School Effectiveness and School Improvement
An International Journal of Research, Policy and Practice

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/nses20

Measuring policymaking capacities of schools: validation of the Policy Making Capacities Questionnaire (PMC-Q)

Joris Van Elsen, Jerich Faddar, Lies Appels, Sven De Maeyer, Jan Vanhoof & Peter Van Petegem

To cite this article: Joris Van Elsen, Jerich Faddar, Lies Appels, Sven De Maeyer, Jan Vanhoof & Peter Van Petegem (2023): Measuring policymaking capacities of schools: validation of the Policy Making Capacities Questionnaire (PMC-Q), School Effectiveness and School Improvement, DOI: 10.1080/09243453.2023.2213219

To link to this article: https://doi.org/10.1080/09243453.2023.2213219

SCHOOL EFFECTIVENESS AND SCHOOL IMPROVEMENT
https://doi.org/10.1080/09243453.2023.2213219

Routledge
Taylor & Francis Group

Measuring policymaking capacities of schools: validation of the Policy Making Capacities Questionnaire (PMC-Q)

Joris Van Elsen [a], Jerich Faddar [b], Lies Appels [a], Sven De Maeyer [a], Jan Vanhoof [a] and Peter Van Petegem [a]

[a]EduBROn, University of Antwerp, Antwerp, Belgium; [b]Department of Educational Sciences, Vrije Universiteit Brussel, Brussels, Belgium

ABSTRACT

In order to support research on school effectiveness, there is a need for valid and reliable instruments to assess policymaking capacities of schools. Increasingly, policymaking is seen as a shared responsibility of the entire pedagogical team of a school. In this article, data were analysed from a sample of 1,696 (care) teachers coordinators and principals from 77 Flemish primary schools to assess critical aspects concerning validity and reliability of the Policy Making Capacities Questionnaire (PMC-Q). Confirmatory factor analysis indicates that the fit for the model with eight factors is limited. The subscales were found to be strongly correlated and difficult to distinguish, and one of the subscales turned out to be not well captured by the questionnaire. The results indicate sufficient interrater reliability and within-group agreement to aggregate responses to school level. Suggestions are made for further development of the PMC-Q and its use for research and practice.

ARTICLE HISTORY
Received 11 July 2022
Accepted 2 May 2023

KEYWORDS
Policymaking capacities; policy effectiveness; questionnaire; confirmatory factor analysis; psychometric properties; validation

Introduction

According to the Organisation for Economic Cooperation and Development, there is a trend towards more decentralisation and more autonomy for schools (OECD, 2021). At the same time, schools are increasingly expected to be accountable to students, parents, and the wider community. Along with the autonomy, the importance of policymaking capacities of schools is increasing. Given the complexity of school policies, policymaking cannot be designed and implemented by principals alone. It is a shared responsibility of principals, teachers, and the other members of the school team (Hoy & Miskel, 2012). Since the beginning of this century, scientific interest in distributed or collaborative leadership increased. Around 2005, distributed leadership became the dominant leadership model in educational research (Gumus et al., 2018). In contrast to more traditional leadership views, distributed leadership does not expect principals to

CONTACT Joris Van Elsen Joris.VanElsen@UAntwerpen.be
This article has been republished with minor changes. These changes do not impact the academic content of the article.

formulate and implement the policy on their own, but rather to facilitate and activate the leadership qualities of others (Harris, 2013). Although the role of the principal in formulating and implementing school policy should not be underestimated, attention should also be paid to the policymaking capacities of the entire team.

Researchers and policymakers assume that the policymaking capacities of schools have an impact on student achievement (for a comprehensive review of the existing literature on the effects of school-level policy on educational outcomes, see Carr-Hill et al., 2016). The exact nature of this relationship is difficult to unravel. A valid and reliable instrument is needed to assess schools' policymaking capacities and enable research into the relationship between policymaking capacities of schools and student achievement or other variables. In addition, such an instrument can help school boards evaluate their policymaking capacities as part of their self-evaluation (Vanhoof et al., 2009).

Policymaking capacities

The concept of policymaking capacities of schools dates back to late 1980s and early 1990s. Sleegers et al. (1994) conceptualised policymaking capacity as "the extent to which schools *can* independently perform their tasks in policy-making" (p. 148, emphasis in original). As this definition suggests, a distinction should be made between policymaking capacities and policymaking. Whereas policymaking refers to coordinated and planned decisions and actions taken by the school board or the broader school team to improve the schools' functioning, policymaking capacities refer to the extent to which schools are capable of developing a policy autonomously. Jongmans et al. (1998) advocate that policymaking capacities are important prerequisites for schools in order to develop their own policy. According to Vanhoof and Van Petegem (2017), policymaking capacities enable schools to successfully use their autonomy to adjust their functioning, improve the quality of their education, and achieve their goals. In this view, policymaking capacities refer to the resources available in the school to develop effective school policy and to the extent to which schools succeed in actually implementing this policy.

That the concept of policymaking capacity emerged in the Netherlands and Flanders (Belgium) is no coincidence. At the turn of the century, schools in the Low Countries had already a relatively high degree of autonomy to develop their own instructional and administrative policy (Stevenson et al., 2021). More recently, the OECD (2018) reported that in the Netherlands, more than 90% of key decisions were taken at the school level. Flanders followed at some distance with 63%, but is still well above the OECD average of 34%. In Belgium, the autonomy of schools is grounded in Article 24 of the constitution. Strengthening the policymaking capacities of schools is therefore an important issue in the Low Countries.

On the basis of the international literature on school effectiveness, Van Petegem et al. (2004) and Vanhoof et al. (2009) distinguish eight indicators of policymaking capacities of schools: (a) effective communication, (b) shared leadership, (c) integrated policy, (d) shared vision, (e) innovative capacity, (f) supportive relations, (g) reflective capacity, and (h) responsive capacity. These eight indicators are not only supported by empirical evidence but can also be influenced by schools and are conceptually distinct from each other. At the same time, Vanhoof and Van Petegem (2017) stress that the indicators are interrelated and that policymaking capacities of schools are best approached as a

whole, taking into account all eight indicators. The eight indicators are relevant for different policymaking activities in schools, such as decision making, improving well-being (Van Gasse et al., 2016), improving instructional quality, and self-evaluation (Vanhoof et al., 2009).

Effective communication

Effective communication refers to the extent to which formal and informal communication initiatives meet information needs and the desire for mutual influence and coordination. Communication is one of the key factors for effective school policy since it connects people, which is crucial for effective operation (Bulgaru, 2019). Open and multi-directional communication is crucial to develop a shared vision, support team members, get people involved, and implement change.

Shared leadership

Shared leadership is another key factor of effective school policy and refers to the participation of the school team in school decision making. Shared or distributed leadership starts from the idea that leadership is a characteristic of an organisation rather than of an individual and makes use of different sources of leadership present throughout the organisation (Hoy & Miskel, 2012). Managing a school is considered too complex to be handled by individual principals. The expertise, involvement, and commitment of team members is needed to develop effective school policies (Harris, 2013). Moreover, shared leadership has a positive effect on teachers' cooperation, sharing of knowledge and experience, and internalisation of common goals (Amels et al., 2021).

