This is the published version of a paper published in *International Journal of Leadership in Education*.

Citation for the original published paper (version of record):

School effectiveness in Sweden: psychometric properties of an instrument to measure pedagogical and social climate (PESOC) focusing on pedagogical leadership
https://doi.org/10.1080/13603124.2019.1623921

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:kau:diva-73565
School effectiveness in Sweden: psychometric properties of an instrument to measure pedagogical and social climate (PESOC) focusing on pedagogical leadership

Lisa Hellström & Curt Hagquist

To cite this article: Lisa Hellström & Curt Hagquist (2019): School effectiveness in Sweden: psychometric properties of an instrument to measure pedagogical and social climate (PESOC) focusing on pedagogical leadership, International Journal of Leadership in Education, DOI: 10.1080/13603124.2019.1623921

To link to this article: https://doi.org/10.1080/13603124.2019.1623921

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 18 Jun 2019.

Submit your article to this journal

Article views: 82

View Crossmark data
Improving school effectiveness is a priority for many countries. The Swedish instrument **Pedagogical and Social Climate in School (PESOC)** has been widely used for measurement of school improvement. Since pedagogical leadership is an important component of school effectiveness, this study aimed to describe the psychometric properties of the PESOC subscale of pedagogical leadership (PESOC-PLP). Participants were 344 teachers from 30 schools in Karlstad, Sweden. Rasch analysis indicated two subdimensions of the scale, corresponding to academic and social objectives. Analysis showed that the instrument worked invariantly across different subgroups and that the response categories functioned as intended. Small, if any, within-school response dependence was noted. PESOC-PLP may be a useful tool for school leaders when evaluating their success in fulfilling academic and social objectives. Given the global demand for measurement of school leadership, also researchers and educators outside Sweden may have interest in translating and adapting the PESOC-PLP scale.

**Introduction**

A positive school climate is of great importance for ensuring a positive school experience, and together with leadership and quality instruction, school climate is often associated with what has been defined as effective schools (Kelley, Thornton, & Daugherty, 2005). Improving school effectiveness is a priority for many countries around the world.

Early research into school effectiveness found that academic results and school adaptation among students could be explained not only by students’ genetic abilities and social factors, but also by factors relating to the organization of the school and leadership (Brookover, Beady, Flood, Schweitzer, & Wisenbaker, 1979; Edmonds, 1979; Rutter, Maugham, Mortimore, Ousten, & Smith, 1979; Sammons, Thomas, & Mortimore, 1997; Weber, 1971). Sammons et al. (1997) showed that student achievement was associated with strong pedagogical leadership, evidenced by a focus on...
pedagogical issues, high visibility to staff and students, interest in monitoring and evaluating students’ progress, and emphasis on staff collaboration and involvement, as well as encouraging involvement of parents and providing feedback to them (Hultin et al., 2016).

The significance of school leadership is strongly connected to changes in the tasks and structures of schools and of the education systems that many countries are witnessing. Besides traditional tasks, new responsibilities are emerging. Being a school leader means fulfilling diverse and complex tasks connected to both professional and personal demands. More and more attention is being given to the role of school leaders in creating the conditions for an effective school (Huber & Hiltmann, 2011). A recent literature review (Preston, Goldring, Guthrie, Ramsey, & Huff, 2017) identified learning-centered leadership, the extent to which leaders hold a vision in the school for learning, as an essential component for effective high schools.

**Pedagogical leadership**

The role of the principal as a learning-centered leader means that he or she is accountable for teaching and learning issues through generating collective learning opportunities for all faculty and students (Orr, 2006; Schechter & Qadach, 2016). The literature on effective leadership has identified many different concepts, styles and models (Male & Palaiologou, 2017) and it has been discussed whether different types of leadership styles associates with student achievement in different ways (Robinson, 2007; Stein & Nelson, 2003). Pedagogical leadership has been identified as including both direct and indirect forms of leadership. Direct pedagogical leadership includes, for example, giving feedback to teachers and analyzing academic results in relation to teaching methods, giving support, and challenging and developing teachers. Indirect pedagogical leadership includes providing opportunities for teaching and learning, such as organizing activities by creating a clear structure for teacher and student schedules, and meeting formats and encouraging collaboration between teachers (The Swedish School Inspectorate, 2012; Törnsén, 2009). Recent research has shown that good pedagogical leaders place trust and responsibility in the hands of people who can make the most impact on the life chances of children within their school (Male & Palaiologou, 2017). Further, pedagogical leaders are expected to be present and close to the teaching and learning processes; involve teachers in quality development; enhance the development of formative assessment; engage in teacher development and develop the internal organization of the school to promote learning (Leo, 2015). There are similarities between pedagogical leadership and instructional leadership as the latter has been summarized to include the defining of a school mission, managing instructional programs and creating a positive school climate (Hallinger, 2005; Purinton, 2013). In the current study, pedagogical leadership refers to the responsibility of the school principal to create conditions where effective teaching and learning as well as possibilities for personal growth are supported (Näsman, 2018).

