

RESEARCH NOTE

Peer Assessment in Parliament

Promises and Pitfalls of a Marginalised Method in Parliamentary Research

Richard Schobess*

Abstract

Peer assessment is a rather marginalised method in political research. This research note argues that the collective expertise of MPs can complement other data to contribute to more comprehensive evaluations of MPs' parliamentary work. Yet, this method is potentially flawed by low survey participation and rater bias among MPs. The experience with a peer assessment survey among members of three Belgian parliaments shows that participation does not necessarily need to be problematic. However, the empirical analysis suggests that scholars should control for various forms of rater bias.

Evaluating the work of colleagues who are active in the same field (peer review) has become a dominant method to judge the quality of scholars' academic work. Despite the general prominence of peer review in academia, the advantages of the use of peer evaluation to measure concepts for which data is otherwise hardly available remain almost entirely unexploited in political research. The lack of

interest in peer evaluation is particularly surprising given the recent surge in scholarly attention to individual MPs' parliamentary performance (e.g. Bouteca, Smulders, Maddens, Devos & Wauters, 2019; Bräuninger, Brunner & Däubler, 2012; Papp & Russo, 2018). This research note argues that peer evaluations among members of parliament (MPs) allow scholars to analyse MPs' parliamentary performance¹ not only with regard to their (formal) parliamentary activity but also with regard to less visible and more qualitative aspects of their parliamentary work.

The scarce use of peer assessments² among MPs by political scientists might be a sign either of lack of awareness or of scepticism towards the methodology in a parliamentary context. Doubts about the suitability of peer assessment in parliament might, for example, be nourished by concerns about participation in political elite surveys (Bailer, 2014) and the attention to rater bias in psychological and educational research (see e.g. Hoyt, 2000; Magin, 2001). The goal of this research note is therefore twofold. First of all, it strives to enhance scholars' familiarity with this unconven-

* Richard Schobess is a PhD candidate at the Department of Political Science of Ghent University. His research focuses on parliaments and elections.

tional (but promising) method for the field of legislative studies. Second, the recent experience with a peer assessment survey among members of three Belgian parliaments allows a more evidence-based debate about the potentials and pitfalls of the still rather marginalised method.

This research note first provides a short overview of existing approaches to evaluate MPs' parliamentary work, before discussing fundamental methodological choices regarding the design of peer assessment surveys as well as its eventual implications. Finally, MPs' survey participation and various empirically identified forms of rater bias among Belgian MPs are presented.

1 Review of Existing Approaches to Evaluate MPs' Parliamentary Work

Previous evaluations of MPs' parliamentary work can be divided into three categories based on the respective data source. The vast majority of recent studies on individual MPs' parliamentary performance relied on behavioural data from official parliamentary repositories analysing MPs' use of parliamentary tools such as parliamentary questions, legislative initiatives or the involvement in parliamentary debates (e.g. Bäck & Debus, 2016; Bowler, 2010; Bräuninger et al., 2012; Papp & Russo, 2018). The rather extensive data availability also enabled cross-country comparisons (e.g. Däubler, Christensen & Linek, 2018) as well as analyses over time (e.g. Wauters, Bouteca & de Vet, 2019). However, that approach typically restricts parliamentary performance to MPs' use of formal parliamentary tools, potentially

neglecting less visible aspects of their parliamentary work (Norton, 2018) or evaluations according to more qualitative criteria (Bouteca et al., 2019).³ Owing to the focus on publicly visible aspects of parliamentary work, data from official parliamentary repositories is particularly well-suited to studying the relationship between parliamentary activity and incumbents' chances of getting re-(s)elected.

