The Predictive Validity of Teacher Candidate Letters of Reference

Richard W. Mason¹ & Mark P. Schroeder²

¹ Curriculum & Instruction Department, University of Wisconsin, Whitewater, USA
² Educational Foundations Department, University of Wisconsin, Whitewater, USA

Correspondence: Richard W. Mason, Educational Leadership Coordinator, Winther Hall 4051, 800 W. Main St., University of Wisconsin-Whitewater, Whitewater, WI 53190, USA. Tel: 1-262-472-4891. E-mail: masonr@uww.edu

Received: May 15, 2014     Accepted: June 24, 2014     Online Published: July 17, 2014
doi:10.5539/jel.v3n3p67          URL: http://dx.doi.org/10.5539/jel.v3n3p67

Abstract

Letters of reference are widely used as an essential part of the hiring process of newly licensed teachers. While the predictive validity of these letters of reference has been called into question it has never been empirically studied. The current study examined the predictive validity of the quality of letters of reference for forty-one student teachers in relation to their attainment of full time employment and performance during their first year of teaching. Results indicated that while letter quality was predictive of whether or not full-time employment was obtained, it was not predictive of performance during the first year of teaching. Findings also suggest that hiring practices should be re-examined and additional measures of teacher quality should be incorporated to increase teacher excellence in schools.

Keywords: letters of reference, reference letters, teacher selection, teacher hiring

1. Introduction

In primary and secondary education, letters of reference serve as one of the main sources of information in the hiring process. Letters of reference are typically one means of reducing the large candidate pool to a manageable number which can lead to a formal interview. Letters of reference are valued for what they do say as well as what they do not say about a candidate. Sometimes candidates look outstanding on paper but disappoint when they are seen face to face. While on other occasions, we are pleasantly surprised when a candidate is hired and performs well despite the low expectations the paper evidence fostered. It is often the case that long-term predictions of professional success are even less accurate, and this begs the question: To what extent do letters of recommendation actually predict future teaching performance? The current study evaluates the predictive validity of teacher letters of reference by comparing teacher candidates’ letters of reference with principal ratings of their performance in their first year of teaching.

2. Review of Literature

In education, few decisions are as impactful as who a principal hires to develop the minds and personalities of the students that inhabit the school walls. The difference between hiring an outstanding teacher and hiring an ineffective teacher has been estimated as being worth up to a year of educational growth for the students in their classroom (Hanushek, 1992, 1997; Hanushek & Rivkin, 2010), and when one considers the potential impact of a succession of excellent or a succession of poor teachers, the gravity of hiring excellent teachers increases. When one also considers additional costs of replacing and training ineffective teachers, professional development opportunities, and the longitudinal deficiencies to students falling behind in a given year, it is of the utmost importance that good hiring decisions be made. However, in spite of numerous candidates competing for teaching vacancies, making the right hiring decision is challenging, as the predictive validity of applicant information is questionable. Given the relative importance and widespread use of letters of reference (Mason & Schroeder, 2010) as a key ingredient in determining who is hired, it is important to evaluate their usefulness in a systematic way. The following section discusses sources of evidence that are commonly used when hiring teachers, with a special focus on letters of reference due to their widespread use and the importance that is attributed to them.
3. The Teacher Hiring Process

In recent years principals looking for prospective hires are typically faced with many teacher applicants, commonly ranging between dozens to hundreds of candidates in both large and small school districts (Gantert, 2012; Hu, 2010; Schmelz, 2010). Previous research on the hiring process in education conceptualized it as a three-step process consisting first of an initial ‘paper’ screening process that focuses on prerequisite credentials, a second in-depth examination of paper-based credentials that go beyond initial prerequisites and results in the identification of candidates for interview, and finally interviews and hiring (Peterson, 2002). This process can be viewed as a balance between the breadth of applicants and the quality of information that can be obtained. The first step is relatively low in costs per teacher and is a simple checklist of necessary and desired qualifications which can eliminate a substantial portion of applicants and requires only a minute or two per application. The second step represents a modest initial cost per applicant that increases as step three draws nearer. In the second step, qualifications are more closely examined. For example, past experiences and letters of recommendations are scrutinized, portfolios and work samples are viewed, and phone calls are made until a short list of candidates is obtained. The third step has the highest cost in terms of administrative hours, which involves one or more interviews of each candidate or possibly an observation of the candidate, typically resulting in the selection of a teacher to hire.