Since strong school leadership is important for the social and pedagogical climate in a school, assessment of effective pedagogical leadership are important (Dake, Price, Telljohann, & Funk, 2004; Höög, Johansson, & Olofsson, 2005; Törnsén, 2009; Yukl, 2012). Empirical results considering the relationship between school leadership and
students’ academic success are, however, ambiguous. School leaders’ ability to make a considerable difference to student outcomes is supported by qualitative research on the impact of leadership on school effectiveness and improvement, while quantitative analyses of the direct and indirect effects of leadership on students’ academic and social outcomes often report no impact (Kyriakides, Creemers, Antoniou, & Demetriou, 2010; Robinson, Lloyd, & Rowe, 2008). Hattie (2003) argues that the principal’s effect on student learning is mainly through factors associated with school climate rather than directly on learning. Rather than focusing on leadership as a unitary construct, Robinson et al. (2008) argue that a leader’s impact on student outcomes will depend on their particular leadership style. Ärlestig (2008) showed that principal leadership styles differed between schools that were successful in achieving both academic and social objectives compared with schools that were only achieving academically.

Robinson et al. (2008) showed good correlation between different teacher responses about their principal whereas the correlation between teacher and principal responses were much lower. The authors concluded that principals seem to perceive and understand leadership traits differently than teachers, illustrating the importance of ensuring valid principal evaluation instruments. The mixed results concerning the associations between school leadership and student outcomes may partly reflect disparities in methodology, for example, what type of leadership is examined. In addition, the scarcity of psychometrically sound principal evaluation instruments may have contributed to the inconsistent results from different studies of school effectiveness (Condon & Clifford, 2010; Tejeda, Scandura, & Pillai, 2001). Many principal evaluation instruments have not been developed from a clear understanding of effective leadership or based on the body of scholarship on school improvement (Murphy, Goldring, & Porter, 2014). Goldring, Porter, Murphy, Elliott, & Cravens (2009) argue that principal evaluation instruments should focus on a learning-centered leadership framework instead of assessing general management skills such as implementing visions. Along the same lines, Schechter and Qadach (2016), show that principals learning mechanisms (PLM) is useful for assessing principals learning-centered leadership through the mechanisms of storing and retrieving, receiving, disseminating and analyzing. Further, a valid measure needs to establish invariance in functioning across different groups to be compared, i.e. the instrument has to work in the same way across groups. Technically, invariance is achieved when the estimates of the item parameters are equivalent across different sample groups, e.g. across genders. In her article, Da’as (2017) showed that the cultural dimension (collectivistic versus individualistic) is important when teachers’ assess their principals’ skills, which have implications for the assessment of principals’ skills in cross-ethnic research. Overall, there is a strong need for standardized tools, which provide systematic feedback to leaders wanting to assess their strengths and individual developmental needs (Huber & Hiltmann, 2011).

**Pedagogical and social climate in school (PESOC)**

Pedagogical and Social Climate in School (PESOC) is a teacher-reporting instrument intended for studies on school effectiveness. PESOC is available in Swedish and consists of 13 sub scales of which the pedagogical leadership of the principal is one (hereafter called the PLP-scale). PESOC was developed by Grosin (2004) and is based on
a summation of the empirical results of research on school effectiveness, adapted for the Swedish school context (Hultin et al., 2016). PESOC rests on the assumption that the school’s social and pedagogical climate includes the expectations and norms of the principal, school management and teachers, as well as perception of the school’s purpose, capabilities and limitations and the consequent actions towards students, colleagues and parents (Grosin, 2002). A Swedish study found that a shorter version of PESOC showed good convergent validity with the Team Climate Inventory (Hultin et al., 2016) supporting the use and further development of the instrument, with promising psychometric properties. The same study using a shorter version of PESOC indicated that the PESOC had a two-factor structure; School leadership including two subscales containing items regarding the support and management of the school board and regarding the pedagogical leadership of the principal. The second factor Pedagogical activities and social relationships included items about teaching activities, perceived support and unity regarding general principles, as well as the social relationships between teachers, students, and families. Furthermore, a study in a Bulgarian sample using the shorter version of PESOC indicated good internal consistency for the overall score and for each domain of PESOC (Dimitrova, Ferrer-Wreder, & Galanti, 2016), indicating cross-cultural applicability of PESOC. Consistent with the study of Hultin et al. (2016), the Bulgarian study supported a bi-dimensional factor model and demonstrated adequate reliability and factor validity.

Since pedagogical leadership of the principal is an important component of school effectiveness, the current study examines the PLP-scale in detail. This scale is measuring one of 13 dimensions of the original full version of PESOC. The scale intends to capture both the structure of the school (i.e. content and form of teaching and education and social interaction between the principal, teacher and students) and the culture of the school (i.e. goal fulfilment, values and theoretical assumptions of the principal and the teachers) (Grosin, 2002, 2004; Höög et al., 2005).