A second (smaller) strand of the literature relied on direct evaluations by relevant stakeholders such as citizens (e.g. Sulkin, Testa & Usry, 2015), journalists (Bouteca et al., 2019; Sheaffer, 2001) or lobby organisations (Miquel & Snyder Jr, 2006).⁴ Relying on data from surveys and interviews notably allows these studies to include more qualitative evaluation criteria. Moreover, evaluations by important stakeholders incorporate the perspective of those actors whose judgments may be most relevant from a normative point of view. However, these external actors are often also unable to observe less visible aspects of MPs' parliamentary work behind the scenes and hence lack access to valuable information (and, potentially, the expertise to evaluate parliamentary work on specific policy issues). Including the (often more general) perceptions by external actors with regard to MPs' parliamentary work can be useful in complementing quantitative measures of parliamentary activity and advancing research that is related to trust and legitimacy.

A third approach to evaluate MPs' parliamentary work relied on the collective expertise of MPs themselves based on survey or interview data. Some of these studies incorporated some form of peer assessment (Fran-

cis, 1962; Humphreys & Weinstein, 2012; Miquel & Snyder Jr, 2006; Sheaffer, 2001).⁵ Owing to MPs' privileged access to information on less visible aspects of parliamentary work (such as the influence within parliamentary party group meetings) and their domain-specific expertise (to evaluate MPs' contributions in parliamentary committees), that approach enabled researchers to assess more diverse facets of MPs' parliamentary work. Integrating the expertise of MPs in measuring their (perceived) qualitative parliamentary performance may be particularly useful to complement quantitative measures of parliamentary activity or investigate topics such as individual MPs' legislative effectiveness in party-centred contexts.⁶ However, some scepticism may be warranted when relying on the perspectives of partisan actors. It might therefore come as a surprise that previous studies that involved peer evaluations among MPs neither reported nor analysed potential patterns of response bias or rater bias.⁷ Moreover, fundamental methodological choices for the survey design that might influence MPs' survey participation and possibilities to control for rater bias have so far remained rather undiscussed. To enable sound methodological choices for future peer evaluations among MPs, this gap will be filled on the basis of the experience with a recent peer assessment survey among Belgian MPs.

2 Methodological Choices for Peer Assessment Surveys Among MPs

Owing to the heterogeneity of previous peer evaluations in educational,

psychological or political research, scholars who are willing to employ peer assessment among MPs will face a variety of methodological choices. This section briefly discusses important questions concerning the design of peer assessment surveys among MPs and its potential implications.

A first choice for the development of a peer assessment survey concerns the content. Scholars may be interested in evaluations of more general or more specific aspects of MPs' parliamentary work. Although the expertise of MPs on particular aspects of parliamentary work might be the central motivation for this methodology, the inclusion of many specific survey questions also has disadvantages. On the one hand, raters may be unable to discriminate between similar questions without having sufficient information or precise evaluation criteria and, consequently, may rely on 'general impressions' of their colleagues instead (Thorndike, 1920).⁸ On the other hand, peer assessments with many specific questions necessarily increase the length of the survey, potentially resulting in lower participation rates. Similarly, this problem may undermine scholars' efforts to include several indicators (survey questions) per concept. While respondents may be unable to discriminate between almost identical evaluation criteria, the inclusion of two indicators per concept results in a doubled survey length. That question is also related to the number of peers every respondent is asked to evaluate. On the one hand, more evaluations per rater will provide more data per respondent and allow more precise identifications of eventual rater biases (Hoyt, 2000), while, on the other hand, longer lists of peers per respondent

ent may cause lower response rates owing to longer surveys with more monotonous survey experiences (long questions).

While often treated less explicitly in previous studies, the design of peer assessment surveys also requires a choice about which peers are evaluated. Respondents may be asked to evaluate a specified number of peers from the entire parliament or a random sample of a subgroup of MPs, e.g. from the same parliamentary party, parliamentary committee or electoral district. Instead of neglecting this question, the choice should be motivated by the content of the survey. If the primary focus is, e.g., on MPs' work within parliamentary committees, evaluations from MPs without any information about the committee work of some colleagues might be less valuable. Similarly, the empirical identification of raters' discrimination based on party characteristics requires that every rater be presented a list of peers with some balance between MPs from the same/different political party as herself.