While this final step is most directly related to who is hired, the number of final candidates is often very small, and there is no guarantee that the best candidates have been interviewed. It is the second step where the greatest number of errors of omission and commission occur: neglecting to bring in who will make the best teachers while at the same time bringing in those who may not be excellent teachers. Mason and Schroeder (2010) investigated the extent to which different sources of information are weighted during the second step of the hiring process and found that the greatest weight was given to verbal references (i.e. when person has actually observed a candidate teaching and makes a positive recommendation), and that letters of reference were the second most valued source of candidate information. However, because verbal recommendations are often not available for any given candidate, letters of reference have the greatest overall impact on the decision of which candidates are granted an interview. Despite their potent impact, letters of reference are not always an accurate representation of the individual, and a variety of issues must be considered.

4. Validity Concerns with Letters of Reference

While research on letters of reference in education is very limited, relevant research from other fields reveals several issues that threaten the predictive validity that letters of reference hold: 1) the inflationary aspect of letters of reference, 2) letter of reference confidentiality, and 3) writer characteristics.

4.1 Inflationary Aspect of Letters of Reference

One seemingly trivial characteristic of letters of recommendation is that they are, by definition, mostly positive. While this seems at first blush innocuous, it represents a bias that is not likely reflective of the sum of all evaluative sources for a potential hire. The awareness writers of letters of recommendation have of this fact likely leads to a further positive distortion on the part of the letter writer to make their letter stand out. Friedman (1983) writes a fanciful piece called Fantasy Land complaining of the inflationary tendency of letters of reference in applications for medical internships, in which about ten percent of applicants are described as ‘the finest I have ever worked with’ and virtually all applicants are in the top 25 percent. Miller and Van Rybroek (1988) echo the same feelings when reading psychology student applications, calling this tendency “letter inflation”. Others, (Ryan & Mortonson, 2000; Schneider, 2000) also complain of letters of reference becoming more and more inflated much like Lake Wobegon where all women are strong, all men good looking and all children are above average (Cannell, 1989). Schneider (2000) speculates that this problem perpetuates itself because a letter writer who is honest, frank, and straightforward will likely put their candidate at a significant disadvantage compared to other candidates. In addition, Morrison (2007) states, “References may not form the basis of a decision [to hire], but they can tip a candidate over the edge, to either failure or success” (p. 32).

4.2 Confidential Aspects of Letters of Reference

The inflationary aspect of letters of reference in education is an echo of the Family Educational Rights and Privacy Act of 1974, also known as the Buckley Amendment of November 1974, which allows students the option to choose whether letters of reference are opened or closed. Two reviews of relevant literature, however, indicate that admission officers and employers view closed letters to be a more accurate portrayal of the candidate (Schaffer & Tomarelli, 1981) and candidates who chose closed letters were favored over those who chose open letters (Shaffer, Mays, & Etheridge, 1976). This is another indication of an awareness of the problem.
and a differentiation of different types of letters among readers, but its impact on validity has not been rigorously investigated. Nevertheless, closed letters have been recommended as a way to potentially increase letters of reference validity (Ceci & Peters, 1984; Shaffer, Mays, & Etheridge, 1976).