Problem statement

Constructing measures for evaluating successful leadership is challenging, as there are no definite answers as to what constitutes effective leadership. Success can also be assessed differently, for example by measuring school climate, bullying incidence, academic success and meeting the needs of every student, collegiality and in relation to state and municipal management. Research suggests that principals influence teaching and learning by creating a safe and supportive school climate and school climate surveys have been proposed as a way to assess principal performance. However, many climate surveys are not explicitly validated for principal evaluation purposes (Clifford, Menon, Gangi, Condon, & Hornung, 2012) and standardized tools for assessing and evaluating leadership aspects are lacking (Huber & Hiltmann, 2011). PESOC is an instrument that is designed for school improvement efforts as it rests on empirically important aspects of school effectiveness research. This study adds to prior evaluation studies of PESOC (Dimitrova, Ferrer-Wreder, & Galanti, 2016; Hultin et al., 2016) by focusing on the pedagogical leadership of the principal, an area in need of assessment. The current study also adds methodological to before mentioned studies by using Rasch Measurement
Theory which enables proper examinations of the extent to which the instrument meet measurement requirements of invariance. The study also addresses challenges caused by the responding teachers being nested within schools implying clustered data. Previous studies on similar instruments have raised concerns about the appropriateness of using traditional reliability measures using clustered data as they treat each teacher’s response independently as if they were rating a different principal (Hallinger, Wang, & Chen, 2013). In the current study, the possible impact of clustered data will be analyzed using Rasch Measurement Theory with a focus on response dependence.

The context of the study

In the 1990’s the Swedish school system was profoundly changed, implying major changes in the roles and responsibilities of school leaders. In the 1990s, an extensive decentralization of responsibility for school operations was made from the state to the municipal level. While school leaders and teachers have been given greater responsibility for managing and improving schools there is still a strong state and local control of the school in terms of curriculum and quality assessments (Ludvigsson, 2009). Another drastic change to the Swedish school system includes a school reform introduced in 1992, enabling companies and organizations to run independent schools with public funding. Also a voucher system was introduced which made students and parents eligible to apply to any school they wanted.

The Swedish Education Act (Svensk författningssamling, SFS, 2010, p. 800) includes both social/democratic and academic goals, and highlights school principals as responsible pedagogical leaders. The Swedish school system has multiple structures, namely the national level, including the curriculum and other steering documents with objectives for all schools, the municipal level and the internal structures at each individual school (Ahlström, 2009).

Since 2010, all newly hired principals are obliged to undergo a comprehensive head teacher training program (Svensk författningssamling, SFS, 2010, p. 1521). The purpose is to improve schools’ goal fulfilment by providing school leaders with the required skills. Public education in Sweden is regulated by the Education Act (SFS 2010, p. 800), which clearly states that schools should work actively with social and civic objectives as well as with academic objectives (Ahlström, 2009). These tasks indicate that a school cannot be perceived as fully successful if it fails to achieve either academically or socially (Ahlström & Höög, 2009; Höög, 2011). The Education Act further emphasizes that the principal is responsible for student achievement and school effectiveness as well as for teachers’ professional development. The Swedish school debate since the beginning of 2000 has increasingly focused on result and goal fulfilment and an important part of efforts to improve schools in a broad sense is to strengthen school leadership (Höög & Johansson, 2011).

Aim of the study

The aim of the study was to analyze psychometric properties of the PESOC-PLP scale for measuring pedagogical leadership. Four research questions were addressed:
Do the items in the instrument work invariantly along the latent scale and among members in different sample groups that are compared, such as gender?

Do the response categories work as intended?

Does the PESOC-PLP constitute one or more dimensions?

Is there any within-school response dependence?

Method

Research design

To facilitate analyses where groups with specific characteristics are to be compared, high-quality measures are required; variables that appropriately describe the concept under investigation (and only that concept) are needed, along with measures that allow the researcher to appropriately compare specific groups. That, in turn, means that the measures need to work consistently across the groups to be compared (such as women and men or different social classifications) (Bergh, 2011). According to a leading methodologist in sociology, Duncan (1984), psychometric analysis based on the Rasch model (Rasch, 1960/1980) provides an attractive example of how these problems can be analyzed. It is thus possible to investigate if a set of items (variables) can be used together to construct a composite measure, describing one single concept, and to investigate if the composite measure works in the same way for different groups of individuals. This is an absolute requirement if we are to compare groups on the specific trait. Similar studies investigating the psychometric properties of different instruments have used the same methodological and analytical approach (e.g. Duncan, Bode, Lai, & Perera, 2003; Hagquist, Bruce, & Gustavsson, 2009; Hagquist & Hellström, 2014; Tennant, McKenna, & Hagell, 2004).

Participants

This study is based on data collected in 2011 as part of the Preventive School Project, which aimed to promote mental health among school children (see Hellström, Beckman, & Hagquist, 2013). The target group was all compulsory-school teachers in the city of Karlstad, Sweden. Some of the teachers reported teaching more than one grade including preschool classes. In Sweden, compulsory school includes grades 1–9. The sample included 344 teachers from thirteen school districts including 30 different schools. The data were collected with a web-based questionnaire concerning teachers’ perception of the social and pedagogical climate at their school. The overall participation rate was 56.7% (n = 344). Sociodemographic characteristics of the participants are presented in Table 1. The proportion of female teachers in the study (78.2%) is the same as in the municipality where the study was conducted (78.3%). In addition, the distribution of teachers in different teaching grades is representative of the population: at the time of the study, the municipality employed 14% of teachers in Kindergarten–3 schools, 34% (K–6), 18% (K–9), 18% (4–9) and 15% (7–9).