Finally, evaluations can take different forms. Depending on the level of complexity deemed acceptable for eventual respondents, evaluations can range from rather simple procedures such as rank-ordering peers based on (electronic) picture cards to more precise estimations of MPs' parliamentary work based on ordinal or continuous scales. While simpler forms may facilitate faster responses and potentially higher response rates, more complex scales entail lower losses of information. However, the use of continuous scales may overestimate respondents' capability to provide infinitely detailed evaluations even though MPs are usu-

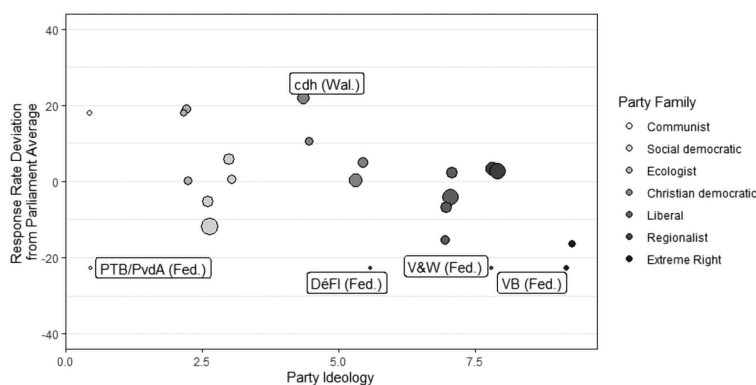
ally highly educated and possess detailed information about the parliamentary work of their (closest) peers.

As the preceding discussion shows, the design of peer assessment surveys is a constant trade-off between measures that may affect the survey participation as well as the ability of researchers to receive more precise measurements, e.g. by controlling for various forms of rater bias. While previous applications of peer assessments in parliamentary research failed to report consistently about survey participation or the control for raters' biases, this research note discusses both aspects for an exemplary peer assessment survey among Belgian MPs. The survey employed here (see Appendix) consisted of six peer assessment questions for each of twelve peers on an ordinal scale ranging from one to five.⁹ In view of our primary interest in assessing MPs' qualitative parliamentary performance within parliamentary committees and their party groups,¹⁰ the lists of peers consisted of 25% randomly sampled MPs from the same parliamentary party as the respondent and 75% randomly sampled MPs that are active in the same parliamentary committees.¹¹

3 Participation in MP Peer Assessment Surveys

This section briefly discusses the participation of MPs in the aforementioned peer assessment survey with a primary focus on the number of participants (response rate) and their representativeness for the population of invited MPs (response bias).

For the purpose of our study we invited 349 members of three Belgian



Note: Circle sizes represent the seat shares of parliamentary party groups. Parliamentary parties with less than two MPs were excluded for ease of interpretation.

Fed. = Federal Parliament; Wal. = Parliament of Wallonia.

Figure 1 Response rates of parliamentary party groups relative to average response rates per parliament

parliaments to participate in an online survey between January and March 2019 at the end of the parliamentary term preceding the general elections on 26 May 2019.¹² Personal invitations for the peer assessment survey were sent by email, outlining the general objectives of the survey and the purely academic purpose. Furthermore, the invitation assured strict confidentiality of all responses as well as full anonymisation of the results before providing a link to the individual survey version. All in all, the peer assessment survey had a response rate of 28.3% and provided 6576 evaluations covering 93.1% of our population of Belgian MPs. Since the response rate is comparable to those of other MP surveys in Europe (Bailer, 2014), the level of participation is rather acceptable – certainly when considering the sensitive topic (evaluating peers) and the limited familiarity with the methodology in European parliaments.

However, the main problem may not be unit non-response but rather response bias given that frontbenchers and MPs from larger parties are typically less likely to participate in MP surveys (Bailer, 2014). Parliamentary parties' deviating response rates from the parliamentary average indicate indeed that the survey participation might not have been completely at random (see Figure 1). Nevertheless, there seem to be no obvious participatory patterns pertaining to the seat share of parliamentary parties or party ideology. The figure may point, however, to slightly lower response rates of MPs from more ideologically extreme parties.