4.3 Letter Writer Competence

A final issue that has not received any attention of note is how the ability of the letter writer to produce a high-quality letter of reference impacts letter validity. While this issue is, understandably, very difficult to investigate empirically, it remains a major weakness in the value of letters of reference. A stellar candidate may very well never receive an interview because they were paired with a poor letter writer, while a relatively poor candidate paired with an excellent letter writer may receive numerous interviews. Mason and Schroeder (2012) outlined four general categories that are used for evaluating the quality of letters of reference: 1) superlatives, 2) teacher traits, 3) testimonials, and 4) interpersonal skills. Superlatives include words that are excessive or exaggerated (e.g., outstanding, excellent), testimonials include phrases that communicate personal observations and judgments relative to other potential candidates (e.g., best student teacher I ever worked with); teacher traits include descriptions of characteristics associated directly with the profession (e.g., highly cooperative, pedagogical knowledge); and interpersonal skills refer to interpersonal traits (e.g., rapport, warmth). Of these, testimonials and superlative were the strongest predictors of overall letter quality. Thus, a letter writer who is concise and to the point, frowns upon the use of superlatives, and fails to support their teacher endorsement with specific details might write a letter that will likely lead to unfavorable judgments towards the candidate they seek to represent. Furthermore, Mason and Schroeder (2012) point out, Given these differences in each potential letter writer, persons receiving outstanding letters of recommendation may sometimes be more about the writer than the candidate. For an excellent teacher candidate, it may be that the writer simply does not write high quality letters, may have extensive interpersonal differences or similarities with the candidate, or may simply be too complimentary or even too honest (p. 5).

5. The Current Study

These issues raise serious doubts about the validity of the inferences that are made based upon letters of reference evidence. The current study is an extension of Mason and Schroeder’s 2012 analysis of student teacher letters of reference and obtains follow-up ratings from the principals of first-year teachers to examine two primary questions related to predictive validity: 1) Does the quality of the letters of reference predict who does and does not get hired, and 2) Do the ratings of superlatives, testimonials, interpersonal skills, and teacher traits predict the parallel ratings provided by principals for teacher candidates who did obtain employment?

6. Methods

6.1 Participants

Participants included forty-one recent graduates of a Midwestern university who obtained their teacher’s license in elementary and secondary education programs, and represent a subset of Mason and Schroeder’s 2012 study.

<table>
<thead>
<tr>
<th>Occupational Status</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full-time teaching positions</td>
<td>17</td>
</tr>
<tr>
<td>Did not have teaching position</td>
<td>13</td>
</tr>
<tr>
<td>Substitute teaching</td>
<td>6</td>
</tr>
<tr>
<td>Unable to make contact</td>
<td>5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>41</strong></td>
</tr>
</tbody>
</table>

6.2 Letters of Reference

The Letters of Reference Evaluation Rubric (Mason & Schroeder, 2012) is an analytic rubric developed to increase inter-rater reliability in student teacher letter of reference evaluations. The rubric employs five rating categories: interpersonal skills, superlatives, testimonials, teacher traits, and overall impression. Each category
uses a five-point scale. Final letter ratings represent the sum of all five categories, and scores range from five to twenty-five points. This rubric was used to generate ratings for all selected letters of reference.

Forty-one letters of reference from Mason and Schroeder’s 2012 study were randomly selected: 11 “poor” letters of reference \((M = 8.27)\), 15 “satisfactory” letters of reference \((M = 13.6)\) and 15 “excellent” student letters of reference \((M = 21.47)\) were selected from the 160 total letters of reference initially analyzed. Attempts were made to contact each of the 41 students to determine if they 1) had a job (see Table 1) and 2) where the job was located. With this information, the principal of those students who had a full-time job was contacted and asked to complete a rating scale survey over the phone.

6.3 Teacher Performance Questionnaire

Questions employed a seven-point scale and reflected each of the five categories: superlatives, testimonials, teacher traits, interpersonal skills, and overall impressions used in the letters of reference evaluation rubric. While the letter of reference evaluation rubric contained a five-point scale, a seven-point scale was employed to allow greater sensitivity to anticipated positive bias on the part of the principals of selected participants. The survey is presented in Appendix A.

6.4 Procedure

Pre-service teachers in a student teacher seminar agreed to submit their letters of reference from student teaching and allow follow-up contact with their administrator in the following year. Forty-one student letters were randomly selected and students were contacted via phone or social media as to their current employment status and employer if they were teaching. Students who had secured a full-time teaching position were considered successfully hired, while those teaching part-time, as a substitute, engaged in a different occupation, or unemployed were considered unsuccessful. The principal of each successfully hired teacher was contacted, briefed about the study and interviewed over the phone using the teacher performance questionnaire.

