

**A META-ANALYSIS OF TEACHER AND SCHOOL EVALUATION STUDIES: A STEP FOR PROPOSING EFFECTIVE TEACHER AND SCHOOL EVALUATION MECHANISMS**

**List of Papers Included in the Meta-Analysis**

**A**

- Adnot, M., Dee, T., Katz, V., & Wyckoff, J. (2017). Teacher turnover, teacher quality, and student achievement in DCPS. *Educational Evaluation and Policy Analysis*, 39(1), 54-76. <https://doi.org/10.3102/0162373716663646>
- Américo, B. L., & Lacruz, A. J. (2017). The school environment and its performance: analysis of the scores obtained by schools from the state of Espírito Santo in the nationwide exam Prova Brasil, using multiple linear regression. *Revista de Administração Pública*, 51(5), 854-878. <https://doi.org/10.1590/0034-7612160483>
- Amrein-Beardsley, A., & Holloway, J. (2019). Value-added models for teacher evaluation and accountability: Commonsense assumptions. *Educational Policy*, 33(3), 516-542. <https://doi.org/10.1177/0895904817719519>
- Amrein-Beardsley, A., Polasky, S., & Holloway-Libell, J. (2016). Validating “value added” in the primary grades: one district’s attempts to increase fairness and inclusivity in its teacher evaluation system. *Educational Assessment, Evaluation and Accountability*, 28(2), 139-159. <https://doi.org/10.1007/s11092-015-9234-5>
- Arar, K., & Oplatka, I. (2011). Perceptions and applications of teachers’ evaluation among elementary school principals in the Arab education system in Israel. *Studies in Educational Evaluation*, 37(2-3), 162-169. <https://doi.org/10.1016/j.stueduc.2011.03.006>

Attinello, J. R., Lare, D., & Waters, F. (2006). The value of teacher portfolios for evaluation and professional growth. *NASSP Bulletin*, 90(2), 132-152.

<https://doi.org/10.1177/0192636506288864>

## **B**

Bacher-Hicks, A., Chin, M. J., Kane, T. J., & Staiger, D. O. (2019). An experimental evaluation of three teacher quality measures: Value-added, classroom observations, and student surveys. *Economics of Education Review*, 73, 101919.

<https://doi.org/10.1016/j.econedurev.2019.101919>

Backes, B., Cowan, J., Goldhaber, D., Koedel, C., Miller, L. C., & Xu, Z. (2018). The common core conundrum: To what extent should we worry that changes to assessments will affect test-based measures of teacher performance?. *Economics of Education Review*, 62, 48-65. <https://doi.org/10.1016/j.econedurev.2017.10.004>

Balch, R., & Koedel, C. (2014). Anticipating and incorporating stakeholder feedback when developing value-added models. *Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas*, 22, 1-14. <https://doi.org/10.14507/epaa.v22.1701>

Barile, J. P., Donohue, D. K., Anthony, E. R., Baker, A. M., Weaver, S. R., & Henrich, C. C. (2012). Teacher–student relationship climate and school outcomes: Implications for educational policy initiatives. *Journal of Youth and Adolescence*, 41(3), 256-267.

<https://doi.org/10.1007/s10964-011-9652-8>

Bastian, K. C., Patterson, K. M., & Pan, Y. (2018). Evaluating teacher preparation programs with teacher evaluation ratings: Implications for program accountability and improvement. *Journal of Teacher Education*, 69(5), 429-447.

<https://doi.org/10.1177/0022487117718182>

Berberick, D. M., Clementson, C. J., Hawkinson, J. K., & Rolandson, D. M. (2016). A comparison of principal practices and music educator perceptions regarding teacher evaluation. *Bulletin of the Council for Research in Music Education*, (209), 43-62.

<https://doi.org/10.5406/bulcouresmusedu.209.0043>

- Bonda, N. (2018). An examination of one factor in school reform: increased teacher accountability in two urban school districts. *Revista Tempos E Espaços Em Educação*, 11(27), 329-344. <https://doi.org/10.20952/revtee.v11i27.9827>
- Borg, S., & Edmett, A. (2019). Developing a self-assessment tool for English language teachers. *Language Teaching Research*, 23(5), 655-679. <https://doi.org/10.1177/1362168817752543>
- Borman, G. D., & Kimball, S. M. (2005). Teacher quality and educational equality: Do teachers with higher standards-based evaluation ratings close student achievement gaps?. *The Elementary School Journal*, 106(1), 3-20. <https://doi.org/10.1086/496904>
- Bouchamma, Y. (2005). Evaluating teaching personnel. Which model of supervision do Canadian teachers prefer?. *Journal of Personnel Evaluation in Education*, 18(4), 289-308. <https://doi.org/10.1007/s11092-007-9025-8>
- Bradford, C., & Braaten, M. (2018). Teacher evaluation and the demoralization of teachers. *Teaching and Teacher Education*, 75, 49-59. <https://doi.org/10.1016/j.tate.2018.05.017>
- Bradshaw, L. K. (2002). Local district implementation of state mandated teacher evaluation policies and procedures: The North Carolina case. *Journal of Personnel Evaluation in Education*, 16(2), 113-127. <https://doi.org/10.1023/A:1020000328622>
- Brown, M., McNamara, G., Ohara, J., O'Brien, S., Young, C., & Faddar, J. (2018). Integrated co-professional evaluation?: Converging approaches to school evaluation across frontiers. *Australian Journal of Teacher Education (Online)*, 43(12), 76-90. <https://doi.org/10.14221/ajte.2018v43n12.6>