Integrated policy

School policy consists of several domains that include education, organisation, personnel, budgets, and dealing with the environment (Vanhoof & Van Petegem, 2017). All these domains need to be addressed and aligned to achieve an effective school policy. Integrated policy entails coordination and the willingness to think beyond one's own functioning and can only exist when the tasks and responsibilities of others are clear to everyone.

Shared vision

Shared vision refers to the extent to which there is a generally accepted vision that guides the work of different stakeholders. A shared vision is essential to get teachers and other stakeholders to take actions to achieve the shared goals (Pineda-Báez et al., 2020). A shared vision requires open communication to explicate values, expectations, and priorities.

Innovative capacity

Innovative capacity refers to the extent to which schools are open to innovations and to the way in which they deal with changes and manage to implement innovations

successfully. Due to the rapid changes in 21st-century societies, schools are forced to constantly adjust their functioning and instruction. The capacity to innovate is necessary in every organisation to remain effective, efficient, and sustainable (Serdyukov, 2017).

Supportive relationships

Supportive relationships refer to the extent to which the professional and personal relationships between members of the school team are perceived as supportive. Members of the school team can only be involved in school policy when they feel sufficiently supported both professionally and personally (Visone, 2020). In line with the self-determination theory, we can state that teachers' motivation to participate in school policy is enhanced when three basic psychological needs are met: autonomy, competence, and relatedness (Martinez & McAbee, 2020). Supportive relationships are mainly related to the latter and require cooperation, collegiality, trust, and openness.

Reflective capacity

Reflective capacity refers to the extent to which schools take initiatives to identify strengths and weaknesses in their functioning. It is about the willingness of team members to question their actions and is linked to a critical attitude that aims to improve one's own functioning. Reflection is an active, proactive, reactive, and action-oriented process in which team members, individually or in a group (Alzayed & Alabdulkareem, 2021), reflect at regular intervals on their own responsibilities and how they deal with them. The capacity of teachers and schools to reflect is required to continuously improve their teaching (York-Barr et al., 2016).

Responsive capacity

Responsive capacity refers to the extent to which schools are open to, and able to respond to external demands and expectations. Schools are not isolated entities but are embedded in a social, economic, political, and cultural environment. Schools must respond to expectations and impulses from their environment. Increased multiculturalism within schools and awareness about differences in gender identity, for example, have reinforced the importance of schools' cultural responsiveness (M. Brown et al., 2022; Luke et al., 2022).

Development and initial validation of the PMC-Q

The Policy Making Capacity Questionnaire (PMC-Q) is developed in the context of a study on policy effectiveness of schools (Van Petegem et al., 2004). The English version of the instrument can be found in Appendix 1. During the development of the PMC-Q, the construct validity and content validity of the instrument and the dimensionality and internal consistency of the scales were assessed. The eight indicators around which the PMC-Q is structured are the result of a thorough analysis of existing models and literature and include the various aspects that influence policymaking (Van Petegem et al., 2004). The theoretical support of this model substantiates the construct validity of the questionnaire.

Subsequently, items were linked to the eight indicators. The scales were constructed based on the existing international scientific literature, semi-structured in-depth interviews with school principals, members of the school board and members of middle management, focus group discussions with teachers, and two panel discussions with a school principal, a teacher, a policy maker, an inspector, a union member, and a researcher (Van Petegem et al., 2006). After a pilot test, the questionnaire was initially completed by a representative sample of 305 Flemish primary and secondary school principals. Based on the results and the feedback, the questionnaire was further adapted and re-submitted to 265 principals (Van Petegem et al., 2004). Factor analysis revealed that all scales were one-dimensional and internally consistent (Cronbach's alpha > .70).

This study

Given the trend towards more decentralisation and more autonomy for schools (OECD, 2021), there is a need for instruments to assess policymaking capacities in schools in a valid and reliable way. Most existing instruments related to policymaking focus on leadership effectiveness or leadership styles. Recently, Shen et al. (2019) developed and validated an instrument to measure integrated leadership (i.e., the combination of principal leadership and teacher leadership). But although intertwined, leadership and policymaking capacities are two different constructs. As an instrument to capture schools' policymaking capacities, the PMC-Q goes beyond leadership and takes into account relevant aspects of organisational culture. During the development and initial validation of the instrument, only principals completed the questionnaire. Because only school leaders were involved in the quantitative part of the initial validation study, measurement invariance, within-group agreement, and interrater reliability have not been examined.

The aim of this study is to investigate critical aspects related to validity and reliability of the PMC-Q, based on a large sample of Flemish primary school teachers, care teachers, policy support staff, coordinators, and principals. The results of this study shed a light on the psychometric qualities of the instrument to assess the policymaking capacities of primary schools in research and practice. The following research questions are being addressed:

(1) How well does the PMC-Q measure the eight dimensions of schools' policymaking capacities?
(2) Are schools' policymaking capacities, as measured by the PMC-Q, equivalent between (care) teachers and staff with coordinating or decision-making roles?
(3) To what extent does the PMC-Q measure schools' policymaking capacities in a reliable way, based on responses from school team members?

Method

Respondents

Data were gathered along with TIMSS 2019 (Trends in International Mathematics and Science Study). On behalf of the Flemish government, schools from the TIMSS sample

were asked to participate in an additional study on schools' policymaking capacities. Linking the two studies allows follow-up research to examine the relationship between schools' policymaking capacities and student performance on the TIMSS tests. The latter research question is not yet reported in this study. Just over half of the schools (77 out of 147) participating in TIMSS also took part in the research on policymaking capacities. Unlike TIMSS, which surveys only principals and Grade 4 teachers, all members of the school's pedagogical team (teachers, care teachers, coordinators, policy support staff, and principals) were invited to complete the online questionnaire (Qualtrics) and received an information letter and personal login code to the questionnaire, distributed by the principals. Participation was voluntary. On average, 21.99 respondents participated per school (Min = 6; Max = 56). A total of 1,693 respondents (86.62% female) from 77 schools filled in the questionnaire. Respondents were between 21 and 65 years ($M = 40.54$; $SD = 10.86$) and had on average 13.98 years of experience in education ($SD = 10.57$). The majority of respondents were teachers ($N = 1{,}206$; 71.18%) or care teachers ($N = 142$; 8.39%). There were 108 (care) coordinators or policy support staff (6.38%) and 67 principals (3.96%).

Materials and procedures

The Policy Making Capacities Questionnaire (PMC-Q) was developed to capture policymaking capacities of schools. The questionnaire used for the current study is based on the questionnaire from the practice book of Vanhoof and Van Petegem (2017) and is structured around the eight indicators of policymaking capacities of schools. Ten statements were formulated for each indicator. For the indicator effective communication, only eight statements were retained. The statements do not refer to the respondents' own actions and attitudes but to their perceptions of the whole school. All statements are formulated in the same way: "In this school … ". Five response options were provided for each item: 1 (*totally disagree*), 2 (*disagree*), 3 (*neither disagree nor agree*), 4 (*agree*), and 5 (*totally agree*). The questions were arranged by indicator. Prior to the statements about the school, respondents were asked to answer some personal questions about their gender, age, education, experience, and position in the school. The items of the questionnaire are included in Appendix 1.