Data collection

First, an email with information about the study was sent out to all compulsory-school teachers in the municipality. Second, the questionnaire was distributed to the teachers
via email and included information about the study such as the aim, type of questions, and voluntary and anonymous participation. Instructions for filling out the questionnaire, a link to the questionnaire and a password were also included. Each password could only be used once. Four reminders were sent to participants.

**Instrument**

The original full version of PESOC consists of 95 questions intended to gather information about 13 domains of the social and pedagogical climate in school (for details, see Grosin, 2004). For the purpose of this study, the domain ‘pedagogical leadership of the principal’ (PLP) was used. The instrument was administered in Swedish but for the purpose of presenting it in this article, the items were translated into English by the first author. The original instrument consists of 19 questions regarding PLP. After consulting with principals at two local secondary schools, and with the approval of the constructor of the instrument (Lennart Grosin), one item was removed because it did not accurately reflect the routines of the schools, viz. ‘The distribution of teachers in classes are made after thorough pedagogical considerations.’ Two original questions were merged and changed, viz. ‘Leaders for different areas in the school essentially agree with the principal concerning basic educational issues and school goals’ and ‘Other leaders in school essentially agree with the principal concerning basic educational issues and school goals’ were merged as follows: ‘Assistant principals and other leaders in the school essentially agree with the principal concerning basic educational issues and school goals.’ These changes were made before any data were collected. Hence, the PESOC-PLP scale included 17 questions intended to capture pedagogical leadership of the principal (Table 2). The number of response categories on the PESOC items is five: ‘strongly disagree,’ ‘disagree,’ ‘agree,’ ‘strongly agree’ and ‘don’t know.’ ‘Don’t know’ responses can be attributed to characteristics of respondents or inadequacies in question design and may be problematic to analyze if not followed up (Fowler & Cannell, 1996; Krosnick et al., 2002). If a ‘don’t know’ response can be interpreted as being on an underlying continuum between agree and

| Table 1. Descriptive statistics of sociodemographic characteristics of participants based on the whole sample (N = 344). |
|-----------------|-----------------|
|                | Frequency | %     |
| Gender          |           |       |
| Male            | 55        | 16.0  |
| Female          | 269       | 78.2  |
| Missing         | 20        | 5.8   |
| Occupation      |           |       |
| Teacher         | 318       | 92.4  |
| Special education teacher | 26 | 7.6   |
| Teaching gradea|           |       |
| Kindergarten    | 7         | 2.0   |
| Preschool       | 29        | 8.4   |
| Primary school  | 122       | 35.5  |
| Middle school   | 158       | 45.9  |
| High school     | 133       | 38.7  |
| Years of employment at current school |       |       |
| ≤ 4 years       | 113       | 32.8  |
| ≥ 5 years       | 231       | 67.2  |

aSome teachers teach more than one grade
disagree, then assigning a value may be reasonable; otherwise, it is problematic (Acock, 2005). Similar to other studies (Hultin et al., 2016), the response option ‘don’t know’ was treated as missing data, because these responses could not be used in a rating where the response categories need to be ordered.

**The Rasch model**

A fundamental feature of the Rasch model is the property of invariance, which means that it is possible to test if items in an instrument work in the same way along the latent trait as well as across different sample groups, e.g. gender (Andrich, 1988, 2011). By analogy, when using a scale to compare body weights of men and women it is required that the scale works in the same way for both genders (Hagquist et al., 2009). Questions on pedagogical leadership also have to work in the same way, i.e. invariantly across individuals and groups of individuals if valid comparisons are to be made. Lack of invariance across sample groups like gender is commonly called Differential Item Functioning (DIF) and can be detected by different procedures, for example analysis of variance (ANOVA) of standardized residuals (Hagquist & Andrich, 2017). If the data meet the criterion of invariance, summation of the person’s responses (raw scores) across the different items is justified.
(Hagquist & Andrich, 2004). The unidimensional Rasch model offers a method for non-linear transformation of ordinal raw scores to interval measures. The model also requires a probabilistic Guttman pattern in the structuring of responses, meaning that for the same person ability, the probability to endorse an easy item has to be higher than the probability to endorse a more difficult item (Andrich, 1985; Hagquist et al., 2009).

The correspondence between the data and the Rasch model can be examined with formal statistics, for example, chi-square based on comparisons between observed and expected values and standardized residuals. Fit residuals above +2.5 and below – 2.5 are considered critical values when judging items (Hagquist et al., 2009). Graphical representations are also valuable tools for examining the operating characteristics of the items. Item characteristic curves (ICC) are graphs that show the expected value curve that predicts the item scores as a function of the item- and person-locations on the latent trait (Hagquist, 2001). The observed values should ideally fit perfectly with the theoretical expected value curve.

Good targeting is a further, essential requirement for good measurement. Mistargeting implies lower reliability and may cause problems in differentiating people along the latent scale (Hagquist et al., 2009). In Rasch modeling, misfit points to potential weakness of the data that cannot be solved by modeling of the data, such as inclusion of an extra parameter that takes account of different discrimination across the items. In contrast, in Rasch analysis misfit points to the need for improvements and revisions of the instrument, which has to be re-examined (Hagquist et al., 2009). In Rasch analysis, reliability may be measured by person separation index (PSI) or Cronbach’s alpha. While Cronbach’s alpha requires complete data, PSI does not. Furthermore, evidence of multidimensionality may imply violation of the principles of local independency (Marais & Andrich, 2008; Tennant & Conaghan, 2007). The Rasch model for polytomous data is also sensitive to the categorization of items, which means that possible problems with the ordering of response categories can be detected (Hagquist & Andrich, 2004).