## C

- Cai, Y., & Lin, C. (2006). Theory and practice on teacher performance evaluation. *Frontiers of Education in China*, 1(1), 29-39. <https://doi.org/10.1007/s11516-005-0004-x>

- Chang, T. C., & Wang, H. (2016). A multi criteria group decision-making model for teacher evaluation in higher education based on cloud model and decision tree. *Eurasia Journal of Mathematics, Science and Technology Education, 12*(5), 1243-1262.  
<https://doi.org/10.12973/eurasia.2016.1510a>
- Cohen, J., Grossman, P., Chambers Schuldt, L., & Brown, L. (2016). Leveraging observation tools for instructional improvement: Exploring variability in uptake of ambitious instructional practices. *Teachers College Record, 118*(11), 1-36.  
<https://doi.org/10.1177/016146811611801105>
- Cohen, J., Loeb, S., Miller, L. C., & Wyckoff, J. H. (2020). Policy implementation, principal agency, and strategic action: Improving teaching effectiveness in New York City middle schools. *Educational Evaluation and Policy Analysis, 42*(1), 134-160.  
<https://doi.org/10.3102/0162373719893338>
- Cohen, J., Ruzek, E., & Sandilos, L. (2018). Does teaching quality cross subjects? Exploring consistency in elementary teacher practice across subjects. *AERA Open, 4*(3), 2332858418794492. <https://doi.org/10.1177/2332858418794492>
- Collins, C. (2014). Houston, We Have a Problem: Teachers Find No Value in the SAS Education Value-Added Assessment System (EVAAS®). *Education Policy Analysis Archives, 22*(98), n98. <https://doi.org/10.14507/epaa.v22.1594>
- Conley, S., Smith, J. L., Collinson, V., & Palazuelos, A. (2016). A small step into the complexity of teacher evaluation as professional development. *Professional Development in Education, 42*(1), 168-170. <https://doi.org/10.1080/19415257.2014.923926>
- Cowan, J., & Goldhaber, D. (2016). National board certification and teacher effectiveness: Evidence from Washington State. *Journal of Research on Educational Effectiveness, 9*(3), 233-258. <https://doi.org/10.1080/19345747.2015.1099768>
- Cruz-Aguayo, Y., Ibararán, P., & Schady, N. (2017). Do tests applied to teachers predict their effectiveness?. *Economics Letters, 159*, 108-111.  
<https://doi.org/10.1016/j.econlet.2017.06.035>

## D

- Daividsdottir, S., & Lisi, P. (2007). Effects of deliberative democracy on school self-evaluation. *Evaluation*, 13(3), 371-386. <https://doi.org/10.1177/1356389007078628>
- Delvaux, E., Vanhoof, J., Tuytens, M., Vekeman, E., Devos, G., & Van Petegem, P. (2013). How may teacher evaluation have an impact on professional development? A multilevel analysis. *Teaching and Teacher Education*, 36, 1-11. <https://doi.org/10.1016/j.tate.2013.06.011>
- Derrington, M. L., & Campbell, J. W. (2015). Implementing new teacher evaluation systems: Principals' concerns and supervisor support. *Journal of Educational Change*, 16(3), 305-326. <https://doi.org/10.1007/s10833-015-9244-6>
- Derrington, M. L., & Campbell, J. W. (2018). High-stakes teacher evaluation policy: US principals' perspectives and variations in practice. *Teachers and Teaching*, 24(3), 246-262. <https://doi.org/10.1080/13540602.2017.1421164>
- Derrington, M. L., & Campbell, J. W. (2018). Teacher evaluation policy tools: Principals' selective use in instructional leadership. *Leadership and Policy in Schools*, 17(4), 568-590. <https://doi.org/10.1080/15700763.2017.1326143>
- Doan, S., Schweig, J. D., & Mihaly, K. (2019). The consistency of composite ratings of teacher effectiveness: evidence from New Mexico. *American Educational Research Journal*, 56(6), 2116-2146. <https://doi.org/10.3102/0002831219841369>
- Donaldson, M. L. (2013). Principals' approaches to cultivating teacher effectiveness: Constraints and opportunities in hiring, assigning, evaluating, and developing teachers. *Educational Administration Quarterly*, 49(5), 838-882. <https://doi.org/10.1177/0013161x13485961>
- Donaldson, M., & Mavrogordato, M. (2018). Principals and teacher evaluation. *Journal of Educational Administration*, 56(6), 586-601. <https://doi.org/10.1108/jea-08-2017-0100>
- Downey, D. B., Von Hippel, P. T., & Hughes, M. (2008). Are "failing" schools really failing? Using seasonal comparison to evaluate school effectiveness. *Sociology of Education*, 81(3), 242-270. <https://doi.org/10.1177/003804070808100302>

## E

Elstad, E., Lejonberg, E., & Christophersen, K. A. (2015). Teaching evaluation as a contested practice: Teacher resistance to teaching evaluation schemes in Norway. *Education Inquiry*, 6(4), 27850. <https://doi.org/10.3402/edui.v6.27850>

