about the extent to which (i.e., *strongly disagree* to *strongly agree*) they had opportunities to engage in activities that the professional development community has identified as critical to increasing teachers' knowledge and skills (Desimone, 2009). For example, the trainees were asked about the extent to which: "*I had opportunities to share knowledge, experiences, and ideas with other participants and the tutor/s*"; "*The program was tailored to my needs – it answered my pressing questions about inclusion*"; "*The inclusion strategies identified are feasible*".

Alpha reliability scores for subscales of the inventory that were used in this study and have been shared by the developers were all ~0.80 (Ingvarson et al., 2005). The inventory was incorporated into the questionnaire administered at T2. A 7-point Likert scale was used for this inventory (with anchors as described for the dependent). PCA confirmed a single component solution (5.77) which compared favorably against other potential components (≤1.09). An alpha reliability was 0.90. Workshop quality perceptions were analyzed as a linear numeric predictor.

Workshop size. Research in education has shown that class size is a meaningful predictor of pupil achievement, and that this is true regardless of the type of content being delivered and whether instructional and emotional supports are being emphasized in the delivery (Allen et al., 2013). There is also a small body of research on the effects of class size in PE contexts which has shown that class size can affect the type and quality of instruction teachers deliver, as well as the amount of time pupils spend on-task (McKenzie et al., 2000). Two factors justify the incorporation of workshop size into our analysis. Firstly, research suggests that the class size within ITT can predict self-efficacy independently of quality perceptions (Barroso et al., 2005). And secondly, more recent research has shown that the effect of class size is both relative to teacher experience and has a meaningful effect on the quality of pre-service teacher's learning – our specific participant group (Maulana, Helms-Lorenz, & Van de Grift, 2017). Workshop size was reported to the researchers by the program design and implementation team at routine intervals when packs of completed questionnaires were sent for data entry and analysis (every ~3 months). Workshop size was analyzed as a linear numeric predictor.

Statistical Analysis

Data were analyzed using linear mixed modelling (Proc mixed) in the Statistical Analysis System (SAS Studio Version 9.4 – University Edition). There are a number of reasons why linear mixed modelling was the appropriate statistical approach for analyzing this data. The most important reason was to account for 'unit of analysis' issues that Li, Xiang, Chen, and Xie (2017) have shown to be regularly violated in the field of research on teaching PE. The basic premise is that, because different groups of trainee teachers were 'nested' together at different workshops delivered by different tutors, their self-efficacy outcomes are more likely to be similar to one another than to trainees who attended other workshops. In other words, because trainees are nested together within workshops, the effect of the workshop on their self-efficacy outcomes cannot be assumed to be independent. Mixed modelling is the only way to account for this violation of the

assumption of independence and to reduce the Type 1 error rate that would be inflated with a simple pre-post t test, or with a repeated measures analysis of variance, or analysis of covariance (Tabachnick & Fidell, 2014). Mixed modelling is also useful because it gives you an outcome statistic (a random effect represented by a standard deviation) summarizing the average difference in the effect of the program between individual workshops. This is a good proxy measure for how effective a tutor was a delivering a given workshop. A high standard deviation would indicate that there was a lot of variability in the effectiveness of workshop delivery. Finally, added to these main benefits, mixed modelling was also useful because it allowed us to control for differences in the baseline (i.e., pre-program) self-efficacy of the trainees. This meant that we could estimate the effect of the workshops without the confounding effect of differences in trainees' baseline self-efficacy.

In this specific study, mixed modelling was used to allow for the clustering of trainees within different workshops and to allow for the effect of the workshops to vary independently across each of the four main subject-characteristic groups: female and male primary and secondary school trainees. The mixed model was specified to estimate mean changes (T2 minus T1) in self-efficacy across each of the four groups, with an adjustment made for baseline differences between groups and with the change score as the dependent variable.

The random-effect that describes the nesting of trainees within workshops is reported here as a standard deviation (SD). This SD represents the amount by which the net mean effect of the intervention differed between workshops (Smith & Hopkins, 2011). Fixed effect solutions for sex and teaching specialization were derived by deriving the mean value for each level of the covariate and calculating the difference between groups (i.e., female – male and secondary – primary school trainees). The effect of teaching practice was also calculated this way, with the effects presented as the difference between levels (teaching practice completed – not yet started). The fixed effects of

age, workshop quality, and workshop size were evaluated as linear numeric predictors; that is, as the change in self-efficacy associated with a two standard deviation (2SD) change in the predictor.