Analysis

A multi-group confirmatory factor analysis (MGCFA) was applied to assess measurement invariance between respondents with an operational position (teachers and care teachers) and respondents with coordinating or decision-making roles (coordinators, policy support staff, and principals). MGCFA allows to investigate the extent to which the measurement model is generalisable across groups and whether we can pool the responses of the respondents with different positions to make statements about the policymaking capacities of schools (Putnick & Bornstein, 2016). We first fitted a (single-level) confirmatory factor analysis (CFA) with unconstrained factor loadings for both groups separately (van de Schoot et al., 2012). Next, a configural, a metric (weak factorial), and a scalar (strong factorial) multi-group model were fitted and compared. Configural invariance was tested by applying the same model to both groups and indicates whether the

structure of the construct is similar across groups. Metric invariance was tested with a model where factor loadings are equal in both groups and indicates the extent to which the item loadings are equivalent across groups. Scalar invariance was tested with a model where factor loadings and intercepts are equal in both groups and indicates the equivalence of item intercepts across groups. A −.01 change in the comparative fit index (CFI) and Tucker–Lewis index (TLI), a .015 change in the root-mean-square error of approximation (RMSEA), and a .030 change in the standardised root-mean-square residual (SRMR) are used as criteria (Cheung & Rensvold, 2002; Putnick & Bornstein, 2016).

Next, we conducted a CFA to assess the factor structure of the instrument. CFA is a hypothesis-driven method that takes into account measurement error and assumes that the latent variables are interrelated. Since the eight dimensions of policymaking capacities are all part of the same construct, CFA allows to check for discriminant validity between the scales (Carter, 2016). Besides a one-level model with all items loading on their respective factors (Model 1), a two-level model (Model 2) was created with all items loading on their respective factors at Level 1 (respondent) and Level 2 (school). Multilevel confirmatory factor analysis (MLCFA) takes into account the clustering of the data (respondents nested within schools) and enables us to investigate the factor structure of the PMC-Q at the school level (T. A. Brown, 2015).

The factor structure of the PMC-Q is investigated in four steps. First, the fit of the model is assessed by the CFI, the TLI, the SRMR, and the RMSEA. We consider the fit of the model to be good if the CFI and TLI are close to .95, the SRMR < .08, and the RMSEA < .06 (T. A. Brown, 2015; Hu & Bentler, 1999). An SRMR and RMSEA < 0.05 would be even better (Lewis, 2017). Normally a significance level (p value) of the maximum-likelihood-based chi-square (χ^2) > .05 is recommended. Since we have a rather large sample and the chi-square is sensitive to the sample size, the p value of the chi-square is a less appropriate fit index for this study.

Second, the consistency and dimensionality of the scales are checked. To assess the internal consistency of the scales, Cronbach's alpha was calculated (Osborne, 2014). Generally, a Cronbach's alpha > .70 is considered good (Taber, 2018). The dimensionality of the scales is verified with an exploratory factor analysis (EFA) on each scale. Scales that turn out to be multidimensional were analysed in terms of content.

Third, the correlation between the scales and the discriminant validity of the PMC-Q was tested. Discriminant validity refers to the extent to which items differentiate between the factor they are theoretically expected to load on and the other factors. T. A. Brown (2015) suggests that correlations between factors above .80 or .85 indicate poor discriminant validity. But only relying on correlations is somewhat simplistic. Carter (2016) showed that correlations ≤ .85 can also be problematic with respect to discriminant validity. In accordance with Fornell and Larcker (1981), we consider the discriminant validity of a scale sufficient if the average variance extracted (AVE) of the scale is larger than the shared variance with all other scales. AVE is the sum of squares of the standardised loadings, divided by the sum of squares of the standardised loadings plus the sum of the observed variable measurement error (Carter, 2016). As such, the AVE is a measure for the amount of variance explained by the construct in relation to the amount of variance due to measurement error (Fornell & Larcker, 1981). AVE > .50 is considered acceptable, which indicates that more than 50% of the variance can be explained by the construct (Carter, 2016). The shared variance is the square of the correlation between two scales.

Fourth, the factor loadings of the items are assessed to identify items with a poor fit. For research that goes beyond the exploratory phase, a factor loading of .70 or higher is often considered desirable (Morrison et al., 2017; Nunkoo et al., 2013). The extent to which an item loads sufficiently on its factor depends on the case. Since we assume that the indicators of policymaking capacities are interrelated, we consider a factor loading $\geq .70$ as good and a loading $\geq .60$ as acceptable. Items with a factor loading $< .60$ will be subjected to a closer examination in terms of content.

The PMC-Q can be categorised as a referent shift consensus composition model (Biemann et al., 2012; Chan, 1998; van Mierlo et al., 2009). This means that respondents were asked to give their perception of their school, rather than evaluate their own policymaking qualities. If there is sufficient agreement between the participants from the respective schools, a score can be calculated for each school. The within-group agreement (r_{WG}) is used to justify aggregation of individual scores to higher level scores (James et al., 1984; Newman & Sin, 2020). Within-group agreement indicates to what extend responses of different respondents are interchangeable (Bliese, 2000). Because each scale consists of several items, we use the derived $r_{WG(j)}$ index. An $r_{WG(j)} > .70$ is usually considered acceptable. However, Lance et al. (2006) demonstrated that $r_{WG(j)}$ increases with the number of items, especially when the number of items per scale ≥ 10. Therefore, we calculate $r^*_{WG(j)}$ as formulated by Lindell et al. (1999).

In addition to the within-group agreement, which tells us something about the similarity of the perceptions of respondents within a school, we estimated interrater reliability to assess the consistency of responses of different respondents (Bliese, 2000). Interrater reliability is calculated with the intra-class correlation coefficient (ICC). Two ICCs can be distinguished: ICC(1) and ICC(2). ICC(1) is a measure of the proportion of the total variance that can be explained by the group (i.e., the school). A value between .05 and .10 is considered as a low to medium effect and indicates that the school to which the participants belong influences their judgement (LeBreton & Senter, 2008). Values $\geq .10$ are considered as medium effects and values $\geq .25$ are considered as strong effects. Bliese (2000) argues that ICC(1) usually falls between .05 and .20 and that values $\geq .30$ are rare. The ICC(2) is a measure of the reliability of the group means and is a function of the ICC(1) and the group size (Bliese, 2000; Bliese et al., 2002). Smaller groups tend to result in lower ICC(2). ICC(2) $> .70$ is generally considered acceptable (LeBreton & Senter, 2008).

All analyses were performed in R 4.1.3 (R Core Team, 2022). The packages *lavaan* (Rosseel, 2012) and *psych* (Revelle, 2021) were used for factor analysis. The analysis of within-group agreement and interrater reliability were conducted with the *multilevel* package (Bliese, 2016).