Elstad, E., Lejonberg, E., & Christophersen, K. A. (2017). Student evaluation of high-school teaching: Which factors are associated with teachers' perception of the usefulness of being evaluated?. *Journal for Educational Research Online*, 9(1), 99-117. <https://doi.org/10.25656/01:12968>

## F

Favero, N., & Meier, K. J. (2013). Evaluating urban public schools: Parents, teachers, and state assessments. *Public Administration Review*, 73(3), 401-412. <https://doi.org/10.1111/puar.12022>

Finster, M., & Milanowski, A. (2018). Teacher perceptions of a new performance evaluation system and their influence on practice: A within-and between-school level analysis. *Education Policy Analysis Archives*, 26, 41. <https://doi.org/10.14507/epaa.26.3500>

Flores, M. A. (2018). Teacher evaluation in Portugal: persisting challenges and perceived effects. *Teachers and teaching*, 24(3), 223-245. <https://doi.org/10.1080/13540602.2018.1425677>

Ford, T. G. (2018). Pointing teachers in the wrong direction: Understanding Louisiana elementary teachers' use of compass high-stakes teacher evaluation data. *Educational Assessment, Evaluation and Accountability*, 30(3), 251-283. <https://doi.org/10.1007/s11092-018-9280-x>

Fresko, B., & Alhija, F. N. A. (2009). When intentions and reality clash: Inherent implementation difficulties of an induction program for new teachers. *Teaching and Teacher Education*, 25(2), 278-284. <https://doi.org/10.1016/j.tate.2008.12.001>

## G

- Gagnon, D. J., Hall, E. L., & Marion, S. (2017). Teacher evaluation and local control in the US: an investigation into the degree of local control afforded to districts in defining evaluation procedures for teachers in non-tested subjects and grades. *Assessment in Education: Principles, Policy & Practice*, 24(4), 489–505.  
<https://doi.org/10.1080/0969594x.2016.1167669>
- Gálvez Suarez, E., & Milla Toro, R. (2018). Teaching Performance Evaluation Model: Preparation for Student Learning within the Framework for Teacher Good Performance. *Journal of Educational Psychology-Propositos y Representaciones*, 6(2), 431-452.  
<http://dx.doi.org/10.20511/pyr2018.v6n2.236>
- Gargani, J., & Strong, M. (2014). Can we identify a successful teacher better, faster, and cheaper? Evidence for innovating teacher observation systems. *Journal of Teacher Education*, 65(5), 389-401. <https://doi.org/10.1177/0022487114542519>
- Garrett, R., & Steinberg, M. P. (2015). Examining teacher effectiveness using classroom observation scores: Evidence from the randomization of teachers to students. *Educational Evaluation and Policy Analysis*, 37(2), 224-242.  
<https://doi.org/10.3102/0162373714537551>
- Geiger, T., & Amrein-Beardsley, A. (2019). Student perception surveys for K-12 teacher evaluation in the United States: A survey of surveys. *Cogent Education*, 6(1), 1602943.  
<https://doi.org/10.1080/2331186x.2019.1602943>
- Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher*, 44(2), 96-104.  
<https://doi.org/10.3102/0013189x15575031>
- Govender, N., Grobler, B., & Mestry, R. (2016). Internal whole-school evaluation in South Africa: The influence of holistic staff capacity. *Educational Management Administration & Leadership*, 44(6), 996-1020. <https://doi.org/10.1177/1741143215595414>

## H

Harris, D. N., Ingle, W. K., & Rutledge, S. A. (2014). How teacher evaluation methods matter for accountability: A comparative analysis of teacher effectiveness ratings by principals and teacher value-added measures. *American Educational Research Journal*, 51(1), 73-112. <https://doi.org/10.3102/0002831213517130>

Hawley, L. R., Bovaird, J. A., & Wu, C. (2017). Stability of teacher value-added rankings across measurement model and scaling conditions. *Applied Measurement in Education*, 30(3), 196-212. <https://doi.org/10.1080/08957347.2017.1316273>

Hewitt, K. K. (2015). Educator evaluation policy that incorporates EVAAS value-added measures: Undermined intentions and exacerbated inequities. *Education Policy Analysis Archives*, 23, 76. <https://doi.org/10.14507/epaa.v23.1968>

## I

Ikin, K., & McClenaghan, P. (2019). Deep diving into values: A way to enhance empowerment evaluation. *Evaluation Journal of Australasia*, 19(4), 204-210. <https://doi.org/10.1177/1035719x19893057>

Ingle, W. K., Willis, C., & Fritz, J. (2015). Collective bargaining agreement provisions in the wake of Ohio teacher evaluation system legislation. *Educational Policy*, 29(1), 18-50. <https://doi.org/10.1177/0895904814559249>

## J

Jiang, J. Y., Sparte, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students. *Educational Researcher*, 44(2), 105-116. <https://doi.org/10.3102/0013189x15575517>

Jones, E., & Bergin, C. (2019). Evaluating teacher effectiveness using classroom observations: A Rasch analysis of the rater effects of principals. *Educational Assessment*, 24(2), 91-118. <https://doi.org/10.1080/10627197.2018.1564272>