The lack of a control group meant that the effect of baseline self-efficacy required special treatment to adjust for the artefactual effects of regression towards the mean[3]. To do this, we derived a standard error of measurement (SEM) for the TSES from a reliability data reported in Tsigilis, Koustelios, and Grammatikopoulos (2010), and we based our adjustment on the assumption that the artefactual effect of regression towards the mean represents a reduction of $2SD*(SEM^2/SD^2)$. Making this adjustment had the additional benefit of enabling us to derive artefact-free estimates that reflected the effect in participants with an above (mean +1SD) and below (mean –1SD) average sense of self-efficacy prior to the intervention.

The magnitudes of the effects were assessed using standardization, by dividing the effect by an appropriate SD at baseline. The harmonic mean of the SDs of the four trainee teacher groups was used to estimate the overall mean these four standardized effects, and an adjustment for small-sample size was used to remove the bias in each of these standardized effects (Becker, 1988). The size of the resulting fixed effects were evaluated according to the following (modification of Cohen's) scale: <0.2, trivial; 0.2-0.6, small; 0.6-1.2, moderate; >1.2, large (Hopkins, Marshall, Batterham, & Hanin, 2009). These thresholds were halved for evaluating random effects: <0.1, trivial; 0.1-0.3, small; 0.3-0.6, moderate; 0.6-1.2, large; >1.2, very large (Smith & Hopkins, 2001).

To evaluate the importance and uncertainty of the effects, we used a reference Bayesian approach with a dispersed uniform prior (Hopkins & Batterham, 2018). This approach is more commonly known in the sport and exercise sciences literature as 'magnitude-based inference' (MBI) (Hopkins et al., 2009). Owing to the number of effects estimated, uncertainty in the estimate is expressed as 99% confidence limits. Probabilities of the true magnitudes of the effects were evaluated according to the following scale: <0.5%, most unlikely; 0.5–5%, very unlikely; 5–25%, unlikely; 25–75%, possibly (*); 75–95%, likely (**); 95–99.5%, very likely (***); >99.5%,

most likely (****) (Hopkins et al., 2009). Clinical MBI criteria were used to derive the inference for the mean effect of the intervention. The effect was deemed implementable if the chance of benefit was at least possible (>25%) and when the risk of harm was most unlikely (<0.5%). The effects of additional moderators were assessed using non-clinical MBI, whereby the effect was deemed clear when the chances a positive or negative magnitude were most unlikely (<0.5%).

Results

Table 2 shows that, on average, trainees were moderately efficacious (4.5 ± 1.0; mean ± standard deviation, SD) to begin with (at T1) and perceived the workshop they attended as being of high quality (6.2 ± 0.6) (at T2). Differences in baseline self-efficacy and workshop-quality perceptions among male and female primary and secondary school specialists were clear and trivial for all comparisons and are therefore not shown.

Pre- (4.5 ± 1.0) and post-workshop (6.1 ± 0.6) self-efficacy scores for trainees, along with a standardized effect statistic that summarizes trainees' pre-post self-efficacy changes (adjusted for differences in trainees' baseline self-efficacy), are shown in Table 3. There was a clear increase in self-efficacy among the trainees (1.6, ±0.2; standardized effect, ±99% confidence limits; most likely substantial). This standardized effect means that, on average, trainees' self-efficacy increased by 1.6 SDs above their baseline or pre-workshop values. That represents a large increase, when referenced against our scale for evaluating effect magnitudes (in the 'Methods' section). The difference in this effect between trainees above (+1SD) and below (–1SD) the mean for self-efficacy at baseline was substantial, but small in terms of effect magnitude (0.31, ±0.27; likely substantial). This indicates that trainees who were more efficacious to begin with were likely to benefit a small bit more from the intervention. Or, phrased a little differently, depending on trainees' baseline values, their increase in self efficacy could have been as low as a 1.3 or as high as a 1.9 SD change. Both of these effects are still within the 'large' range for evaluating effect magnitudes, however.

Our results also indicated that the net mean effect of the intervention varied substantially between workshops. This outcome was summarized in the random effect from the mixed model. It represents the variation in average self-efficacy change from workshop to workshop expressed as a SD. The SD was moderate-sized (0.30, ±0.30; SD representing the random effect, ±99% confidence limits; likely substantial), which suggests that (though it does not directly test whether) an individual workshop tutor's delivery on the day might have been a substantial moderator of beneficial effects for trainees. Or, more broadly, it suggests that there were some workshop-level factors that explained trainee teachers' self-efficacy change.

