Data-Driven Approaches to Discovering Knowledge Gaps Related to Factors Affecting Construction Labor Productivity

Abraham Assefa TSEHAYAE\(^1\) and Aminah Robinson FAYEK\(^2\)

\(^1\) Student Member ASCE, PhD Candidate and Graduate Research Assistant, Hole School of Construction Engineering, Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta, T6G 2W2, Canada; PH (780) 492-9131; email: tsehayae@ualberta.ca

\(^2\) Member ASCE, Professor, NSERC Industrial Research Chair in Strategic Construction Modeling and Delivery, Ledcor Professor in Construction Engineering, Hole School of Construction Engineering, Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta, T6G 2W2, Canada; PH (780) 492-1205; email: aminah.robinson@ualberta.ca

ABSTRACT

Construction labor productivity remains of great importance due its direct effect on project costs. Numerous parameters (factors and practices) that critically affect labor productivity have been identified in past studies through expert knowledge obtained from surveys. The objective of this paper is to explore whether there is a gap in experts’ knowledge in identifying the critical parameters by comparing their perspectives to the results of data-driven analysis of the parameters and labor productivity field data. This paper presents a methodology for identifying critical parameters using both a factor survey and a data-driven approach. The factor survey approach ranks the critical parameters based on the responses of both project management and trade level personnel on a project. The data-driven approach ranks the parameters based on their degree of influence on productivity through filter feature selection on data collected from the actual project. Results of the comparison of factor rankings from the project management perspective, trade perspective, and data-driven approach indicate a major discrepancy between the expert perspectives and the data-driven results, suggesting a need for verification of expert-based results with additional field studies of factors affecting labor productivity.

INTRODUCTION

Construction project execution can be viewed as a conversion process whereby sets of inputs (manpower, material, equipment, and information/knowledge) are converted into an output (project components, such as concrete poured per structural element). Construction labor productivity, the focus of this paper, deals with the efficiency of the labor component of this conversion process and is computed as a ratio of output (units produced) to input (total manhours). The conversion process and its environment are influenced by a number of parameters (factors and practices), which exist at different levels of the project hierarchy (activity, project, organizational, provincial, national, and global levels). The parameters, which consist of various quantitative and qualitative factors and practices, are known to define the cause and effect relation of the environment with the efficiency measures of the process in terms of construction labor productivity.
Construction labor productivity is a major source of project risk and exhibits the highest variability among project resources. Extensive research has been carried out to understand, model, and formulate productivity improvement techniques (Panas and Pantouvakis 2010); however, despite the existence of these research studies, labor productivity levels remain low (Moselhi and Khan 2012).

This paper explores whether a gap exists in experts’ (project management and trade groups’) knowledge in identifying the critical parameters. It evaluates the knowledge gap by comparing each of the construction expert (project management and trade) groups’ perspectives on the critical parameters with the results of a data-driven analysis based on filter feature selection. The paper begins with a literature review of previous research and approaches used to establish critical parameters, and then proposes a methodology to: (1) collect expert knowledge through a factor survey to identify a list of critical parameters, (2) identify various quantitative and qualitative parameters and associated productivity values by using a standardized data collection strategy, and (3) identify the critical parameters by applying a data-driven approach to field data related to the parameters and labor productivity. The paper then presents a comparison of the critical parameter lists resulting from the factor surveys and the data-driven approach before conclusions and future research are discussed.

LITERATURE REVIEW

Construction labor productivity is of great importance to a construction project as it directly affects the profitability and success of the project. Thus, construction labor productivity continues to be an area of study and numerous parameters that critically affect labor productivity have been already been identified (Liberda et al. 2003; Dai et al. 2009). Past studies have relied on expert knowledge that was mainly collected through factor surveys and a group of experts to establish critical parameters. The established critical parameters were then used to either suggest further improvements or to carry out further data collection for analysis and modeling. Very few studies have attempted to identify the relative importance of the individual parameters through the use of a data-driven approach (Moselhi and Khan 2012).