7. Results

The predictive validity of letters of reference was examined in two major ways: 1) initial hiring and 2) first-year performance. The relationship between letters of recommendation and whether or not an individual was hired was examined using a point-biserial correlation between the dichotomous variable of whether or not the individual was hired and both the overall score and the component scores for interpersonal skills, superlative use, teacher traits, and testimonials. Results indicated that employment outcomes were predicted by overall scores, \(r(39) = .35, p < .05\), the use of superlatives, \(r(39) = .34, p < .05\), and the use of testimonials, \(r(39) = .39, p < .05\).

Correlations, means, and standard deviations are presented in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>Overall Score</th>
<th>Interpersonal Skills</th>
<th>Superlatives</th>
<th>Testimonials</th>
<th>Teacher Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hired</td>
<td>.35*</td>
<td>-.10</td>
<td>.34*</td>
<td>.39*</td>
<td>.09</td>
</tr>
<tr>
<td>Mean</td>
<td>3.17</td>
<td>2.00</td>
<td>2.54</td>
<td>3.02</td>
<td>4.32</td>
</tr>
<tr>
<td>SD</td>
<td>1.55</td>
<td>1.40</td>
<td>1.66</td>
<td>1.70</td>
<td>1.12</td>
</tr>
</tbody>
</table>

*Note. * \(p < .05\)

An analysis of the predictive power of letters of reference relative to employment revealed no significant correlations between the component portions of the letters of reference and corresponding job performance. There was also no relationship between overall scores on letters of reference and overall principal impressions (See Table 3).
Table 3. Correlations between letter of reference ratings and principal ratings

<table>
<thead>
<tr>
<th>Letters of reference ratings</th>
<th>Principal ratings of teacher performance (n = 17)</th>
<th>IP</th>
<th>S</th>
<th>T</th>
<th>TT</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpersonal Skills (IP)</td>
<td></td>
<td>.10</td>
<td>-.03</td>
<td>-.01</td>
<td>-.24</td>
<td>-.09</td>
</tr>
<tr>
<td>Superlatives (S)</td>
<td></td>
<td>.02</td>
<td></td>
<td>-.13</td>
<td>-.07</td>
<td>-.14</td>
</tr>
<tr>
<td>Testimonials (T)</td>
<td></td>
<td></td>
<td>.18</td>
<td></td>
<td>.14</td>
<td>.14</td>
</tr>
<tr>
<td>Teacher Traits (TT)</td>
<td></td>
<td></td>
<td></td>
<td>.19</td>
<td>.17</td>
<td></td>
</tr>
<tr>
<td>Overall Impression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.17</td>
</tr>
</tbody>
</table>

| Mean | 6.53 | 6.28 | 6.62 | 6.44 | 6.47 |
| SD   | .49  | .57  | .47  | .38  | .44  |

8. Discussion

While some aspects of letters of recommendation were predictive of whether or not an individual teacher was hired, they were not related to principal ratings of performance in the first year of teaching. Regarding the former, overall impression, use of superlatives, and testimonials were each independently related to hiring, while interpersonal skills and teacher traits were not. This is similar to the findings of Mason and Schroeder (2012) who found that testimonials and superlatives were the strongest overall predictors of overall letter of reference ratings, and go the furthest towards leaving a positive, lukewarm, or negative impression with principals and hiring committees. This also reflects literature on letters of reference that suggests letters which are not filled with excessive praise, regardless of whether the praise is warranted or not, are viewed negatively and have a direct impact on hiring outcomes (Friedman, 1983; Miller & Rybroek, 1988; Ryan & Mortinson, 2000). The findings from the current study include the importance of personal testimonials, and the relative unimportance in directly referencing teacher traits and interpersonal skills in letters of reference.