Jung Peng, W., Thomas, S. M., Yang, X., & Li, J. (2006). Developing school evaluation methods to improve the quality of schooling in China: a pilot 'value added' study. *Assessment in Education: Principles, Policy & Practice*, 13(2), 135-154.  
<https://doi.org/10.1080/09695940600843252>

## K

Katsuno, M. (2010). Teacher evaluation in Japanese schools: An examination from a micro-political or relational viewpoint. *Journal of Education Policy*, 25(3), 293-307.  
<https://doi.org/10.1080/02680931003700431>

Kember\*, D., Jenkins, W., & Chi Ng, K. (2004). Adult students' perceptions of good teaching as a function of their conceptions of learning—Part 2. Implications for the evaluation of teaching. *Studies in Continuing Education*, 26(1), 81-97.  
<https://doi.org/10.1080/158037042000199461>

Kettler, R. J., Arnold-Berkovits, I., Rutgers, L. A. R., Kurz, A., Dudek, C. M., Hua, A. N., & Lekwa, A. (2018). Multi-method teacher evaluation for high poverty schools: Observations and self-ratings of instructional and behavioral management. *Studies in Educational Evaluation*, 59, 224-2. <https://doi.org/10.1016/j.stueduc.2018.08.004>

Kim, J., Sun, M., & Youngs, P. (2019). Developing the " Will": The Relationship between Teachers' Perceived Policy Legitimacy and Instructional Improvement. *Teachers College Record*, 121(3), 1-44. <https://doi.org/10.1177/016146811912100301>

Kimball, S. M. (2002). Analysis of feedback, enabling conditions and fairness perceptions of teachers in three school districts with new standards-based evaluation systems. *Journal of Personnel Evaluation in Education*, 16(4), 241-268.  
<https://doi.org/10.1023/A:1021787806189>

Kimball, S. M., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, 45(1), 34-70. <https://doi.org/10.1177/0013161x08327549>

- Koedel, C., Li, J., Springer, M. G., & Tan, L. (2017). The impact of performance ratings on job satisfaction for public school teachers. *American Educational Research Journal*, 54(2), 241-278. <https://doi.org/10.3102/0002831216687531>
- Koedel, C., Li, J., Springer, M. G., & Tan, L. (2019). Teacher performance ratings and professional improvement. *Journal of Research on Educational Effectiveness*, 12(1), 90-115. <https://doi.org/10.1080/19345747.2018.1490471>
- Kraft, M. A., & Gilmour, A. F. (2017). Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher*, 46(5), 234-249. <https://doi.org/10.3102/0013189x17718797>
- Kyrgiridis, P., Derri, V., Emmanouilidou, K., Chlapoutaki, E., & Kioumourtzoglou, E. (2014). Development of a questionnaire for self-evaluation of teacher effectiveness in physical education (SETEQ-PE). *Measurement in Physical Education and Exercise Science*, 18(2), 73-90. <https://doi.org/10.1080/1091367x.2013.866557>
- Kyriakides, L., & Demetriou, D. (2007). Introducing a teacher evaluation system based on teacher effectiveness research: An investigation of stakeholders' perceptions. *Journal of Personnel Evaluation in Education*, 20(1-2), 43-64. <https://doi.org/10.1007/s11092-007-9046-3>

## L

- Lavigne, A. L., & Chamberlain, R. W. (2017). Teacher evaluation in Illinois: School leaders' perceptions and practices. *Educational Assessment, Evaluation and Accountability*, 29(2), 179-209. <https://doi.org/10.1007/s11092-016-9250-0>
- LeChasseur, K., Donaldson, M., Fernandez, E., & Fenc-Bagwell, M. (2018). Brokering, buffering, and the rationalities of principal work. *Journal of Educational Administration*, 56(3), 262-276. <https://doi.org/10.1108/jea-10-2016-0129>

Lejonberg, E., Elstad, E., & Christophersen, K. A. (2018). Teaching evaluation: antecedents of teachers' perceived usefulness of follow-up sessions and perceived stress related to the evaluation process. *Teachers and Teaching*, 24(3), 281-296.

<https://doi.org/10.1080/13540602.2017.1399873>

Lenhoff, S. W., Pogodzinski, B., Mayrowetz, D., Superfine, B. M., & Umpstead, R. R. (2018). District stressors and teacher evaluation ratings. *Journal of Educational Administration*, 56(2).

<https://doi.org/10.1108/jea-06-2017-0065>

Lillejord, S., Elstad, E., & Kavli, H. (2018). Teacher evaluation as a wicked policy problem. *Assessment in Education: Principles, Policy & Practice*, 25(3), 291-309.

<https://doi.org/10.1080/0969594x.2018.1429388>

Liu, S., & Teddlie, C. (2005). A follow-up study on teacher evaluation in China: Historical analysis and latest trends. *Journal of Personnel Evaluation in Education*, 18(4), 253-272.

<https://doi.org/10.1007/s11092-007-9029-4>

Liu, S., & Zhao, D. (2013). Teacher evaluation in China: latest trends and future directions. *Educational Assessment, Evaluation and Accountability*, 25(3), 231-250.

<https://doi.org/10.1007/s11092-013-9168-8>

Liu, Y., Visone, J., Mongillo, M. B., & Lisi, P. (2019). What matters to teachers if evaluation is meant to help them improve?. *Studies in Educational Evaluation*, 61, 41-54.