[INSERT TABLE 3 HERE]

Finally, the effects of potential moderators of trainees' self-efficacy change are shown in Table 4. Trainees' perceptions about the quality of workshop delivery was a clear and substantial moderator (0.39, ±0.19; most likely substantial). This effect indicates that a 2SD difference in trainees' workshop quality perceptions was associated with a small increase in their self-efficacy change. In other words, and naturally enough, trainees benefitted more when the quality of the workshop was perceived as being above average. The number of participants attending a given workshop was also a clear moderator. However, when we interacted this moderator with trainee specialization, further analysis revealed that the size of workshop was only a substantial moderator for primary-school trainees. We found that a 2SD difference in workshop size was associated moderate increase (0.76, ±0.58; very likely substantial) in primary-school trainees self-efficacy. That is to say, primary-school specialists benefitted more when the size of the workshop was above average. Whether or not trainees had completed teaching practice was only a possible small moderator (0.29, ±0.45; possibly substantial). Age was a trivial moderator (-0.14, ±0.30), and the effect of sex was unclear (0.03, ±0.30).

[INSERT TABLE 4 HERE]

Discussion

The present study explored the effects of an inclusive PE (IPE) training workshop on trainee teachers' self-efficacy, and estimated the extent to which these effects were moderated by workshop- and trainee-related factors. Building on research on the *beliefs* and *attitudes* of trainee teachers towards IPE (Rekaa et al., 2018), our study revealed that IPE training can have a beneficial effect on perceived *competencies*. The study therefore speaks to the viability of embedding IPE workshops within initial teacher training (ITT) course, particularly given the size and likelihood of the beneficial effect.

Unique in respect of other studies, our analysis showed that a trainee's starting point when they are undertaking IPE training is an important moderating factor. We found that trainee's with an above average sense of self-efficacy prior to taking the course were likely to benefit more. Moreover, whether or not trainees had undertaken their school placement was also a possible moderator. These outcomes suggest that trainees benefitted more when they had at least some exposure to practical, or 'hands-on', experience prior to attending the workshop. The ability to relate new information back to a concrete practical experience and visualize how one could have use this knowledge in the context of one's own practice is a well-established tenet of situated learning and other social-psychological perspectives. This research evidence therefore both supports existing theoretical frameworks whilst also providing a clear practical message: context matters to the acquisition of new knowledge and skills, therefore trainee's prior practical experiences should be factored into decisions about when they receive IPE training.

There were other contextual factors that moderated IPE outcomes. Most notably, trainees' perceptions about the quality of the workshops they attended was an important predictor of the degree to which self-efficacy changed. One might be inclined to interpret this

outcome with caution, owing to the fact that participants made quality assessments immediately post-workshop (and were therefore possibly susceptible to acquiescence response bias). However, this importance of workshop quality was also corroborated by the random effects, which showed that a substantial proportion of the variability in trainee's self-efficacy change scores was attributable to the specific workshops they attended. This implies that the benefits that accrued to participants across IPE workshops were relative to the quality of the specific tutors' behavior, workshop design, and overall delivery.

Workshop size was also an important contextual factor. Recall that the number of participants attending workshops was a likely predictor, and that it was an especially important predictor for primary-school trainees, in particular. Our results showed that a 2SD difference in the size of a given workshop was associated with a moderate–sized difference in the self-efficacy changes reported by primary-school trainees'. What does this mean in practical terms? In justifying the inclusion of potential moderators, we noted that primary-school trainees lack opportunities to learn about PE in their ITT courses (Ofsted, 2013). Because of this, we would have speculated that a smaller the class size would have been more beneficial to primary-school trainees. It stood to reason that smaller classes would have increased the likelihood of direct contact time with the workshop tutor, and therefore greater consideration could be given to the specific context within which they intend to work. Our analysis shows that the opposite is more likely to be true, however. Primary-school trainees reported benefitting more from larger workshops. A possible explanation for this is that larger workshops led to primary-school trainees being exposed to more IPE ideas, which in turn meant greater scope for discussion of IPE principles. For example, one of the main aspects of the IPE workshop was the time devoted to participant-led design activities, whereby smaller groups of trainees worked together to design an appropriate learning episode using a specific inclusion scenario, or with a specific SEND child in mind. Once designed, each subgroup

would be given time demonstrate and discuss the activities they had designed in front of the whole class. At larger workshops, there would have been more groups working on more scenarios and showcasing a greater variety of learning episodes across different kinds of SEND children. Perhaps it was exposure to these types of practical activities that factored into primary-school trainees' perceptions of enhanced self-efficacy, since all knowledge being gained during these activities was most likely resulting in new knowledge.