Results

Measurement invariance

The fit of the model for the first group representing the teachers and care teachers (CFI = .829, TLI = .823, RMSEA = .053, SRMR = .055) and the second group representing the coordinators, policy support staff, and principals (CFI = .789, TLI = .782, RMSEA = .064, SRMR = .069) is limited. Although the RMSEA and SRMR fall within the criteria, the CFI and TLI are well under the prescribed threshold for both groups.

Table 1. Fit indices of the multi-group confirmatory factor analysis distinguishing between operational staff (teachers and care teachers) and decision-making staff (coordinators, policy support staff, and principals).

Invariance level	CFI	TLI	SRMR	RMSEA	AIC	χ^2 (*df*)	*df*
Configural invariance	.820	.813	.057	.055	217212	19151	5794
Metric invariance	.820	.815	.058	.055	217179	19258	5864
	Δ .000	Δ .002	Δ .001	Δ .000		Δ 106.73***	
Scalar invariance	.818	.816	.059	.055	217234	19454	5934
	Δ .002	Δ .001	Δ .001	Δ .000		Δ 195.73***	
Cut-off	≥ .95	≥ .95	< .08	< .06			
	Δ −.01	Δ −.01	Δ .030	Δ .015			

Note: CFI = comparative fit index; TLI = Tucker–Lewis index; SRMR = standardised root-mean-square residual; RMSEA = root-mean-square error of approximation; AIC = Akaike information criterion; *df* = degrees of freedom; Δ = difference with previous model. Cut-offs according to Hu and Bentler (1999), Putnick and Bornstein (2016), and Cheung and Rensvold (2002).
***$p < 0.01$.

The fit indices of the MGCFA models are presented in Table 1. The RMSEA and SRMR of the unconstrained model (configural invariance) are below the threshold, but the CFI and TLI are too low. The χ^2 of the models differ significantly from each other. However, since we used a rather large sample, this is not very informative. Comparison of the fit indices of the configural, metric, and scalar models indicate metric invariance (ΔCFI < .001, ΔTLI = .002, ΔRMSEA < .001) and scalar invariance (ΔCFI = .002, ΔTLI = .001, ΔRMSEA = .001).

Model fit

The fit of the single-level model is limited. As shown in Table 2, the SRMR (.055) and RMSEA (.053) are well below the threshold, but the CFI (.834) and TLI (.828) are somewhat too low. The model with the same factor structure on both levels (Model 2) did not converge. This model with eight factors and 78 items at both levels is presumably too complex for the data. Therefore, no fit measures were obtained and estimates are unreliable.

Consistency and dimensionality of the scales

Table 3 shows Cronbach's alphas and dimensionality per scale. All Cronbach's alphas lie between .86 and .93. This is well above the prescribed .70. Except for responsive capacity, all scales are one-dimensional. Analysis of the factor loadings of the EFA and content analysis confirms this dichotomy in the responsive capacity scale. The first five questions relate explicitly to responsiveness towards parents, the following five items relate to responsiveness towards the wider society.

Table 2. Fit indices of the models.

Index	CFI	TLI	SRMR	RMSEA
Model 1 (Single level)	.834	.828	.055	.053
Model 2 (Multilevel)	NA	NA	NA	NA
Cut-off	≥ .95	≥ .95	< .08	< .06

Note: CFI = comparative fit index; TLI = Tucker–Lewis index; SRMR = standardised root-mean-square residual; RMSEA = root-mean-square error of approximation. Cut-offs according to Hu and Bentler (1999). Model 2 did not converge. Therefore, fit measures are not provided.

Table 3. Cronbach's alpha and dimensionality of the scales.

Scale	α	Dimensionality
Effective communication	.88	1
Shared leadership	.93	1
Integrated policy	.86	1
Shared vision	.91	1
Innovative capacity	.89	1
Supportive relations	.92	1
Reflective capacity	.91	1
Responsive capacity	.88	2

Table 4. AVE (diagonal), shared variance (below diagonal), and standardised covariances (r^2) (above diagonal) of the factors (Model 1).

	Com	Shl	Int	Vis	Inn	Rel	Ref	Res
Effective communication (Com)	**.49**	.85	.77	.74	.65	.74	.60	.51
Shared leadership (Shl)	.73	**.57**	.82	.76	.65	.68	.58	.55
Integrated policy (Int)	.60	.67	**.37**	.86	.81	.76	.75	.65
Shared vision (Vis)	.55	.58	.74	**.51**	.82	.70	.74	.64
Innovative capacity (Inn)	.42 [a]	.42 [a]	.65	.68	**.45**	.69	.80	.69
Supportive relations (Rel)	.54	.46 [a]	.57	.49 [a]	.48	**.54**	.71	.57
Reflective capacity (Ref)	.37 [a]	.34 [a]	.57	.54	.64	.51	**.49**	.67
Responsive capacity (Res)	.26 [a]	.30 [a]	.43	.41 [a]	.47	.32 [a]	.45	**.34**

Note: Below diagonal: average variance extracted (AVE); above diagonal: standardised covariances (r^2); diagonal and bold: Shared variance (SV).
[a] SV < AVE (for both factors).

Discriminant validity

The AVE, shared variance, and standardised correlation between all factors are presented in Table 4. There are six correlations ≥ .80. The strongest correlations are between Shared leadership and Integrated policy and between Integrated policy and Shared vision ($r^2 \geq .85$). Most AVEs are < .50, meaning that less than 50% of the variation can be attributed to the respective scales. Only for Shared leadership, Shared Vision, and Supportive relations, the AVE exceeds the threshold. In most cases, the shared variance exceeds the AVE, indicating rather poor discriminant validity between the factors. Shared leadership and responsive capacity can be fairly distinguished from four other factors. By contrast, Integrated policy cannot be distinguished from any of the other factors. Content analysis reveals that some of the items can be linked to other scales. For example, the item "the leadership informs the team about the policymaking activities" can be seen as part of Effective communication or Shared leadership.

Factor loadings

Table 5 shows the standardised factor loadings and residual variances for all items. All loadings are significant ($p < 0.001$). Most items load sufficiently on their scale (loading ≥ .60). All but one item of Shared leadership load ≥ .70, making this factor look most coherent. Two items from Effective communication have a factor loading < .60. Whereas most items of this scale focus on managing organisational communication in general, those items are somewhat more specific. One item focuses on professional development (" … communication skills are an important focus of professional development"), the

other on the availability of information (" … the necessary sources of information are available to carry out their own tasks properly"). The Integrated policy scale seems least coherent. Up to six out of 10 items have a load < .60. Content analysis reveals that the items refer to knowledge about the school organisation (Int1, Int3, Int7), commitment (Int2, Int5), vision (Int4), communication and shared leadership (Int6), interpersonal relations (Int7), and coordination (Int8, Int9, Int10).

Within-group agreement

Descriptive statistics of the $r^*_{wg(j)}$ of all scales are presented in Table 6. The within-group agreement of all scales is good. For all scales, the mean of the $r^*_{wg(j)}$ lies above or close to

Table 5. Factor loadings and residual variances (between brackets) ($N = 1{,}500$).