<https://doi.org/10.1016/j.stueduc.2019.01.006>

Lo, L. N. K., Lai, M., & Wang, L. (2014). The impact of reform policies on teachers' work and professionalism in the Chinese Mainland. *Asia-Pacific Journal of Teacher Education*, 41(3), 239–252. <https://doi.org/10.1080/1359866x.2013.809054>

## M

Maier, U. (2010). Accountability policies and teachers' acceptance and usage of school performance feedback—a comparative study. *School Effectiveness and School Improvement*, 21(2), 145-165.

<https://doi.org/10.1080/09243450903354913>

- Malakolunthu, S., & Vasudevan, V. (2012). Teacher evaluation practices in Malaysian primary schools: Issues and challenges. *Asia Pacific Education Review, 13*(3), 449-456. <https://doi.org/10.1007/s12564-012-9207-z>
- Martínez, J. F., Schweig, J., & Goldschmidt, P. (2016). Approaches for combining multiple measures of teacher performance: Reliability, validity, and implications for evaluation policy. *Educational Evaluation and Policy Analysis, 38*(4), 738-756. <https://doi.org/10.3102/0162373716666166>
- Master, B. (2014). Staffing for success: Linking teacher evaluation and school personnel management in practice. *Educational Evaluation and Policy Analysis, 36*(2), 207-227. <https://doi.org/10.3102/0162373713506552>
- Mercer, J. (2005). Challenging appraisal orthodoxies: Teacher evaluation and professional development in the United Arab Emirates. *Journal of Personnel Evaluation in Education, 18*(4), 273-287. <https://doi.org/10.1007/s11092-007-9024-9>
- Milanowski, A. T., & Heneman, H. G. (2001). Assessment of teacher reactions to a standards-based teacher evaluation system: A pilot study. *Journal of Personnel Evaluation in Education, 15*(3), 193-212. <https://doi.org/10.1023/A:1012752725765>
- Miloş, L. R., Chidmi, B., & Miloş, M. C. (2019). Does data matter? Statewide longitudinal data systems and student performance in the US. *Applied Economics Letters, 26*(16), 1369-1372. <https://doi.org/10.1080/13504851.2018.1558342>
- Montecinos, C., Madrid, R., Fernández, M. B., & Ahumada, L. (2014). A goal orientation analysis of teachers' motivations to participate in the school self-assessment processes of a quality assurance system in Chile. *Educational Assessment, Evaluation and Accountability, 26*(3), 241-261. <https://doi.org/10.1007/s11092-014-9190-5>
- Moran, R. M. (2017). The impact of a high stakes teacher evaluation system: Educator perspectives on accountability. *Educational Studies, 53*(2), 178-193. <https://doi.org/10.1080/00131946.2017.1283319>

Namaghi, S. A. O. (2010). A Data-Driven Conceptualization of Teacher Evaluation.

*Qualitative Report*, 15(6), 1504-1522. <https://doi.org/10.46743/2160-3715/2010.1358>

Norris, J. M., Kwon, J., van der Mars, H., & Kulinna, P. H. (2016). A document analysis of physical education teacher evaluation systems. *Research Quarterly for Exercise and Sport*, 87(S2), A130. <https://www.proquest.com/scholarly-journals/document-analysis-physical-education-teacher/docview/1817495191/se-2?accountid=17200>

Norris, J. M., van der Mars, H., Kulinna, P., & Amrein-Beardsley, A. (2017). Administrators' perceptions of physical education teacher evaluation. *The Physical Educator*, 74(4), 730–756. <https://doi.org/10.18666/tpe-2017-v74-i4-7468>

## O

Obreque, A. S., Hernández-Mosqueira, C., Peña-Troncoso, S., Agredo, M. A. T., & Salvatierra, M. O. (2019). Evaluation of teacher performance in Chile: Perception of poorly evaluated teachers. *Cadernos de Pesquisa*, 49(172), 144-163. <https://doi.org/10.1590/198053145792>

Omona, J. (2003). Design that supports service quality: New Zealand's model. *Quality Assurance in Education*. <https://doi.org/10.1108/09684880310488472>

Orphanos, S. (2014). What matters to principals when they evaluate teachers? Evidence from Cyprus. *Educational Management Administration & Leadership*, 42(2), 243-258. <https://doi.org/10.1177/1741143213499262>

Ovando, M. N. (2001). Teachers' perceptions of a learner-centered teacher evaluation system. *Journal of Personnel Evaluation in Education*, 15(3), 213-231. <https://doi.org/10.1023/A:1012704809835>

## P

Papay, J. P., & Johnson, S. M. (2012). Is PAR a good investment? Understanding the costs and benefits of teacher peer assistance and review programs. *Educational Policy*, 26(5), 696-729. <https://doi.org/10.1177/0895904811417584>