Taken in conjunction with the importance trainees' attribute to workshop quality, it is clear that the extent to which a tutor delivers and tailors content for individual trainees is a crucial factor in self-efficacy change. This is an especially important aspect of tutor delivery within a short-course context, since short course professional learning has often been criticized for prescribing standardized content without due consideration of individual contexts and diversity in professional learning needs (Kennedy, 2016). Tailoring course delivery requires the ability to adapt, rather than standardize, and the ability to facilitate participant-led activities and learning, rather than simply transmit information. There were many examples of this type of facilitation during IPE workshops. That a substantial proportion of the time was devoted to translating inclusive principles into concrete practical activities is undoubtedly one example. Similarly, the extent to which trainees worked in groups, designed their own bespoke activities and learning episodes, shared outcomes with peers, engaged in discussion, group questioning, and reflection. Such cooperative learning principles have been shown to enhance higher-level efficacy outcomes which could underpin future effective IPE, such as persistence and perseverance, transfer ability, and even mastery (see for example Pai, Sears, & Maeda, 2015).

Limitations

There are a number of limitations which should be stated, and which can inform future research. Two are particularly noteworthy.

Firstly, the lack of controlled conditions is an obvious limitation which restricts, in part, what we can say about the impact of the IPE Program overall. Our results should therefore be taken as indicative of the factors affecting trainee teacher self-efficacy in an IPE context, rather than as definitive evidence of the efficacy of IPE interventions. We place emphasis on the phrase 'in part', however, since we have done a great deal to account for limitations in our design – most notably, by adjusting the dependent variable for the artefactual effects of regression towards the mean, and by providing estimates of the effect of the intervention for participants at 1SD above and below the mean self-efficacy at baseline. We worked with the program designers on this study and sought to embed it within the program delivery so as to enhance ecological validity. Despite the lack of control conditions, we would therefore reiterate that the size and scope are major strengths of the present study. We have contributed substantively to the existing literature base; however, future studies could benefit by employing a cluster-randomized control trial design.

Secondly, and more specifically, whilst our results indicated that the net mean effect of the program differed substantially between workshops, we are limited in what we can say about how this variability is distributed. For example, since data was not recorded to account for tutors who delivered multiple workshops, we have been unable to tease apart the exact proportion of variability that is specifically owing to tutor delivery. A three-level hierarchical mixed model design, with trainees nested within tutors delivering across workshops, is our preferred model to answer this question more precisely in future designs. The importance of balancing research design and real-world consideration is a key take-home point here. Future research should seek to approximate the ecological validity that we have obtained by embedding the study within the process of program delivery; however, the randomization of tutors to workshops and trainees to tutors would certainly enhance the robustness of the research design.

Future Research Directions

Methodological limitations which can inform future research design are stated above. However, a crucial additional consideration for studies like this in the future would be to include a meaningful follow-up. Trainees soon become newly qualified teachers who are faced with challenges of implementing new subject knowledge in the context of a real-life professional setting, with a new role and its accompanying practice expectations. To be meaningful, pre-service teacher training of the kind delivered in IPE must therefore be sustainable over time, actionable in concrete school settings and connected to in-service professional learning and development trajectories. Follow-up research can and should be both quantitative and qualitative, as this will enable a more rounded understanding of both behavioral change and the quality of teaching and learning experiences – for newly qualified teachers and their pupils alike.

What Does This Article Add?

In this article we provided evidence which suggests that targeted training can enhance the competence of trainee teachers to include children with SEND in PE. Our data shows that both individual trainee and contextual factors play an important moderating role in training efficacy.

The article builds upon previous research, which has often focused on trainees' belief systems and attitudes towards inclusion. Our study is novel because it focuses on self-efficacy in an inclusive PE context, and because it highlights the importance of trainees' self-efficacy prior to receiving training. Our study shows that trainees with an above average sense of self-efficacy prior to training were likely to benefit more. It also suggests that trainees who have completed their school-based placements could possibly benefit more. This finding has implications for practitioners: whether scheduling IPE workshops after trainees have undertaken a placement could more fully maximize impact.

As regards the importance of contextual factors, the evidence provided in this article reiterates the importance of quality delivery, and of quality tutors facilitating meaningful learning experiences for trainees in particular. Such an emphasis on quality might seem self-evident. However, research continues to show that the delivery of professional learning continues to be both highly variable and overly focused on the passive transmission of intellectually superficial content (Higgins et al., 2016). When it comes to educating trainee teachers for the demands of high-quality inclusive PE in the future, we suspect that more of the same rote learning is not enough.

References

Allen, J., Gregory, A., Mikami, A., Lun, J., Hamre, B., & Pianta, R. (2013). Observations of effective teacher–student interactions in secondary school classrooms: Predicting student achievement with the classroom assessment scoring system—secondary. *School Psychology Review*, *42*, 76–98.