Item	Effective communication		Innovative capacity
Com1	.83 (.31)	Inn1	.68 (.54)
Com2	.83 (.31)	Inn2	.65 (.57)
Com3	.71 (.49)	Inn3	.64 (.60)
Com4	.81 (.35)	Inn4	.74 (.46)
Com5	.64 (.59)	Inn5	.61 (.63)
Com6	.65 (.58)	Inn6	.64 (.59)
Com7	.47 (.78)	Inn7	.74 (.45)
Com8	.52 (.73)	Inn8	.73 (.47)
		Inn9	.70 (.51)
		Inn10	.68 (.53)
	Shared leadership		Supportive relations
Shl1	.77 (.41)	Rel1	.79 (.37)
Shl2	.77 (.41)	Rel2	.61 (.63)
Shl3	.75 (.44)	Rel3	.82 (.34)
Shl4	.79 (.37)	Rel4	.80 (.37)
Shl5	.78 (.39)	Rel5	.74 (.45)
Shl6	.70 (.50)	Rel6	.83 (.31)
Shl7	.73 (.47)	Rel7	.61 (.63)
Shl8	.65 (.58)	Rel8	.67 (.55)
Shl9	.79 (.37)	Rel9	.74 (.45)
Shl10	.79 (.37)	Rel10	.76 (.43)
	Integrated policy		Reflective capacity
Int1	.59 (.66)	Ref1	.66 (.56)
Int2	.59 (.65)	Ref2	.73 (.47)
Int3	.55 (.70)	Ref3	.73 (.47)
Int4	.54 (.71)	Ref4	.75 (.44)
Int5	.62 (.61)	Ref5	.75 (.45)
Int6	.61 (.63)	Ref6	.65 (.58)
Int7	.70 (.51)	Ref7	.62 (.62)
Int8	.74 (.46)	Ref8	.67 (.56)
Int9	.59 (.65)	Ref9	.74 (.46)
Int10	.57 (.68)	Ref10	.76 (.42)
	Shared vision		Responsive capacity
Vis1	.70 (.52)	Res1	.60 (.64)
Vis2	.76 (.43)	Res2	.65 (.58)
Vis3	.69 (.53)	Res3	.63 (.61)
Vis4	.68 (.54)	Res4	.67 (.56)
Vis5	.65 (.58)	Res5	.67 (.55)
Vis6	.75 (.44)	Res6	.64 (.59)
Vis7	.71 (.50)	Res7	.67 (.55)
Vis8	.70 (.51)	Res8	.69 (.52)
Vis9	.74 (.45)	Res9	.72 (.48)
Vis10	.76 (.43)	Res10	.61 (.62)

Note: All loadings and residual variances are significant ($p < .001$).

Table 6. Descriptive statistics of the $r_{wg(j)}$, ICC(1) and ICC(2) of the scales.

	$r_{wg(j)}$				ICC(1)				ICC(2)			
Indicator	Min	Max	*M*	*SD*	Min	Max	*M*	*SD*	Min	Max	*M*	*SD*
Effective communication	.32	.87	.67	.12	.07	.24	.17	.06	.62	.88	.79	.10
Shared leadership	.49	.88	.70	.09	.14	.17	.16	.01	.79	.82	.80	.01
Integrated policy	.43	.88	.71	.09	.09	.16	.12	.02	.68	.80	.74	.04
Shared vision	.45	.90	.75	.09	.12	.23	.17	.03	.75	.87	.81	.04
Innovative capacity	.41	.85	.72	.09	.07	.21	.15	.04	.64	.85	.78	.06
Supportive relations	.08	.85	.67	.12	.09	.21	.15	.04	.69	.86	.78	.05
Reflective capacity	.37	.86	.71	.09	.09	.19	.14	.03	.69	.84	.77	.05
Responsive capacity	.47	.89	.78	.08	.08	.19	.11	.04	.64	.83	.71	.06

Note: ICC = intra-class correlation coefficient.

the predefined threshold of .70. However, the minima and maxima in Table 6 indicate that the within-group agreement differs from school to school. In some schools, there seems to be broad agreement ($r^*_{wg(j)} \geq .90$), in other schools, agreement on some scales is lacking (< .50). The mean $r^*_{wg(j)}$ for Supportive relations is strongly influenced by one school which scores extremely low (.08).

Interrater reliability

The mean ICC(1) of all scales is above .10 (see Table 6). This indicates that the school to which the teachers belong has a medium effect on their judgement and that there is a reasonable consensus between respondents from the same school. Since the ICC(1) of even the lowest scoring items is ≥ .05, we can state that group membership influences respondents' perceptions of their school's policymaking capacities.

For all items of Shared leadership and Shared vision, the ICC(2) exceeds the threshold of .70. The reliability of the group means for the items of the Responsive capacities scale is < .70 for six items. In the other scales, group means of one or two items are somehow less reliable.

Discussion

Given the ongoing trend towards more decentralisation and autonomy for schools (OECD, 2021), the importance of schools' policymaking capacities is gaining importance. In this study, data of 1,693 teachers, care teachers, coordinators, policy support staff, and principals from 77 Flemish primary schools were used to analyse critical aspects concerning validity and reliability of the PMC-Q. Therefore, measurement invariance, model fit, internal consistency and dimensionality of the scales, correlations between the scales and discriminant validity, factor loadings, within-group agreement, and interrater reliability were assessed. Results of a CFA indicate that the overall model fit is limited and that not all dimensions of a schools' policymaking capacities are equally well measured. However, we found some evidence that responses of staff with different positions can be pooled and that within-group agreement and interrater reliability is sufficient to aggregate responses of staff members to the school level.

The first question of this study was whether the eight-dimensional model is supported by the data. The CFA indicates limited model fit. Since a full two-level model did not

converge, we were not able to assess the construct validity of the PMC-Q at the school level. The two-level model turned out to be too complex to fit, based on the data. The internal consistency of all scales is good and, except for Responsive capacities, all scales are one-dimensional. Content analysis revealed that half of the items relate to responsiveness towards parents and the other half relate to responsiveness towards the wider society. This dichotomy is not surprising. In the first part of the original research in which the instrument was developed, Van Petegem et al. (2004) made a distinction between responsiveness towards parents and responsiveness towards the environment. Later, both scales were combined. The discriminant validity of the scales is rather weak. All factors correlate strongly with at least three other factors. Shared leadership and Responsive capacity can be fairly distinguished from four other scales. Integrated policy cannot be distinguished from any other scale. Except for Integrated policy and the two last items of Effective communication, the items load sufficiently on their scales, with Shared leadership being the most coherent of the scales.

Concerning the second research question, measurement invariance across groups was established, indicating that the factor structure of the PMC-Q is equivalent for staff members with different positions.

Finally, we investigated to what extent individual responses can reliably be aggregated to the school level. The results indicate that, on average, within-group agreement and interrater reliability are sufficient to make reliable statements about the school's policymaking capacities. However, there are differences between schools in the degree of agreement and interrater reliability between staff members.