- Paufler, N. A., & Amrein-Beardsley, A. (2014). The random assignment of students into elementary classrooms: Implications for value-added analyses and interpretations. *American Educational Research Journal*, *51*(2), 328-362.  
<https://doi.org/10.3102/0002831213508299>
- Paufler, N. A., & Clark, C. (2019). Reframing conversations about teacher quality: school and district administrators' perceptions of the validity, reliability, and justifiability of a new teacher evaluation system. *Educational Assessment, Evaluation and Accountability*, *31*(1), 33-60. <https://doi.org/10.1007/s11092-019-09292-w>
- Peterson, K. D., Wahlquist, C., & Bone, K. (2000). Student surveys for school teacher evaluation. *Journal of Personnel Evaluation in Education*, *14*(2), 135-153.  
<https://doi.org/10.1023/A:1008102519702>
- Piezunka, A. (2019). Struggle for acceptance—maintaining external school evaluation as an institution in Germany. *Historical Social Research/Historische Sozialforschung*, *44*(2) (168), 270-287. <https://doi.org/10.12759/hsr.44.2019.2.270-287>
- Pogodzinski, B., Umpstead, R., & Witt, J. (2015). Teacher evaluation reform implementation and labor relations. *Journal of Education Policy*, *30*(4), 540-561.  
<https://doi.org/10.1080/02680939.2014.999827>
- Pressley, T., Roehrig, A. D., & Turner, J. E. (2018). Elementary teachers' perceptions of a reformed teacher-evaluation system. *The Teacher Educator*, *53*(1), 21-43.  
<https://doi.org/10.1080/08878730.2017.1391362>
- Pritchard, R. E., Saccucci, M. S., & Potter, G. C. (2010). Evaluating a program designed to demonstrate continuous improvement in teaching at an AACSB-accredited college of business at a regional university: A case study. *Journal of Education for Business*, *85*(5), 280-283. <https://doi.org/10.1080/08832320903449568>

## Q

- Quinn, D. M., Kane, T. J., Greenberg, M., & Thal, D. (2018). Effects of a video-based teacher observation program on the de-privatization of instruction: Evidence from a randomized

experiment. *Educational Administration Quarterly*, 54(4), 529-558.

<https://doi.org/10.1177/0013161x18764333>

Quiroga, M., & Aravena, F. (2018). The new teaching career policy in Chile: Perspectives from school principals. *KEDI Journal of Educational Policy*, 15(2).

<https://www.proquest.com/scholarly-journals/new-teaching-career-policy-chile-perspectives/docview/2163375845/se-2?accountid=17200>

## R

Ramirez, A., Clouse, W., & Davis, K. W. (2014). Teacher evaluation in Colorado: How policy frustrates practice. *Management in Education*, 28(2), 44-51.

<https://doi.org/10.1177/0892020613511264>

Ramirez, A., Lamphere, M., Smith, J., Brown, S., & Pierceall-Herman, J. (2011). Teacher development and evaluation: A study of policy and practice in Colorado. *Management in Education*, 25(3), 95-99. <https://doi.org/10.1177/0892020610387956>

Räty, H. (2003). The significance of parents' evaluations of their own school for their educational attitudes. *Social Psychology of Education*, 6(1), 43-60.

<https://doi.org/10.1023/A:1021798713525>

Reid, D. B. (2019). What information do principals consider when evaluating teachers? *School Leadership & Management*, 39(5), 457-477.

<https://doi.org/10.1080/13632434.2019.1576167>

Resch, A., & Isenberg, E. (2018). How do test scores at the ceiling affect value-added estimates? *Statistics and Public Policy*, 5(1), 1-6.

<https://doi.org/10.1080/2330443x.2018.1460226>

Robertson-Kraft, C., & Zhang, R. S. (2018). Keeping great teachers: A case study on the impact and implementation of a pilot teacher evaluation system. *Educational Policy*, 32(3), 363-394. <https://doi.org/10.1177/0895904816637685>

Roth McDuffie, A., Drake, C., Choppin, J., Davis, J. D., Magaña, M. V., & Carson, C. (2017). Middle school mathematics teachers' perceptions of the Common Core State

- Standards for Mathematics and related assessment and teacher evaluation systems. *Educational Policy*, 31(2), 139-179. <https://doi.org/10.1177/0895904815586850>
- Rovai, A. P., Ponton, M. K., Derrick, M. G., & Davis, J. M. (2006). Student evaluation of teaching in the virtual and traditional classrooms: A comparative analysis. *The Internet and Higher Education*, 9(1), 23-35. <https://doi.org/10.1016/j.iheduc.2005.11.002>
- Rutkowski, D., & Wild, J. (2015). Stakes matter: Student motivation and the validity of student assessments for teacher evaluation. *Educational Assessment*, 20(3), 165-179. <https://doi.org/10.1080/10627197.2015.1059273>
- S**
- Sampaio, M., & Leite, C. (2017). From curricular justice to educational improvement: What is the role of schools' self-evaluation?. *Improving Schools*, 20(1), 62-75. <https://doi.org/10.1177/1365480216688553>
- Santelices, M. V., & Taut, S. (2011). Convergent validity evidence regarding the validity of the Chilean standards-based teacher evaluation system. *Assessment in Education: Principles, Policy & Practice*, 18(1), 73-93. <https://doi.org/10.1080/0969594x.2011.534948>
- Santelices, M. V., Valencia, E., Gonzalez, J., & Taut, S. (2017). Two teacher quality measures and the role of context: evidence from Chile. *Educational Assessment, Evaluation and Accountability*, 29(2), 111-146. <https://doi.org/10.1007/s11092-016-9247-8>
- 8**
- Sarrico, C. S., & Rosa, M. J. (2009). Measuring and comparing the performance of Portuguese secondary schools. *International Journal of Productivity and Performance Management*. <https://doi.org/10.1108/17410400911000408>
- Schildkamp, K., & Visscher, A. (2009). Factors influencing the utilisation of a school self-evaluation instrument. *Studies in Educational Evaluation*, 35(4), 150-159. <https://doi.org/10.1016/j.stueduc.2009.12.001>