In order to examine the participatory patterns more closely, the survey participation of MPs has been analysed empirically. Table 1 presents the results for three probit models with a dichotomous dependent variable (sur-

vey participation = yes/no) and several potential explanatory factors that might be associated with MPs' participation in the peer assessment survey.¹³ As the results for Model 1 show, none of the variables are significantly associated with MPs' survey participation (with the potential exception that MPs from regional parliaments might be slightly more likely to participate, $p < 0.1$). Furthermore, MPs from ideologically more extreme parties did not participate significantly less often than more moderate MPs (Model 2). Finally, the results show that MPs who are generally more active in parliament were significantly more likely to participate in the peer assessment survey. In contrast, those MPs who were characterised by more qualitative parliamen-

tary work (rated by their peers) did not participate significantly more often.¹⁴

While MPs' participation in the peer assessment survey might be rather independent of party characteristics, these findings indicate that more parliamentary active MPs might be overrepresented among survey respondents. Such a self-selection mechanism could eventually result in respondents assigning more importance to quantitative aspects of parliamentary work (potentially inflating the correlation between measures of quantity and quality of parliamentary work). Additionally, several spontaneous reactions from invited MPs showed that lack of time is a repeatedly mentioned reason for non-participation, underlining the importance of short peer assessment surveys.

4 Rater Bias: Patterns of Systematically Deviating Evaluations

Although MPs' participation in peer assessment surveys may be a common cause of concern, scholars might be even more sceptical about whether MPs will actually assign honest evaluations. To facilitate a more evidence-based discussion on whether this scepticism is warranted, this section presents a brief overview of empirically identified forms of rater bias for respondents of a peer assessment survey among Belgian MPs (see foregoing discussion).

In the absence of valid alternative measures for the six indicators of qualitative parliamentary work employed here, rater biases have been identified on the basis of systematic patterns

Table 1 *Peer Assessment Survey Participation: Probit Models with Individual MPs' Survey Participation as Dependent Variable (yes = 1, no = 0) to Identify Systematic Forms of Response Bias.*

	Dependent Variable:		
	Survey Participation		
	(Model 1)	(Model 2)	(Model 3)
PPG Size	-0.003 (0.01)	-0.003 (0.01)	0.003 (0.01)
Opposition	-0.34 (0.24)	-0.33 (0.29)	-0.04 (0.36)
Ideology	-0.07 (0.05)	-0.07 (0.06)	0.02 (0.07)
Regional	0.30 (0.15)	0.30 (0.15)	0.30 (0.17)
Frontbencher	-0.03 (0.20)	-0.03 (0.20)	-0.04 (0.22)
Female	0.001 (0.15)	-0.0005 (0.15)	0.06 (0.16)
Dutch	0.08 (0.16)	0.08 (0.16)	0.11 (0.17)
Ideology ²		-0.002 (0.02)	-0.04 (0.03)
Activity			0.46** (0.16)
Quality			0.01 (0.27)
Constant	-0.55* (0.26)	-0.55* (0.26)	-0.80** (0.29)
Observations	349	349	325
Log Likelihood	-204.71	-204.71	-180.51
AIC	425.42	427.41	383.01

Note: *p < 0.05 **p < 0.01 ***p < 0.001

among evaluations with various dyadic characteristics between raters (respondents) and targets (their evaluated peers). Taking previous findings in educational settings and the particularities of the parliamentary context into account, we tested for (dyadic) rater bias deriving from characteristics of MPs' parliamentary parties, institutional factors or individual characteristics.¹⁵ Systematic deviations have been identified with a Bayesian ordered probit varying-intercepts, varying-slopes model.¹⁶ The results show that MPs were generally more likely to assign higher scores to members of their own parliamentary party as well as to those MPs with higher political positions than their own (see Table 2).¹⁷

Importantly, the impact of MPs' rater biases can be quite substantial. Figure 2 summarises several important findings pertaining to select forms of rater bias. First of all, the black elements of the figure show the expected *average* difference between peer evaluations resulting exclusively from both MPs belonging to the same/different parliamentary party (above) or the same/different gender (below). While raters can be expected to have a 99.2% probability of assigning an above-medium score for MPs of their own parliamentary party, this probability drops to only 8.6% for MPs from other parliamentary parties.¹⁸ In contrast, there appears to be no *general* gender effect for MPs' peer evaluations. However, a second important finding is the sub-