**Data analysis**

The psychometric properties of the PESOC-PLP scale were examined using Rasch analysis (Rasch, 1960/1980). All analyses were performed using the item analysis program RUMM2030 (Andrich, Sheridan, & Luo, 2014). In the current study, different aspects of the PESOC-PLP scale were examined. The Rasch model was used to judge whether the PESOC-PLP scale meets the requirements of unidimensionality, invariance and proper item categorization. First, threshold ordering was examined to analyze the response categories of all 17 items. Second, by comparing the items and person locations, targeting was examined. Mistargeting can be detected by examining the proportion of responses in all categories for all items, by examining graphical representations of the person-item threshold distribution, or by examining the mean values of the persons’ estimates (Hagquist, 2009). Third, the fit of items to the Rasch model was examined with formal test statistics as well as graphical representations. Fourth, unidimensionality was examined by principal component analysis (PCA) of the item residuals and independent t-tests of differences between person location values from different subsets of items. The proportion of statistically significant t-tests should not exceed 5% for the scale to be unidimensional (Smith, 2002; Tennant & Conaghan, 2007). The results of the PCA and qualitative content
analysis of the items indicated two subdimensions of leadership. Fifth, new analyses of the two subsets of items were performed and included: item fit analysis, DIF analysis (ANOVA), and targeting. Sixth, the possible impact of the nested data on measurement was examined by Rasch analysis of response dependence between persons. Each dimension was examined separately. Only schools with 10 or more respondents were included in this analysis. Complete data were used from 161 (subdimension 1) and 106 participants (subdimension 2) from 15 schools. Because the number of responses coded as missing (including 'don’t know' responses) varied across items, responses included in the analysis of response dependence were different for the two subdimensions.

To examine whether there was any within-school dependence in the responses, each of the two dimensions was analyzed applying the following steps:

1. A person-by-item residual matrix was generated by the Rasch analysis, including a record for every person. The person-by-item matrix was transposed into a person-by-person matrix, enabling calculations of correlations between persons. The mean value of all correlations between persons was calculated, excluding correlations between one person and another person at the same school.
2. For each school a person-by-item residual matrix was generated by the Rasch analysis, including a record for every person in the school. The person-by-item matrix was transposed into a person-by-person matrix. Correlations between persons and the mean of these correlations were calculated for each school. Finally, the overall mean value across all schools of the within-school mean values was calculated.
3. The mean value of the correlations between persons calculated in step 1 was compared with the overall mean value across schools for the within-school mean values of the correlations between persons calculated in step 2.

**Results**

**General level of analysis (17 items)**

**Reliability and frequency distribution**

The PSI for the 17 items was 0.90, indicating good separation of persons and high power of the test statistics. The frequency distribution of the 17 items is reported in Table 2.

**Targeting**

Targeting was examined by person-item threshold distribution, showing the locations of the item threshold parameter estimates relative to the distribution of the persons for the items. The persons were relatively well targeted as a whole, but the person locations were skewed, with a positive mean (1.17). This shows that the general targeted population of teachers as a whole reported a relatively high level of PLP. The person-item threshold distribution was spread along the latent variable, but at the positive end of the scale, there were no thresholds matching the person locations. This implies that item thresholds capturing persons located further to the right (higher levels of PLP) of the scale were missing.
Threshold ordering and item fit
None of the 17 items showed reversed thresholds. Table 3 shows item characteristics and threshold values for all items. The spread of the item location values corresponds to different severity of the items with respect to the principal as a pedagogical leader. That is, there are items corresponding to the distribution (ranging from low to high values) of the teachers’ assessment of what constitutes a pedagogical leader. Item 7 was the ‘easiest’ whereas items 1 and 17 represent the two most ‘difficult’ items. Item 14 showed misfit according to the chi-square statistics ($p < 0.001$) and the residual fit values. In Figure 1, the ICCs show under discrimination for item 14, which means that individuals located at the beginning of the scale (lower levels of PLP) tended to respond higher than expected, and individuals located at the right end of the continuum (higher levels of PLP) responded lower than expected on this item. Deleting item 14 increased the PSI value slightly (from 0.902 to 0.904) without affecting targeting. This indicates that this item was redundant in relation to the other 16 items and it was removed given no loss in content validity.

Test for multidimensionality
PCA of the item residuals was conducted to examine possible violation of local independence. The set comprised two subsets. One (subdimension 1) included items with positive component-item residual correlations and the other (subdimension 2) included items with negative component-item residual correlations. The person location values from the two subsets were compared and differences assessed with independent t-tests. The proportion of statistically significant t-tests (11.52% at 0.05; 3.03% at 0.01) indicated multidimensionality. Based on the PCA of residuals, t-tests, and explorative content analysis, the entire set was split into two subdimensions of PLP that were subjected to further analyses.