Ammah, J. O., & Hodge, S. R. (2005). Secondary physical education teachers' beliefs and practices in teaching students with severe disabilities: A descriptive analysis. *The High School Journal*, *89*, 40–54.

An, J., & Goodwin, D. L. (2007). Physical education for students with spina bifida: Mothers' perspectives. *Adapted Physical Activity Quarterly*, *24*, 38–58.

Avramidis, E., & Norwich, B. (2002). Teachers' attitudes towards integration/inclusion: A review of the literature. *European Journal of Special Needs Education*, *17*, 129–147.

Azzarito, L., & Ennis, C. D. (2003). A sense of connection: Toward social constructivist physical education. *Sport, Education and Society, 8*, 179–197.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W. H. Freeman & Company.

Barroso, C. S., McCullum–Gomez, C., Hoelscher, D. M., Kelder, S. H., & Murray, N. G. (2005). Self-reported barriers to quality physical education by physical education specialists in Texas. *Journal of School Health, 75*, 313–319.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41*, 257–278.

Blinde, E. M., & McCallister, S. G. (1998). Listening to the voices of students with physical disabilities. Experiences in the physical education classroom. *Journal of Physical Education, Recreation & Dance, 69*, 64–68.

Block, M. E., Hutzler, Y., Barak, S., & Klavina, A. (2013). Creation and validation of the self-efficacy instrument for physical education teacher education majors toward inclusion. *Adapted Physical Activity Quarterly, 30*, 184–205.

Block, M., Grenier, M., & Hutzler, Y. (2017). Strategies to maximize social participation and inclusive of students with disabilities in physical education. In A. J. S Morin, C. Maïano, D. Tracey, & R. G. Craven (Eds.), *Inclusive physical activities: International perspectives* (pp. 109–132). Charlotte, NC: Information Age Publishing.

Block, M., Taliaferro, A., Harris, N., & Krause, J. (2010). Using self-efficacy theory to facilitate inclusion in general physical education. *Journal of Physical Education, Recreation & Dance, 81*, 43–46.

Bloemen, M. A., Backx, F. J., Takken, T., Wittink, H., Benner, J., Mollema, J., & De Groot, J. F., (2015). Factors associated with physical activity in children and adolescents with a physical disability: A systematic review. *Developmental Medicine & Child Neurology, 57*, 137–148.

Chao, C. N. G., Chow, W. S. E., Forlin, C., & Ho, F. C. (2017). Improving teachers' self-efficacy in applying teaching and learning strategies and classroom management to

students with special educational needs in Hong Kong. *Teaching and Teacher Education, 66*, 360–369.

Coates, J. K. (2012). Teaching inclusively: Are secondary physical education student teachers sufficiently prepared to teach in inclusive environments? *Physical Education and Sport Pedagogy, 17*, 349–365.

Coates, J., & Vickerman, P. (2010). Empowering children with special educational needs to speak up: Experiences of inclusive physical education. *Disability and Rehabilitation, 32*, 1517–1526.

Depaepe, F., & König, J. (2018). General pedagogical knowledge, self-efficacy and instructional practice: Disentangling their relationship in pre-service teacher education. *Teaching and Teacher Education, 69*, 177–190.

DePauw, K. P., & Goc Karp, G. (1994). Integrating knowledge of disability throughout the physical education curriculum: An infusion approach. *Adapted Physical Activity Quarterly, 11*, 3–13.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Towards better conceptualizations and measures. *Educational Researcher, 38*, 181–199.

Fitzgerald, H., & Stride, A. (2012). Stories about physical education from young people with disabilities. *International Journal of Disability, Development and Education, 59*, 283–293.

Goodwin, D., & Watkinson, E. (2000). Inclusive physical education from the perspective of students with physical disabilities. *Adapted Physical Activity Quarterly, 17*, 144–160.

Gurvitch, R., & Metzler, M. W. (2009). The effects of laboratory-based and field-based practicum experience on pre-service teachers' self-efficacy. *Teaching and Teacher Education, 25*, 437–443.

Hardin, B. (2005). Physical education teachers' reflections on preparation for inclusion. *Physical Educator, 62,* 44–55.

Higgins, S., Cordingley, P., Greany, T., & Coe, R. (2016). *Developing great teaching: Lessons from the international reviews into effective professional development*. London, UK: The teacher development trust. Retrieved from http://tdtrust.org/wp-content/uploads/2015/10/DGT-Full-report.pdf

Hodge, S. R., Davis, R., Woodward, R., & Sherrill, C. (2002). Comparison of practicum types in changing preservice teachers' attitudes and perceived competence. *Adapted Physical Activity Quarterly, 19,* 155–171.