Table 2 *Rater Bias in Peer Assessment: Bayesian Multilevel Ordered Probit Model (Varying Intercepts and Varying Slopes) with Peer Assessment Scores as Dependent Variable (Ordinal Scale from One to Five).*

	Dependent Variable:		
	Peer Evaluation		
	5%	50%	95%
Same Party	0.28	0.5	0.72
Same Coalition	-0.05	0.13	0.32
Ideol. Distance	-0.13	-0.05	0.03
Hierarchy	0.11	0.23	0.36
Same Gender	-0.08	0.02	0.12
Same Language	-0.1	0.12	0.33
Question 2	-0.37	-0.25	-0.13
Question 3	-0.52	-0.41	-0.31
Question 4	0.29	0.46	0.63
Question 5	-0.33	-0.19	-0.06
Question 6	0	0.1	0.2
Constant	1.81	2.15	2.48
Observations		6576	
Groups (Raters)		99	

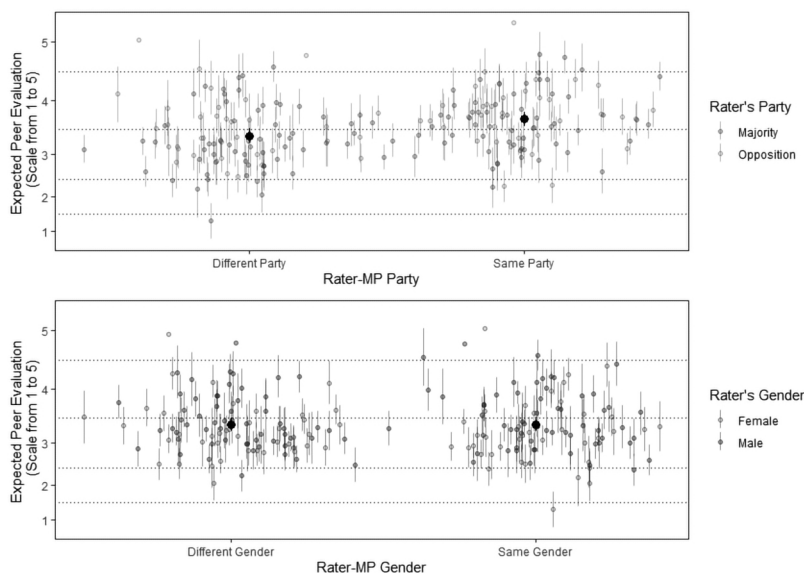
Note: Coefficients' percentiles of the posterior distribution shown (the same sign of a coefficient in all three columns indicates a 95 percent posterior probability that the coefficients is positive/negative). Threshold estimates and variance terms not reported here.

stantial difference between individual raters (grey elements in Figure 2). In fact, individual raters tend to assign *generally* higher/lower evaluations (rater severity) captured by varying intercepts (left part of Figure 2). Furthermore, individual raters may also differ in their *strength* of various forms of dyadic rater bias (varying slopes, right part of Figure 2). While MPs' same party bias might apply to almost all raters independent of, e.g., government party status, same gender bias may be observed only for some respondents but not for others. As such, same gender bias is more pronounced among male respondents but is much more limited among female

respondents.¹⁹ Taken together, these findings highlight the importance of considerable efforts into the control for various forms of rater bias when relying on peer assessments among MPs.

5 Conclusion

This research note argues that evaluations of individual MPs' parliamentary work based on the collective expertise of MPs can enrich political analyses by complementing other sources of data. While the growing literature on (aspects of) MPs' parliamentary performance relies largely on publicly



Note: Expected average effects (black) and individual effects for 99 raters (grey) based on a Bayesian multilevel ordered probit model. Posterior medians and 90% confidence intervals shown.