- Schrodt, P., Witt, P. L., & Turman, P. D. (2007). Reconsidering the measurement of teacher power use in the college classroom. *Communication Education, 56*(3), 308-332.  
<https://doi.org/10.1080/03634520701256062>
- Spooren, P. (2010). On the credibility of the judge: A cross-classified multilevel analysis on students' evaluation of teaching. *Studies in Educational Evaluation, 36*(4), 121-131.  
<https://doi.org/10.1016/j.stueduc.2011.02.001>
- Spooren, P., & Mortelmans, D. (2006). Teacher professionalism and student evaluation of teaching: will better teachers receive higher ratings and will better students give higher ratings?. *Educational Studies, 32*(2), 201-214.  
<https://doi.org/10.1080/03055690600631101>
- Stamelos, G., Vassilopoulos, A., & Bartzakli, M. (2012). Understanding the difficulties of implementation of a teachers' evaluation system in Greek primary education: From national past to European influences. *European Educational Research Journal, 11*(4), 545-557. <https://doi.org/10.2304/eej.2012.11.4.545>
- Steinberg, M. P., & Donaldson, M. L. (2016). The new educational accountability: Understanding the landscape of teacher evaluation in the post-NCLB era. *Education Finance and Policy, 11*(3), 340-359. [https://doi.org/10.1162/edfp\\_a\\_00186](https://doi.org/10.1162/edfp_a_00186)
- Steinberg, M. P., & Kraft, M. A. (2017). The sensitivity of teacher performance ratings to the design of teacher evaluation systems. *Educational Researcher, 46*(7), 378-396.  
<https://doi.org/10.3102/0013189x17726752>
- Strong, M., Gargani, J., & Hacifazlıoğlu, Ö. (2011). Do we know a successful teacher when we see one? Experiments in the identification of effective teachers. *Journal of Teacher Education, 62*(4), 367-382. <https://doi.org/10.1177/0022487110390221>
- Strunk, K. O., Weinsten, T. L., & Makkonen, R. (2014). Sorting out the signal: Do multiple measures of teachers' effectiveness provide consistent information to teachers and principals?. *Education Policy Analysis Archives, 22*, 100.  
<https://doi.org/10.14507/epaa.v22.1590>

## T

- Tarhan, H., Karaman, A. C., Kemppinen, L., & Aerila, J. A. (2019). Understanding teacher evaluation in Finland: A professional development framework. *Australian Journal of Teacher Education*, 44(4), 33–50. <https://doi.org/10.14221/ajte.2018v44n4.3>
- Taut, S., & Sun, Y. (2014). The Development and Implementation of a National, Standards-based, Multi-method Teacher Performance Assessment System in Chile. *Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas*, 22, 1-30. <https://doi.org/10.14507/epaa.v22n71.2014>
- Taut, S., Jiménez, D., Puente-Duran, S., Palacios, D., Godoy, M. I., & Manzi, J. (2019). Evaluating the quality of teaching: can there be valid differentiation in the middle of the performance distribution? *School Effectiveness and School Improvement*, 30(3), 328-348. <https://doi.org/10.1080/09243453.2018.1510842>
- Taut, S., Santelices, V., Araya, C., & Manzi, J. (2010). Theory underlying a national teacher evaluation program. *Evaluation and Program Planning*, 33(4), 477-486. <https://doi.org/10.1016/j.evalprogplan.2010.01.002>
- Taut, S., Santelices, M. V., Araya, C., & Manzi, J. (2011). Perceived effects and uses of the national teacher evaluation system in Chilean elementary schools. *Studies in Educational Evaluation*, 37(4), 218-229. <https://doi.org/10.1016/j.stueduc.2011.08.002>
- Taut, S., Santelices, M. V., & Stecher, B. (2012). Validation of a National Teacher Assessment and Improvement System. *Educational Assessment*, 17(4), 163–199. <https://doi.org/10.1080/10627197.2012.735913>
- Taut, S., Valencia, E., Palacios, D., Santelices, M. V., Jimenez, D., & Manzi, J. (2016). Teacher performance and student learning: linking evidence from two national assessment programmes. *Assessment in education: principles, policy & practice*, 23(1), 53-74. <https://doi.org/10.1080/0969594x.2014.961406>
- Tigelaar, D. E., Dolmans, D. H., Wolfhagen, I. H., & Van Der Vleuten, C. P. (2004). The development and validation of a framework for teaching competencies in higher