Hodge, S., Ammah, J. O. A., Casebolt, K. M., LaMaster, K., Hersman, B., Samalot-Rivera, A., & Sato, T. (2009). A diversity of voices: Physical education teachers' beliefs about inclusion and teaching students with disabilities. *International Journal of Disability Development and Education, 56,* 401–419.

Hodge, S., Ammah, O., Casebolt, K., LaMaster, K., & O'Sullivan, M. (2004). High school general physical education teachers behaviours associated with inclusion. *Sport, Education and Society, 9,* 395–419.

Hopkins, W. G., & Batterham, A. M. (2018). Advice on the use of MBI: A comment on *The Vindication of Magnitude-Based Inference*. *Sportscience, 22*. Retrieved from http://sportsci.org/2018/CommentsOnMBI/wghamb.htm

Hopkins, W. G., Marshall, S., Batterham, A., Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine & Science in Sports & Exercise, 41,* 3–13.

Humphries, C. A., Hebert, E., Daigle, K., & Martin, J. (2012). Development of a physical education teaching efficacy scale. *Measurement in Physical Education and Exercise Science, 16,* 284–299.

Hutzler, Y. (2003). Attitudes toward the participation of individuals with disabilities in physical activity: A review. *Quest, 55*, 347–373.

Hutzler, Y., & M. Bar-Eli. (2013). How to cope with bias while adapting for inclusion in physical education and sports: A judgment and decision-making perspective. *Quest, 65*, 57–71.

Hutzler, Y., Meier, S., Reuker, S., & Zitomer, M. (2019). Attitudes and self-efficacy of physical education teachers towards inclusion of children with disabilities: A narrative review of international literature. *Physical Education and Sport Pedagogy*, *24*, 249–266.

Ingvarson, L., Meiers, M., & Beavis, A. (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes and efficacy. *Education Policy Analysis Archives*, *13*. Retrieved from https://research.acer.edu.au/cgi/viewcontent.cgi?referer=https://scholar.google.com/&httpsredir=1&article=1000&context=professional_dev

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, *86*, 945–980.

Klavina, A., Block, M., & Larins, V. (2007). General physical educators' perceptions of including students with disabilities in general physical education in Latvia. *Palaestra, 23*, 26–31.

Kowalski, E. M. (1995). The infusion approach to teacher development. *Journal of Physical Education, Recreation & Dance*, *4*, 49–53.

Kowalski, E. M., & Rizzo, T. L. (1996). Factors influencing preservice student attitudes towards individuals with disabilities. *Adapted Physical Activity Quarterly*, *13*, 180–196.

Li, W., Xiang, P., Chen, Y. J., Xie, X., & Li, Y. (2017). Unit of analysis: Impact of Silverman and Solmon's article on field-based intervention research in physical education in the USA. *Journal of Teaching in Physical Education*, *36*, 131–141.

Lienert, C., Sherrill, C., & Myers, B. (2001). Physical educators' concerns about integration children with disabilities: A cross-cultural comparison. *Adapted Physical Activity Quarterly, 18*, 1–17.

Maher, A. J. (2017). 'We've got a few who don't go to PE': Learning support assistant and special educational needs coordinator views on inclusion in physical education in England. *European Physical Education Review*, *23*, 257–270.

Maher, A. J., & Macbeth, J. (2014). Physical education, resources and training: The perspective of special educational needs coordinators working in secondary schools in North-West England. *European Physical Education Review*, *20*, 90–103.

Makopoulou, K., & Thomas, G. (2016). Educating teachers for effective inclusive pedagogies. In K. Armour (Ed.), *Routledge handbook of physical education pedagogies* (pp. 491–502). London, UK: Routledge.

Maulana, R., Helms-Lorenz, M., & Van de Grift, W. (2017). Validating a model of effective teaching behaviour of pre-service teachers. *Teachers and Teaching*, *23*, 471–493.

McKenzie, T. L., Marshall, S. J., Sallis, J. F., & Conway, T. L. (2000). Student activity levels, lesson context, and teacher behavior during middle school physical education. *Research Quarterly for Exercise and Sport*, *71*, 249–259.

Morris, D. B., Usher, E. L., & Chen, J. A. (2017). Reconceptualizing the sources of teaching self-efficacy: A critical review of emerging literature. *Educational Psychology Review*, *29*, 795–833.

Murata, N. M., & Jansma, P. (1997). Influence of support personnel on students with and without disabilities in general physical education. *Clinical Kinesiology, 51*, 37–46.