Figure 2 *Expected peer assessment scores for MPs of the same/different parliamentary party as the rater (above) and the same/different gender as the rater (below)*

available data on MPs' use of formal parliamentary tools, these studies may neglect other important details of the work inside parliaments. Notably, that approach often leaves differences *between* individual parliamentary questions or legislative initiatives unattended. Yet one single parliamentary question revealing a major government scandal may outweigh 100 questions simply reiterating publicly available statistics in many respects. Moreover, the exclusive focus on formal parliamentary tools largely disregards MPs' parliamentary work behind closed doors such as their activities aimed at representing voters within their parliamentary party group or seeking support for legislative initiatives in the informal space. Therefore,

peer assessment among MPs provides a promising approach to complement parliamentary activity data with more qualitative aspects of MPs' parliamentary work, thereby also taking activities in less visible areas (such as parliamentary party groups) into account.

However, MPs' survey participation and bias among raters are potential pitfalls that might discourage scholars from employing this method. The experience with a peer assessment survey among members of three Belgian parliaments shows that participation does not necessarily need to be problematic aside from the eventual over-representation of more parliamentary active MPs. However, the empirical identification of systematically deviating evaluations suggests

that future applications of this method should be careful to control for theoretically expected forms of rater bias.²⁰ In the Belgian context, characterised by hierarchically organised parliaments, strong political parties and a linguistic divide, scholars may need to control for potential sources of dyadic rater bias that are based on MPs' party characteristics, linguistic groups and hierarchical relations between MPs in addition to personal characteristics such as gender. Only when potential pitfalls such as low/unbalanced participation and rater bias are taken into account may scholars fully benefit from the advantages of peer assessment among MPs to complement other data on MPs' parliamentary performance, allowing them to investigate new research questions.

This research note facilitates a discussion about potential risks and benefits of peer assessment in parliament. While this study is only a first step towards a more evidence-based debate, we strongly encourage other scholars to report systematically about methodological choices as well as about participation and rater bias in peer assessment surveys among MPs.

Notes

- 1 For a conceptual discussion and normative concerns see (Schobess, 2021).
- 2 Although 'peer evaluation' and 'peer assessment' are often used interchangeably, this research note considers peer assessment as a subtype of peer evaluation methods that strives to collect quantitative data.
- 3 Notable exceptions are, e.g., Martin (2011) and Solvak (2013), allowing the inclusion of specific qualitative evaluation criteria for selected formal parliamentary activities.
- 4 Some of these studies actually relied on a combination of several types of actors.
- 5 Other studies relied on MPs' self-reported activities (e.g. Deschouwer, Depauw & André, 2014).
- 6 That is because previous measures of legislative effectiveness based on bill passage (Volden & Wiseman, 2014) are considerably flawed under very high levels of party unity.
- 7 A notable exception is the study of Humphreys and Weinstein (2012) that mean-standardised raw peer assessment scores for MPs from majority vs. opposition parties.
- 8 This effect is also called halo error.
- 9 Previous approaches in parliamentary research ranged from one to six questions (concepts) and a list of peers to be evaluated per respondent ranging from 15 to all MPs in parliament. Furthermore, these studies did not specify subgroups of (more closely related) MPs and made use of rank-ordering or ordinal scales (see Francis, 1962; Humphreys & Weinstein, 2012; Miquel & Snyder Jr, 2006; Sheaffer, 2001).
- 10 For a general description and the precise survey questions see (Schobess, 2021).
- 11 In practice, this approach may require additional steps of random sampling for exceptional cases such as MPs from political parties with fewer than three MPs.
- 12 The following MPs have been invited: all members of the Belgian Chamber of Representatives (150 MPs) as well as two regional parliaments: Flemish Parliament (124) and the Parliament of Wallonia (75). The lists of MPs have been created in October 2018 before