Finer level of analysis (two subdimensions)
Threshold ordering and item fit for the two subdimensions
Table 4 shows item characteristics for subdimensions 1 and 2. None of the items showed reversed thresholds. The PSI for subdimension 1 (6 items) was 0.80 and for

<table>
<thead>
<tr>
<th>Item</th>
<th>Location</th>
<th>Fit residual</th>
<th>Probability</th>
<th>Thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.788</td>
<td>2.999</td>
<td>0.086021</td>
<td>−1.468</td>
</tr>
<tr>
<td>2</td>
<td>0.686</td>
<td>−0.903</td>
<td>0.081552</td>
<td>−1.719</td>
</tr>
<tr>
<td>3</td>
<td>0.706</td>
<td>1.014</td>
<td>0.425889</td>
<td>−1.882</td>
</tr>
<tr>
<td>4</td>
<td>−0.638</td>
<td>−1.398</td>
<td>0.513740</td>
<td>−1.324</td>
</tr>
<tr>
<td>5</td>
<td>1.109</td>
<td>−0.330</td>
<td>0.806112</td>
<td>−2.160</td>
</tr>
<tr>
<td>6</td>
<td>0.038</td>
<td>−0.913</td>
<td>0.338563</td>
<td>−1.609</td>
</tr>
<tr>
<td>7</td>
<td>−3.362</td>
<td>0.846</td>
<td>0.362164</td>
<td>−4.638</td>
</tr>
<tr>
<td>8</td>
<td>−0.616</td>
<td>−0.387</td>
<td>0.209334</td>
<td>−1.527</td>
</tr>
<tr>
<td>9</td>
<td>0.138</td>
<td>1.189</td>
<td>0.801742</td>
<td>−1.763</td>
</tr>
<tr>
<td>10</td>
<td>0.133</td>
<td>2.038</td>
<td>0.190982</td>
<td>−1.887</td>
</tr>
<tr>
<td>11</td>
<td>−0.544</td>
<td>−0.533</td>
<td>0.056578</td>
<td>−1.620</td>
</tr>
<tr>
<td>12</td>
<td>−0.460</td>
<td>−1.321</td>
<td>0.155519</td>
<td>−2.280</td>
</tr>
<tr>
<td>13</td>
<td>−0.332</td>
<td>−1.996</td>
<td>0.018239</td>
<td>−2.199</td>
</tr>
<tr>
<td>14</td>
<td>−0.782</td>
<td>2.724</td>
<td>0.000016</td>
<td>−1.678</td>
</tr>
<tr>
<td>15</td>
<td>0.701</td>
<td>−1.559</td>
<td>0.030420</td>
<td>−2.686</td>
</tr>
<tr>
<td>16</td>
<td>0.129</td>
<td>−1.597</td>
<td>0.041595</td>
<td>−2.474</td>
</tr>
<tr>
<td>17</td>
<td>1.306</td>
<td>0.257</td>
<td>0.634063</td>
<td>−1.900</td>
</tr>
</tbody>
</table>
subdimension 2 (10 items) 0.85. None of the items in subdimension 1 showed misfit. One item in subdimension 2 showed a fit residual above +2.5 (item 10). In Figure 2, an ICC shows that item 10 is slightly under discriminating. However, deleting item 10 decreased the PSI value, and it was retained.

**Differential item functioning (DIF) for the two subdimensions**

In Table 5, ANOVA based on the standardized residuals for the two subdimensions is shown. Two separate DIF analyses based on gender and years of employment were conducted. At the Bonferroni adjusted significance level (0.001) one item in

---

**Table 4. Item location, item fit, and threshold values for the two subdimensions.**

<table>
<thead>
<tr>
<th>Subdimension 1</th>
<th>Location</th>
<th>Fit residual</th>
<th>Probability</th>
<th>Thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0.900</td>
<td>−0.228</td>
<td>0.563981</td>
<td>−1.652</td>
</tr>
<tr>
<td>2</td>
<td>−0.325</td>
<td>−0.814</td>
<td>0.392079</td>
<td>−1.954</td>
</tr>
<tr>
<td>3</td>
<td>−0.261</td>
<td>0.417</td>
<td>0.899995</td>
<td>−2.059</td>
</tr>
<tr>
<td>5</td>
<td>0.174</td>
<td>0.605</td>
<td>0.436804</td>
<td>−2.298</td>
</tr>
<tr>
<td>9</td>
<td>0.848</td>
<td>1.262</td>
<td>0.649005</td>
<td>−1.893</td>
</tr>
<tr>
<td>17</td>
<td>0.361</td>
<td>0.503</td>
<td>0.649005</td>
<td>−1.977</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subdimension 2</th>
<th>Location</th>
<th>Fit residual</th>
<th>Probability</th>
<th>Thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>−0.165</td>
<td>−0.791</td>
<td>0.855466</td>
<td>−1.511</td>
</tr>
<tr>
<td>6</td>
<td>0.560</td>
<td>0.110</td>
<td>0.737295</td>
<td>−1.844</td>
</tr>
<tr>
<td>7</td>
<td>−3.123</td>
<td>0.194</td>
<td>0.356637</td>
<td>−4.973</td>
</tr>
<tr>
<td>8</td>
<td>−0.139</td>
<td>0.185</td>
<td>0.768685</td>
<td>−1.758</td>
</tr>
<tr>
<td>10</td>
<td>0.711</td>
<td>2.978</td>
<td>0.125338</td>
<td>−2.038</td>
</tr>
<tr>
<td>11</td>
<td>−0.158</td>
<td>−0.803</td>
<td>0.615566</td>
<td>−2.001</td>
</tr>
<tr>
<td>12</td>
<td>0.091</td>
<td>−0.528</td>
<td>0.530959</td>
<td>−2.368</td>
</tr>
<tr>
<td>13</td>
<td>0.186</td>
<td>−1.330</td>
<td>0.273510</td>
<td>−2.397</td>
</tr>
<tr>
<td>15</td>
<td>1.352</td>
<td>−0.458</td>
<td>0.135655</td>
<td>−2.910</td>
</tr>
<tr>
<td>16</td>
<td>0.684</td>
<td>−1.051</td>
<td>0.286170</td>
<td>−2.749</td>
</tr>
</tbody>
</table>