690 Murphy, T. R., & de Paor, C. (2017). Teachers' CPD and sectoral interests: Opportunities for
691 convergence and divergence. *Teaching and Teacher Education, 66,* 242–249.
692 OECD (2009). *Creating effective teaching and learning environments: First results from*
693 *TALIS*. Luxenburg: OECD publishing. Available from:
694 https://www.oecd.org/edu/school/43023606.pdf
695 Ofsted (2013). *Beyond 2012: Outstanding physical education for all*. Manchester, UK.
696 Retrieved from
697 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachme
698 nt_data/file/413187/Beyond_2012_-_outstanding_physical_education_for_all.pdf
699 Gore, J., Lloyd, A., Smith, M., Bowe, J., Ellis, H., & Lubans, D. (2017). Effects of
700 professional development on the quality of teaching: Results from a randomised
701 controlled trial of Quality Teaching Rounds. *Teaching and Teacher Education, 68,*
702 99-113.
703 Pai, H. H., Sears, D. A., & Maeda, Y. (2015). Effects of small-group learning on transfer: A
704 meta-analysis. *Educational Psychology Review, 27,* 79–102.
705 Place, K., & Hodge, S. R. (2001). Social inclusion of students with physical disabilities in
706 general physical education: A behavioral analysis. *Adapted Physical Activity*
707 *Quarterly, 18,* 389–404.
708 Qi, J., & Ha, A. S. (2012). Inclusion in physical education: A review of
709 literature. *International Journal of Disability, Development and Education*, *59*, 257–
710 281.
711 Rekaa, H., Hanisch, H., & Ytterhus, B. (2018). Inclusion in physical education: Teacher
712 attitudes and student experiences. A systematic review. *International Journal of*
713 *Disability, Development and Education, 66,* 36–55.

Rieser, R. (2013). *Educating teachers for children with disabilities: Mapping, scoping and best practices exercise in the context of developing inclusive education*. Paris, France: UNESCO.

Schwab, S., Hellmich, F., & Görel, G. (2017). Self-efficacy of prospective Austrian and German primary school teachers regarding the implementation of inclusive education. *Journal of Research in Special Educational Needs*, *17*, 205–217.

Seymour, H., Reid, G., & Bloom, G. A. (2009). Friendship in inclusive education. *Adapted Physical Activity Quarterly*, *26*, 201–219.

Smith, T. B., & Hopkins, W. G. (2011). Variability and predictability of finals times of elite rowers. *Medicine & Science in Sports & Exercise*, *43*, 2155–2160.

Spencer-Cavaliere, N., & Watkinson, E.(2010). Inclusion understood from the perspectives of children with disability. *Adapted Physical Activity Quarterly*, *27*, 275–293.

Stevenson P. (2009). The pedagogy of inclusive youth sport: Working towards real solutions. In H. Fitzgerald (Ed.), *Disability and youth sport* (pp. 119–131). London, UK: Routledge.

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6th ed.). Harlow, Essex: Pearson.

Taliaferro, A. R., Hammond, L., & Wyant, K. (2015). Preservice physical educators' self-efficacy beliefs toward inclusion: The impact of coursework and practicum. *Adapted Physical Activity Quarterly*, *32*, 49–67.

Taliaferro, A., & Harris, N. P. (2014). The effects of a one-day workshop on physical educators' self-efficacy toward inclusion of students with autism. *Palaestra*, *28*, 38–43.

Tripp, A., & Rizzo, T. L. (2006). Disability labels affect physical educators. *Adapted Physical Activity Quarterly*, *23*, 310–326.

739 Tripp, A., Rizzo, T. L., & Webbert, L. (2007). Inclusion in physical education: Changing the
740 culture. *Journal of Physical Education, Recreation & Dance*, *78*, 32–48.
741 Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its
742 meaning and measure. *Review of Educational Research, 68*, 202–248.
743 Tschannen–Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive
744 construct. *Teaching and Teacher Education*, *17*, 783–805.
745 Tsigilis, N., Koustelios, A., & Grammatikopoulos, V. (2010). Psychometric properties of the
746 Teachers' Sense of Efficacy Scale within the Greek educational context. *Journal of*
747 *Psychoeducational Assessment*, *28*, 153–162.
748 UNESCO (1994). *Salamanca statement and framework for action on special education*
749 *needs*. Paris, France: UNESCO.
750 Van Laarhoven, T. R., Munk, D. D., Lynch, K., Bosma, J., & Rouse, J. (2007). A model for
751 preparing special and general education preservice teachers for inclusive
752 education. *Journal of Teacher Education*, *58*, 440–455.
753 Vickerman, P. (2007). Training physical education teachers to include children with special
754 educational needs: Perspectives from physical education initial teacher training
755 providers. *European Physical Education Review*, *13*, 385–402.
756 Youth Sport Trust [organizational webpage]. (2019). Retrieved August 16, 2019, from
757 https://www.inclusivepe.org.uk
758 Vickerman, P., & Coates, J. K. (2009). Trainee and recently qualified teachers' perspectives
759 on including students with special educational needs. *Physical Education and Sport*
760 *Pedagogy*, *14*, 137–153.
761 Zach, S., Harari, I., & Harari, N. (2012). Changes in teaching efficacy of pre-service teachers
762 in physical education. *Physical Education and Sport Pedagogy*, *17*, 447–462.