Implications, limitations, and further research

In this study, we revealed some critical issues concerning validity and reliability of the PMC-Q. Our findings can help researchers who want to use the PMC-Q to assess schools' policymaking capacities in consolidating validity and reliability of their research. After all, validity is not a characteristic of an instrument but has to do with its use and the interpretation of the results (Kane, 2016). It is up to the researchers who want to use the instrument to verify the validity and reliability of their research and claims (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Knekta et al., 2019). At the same time, we argue that our findings can provide an impetus to further develop the PMC-Q. The results demonstrate that the eight-factor solution is probably not the best. Reducing the number of factors and the number of items has the potential of improving the instrument.

Besides using the PMC-Q for scientific purposes, school boards can use the instrument to strengthen their policymaking capacities through self-reflection or self-evaluation (Vanhoof et al., 2009). The instrument can be used quantitatively as well as qualitatively (Vanhoof & Van Petegem, 2017). In the latter case, the statements can be discussed by members of the school team. Caution is needed when school boards choose the quantitative approach to assess the policymaking capacities of their school. We found that schools differ in the extent to which staff members agree on the statements. However, assessing (dis)agreement between staff members can be informative as well.

The PMC-Q is a valuable instrument to assess policymaking capacities of schools. However, caution is needed when using the PMC-Q for research and practice. Our analysis

demonstrates that the eight dimensions are strongly intertwined and difficult to distinguish. One can question whether it is necessary to distinguish between eight dimensions of schools' policymaking capacities. By means of cross-validation, further research can explore possibilities to reduce the number of dimensions of the PMC-Q.

Our analysis showed that the PMC-Q fails to distinguish Integrated policy from the other dimensions. Factor loadings of most items are too low, and the interrater reliability is rather low. Integrated policy refers to the extent to which schools succeed in aligning different policy domains, starting from a vision. Content analysis of the items revealed that the items refer to knowledge about the school organisation, commitment, vision, communication, and shared leadership, interpersonal relations, and coordination. This operationalisation of Integrated policy makes the scale less coherent and results in strong correlations with other scales. In addition, assessing Integrated policy in a questionnaire assumes not only that respondents have knowledge of the different policy domains but also that they can assess the alignment of these policy domains. Faddar et al. (2017) demonstrated that teachers and principals have difficulty answering these questions, resulting in cognitively invalid answers (i.e., not in line with how the item is intended). This makes Integrated policy difficult to measure with a questionnaire. However, we advocate that Integrated policy is an important condition to implement an effective school policy (Vanhoof & Van Petegem, 2017). Although well studied in the domain of public policy (Trein et al., 2023), policy integration at the school level has hardly been investigated. More research is needed on how to operationalise and measure schools' Integrated policy in a valid and reliable way.

A limitation of our research is that we did not make use of other instruments to assess concurrent validity. It can be useful to compare the results of the PMC-Q with other validated instruments related to policymaking or the subscales, such as instruments on shared decision making, distributed leadership, professional (formal and informal) communication, or school culture. Investigating the correlation between the policymaking capacities of schools as measured by the PMC-Q and other outcome measures (e.g., achieving goals, students' academic achievement) can shed a light on the predictive validity of the instrument.

The study was conducted in Flemish primary schools. It can be relevant to implement the PMC-Q in secondary education or other educational systems to assess the external validity of the instrument. Given the peculiarities of the Flemish education system, in which schools enjoy a large degree of autonomy, the findings of this study might not be generalisable to all other education systems.

Finally, we were not able to fit a two-level model that enables us to assess the eight-factor structure on the school level. More data are needed to fit such a complex model. Alternatively, MLCFA on a reduced model can provide a solution here.

Assessing schools' policymaking capacities is an important step in evaluating and improving school effectiveness, especially given the current trend towards more school autonomy in many education systems. In this study, we evaluated the psychometric properties of the PMC-Q and made the instrument available for researchers and practitioners worldwide.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Joris Van Elsen is a PhD candidate at the Department of Training and Education Sciences of the University of Antwerp. His research focuses on Assessment for Learning and educational measurement.

Jerich Faddar is an assistant professor at the Department of Educational Sciences of the Vrije Universiteit Brussel. His research focuses on educational governance, educational innovations, and educational measurements. He teaches courses on educational effectiveness and curriculum development.

Lies Appels is a PhD candidate at the Department of Training and Education Sciences of the University of Antwerp.

Sven De Maeyer is a full professor at the Department of Training and Education Sciences of the University of Antwerp. He is a member of the EduBROn Research group. His research focuses on educational measurement and methodological issues in educational sciences.

Jan Vanhoof is a full professor at the Department of Training and Education Sciences of the University of Antwerp. He is a member of the EduBROn Research group. His current research activities focus on school policy and quality care in general and on school self-evaluation and data-driven school policy in particular.

Peter Van Petegem is a full professor at the Department of Training and Education Sciences of the University of Antwerp. He is a member of the EduBROn Research group. His research focuses on evaluation research, educational policy, international comparative research, educational innovation, and quality assurance.

ORCID

Joris Van Elsen http://orcid.org/0000-0002-9674-2927
Jerich Faddar http://orcid.org/0000-0001-5465-8615
Lies Appels http://orcid.org/0000-0001-9610-5509
Sven De Maeyer http://orcid.org/0000-0003-2888-1631
Jan Vanhoof http://orcid.org/0000-0001-8736-4509
Peter Van Petegem http://orcid.org/0000-0002-4078-7800

References

Alzayed, Z. A., & Alabdulkareem, R. H. (2021). Enhancing cognitive presence in teachers' professional learning communities via reflective practice. *Journal of Education for Teaching*, *47*(1), 18–31. https://doi.org/10.1080/02607476.2020.1842134

Amels, J., Krüger, M. L., Suhre, C. J., & van Veen, K. (2021). The relationship between primary school leaders' utilization of distributed leadership and teachers' capacity to change. *Educational Management Administration & Leadership*, *49*(5), 732–749. https://doi.org/10.1177/1741143220915921

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Biemann, T., Cole, M. S., & Voelpel, S. (2012). Within-group agreement: On the use (and misuse) of r_{WG} and $r_{WG(J)}$ in leadership research and some best practice guidelines. *The Leadership Quarterly*, *23*(1), 66–80. https://doi.org/10.1016/j.leaqua.2011.11.006

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). Jossey-Bass.

Bliese, P. D. (2016). *multilevel: Multilevel functions. R package* (Version 2.6) [Computer software]. https://CRAN.R-project.org/package=multilevel

Bliese, P. D., Halverson, R. R., & Schriesheim, C. A. (2002). Benchmarking multilevel methods in leadership: The articles, the model, and the data set. *The Leadership Quarterly*, *13*(1), 3–14. https://doi.org/10.1016/S1048-9843(01)00101-1

Brown, M., Altrichter, H., Shiyan, I., Rodríguez Conde, M. J., McNamara, G., Herzog-Punzenberger, B., Vorobyeva, I., Vangrando, V., Gardezi, S., O'Hara, J., Postlbauer, A., Milyaeva, D., Sergeevna, N., Fulterer, S., García, A. G., & Sánchez, L. (2022). Challenges and opportunities for culturally responsive leadership in schools: Evidence from four European countries. *Policy Futures in Education*, *20* (5), 580–607. https://doi.org/10.1177/14782103211040909

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.