- education. *Higher education*, 48(2), 253-268.  
<https://doi.org/10.1023/b:high.0000034318.74275.e4>
- Tornero, B., & Taut, S. (2010). A mandatory, high-stakes national teacher evaluation system: Perceptions and attributions of teachers who actively refuse to participate. *Studies in Educational Evaluation*, 36(4), 132-142. <https://doi.org/10.1016/j.stueduc.2011.02.002>
- Tucker, P. D., Stronge, J. H., Gareis, C. R., & Beers, C. S. (2003). The efficacy of portfolios for teacher evaluation and professional development: Do they make a difference?. *Educational Administration Quarterly*, 39(5), 572-602.  
<https://doi.org/10.1177/0013161x03257304>
- Tuytens, M., & Devos, G. (2010). The influence of school leadership on teachers' perception of teacher evaluation policy. *Educational Studies*, 36(5), 521-536.  
<https://doi.org/10.1080/03055691003729054>
- Tuytens, M., & Devos, G. (2014). How to activate teachers through teacher evaluation?. *School Effectiveness and School Improvement*, 25(4), 509-530.  
<https://doi.org/10.1080/09243453.2013.842601>
- Tuytens, M., & Devos, G. (2014). The problematic implementation of teacher evaluation policy: School failure or governmental pitfall?. *Educational Management Administration & Leadership*, 42(4\_suppl), 155-174. <https://doi.org/10.1177/1741143213502188>
- Tuytens, M., & Devos, G. (2017). The role of feedback from the school leader during teacher evaluation for teacher and school improvement. *Teachers and Teaching*, 23(1), 6-24.  
<https://doi.org/10.1080/13540602.2016.1203770>
- Tuytens, M., & Devos, G. (2018). Teacher evaluation policy as perceived by school principals: the case of Flanders (Belgium). *Teachers and teaching*, 24(3), 209-222.  
<https://doi.org/10.1080/13540602.2017.1397508>

U

V

van der Lans, R. M., van de Grift, W. J., & van Veen, K. (2015). Developing a teacher evaluation instrument to provide formative feedback using student ratings of teaching acts. *Educational Measurement: Issues and Practice*, 34(3), 18-27.

<https://doi.org/10.1111/emip.12078>

van der Steeg, M., & Gerritsen, S. (2016). Teacher evaluations and pupil achievement gains: Evidence from classroom observations. *De Economist*, 164(4), 419-443.

<https://doi.org/10.1007/s10645-016-9280-5>

van Elk, R., & Kok, S. (2016). The impact of a comprehensive school reform policy for weak schools on educational achievement; results of the first 4 years. *De Economist*, 164(4), 445-476. <https://doi.org/10.1007/s10645-016-9281-4>

## W

Walder, A. M. (2017). Pedagogical Innovation in Canadian higher education: Professors' perspectives on its effects on teaching and learning. *Studies in Educational Evaluation*, 54, 71-82. <https://doi.org/10.1016/j.stueduc.2016.11.001>

Walsh, E., & Isenberg, E. (2015). How does value added compare to student growth percentiles?. *Statistics and Public Policy*, 2(1), 1-13.

<https://doi.org/10.1080/2330443x.2015.1034390>

Wang, J. H., & Chen, Y. T. (2016). Which way is better for teacher evaluation? The discourse on teacher evaluation in Taiwan. *Policy Futures in Education*, 14(7), 910-925.

<https://doi.org/10.1177/1478210316652049>

Wang, W. (2007). Evaluation of 2+ 2 alternative teacher performance appraisal program in Shanxi, People's Republic of China. *Teaching and Teacher Education*, 23(7), 1012-1023.

<https://doi.org/10.1016/j.tate.2006.05.005>

Warren, A. N., & Ward, N. A. (2019). "It didn't make me a better teacher": Inservice teacher constructions of dilemmas in high-stakes teacher evaluation. *School Effectiveness and School Improvement*, 30(4), 531-548. <https://doi.org/10.1080/09243453.2019.1619185>

- Wei, X., Patel, D., & Young, V. M. (2014). Opening the “black box”: Organizational differences between charter schools and traditional public schools. *Education Policy Analysis Archives*, 22, 3. <https://doi.org/10.14507/epaa.v22n3.2014>
- Welsh, R., Graham, J., & Williams, S. (2019). Acing the test: an examination of teachers’ perceptions of and responses to the threat of state takeover. *Educational Assessment, Evaluation and Accountability*, 31(3), 315-347. <https://doi.org/10.1007/s11092-019-09301-y>
- Wilkerson, D. J., Manatt, R. P., Rogers, M. A., & Maughan, R. (2000). Validation of student, principal, and self-ratings in 360 feedback® for teacher evaluation. *Journal of Personnel Evaluation in Education*, 14(2), 179-192. <https://doi.org/10.1023/A:1008158904681>

## X

- Xu, X., Grant, L. W., & Ward, T. J. (2016). Validation of a statewide teacher evaluation system: Relationship between scores from evaluation and student academic progress. *NASSP Bulletin*, 100(4), 203-222. <https://doi.org/10.1177/0192636516683247>

## Y

- Yeung, S. Y. S. (2012). A school evaluation policy with a dual character: Evaluating the school evaluation policy in Hong Kong from the perspective of curriculum leaders. *Educational Management Administration & Leadership*, 40(1), 37-68. <https://doi.org/10.1177/1741143211420616>
- Yoo, J. (2018). Evaluating the new teacher evaluation system in South Korea: Case studies of successful implementation, adaptation, and transformation of mandated policy. *Policy Futures in Education*, 16(3), 277–290. <https://doi.org/10.1177/1478210317751274>

## Z