While the importance of testimonials may convey a high quality, personal relationship that demonstrates a relatively high degree of conviction on the part of the letter writer, the unimportance of teacher traits and interpersonal skills perhaps has more meaningful implications. It is possible to interpret these two areas as the most informative elements that a teacher letter of reference can contain: information regarding skills and traits specific to the teaching profession and information regarding the ability to build rapport with others (e.g., students, parents, and peers). Despite this, they do not seem to influence or sway our perception and judgments about who should be hired either because they are sparsely mentioned or because they are largely ignored. What this may represent is a phenomenon similar to what Ambady and Rosenthal (1993) found regarding our tendency and consistency in focusing both our efforts to convey and interpret judgments on a limited, yet superficial, number of facets such as physical attractiveness and nonverbal behavior. However, it is notable to mention that, similar to Ambady and Rosenthal’s focus, there is little relation made to actual effectiveness – only perceptions of indirect effectiveness. Someone who appears attractive, comfortable, and commanding can go very far – what Malcolm Gladwell (2005) termed a “Warren Harding Error” to represent how superficial attributes can cover up a lack of competency and essential skills (in this case, how Warren Harding was elected president of the United States of America, despite his shortcomings in relevant competencies). These types of errors appeal to our common sense and are supported by a fair amount of literature (Bolino & Turnley, 2003; Heneman, Greenberger, & Anonyuo, 1989; Lefkowitz, 2000; Varma & Stroh, 2001). It may be that when we make our limited judgments we need to overhaul our evaluative criteria and ask ourselves about the potential validity in the criteria we employ.

This issue is reinforced by the findings that the ratings of the letters of reference were not related to the ratings of principals in their first year. While one might argue that these findings were due to a restriction of range issue in the data (averages ranged from 6.28 to 6.53 on a 7-point scale across all principal ratings with a relatively small standard deviations ranging from .38 to .57) or a small sample size (n = 17), they would be missing the larger issue at hand: that principals did not distinguish between levels of teacher quality. While possible, it is unlikely that all first year teachers in the sample were actually excellent at their job, rating 6.5 out of 7 possible points on
the average. The likelihood of these assessments being accurate is lessened all the more by research that suggests the effect of a first year teachers on student growth is largely negative (Hanushek, 1986, 1997; Rockoff, 2004).

One might argue that principal ratings have two primary flaws: 1) limited exposure to quality sources of information, and 2) shifting frames of reference for providing ratings of effectiveness. In the case of the former, despite daily interactions with their teachers, principals are likely not able to directly observe the amount of content that has been covered and learned in a day, week, or quarter; the total amount of instructional time during each day, the clarity and effectiveness of execution; the creation, compilation, and use of assessment data, or the match between particular student needs and differentiated delivery, as these would require time and attention that far surpasses a principal’s ability to give. Regarding the latter, that the average first-year teacher ratings was near the ceiling of the scale suggests that principals were not rating participants on the general construct of “teacher quality”, but rather on the more specific (and more forgiving) construct of “first year teacher quality”, and even then were demonstrating some type of positive bias.

Thus, it seems that both letters of recommendation and principal ratings have rather serious flaws, and we must develop better measures of teacher quality if we want to improve education in the United States. Efforts to increase the rigor with which we measure teacher effectiveness have often been met with some, oftentimes justified, resistance. The passing of No Child Left Behind, Race to the Top, the adoption of the Common Core Standards, the implementation of the Educational Teacher Performance Assessment for initial licensure in 28 states, and widespread legislative action to adopt consequential annual teacher effectiveness measures have not been unopposed. It seems, however, prudent to acknowledge these objections while accepting that our traditional systems are deeply flawed and that the new movement in assessment is likely needed.

9. Limitations

It should be noted that the conclusions drawn from this study are based on a small sample with a relatively homogenous demographic makeup of schools, confined to a relatively small geographic area. As such, some concern should exist over the generalizability of these findings. However, given the existing body of research, there is little reason to believe that the issues inherent in letters of recommendation and principal judgments of teaching quality vary widely across states, school sizes, school locations, and ethnicities. Nevertheless, future studies might assess the predictive validity of letters of recommendation across these demographic lines.