Location values with significant deviations (p < 0.001) and item fit residuals ±2.5 are shown in bold.

---

**Figure 1.** Item characteristic curves for item 14: ‘Teaching methods are first of all chosen and judged based on whether they lead to fulfillment of students’ learning.’
subdimension 1 (item 5) and one item in subdimension 2 (item 7) showed evidence of DIF with respect to years of employment at current school. Figure 3a,b shows graphical comparisons between teachers working ≤4 years and teachers working ≥5 years at their current school for items 5 and 7. The figure shows that teachers who have been working 4 years or less at their current school score higher on item 5 and lower on item 7 compared with teachers who have been at their current school for longer than 5 years, given the same location on the latent trait.
Targeting for the two subdimensions

For subdimension 1, the person-item threshold distribution showed that at the negative and positive end of the scale, there were no thresholds matching the person locations (Figure 4a). The mean value of the person’s location values (0.21) indicated a well-targeted measure. For subdimension 2, the person locations were positively skewed with a mean value of 1.81 (Figure 4b). There were areas where there were no thresholds matching the persons, indicating that there was information lacking about these persons. Although the persons were located towards the upper end of the continuum representing better pedagogical leadership, the item thresholds were spread along the latent variable. The thresholds at the lower end of the continuum matched the persons well, indicating that the items were good at targeting persons with lower levels of PLP.

Figure 3. Item characteristic curves for (a) item 5: ‘The principal places high priority on pedagogical activities over administrative tasks and contacts outside school,’ and (b) item 7: ‘The principal’s expectations of me as a teacher are high,’ divided by years of employment at current school.
Response dependence for the two subdimensions

The analysis of response dependence showed that the difference between the overall mean value of the means of the within-school inter-person correlations [Dim 1: 0.079/Dim 2: 0.009] and the mean value of all inter-person correlations [Dim 1: −0.011/Dim2: −0.009] were small for both dimensions. For dimension 1 the difference was 0.090 and for dimension 2 0.018, indicating small, if any, within-school response dependence.

Discussion

Murphy et al. (2014) noted that there are many instruments aimed at principal performance evaluation, but most have not been developed from a clear understanding of effective leadership or based on the body of scholarship on school improvement. Furthermore, many climate surveys are used for principal evaluation but are not
explicitly validated for these purposes (Clifford et al., 2012). This study contributes to the call for standardized tools that provide systematic feedback to leaders wanting to assess their strengths and individual developmental needs (Höög, 2011; Huber & Hiltmann, 2011). While the psychometric properties of PESOC as a school climate measure have been examined by earlier Swedish (Hultin et al., 2016) and Bulgarian (Dimitrova et al., 2015) studies using confirmatory factor analysis, this study focused on the important aspect of pedagogical leadership of the principal applying Rasch Measurement. At a general level of analysis, the PESOC-PLP scale seemed to work well. The analysis of the original set of 17 items showed a PSI of 0.90, indicating that the scale can effectively discriminate between persons. Categorization of items into response categories was possible and no reversed thresholds appeared. Item 14, stating ‘Teaching methods are first of all chosen and judged based on whether they lead to fulfiment of students’ learning’ showed severe misfit according to formal test statistics and the ICC curve. This item may also be questioned as regards its content validity because ‘Teaching methods’ is a broad concept and its meaning could be interpreted differently by different teachers. The misfit could be due to the phrasing of this item as it does not follow the pattern of the other items, that is, no person subject is included.

At a finer level of analysis, PCA of the residuals indicated multidimensionality. Based on these results and on an analysis of the content of items, items were split into two subdimensions. The two subsets correspond to the dual mission of the principal: achieving academic objectives, interpreted as responsibility for teaching methods and quality; and achieving social objectives, the ability to lead staff and build relationships. Rasch analysis of the two subdimensions showed sound psychometric properties for measuring pedagogical leadership, for example, proper item categorization, good item fit and ability to discriminate between persons along the latent scale. When teachers assess their principals’ skills, it is important to check for measurement equivalence across different groups to be compared (Da’as, 2017). Analysis of DIF showed that most items worked invariantly across two different sample groups except for two items that showed DIF with respect to years of employment at current school.