Bulgaru, I. (2019). Managerial communication in school organisations. In E. Soare & C. Langa (Eds.), *European proceedings of social and behavioural sciences: Vol. 67*. *Education facing contemporary world issues* (pp. 2128–2134). Future Academy. https://doi.org/10.15405/epsbs.2019.08.03.264

Carr-Hill, R., Rolleston, C., & Schendel, R. (2016). The effects of school-based decision-making on educational outcomes in low- and middle-income contexts: A systematic review. *Campbell Systematic Reviews*, *12*(1), 1–169. https://doi.org/10.4073/csr.2016.9

Carter, S. R. (2016). Using confirmatory factor analysis to manage discriminant validity issues in social pharmacy research. *International Journal of Clinical Pharmacy*, *38*(3), 731–737. https://doi.org/10.1007/s11096-016-0302-9

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology*, *83*(2), 234–246. https://doi.org/10.1037/0021-9010.83.2.234

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Faddar, J., Vanhoof, J., & De Maeyer, S. (2017). Instruments for school self-evaluation: Lost in translation? A study on respondents' cognitive processing. *Educational Assessment, Evaluation and Accountability*, *29*(4), 397–420. https://doi.org/10.1007/s11092-017-9270-4

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/10.2307/3151312

Gumus, S., Bellibas, M. S., Esen, M., & Gumus, E. (2018). A systematic review of studies on leadership models in educational research from 1980 to 2014. *Educational Management Administration & Leadership*, *46*(1), 25–48. https://doi.org/10.1177/1741143216659296

Harris, A. (2013). Distributed leadership: Friend or foe? *Educational Management Administration & Leadership*, *41*(5), 545–554. https://doi.org/10.1177/1741143213497635

Hoy, W. K., & Miskel, C. G. (2012). *Educational administration: Theory, research, and practice* (9th ed.). McGraw-Hill.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, *69*(1), 85–98. https://doi.org/10.1037/0021-9010.69.1.85

Jongmans, C. T., Beijaard, D., & Biemans, H. J. A. (1998). Teachers' involvement in school policy-making and the effectiveness of schools' in-service training policy: Results of a Dutch study. *Teacher Development*, *2*(1), 59–73. https://doi.org/10.1080/13664539800200037

Kane, M. T. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 198–211. https://doi.org/10.1080/0969594X.2015.1060192

Knekta, E., Runyon, C., & Eddy, S. (2019). One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research. *CBE—Life Sciences Education*, *18*(1). https://doi.org/10.1187/cbe.18-04-0064

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods*, *9*(2), 202–220. https://doi.org/10.1177/1094428105284919

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*(4), 815–852. https://doi.org/10.1177/1094428106296642

Lewis, T. F. (2017). Evidence regarding the internal structure: Confirmatory factor analysis. *Measurement and Evaluation in Counseling and Development*, *50*(4), 239–247. https://doi.org/10.1080/07481756.2017.1336929

Lindell, M. K., Brandt, C. J., & Whitney, D. J. (1999). A revised index of interrater agreement for multi-item ratings of a single target. *Applied Psychological Measurement*, *23*(2), 127–135. https://doi.org/10.1177/01466219922031257

Luke, M., Goodrich, K. M., & Brammer, M. K. (2022). LGBTQI+ responsive school counseling: Exemplary school counselor educators' curricular integration. *Counselor Education and Supervision*, *61*(3), 230–246. https://doi.org/10.1002/ceas.12240

Martinez, J., & McAbee, S. (2020). School administrator support of teachers: A systematic review (2000-2019). *Education Leadership Review*, *21*(1), 230–254. https://www.icpel.org/ed-leadership-review.html

Morrison, T. G., Morrison, M. A., & McCutcheon, J. M. (2017). Best practice recommendations for using structural equation modelling in psychological research. *Psychology*, *8*(9), 1326–1341. https://doi.org/10.4236/psych.2017.89086

Newman, D. A., & Sin, H.-P. (2020). Within-group agreement (r_{WG}): Two theoretical parameters and their estimators. *Organizational Research Methods*, *23*(1), 30–64. https://doi.org/10.1177/1094428118809504

Nunkoo, R., Ramkissoon, H., & Gursoy, D. (2013). Use of structural equation modelling in tourism research: Past, present, and future. *Journal of Travel Research*, *52*(6), 759–771. https://doi.org/10.1177/0047287513478503

Organisation for Economic Co-operation and Development. (2018). *Education at a glance 2018: OECD Indicators*. https://doi.org/10.1787/eag-2018-en

Organisation for Economic Co-operation and Development. (2021). *Education at a glance 2021: OECD Indicators*. https://doi.org/10.1787/b35a14e5-en

Osborne, J. W. (2014). *Best practices in exploratory factor analysis*. Createspace publishing.

Pineda-Báez, C., Bauman, C., & Andrews, D. (2020). Empowering teacher leadership: A cross-country study. *International Journal of Leadership in Education*, *23*(4), 388–414. https://doi.org/10.1080/13603124.2018.1543804

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

R Core Team. (2022). *R: A language and environment for statistical computing* (Version 4.2.1) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Revelle, W. (2021). *psych: Procedures for personality and psychological research* (Version 2.1.6) [Computer software]. Northwestern University. https://CRAN.R-project.org/package=psych

Rosseel, Y. (2012). lavaan: An R package for structural equation modelling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Serdyukov, P. (2017). Innovation in education: What works, what doesn't, and what to do about it? *Journal of Research in Innovative Teaching & Learning*, *10*(1), 4–33. https://doi.org/10.1108/JRIT-10-2016-0007

Shen, J., Ma, X., Gao, X., Bierlien Palmer, L., Poppink, S., Burt, W., Leneway, R., McCrumb, D., Pearson, C., Rainey, M., Reeves, P., & Wegenke, G. (2019). Developing and validating an instrument measuring school leadership. *Educational Studies*, *45*(4), 402–421. https://doi.org/10.1080/03055698.2018.1446338

Sleegers, P., Bergen, T., & Giesbers, J. (1994). The policy-making capacity of schools: Results of a Dutch study. *Educational Management & Administration*, *22*(3), 147–159. https://doi.org/10.1177/0263211X9402200302

Stevenson, L., Honingh, M., & Neeleman, A. (2021). Dutch boards governing multiple schools: Navigating between autonomy and expectations. *School Leadership & Management*, *41*(4–5), 370–386. https://doi.org/10.1080/13632434.2021.1945024

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Trein, P., Fischer, M., Maggetti, M., & Sarti, F. (2023). Empirical research on policy integration: A review and new directions. *Policy Sciences*, *56*(1), 29–48. https://doi.org/10.1007/s11077-022-09489-9

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740

Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2016). The impact of school culture on schools' pupil well-being policy-making capacities. *Educational Studies*, *42*(4), 340–356. https://doi.org/10.1080/03055698.2016.1195718

Vanhoof, J., & Van Petegem, P. (2017). *Doeltreffend schoolbeleid: Praktijkboek beleidsvoerend vermogen in scholen* [Effective school policy: Practice book on policymaking capacity in schools]. Acco.