- the expected reshuffle following the 2018 local elections.
- 13 In addition to expected lower participation rates in MP surveys for front-benchers and MPs from larger parties, it has been tested for potential effects of MPs' government party status, party ideology (general left-right ideology, see Polk et al., 2017), squared party ideology (extremism), type of parliament (regional vs. federal), gender and language.
 - 14 The measure of parliamentary activity included MPs' use of six parliamentary tools comprising parliamentary speeches, parliamentary questions and legislative initiatives. For a more detailed description of the measures of parliamentary activity and quality of parliamentary work see Schobess (2021).
 - 15 The peer assessment literature identified various forms of rater bias in educational contexts: bias for members of the same group, dominant members and based on friendship (Pond & ul-Haq, 1997; Strijbos, Ochoa, Sluijsmans, Segers & Tillema, 2009).
 - 16 For more details about the empirical approach see Schobess (2021).
 - 17 The presidents of parliamentary party groups, parliaments and political parties have been counted as holding higher-level positions.
 - 18 All calculations are based on posterior probabilities for predictions, with all other explanatory variables held constant at their median.
 - 19 Paired t-tests for male and female respondents' predicted same gender bias (posterior medians for same gender vs. different gender evaluations) show a positive effect for male respondents ($p = 0.059$) but not for female respondents (negative sign, $p = 0.32$).
 - 20 Based on the peer assessment literature and specific characteristics of the selected case and the respective research objective.

References

- Bäck, H. & Debus, M. (2016). *Political Parties, Parliaments and Legislative Speech-making*. Springer.
- Bailer, S. (2014). Interviews and surveys in legislative research. In S. Martin, T. Saalfeld & K. Strøm (Eds.), *The Oxford Handbook of Legislative Studies* (pp. 167-193). Oxford University Press.
- Bouteca, N., Smulders, J., Maddens, B., Devos, C. & Wauters, B. (2019). 'A Fair Day's Wage for a Fair Day's Work'? Exploring the Connection between the Parliamentary Work of MPs and Their Electoral Support. *The Journal of Legislative Studies*, 25(1), 44-65. doi: 10.1080/13572334.2019.1570602.
- Bowler, S. (2010). Private Members' Bills in the UK Parliament: Is There An 'Electoral Connection'? *The Journal of Legislative Studies*, 16(4), 476-494.
- Bräuninger, T., Brunner, M. & Däubler, T. (2012). Personal Vote-Seeking in Flexible List Systems: How Electoral Incentives Shape Belgian MPs' Bill Initiation Behaviour. *European Journal of Political Research*, 51(5), 607-645. doi: 10.1111/j.1475-6765.2011.02047.x.
- Däubler, T., Christensen, L. & Linek, L. (2018). Parliamentary Activity, Re-Selection and the Personal Vote. Evidence from Flexible-List Systems. *Parliamentary Affairs*, 71(4), 930-949. doi: 10.1093/pa/gsx048.
- Deschouwer, K., Depauw, S. & André, A. (2014). Representing the people in parliaments. *Representing the People. A Survey Among Members of Statewide and Substate Parliaments* (pp. 1-18). Oxford University Press.
- Francis, W. L. (1962). Influence and Interaction in a State Legislative Body.