Items in the PESOC-PLP scale are intended to capture both the structure and culture of a school (Grosin, 2004). In the current study, one subdimension corresponded to leadership variables such as regular observations of teachers’ teaching methods, giving feedback, responsibility for teaching quality and quality instruction. These are variables that have been described as important for the structure of the school and academic goal fulfilment (Höög, Johansson, & Olofsson, 2009). It has been suggested that leaders make a difference to student achievement through a clear focus on academic and learning goals (Robinson, 2007; Stronge, Richard, & Catano, 2008), which could be accomplished by a leader taking responsibility for teaching methods and quality. Furthermore, the second subdimension includes leadership items such as agreement among school personnel concerning goals and visions, high availability to resolve conflict, and collaboration with teachers. These items rather seem to correspond to the culture of the school and could be seen as important for social and democratic goal fulfillment (Höög et al., 2009).

Leadership is important for the structure and culture of a school, which in turn is important for reaching social, democratic and academic goals. The results and content of the two subdimensions of the PESOC-PLP scale correspond well with the intention to capture the structure and culture of the school and embrace social, democratic and
academic goals. The PESOC-PLP scale may be a useful feedback tool for school leaders concerning strengths and weaknesses when evaluating their success in creating a favorable structure and culture.

Limitations of this study include the relatively small regional sample of data which affects the possibilities to generalize the results, which should be considered tentative. Previous studies of similar instruments have raised concerns about the appropriateness of using traditional reliability measures with clustered data (Hallinger et al., 2013). In the present study, the possible impact of clustered data was analyzed using the Rasch model with a focus on response dependence. The results indicated small, if any, within-school response dependence. Larger data sets would enable further analysis of possible response dependence as well as the dimensionality of the PESOC-PLP scale. While labeling individual items as characteristics corresponding to the school leaders’ two different roles (academic and social) may be hard solely based on content analysis, the PCA adds empirical evidence for the alignment of items within different subdimensions. More research is needed on the content validity of the two subdimensions to confirm their adequacy as measures of this dual responsibility. Some items may need to be revisited due to a large number of ‘don’t know’ responses that we treated as missing data.

The reasons for the high number of ‘don’t know’ responses for the item ‘Assistant principals and other leaders in the school essentially agree with the principal concerning basic educational issues and school goals’ needs to be examined. Was the number of non-responses high because the schools do not have assistant principals? Or was it because the teachers are unsure who to consider as ‘other leaders in school’? Researchers have cautioned that a ‘don’t know’ response does not necessarily mean that the person does not have an opinion (Krosnick et al., 2002). Such responses can be attributed to characteristics of respondents or inadequacies in question design (Fowler & Cannell, 1996; Krosnick et al., 2002). While offering a no-opinion option may reduce the amount of random variance in the attitude reports obtained, it may also preclude meaningful responses. Finally, while the current study as well as more recent studies (Dimitrova et al., 2016; Hultin et al., 2016) clearly indicate that PESOC is a valid measure of different aspects of school effectiveness, it may be noteworthy to acknowledge that differences and changes in the educational systems may affect its applicability.

Conclusions

Schools cannot be considered successful if both academic and social goals are not reached. Research has shown that principal leadership differs in schools that succeed academically and schools that succeed both academically and socially. This study adds to research on the development of pedagogical leadership as it shows that the PESOC-PLP scale may be used as a feedback tool for school leaders on how their leadership is perceived to relate to academic and social objectives. Considering global changes in the role of the school principal and the strong international focus on maximizing school effectiveness and quality education, there is a need for valid measurement instruments of leadership. Researchers and educators outside Sweden may have interest in further adapting the PESOC-PLP scale using a forward/backward translation process and expert panels to approve translations and cross-cultural adaptation followed by psychometric analyses, ultimately using the scale to evaluate pedagogical leadership as a means to achieve more effective schools.
Note

1. (1) Teachers’ expectations for students’ behavior and academic performance; (2) Perceived teacher agreement about school goals, norms, and rules; (3) Social rules regarding students’ behavior and achievement; (4) Student focus; (5) Basic assumptions about students’ ability to learn; (6) Communication between school and home; (7) Teacher interaction and cooperation; (8) Teachers’ work environment; (9) Teachers’ confidence and professional development; (10) Teaching activities; (11) Evaluation of students’ academic progress; (12) Principal’s pedagogical leadership; (13) Teachers’ perception of the school management’s involvement and support.

Acknowledgments

We thank Michelle Pascoe, PhD, from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was financially supported by the Swedish National Institute of Public Health, formalised as an agreement between Karlstad University and the City of Karlstad, and by a program grant from the Swedish Research Council for Health, Working Life and Welfare (Forte, 2012-1736) as well as by Malmö University, Department of School Development and Leadership; Forskningsrådet för Hälsa, Arbetsliv och Vålfärd.

Notes on contributors

Lisa Hellström, PhD in Public Health, works as a senior lecturer at Malmö university, at the institution for school development and leadership. Her research interests involves methodology, bullying and mental health and research involving children with special education needs.

Curt Hagquist is a professor of Public Health and the director for the Centre for Research on Child and Adolescent Mental Health (CFBUPH) at Karlstad University. Curt’s research concerns improving mental health and wellbeing of young people through excellent and interdisciplinary research focused on broader social and economic living conditions.

ORCID

Lisa Hellström http://orcid.org/0000-0002-9326-1175

References