Vanhoof, J., Van Petegem, P., Verhoeven, J. C., & Buvens, I. (2009). Linking the policymaking capacities of schools and the quality of school self-evaluations: The view of school leaders. *Educational Management Administration & Leadership*, *37*(5), 667–686. https://doi.org/10.1177/1741143209339653

van Mierlo, H., Vermunt, J. K., & Rutte, C. G. (2009). Composing group-level constructs from individual-level survey data. *Organizational Research Methods*, *12*(2), 368–392. https://doi.org/10.1177/1094428107309322

Van Petegem, P., Mahieu, P., Kim, T. D., Devos, G., & Warmoes, V. (2006). *Beleidsvoerend vermogen van Vlaamse basis- en secundaire scholen: Onderwijskundig onderzoek in opdracht van de Vlaamse minister van Werk, Onderwijs en Vorming* [Policymaking capacity of Flemish primary and secondary schools: Educational research commissioned by the Flemish minister for Work, Education and Training]. Vlaams Ministerie van Onderwijs en Vorming. https://www.vlaanderen.be/publicaties/beleidsvoerend-vermogen-van-vlaamse-basis-en-secundaire-scholen

Van Petegem, P., Verhoeven, J. C., Buvens, I., & Vanhoof, J. (2004). *Zelfevaluatie en beleidseffectiviteit in Vlaamse scholen: Het gelijke onderwijskansenbeleid als casus* [Self-evaluation and policy effectiveness in Flemish schools: The equal educational opportunities policy as a case study]. Academia Press.

Visone, J. D. (2020). Teacher leadership for excellence in US National Blue Ribbon Schools. *International Journal of Leadership in Education*. Advance online publication. https://doi.org/10.1080/13603124.2020.1811897

York-Barr, J., Sommers, W. A., Ghere, G. S., & Montie, J. (2016). *Reflective practice for renewing schools: An action guide for educators* (3rd ed.). Corwin. https://doi.org/10.4135/9781506350530

Appendix 1. Items of the Policy Making Capacities Questionnaire (PMC-Q)

Item.	In this school …
Effective communication	
Com1	… communication is stimulating and supportive.
Com2	… people have good communication skills.
Com3	… information is made easily available.
Com4	… communication between the leadership and the team takes place in both directions.
Com5	… there is sufficient space for informal communication.
Com6	… there is someone responsible for monitoring good communication.
Com7	… communication skills are an important focus of professional development.
Com8	… the necessary sources of information are available to carry out their own tasks properly.
Com9	… *the internal consultative bodies are used in a constructive way.* [a]
Com10	… *people communicate openly about motives, ideas, concerns and uncertainties.* [a]
Shared leadership	
ShI1	… decision-making procedures are transparent.
ShI2	… there are sufficient opportunities to get involved in decision-making processes.
ShI3	… people encourage others to participate in decision-making processes.
ShI4	… all stakeholders are involved in making decisions.
ShI5	… the expertise present within the school is called upon when making decisions.
ShI6	… there is a willingness to take responsibility in policymaking.
ShI7	… one is involved in the evaluation of the school policy.
ShI8	… people are encouraged to take initiatives.
ShI9	… people consult with others in the school when making important decisions.
ShI10	… decisions are made taking into account the extent to which others in the school support the decision.
Integrated policy	
Int1	… the responsibilities of others in the school are understood.
Int2	… there is a strong commitment to teaching practice.
Int3	… there is a clear view of the job descriptions of others in the school.
Int4	… one has an ideal image of education that goes beyond one's own functioning.
Int5	… people think about issues without having an immediate interest in them.
Int6	… the leadership informs the team about the policymaking activities.
Int7	… one takes into account the activities, ambitions, and wishes of others in the school in their own actions.
Int8	… one believes in the added value of coordination.
Int9	… there are clear agreements regarding the responsibilities of existing working groups.
Int10	… different working groups work together in a complementary way.
Shared vision	
Vis1	… clear priorities are set.
Vis2	… one works towards a clear goal.
Vis3	… the vision is recorded in a policy plan.
Vis4	… an atmosphere of solidarity is created.
Vis5	… policymaking is based on an overarching vision.
Vis6	… the school's vision is evident in everyday activities.
Vis7	… there is mutual agreement on the aims to be achieved.
Vis8	… there is a consensus on the approach to work.
Vis9	… a long-term vision is promoted.
Vis10	… people make explicit what they think is important.
Innovative capacity	
Inn1	… there is a positive attitude towards innovation.
Inn2	… people are encouraged to experiment.
Inn3	… resistance to innovation is avoided.
Inn4	… the focus is on continuous improvement.
Inn5	… there is a wide dissemination of the insights gained from in-service training.
Inn6	… one dares to question the obvious.
Inn7	… people try to foster initiatives for innovations from within.
Inn8	… one examines whether innovations fit with the school vision.
Inn9	… new ways of working and approaches are tried out.
Inn10	… in the job descriptions, attention is paid to innovation.
Supportive relations	
Rel1	… there is a good relationship.
Rel2	… the leadership is experienced as supportive.
Rel3	… people are skilled at working in teams.
Rel4	… people trust each other.
Rel5	… the expertise of others is used well.

(*Continued*)

Appendix 1. Continued.

Item.	In this school …
Rel6	… one works together as a close team.
Rel7	… colleagues are coached when they are just starting out.
Rel8	… duplication of work is avoided through collaboration.
Rel9	… one uses each other's skills optimally.
Rel10	… people find support in each other.
Reflective capacity	
Ref1	… the leadership encourages regular self-reflection of the team.
Ref2	… people are convinced of the importance of reflection.
Ref3	… identifying areas for improvement is not seen as threatening.
Ref4	… there is a critical attitude towards one's own actions.
Ref5	… there is a positive attitude towards joint reflection.
Ref6	… one observes the work of others in order to learn from it.
Ref7	… one receives regular feedback about one's own work and functioning.
Ref8	… there are no problems with being evaluated.
Ref9	… one takes initiatives to collect data about one's own functioning.
Ref10	… one is prepared to question one's own functioning.
Responsive capacity	
Res1	… they actively seek the opinion of parents.
Res2	… the contacts with parents are organised in such a way that they are experienced as supportive.
Res3	… parents experience the contacts with the team as supportive.
Res4	… parents are encouraged to have their say.
Res5	… one is open to impulses from parents.
Res6	… external help is sought in the event of problems.
Res7	… one responds to developments in society.
Res8	… one tests one's own convictions against those of others.
Res9	… they are open to the input of external people.
Res10	… they respond to current events.

[a] Items that were not included in the questionnaire of the present study.