- American Political Science Review*, 56(4), 953-960.
- Hoyt, W. T. (2000). Rater Bias in Psychological Research: When Is It a Problem and What Can We Do About It? *Psychological methods*, 5(1), 64.
- Humphreys, M. & Weinstein, J. (2012). *Policing Politicians: Citizen Empowerment and Political Accountability in Uganda*. Unpublished manuscript. Working Paper. International Growth Centre.
- Magin, D. (2001). Reciprocity as a Source of Bias in Multiple Peer Assessment of Group Work. *Studies in Higher Education*, 26(1), 53-63. doi: 10.1080/03075070020030715.
- Martin, S. (2011). Using Parliamentary Questions to Measure Constituency Focus: An Application to the Irish Case. *Political Studies*, 59(2), 472-488. doi: 10.1111/j.1467-9248.2011.00885.x.
- Miquel, G. P. I. & Snyder Jr, J. M. (2006). Legislative Effectiveness and Legislative Careers. *Legislative Studies Quarterly*, 31(3), 347-381.
- Norton, P. (2018). Power Behind the Scenes: The Importance of Informal Space in Legislatures1. *Parliamentary Affairs*, 72(2), 245-266. doi: 10.1093/pa/gsy018.
- Papp, Z. & Russo, F. (2018). Parliamentary Work, Re-Selection and Re-Election: In Search of the Accountability Link. *Parliamentary Affairs*, 71(4), 853-867. doi: 10.1093/pa/gsx047.
- Polk, J., Rovny, J., Bakker, R., Edwards, E., Hooghe, L., Jolly, S., . . . Schumacher, G. (2017). Explaining the salience of anti-elitism and reducing political corruption for political parties in Europe with the 2014 Chapel Hill Expert Survey data. *Research & Politics*, 4(1), 2053168016686915.
- Pond, K. & ul-Haq, R. (1997). Learning to Assess students Using Peer Review. *Studies in Educational Evaluation*, 23(4), 331-348. doi: 10.1016/S0191-491X(97)86214-1.
- Schobess, R. (2021). Behind the Scenes: What is Parliamentary Performance and How Can We Measure It? *Parliamentary Affairs*. doi: 10.1093/pa/gsab024.
- Sheafer, T. (2001). Charismatic Skill and Media Legitimacy: An Actor-Centered Approach to Understanding the Political Communication Competition. *Communication Research*, 28(6), 711-736. doi: 10.1177/009365001028006001.
- Solvak, M. (2013). Private Members' Bills and the Personal Vote: Neither Selling nor Shaving. *The Journal of Legislative Studies*, 19(1), 42-59. doi: 10.1080/13572334.2013.736786.
- Strijbos, J.-W., Ochoa, T. A., Sluijsmans, D. M., Segers, M. S. & Tillema, H. H. (2009). Fostering Interactivity Through Formative Peer Assessment in (Web-Based) Collaborative Learning Environments. In C. Mourlas, N. Tsianos & P. Germanakos (Eds.), *Cognitive and Emotional Processes in Web-Based Education: Integrating Human Factors and Personalization* (pp. 375-395). IGI Global.
- Sulkin, T., Testa, P. & Usry, K. (2015). What Gets Rewarded? Legislative Activity and Constituency Approval. *Political Research Quarterly*, 68(4), 690-702. doi: 10.1177/1065912915608699.
- Thorndike, E. L. (1920). A Constant Error in Psychological Ratings. *Journal of Applied Psychology*, 4(1), 25-29.
- Volden, C. & Wiseman, A. E. (2014). *Legislative Effectiveness in the United States Congress: The Lawmakers*. Cambridge University Press.
- Wauters, B., Bouteca, N. & de Vet, B. (2019). Personalization of Parliamentary Behaviour: Conceptualization and Empirical Evidence from Belgium (1995-2014). *Party Politics*, doi: 10.1177/1354068819855713.

Appendix

Tabel A1 *Peer assessment survey questionnaire for the operationalisation of qualitative aspects of individual MPs' parliamentary performance*

Aspect of Parliamentary Performance	Peer Assessment Survey Statement (Disagree/Agree, Five-Point Scale)
Representation Quality	He/she is very loyal towards his/her voters (e.g. he/she keeps his/her electoral promises).
Legislative Quality	He/she is very competent in developing legislative initiatives to solve current problems in society.
Control Quality	Controlling the government with his/her parliamentary work, he/she focuses on relevant problems in society (instead of insignificant questions).
Representation Effectiveness	In comparison with other MPs, he/she is very successful in representing the interests of his/her voters, attracting attention to topics that are important to them.
Legislative Effectiveness	He/she is very successful in building support among other MPs for his/her legislative initiatives.
Control Effectiveness	In comparison with other MPs, he/she has more policy impact with his/her parliamentary control work (parliamentary questions, committee work, budgetary control).

Note: Statements presented to MPs (disagree/agree, five-point scale) with regard to the parliamentary work of colleagues during the current legislative term. Source: (Schobess, 2021)