763 Zee, M., & Koomen, H. M. Y. (2016). Teacher self-efficacy and its effects on classroom
764 processes, student academic adjustment, and teacher well-being: A synthesis of 40
765 years of research. *Review of Educational Research*, *86*, 981–1015.
766

Table 1. Descriptive statistics for workshop characteristics assessed by systematic observation

	Mean ± SD
Duration	
Total workshop duration (mins)	219 ± 37
Workshop Delivery Aspects[a]	
Theory	35 ± 10
Practical	65 ± 11
Tutor-Led Activities	57 ± 8
Theory of instruction	44 ± 4
Examples	5 ± 1
Tutor-led practical	8 ± 3
Participant Led Activities	15 ± 2
Group task	9 ± 2
Sharing outcomes	2 ± 2
Discussions	4 ± 2
Active learning (practical)	26 ± 5
Design activities	17 ± 1
Sharing outcomes	7 ± 6
Sharing rationale	2 ± 1

[a]Data are percentages of total time devoted to each individual workshop delivery aspect. Data were derived by dividing each variable by the respective duration and multiplying by 100. For example the proportion dedicated to Theory=(Theory/Duration)*100. Proportions do not add up to 100. This is due to workshop tutor preparation time.

Abbreviations: SD, standard deviation.

Table 2 Simple statistics for the dependent variable and potential moderators of the 366 trainee teachers

Dependent	
Baseline self-efficacy (1 to 7)	4.5 ± 1.0
Moderators	
Age (y)	23 ± 5
Females (%)	63
Primary-school specialists (%)	57
Teaching practice completed (%)	62
Quality perceptions (1 to 7)	6.2 ± 0.6

Data are raw means ± SD or proportions.
Abbreviations: SD, standard deviation.

Table 3 Trainee teachers pre- and post-workshop self-efficacy, with the standardized adjusted pre-post change magnitude-based inference for this change.

	Pre-workshop	Post-workshop	Standardized Effect[a]	
	Mean ± SD	Mean ± SD	Mean, ±99%CL	Inference[b]
Self-efficacy	4.5 ± 1.0	6.1 ± 0.6	1.6, ±0.2	large ↑****

[a]Effect represents a standardized pre-post change score with an adjustment made for baseline differences between groups (i.e., baseline differences between male and female trainees with a primary and secondary school specialization).

[b]Magnitude thresholds: <0.2, trivial; 0.2-0.6, small; 0.6-1.2, moderate; >1.2 large. Asterisks indicate effects clear at the 99% level and likelihood that the true effect is clear, as follows: *possible, **likely, ***very likely, ****most likely.

Abbreviations: SD, standard deviation; CL, confidence limits.

Table 4 Effect of moderators on changes in the dependent, with magnitude-based inferences.

	Mean[a], ±99%CL	Inference[b]
Baseline self-efficacy (above – below average)[c]	0.31, ±0.27	small ↑**
Age	-0.14, ±0.30	trivial*
Sex (female – male)	0.03, ±0.30	trivial
Specialization (secondary – primary)	0.12, ±0.30	trivial**
Teaching practice (completed – not yet started)	0.29, ±0.45	small ↑*
Workshop size	0.35, ±0.42	small ↑**
Workshop quality	0.39, ±0.19	small ↑****

[a]Standardized effects shown for moderators represent the changes in the self-efficacy associated with either a 2 standard deviation difference in a numeric moderator or the difference in the levels shown for sex, specialization, and teaching practice.

[b]Magnitude thresholds: <0.2, trivial; 0.2-0.6, small; 0.6-1.2, moderate; >1.2 large. Asterisks indicate effects clear at the 99% level and likelihood that the true effect is clear, as follows: *possible, **likely, ***very likely, ****most likely.

[c]The effect of baseline represents the difference in the changes in self-efficacy for trainees with above (mean +1 standard deviation) and below (mean –1 standard deviation) average self-efficacy at baseline.