A second limitation is that no other measures of first-year teaching effectiveness were obtained. It would have been beneficial to include student test scores, student and parent perceptions, peer appraisals, and self-evaluations to arrive at a more robust representation of first-year performance, but these additional measures were beyond the scope and resources of the current study. It remains for future research to address this issue and to provide a more robust documentation of predictive validity.

10. Conclusion

Despite the aforementioned limitations, the results of the current study imply that current hiring practices should be reevaluated and reconsidered, and that measures of teacher quality should be included in this process. An ever-increasing system of new structures is supplying options to aid in this endeavor, but each new assessment tool or process should be met with the same critical eye towards predictive validity to ensure that our educational system maximizes its resources and supplies our children with the highest quality teachers possible.

References


Appendix

Questionnaire for principal’s letters of reference follow-up

1. Name: ____________________________________
2. Gender: M F
3. School Type Elem. Middle High
4. School Size _________________________
5. Highest Degree Bachelor Masters PhD
6. Years of Experience Teaching:
7. Years of Experience in Administration:

For the next few questions, I will be asking about the teacher’s Interpersonal skills. For each of the statements, please respond on a scale of 1 to 7 whether you agree (7) or disagree (1) as each statement pertains to (the teacher)

8. Mr. / Ms. _____ is understanding when interacting with both students, teachers, and parents
   1 2 3 4 5 6 7
9. Mr. / Ms. _____ has a good rapport with students, teachers, and parents
   1 2 3 4 5 6 7
10. Mr. / Ms. _____ has excellent cooperation skills with students, teachers, and parents
   1 2 3 4 5 6 7
11. Mr. / Ms. _____ displays a high level of enthusiasm in all aspects of his/her duties
   1 2 3 4 5 6 7
12. Mr. / Ms. _____ establishes strong relationships with students, teachers, and parents
   1 2 3 4 5 6 7

Is there anything else you would like to add about _____’s interpersonal skills?

For the next few questions, I will be asking you to rate the extent to which you think that the following words reflect Mr./ Ms. ____________ on the same 1 to 7 scale. Each word represents an adjective that could be used to describe Mr./ Ms. ____________.

13. Excellent
   1 2 3 4 5 6 7
14. Outstanding
   1 2 3 4 5 6 7
15. Effective
   1 2 3 4 5 6 7
16. Special
   1 2 3 4 5 6 7
17. Successful
   1 2 3 4 5 6 7
Are there any other words you would use to describe ____________?

The next set of questions deal with Mr./Ms. ____________ teacher traits. For each statement, using the same 7 point scale, indicate the extent to which you agree with the statements made in reference to Mr./Ms. ____________.

18 Mr. / Ms. _____ is a positive force and role model for students and other teachers
   1  2  3  4  5  6  7

19 Mr. / Ms. _____ demonstrates the highest level of teaching ability
   1  2  3  4  5  6  7

20 Mr. / Ms. _____ is an consummate professional in all aspects of the job
   1  2  3  4  5  6  7

21 Mr. / Ms. _____ is highly team-oriented in all aspects of the job
   1  2  3  4  5  6  7

22 Mr. / Ms. _____ is always extremely well-prepared to complete all his/her duties
   1  2  3  4  5  6  7

Is there anything else you would like to add about ____________’s teacher traits?

For the final set of questions, I will be asking about testimonials that you would give regarding ____________. For each statement, please use the same 1 to 7 scale.

23 You would give ____________ your highest recommendation
   1  2  3  4  5  6  7

24 It has been a true pleasure to work with ____________. 
   1  2  3  4  5  6  7

25 ____________ has been an incredible asset to the school?
   1  2  3  4  5  6  7

26 I have complete and total confidence in ____________. 
   1  2  3  4  5  6  7

Is there any other testimonial you would like to make about ____________?

Thank you for your time!

Copyrights
Copyright for this article is retained by the author(s), with first publication rights granted to the journal.
This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).