Liberda et al. (2003) identified 51 factors grouped under three categories—human, external, and management—and carried out an interview survey with 20 experts to identify the relative importance of the factors and establish the top 15 critical factors. In the most extensive and detailed study of factors affecting labor productivity, Dai et al. (2009) identified 83 factors grouped under eleven categories including supervisor direction, communication, safety, tools and consumables, materials, engineering drawing management, labor, foreman, superintendent, project management, and construction equipment; they then carried out a factor survey with 1,996 craftspeople on 28 U.S. industrial construction projects to identify the top 10 critical factors and the relationship between these factors. While the above studies mainly relied on experts, Thomas et al. (1990) relied on a multivariate labor productivity factor model approach, based on field data collection of more than 50 factors that were grouped into six categories—manpower-labor, design features-work content, environmental-site conditions, management practices-control, construction methods, and project organization structure—and developed prediction models based
The parameters affecting labor productivity are numerous, complex, interlinked, and dynamic thus making data collection a challenging task. Additionally, documentation of the parameters is complicated as the factors and practices are a mix of quantitative and qualitative concepts and require the development of an appropriate measurement scheme (Thomas et al. 1990). Parameters having qualitative concepts like fairness of foreman in work assignment or uniformity of safety procedures require detailing of the parameter to a level that accurate data can be collected. Though measurement of quantitative parameters, such as temperature and crew size, has been easy to carry out, measurement of qualitative parameters like supervision skill of superintendent has presented challenges that researchers have attempted to address through the use of simple rating scales without calibration of each measurement scale (Oduba 2002; Thomas et al. 1990). As a result, past studies have tried to first identify the critical parameters based on expert knowledge before completing detailed measurements so as to simplify the data collection process (Thomas et al. 1990; AbouRizk et al. 2001; Chan et al. 2004; Dai et al. 2009). Unfortunately, this deductive approach has not improved understanding of the parameters and their impact on the complex construction process (Panas and Pantouvakis 2010).

In rare cases where detailed parameter documentation was carried out together with documentation of the output parameter (achieved labor productivity), data-driven methods have been employed to identify critical parameters. Data-driven methods like correlation analysis, feature reduction, and principal component analysis have been useful in identifying critical parameters, resulting in better prediction ability (Gray and MacDonell 1997). In a recent publication, Moselhi and Khan (2012) compared various parameter ranking approaches using actual data from formwork installation and based on nine parameters; they identified temperature as the top ranking parameter. However, no past study has verified the expert’s knowledge regarding parameters affecting labor productivity on a given project. Such verification will require the development and testing of a data-driven approach capable of revealing which parameters affect labor productivity for the same projects for which the experts’ knowledge was collected. Unfortunately, documenting the number of parameters known to affect labor productivity is not an easy task so it has rarely been tried let alone achieved to a level at which the actual parameters could be determined using data-driven techniques (Moselhi and Khan 2012).

**METHODOLOGY**

This paper is part of a larger research study that explores the development of construction labor productivity prediction models. The methodology proposed here involves the identification and quantification of parameters together with productivity data and the establishment of critical parameters using expert-based and data-driven approaches. The methodology first identifies the critical parameters for a given project activity using the expert-based approach, and then verifies the critical parameters through the data-driven approach using field data from the same project.
Parameter identification

We prepared an initial list of 169 parameters (factors and practices) affecting labor productivity by analyzing existing literature related to Albertan and North American construction projects (Knight and Fayek 2000; AbouRizk et al. 2001; Oduba 2002; Park 2002; Liberda et al. 2003; Chan et al. 2004; PMI 2004; CII 2006; Jergeas 2009; Dai et al. 2009). The identified practices related to project management practices specified in the Project Management Institute’s Project Management Body of Knowledge and the Construction Industry Institute’s best practices (PMI 2004, Park 2002). We categorized the initial parameters (factors and practices) into a hierarchal structure of activity, project, organizational, provincial, national, and global levels, which we adopted for its capacity to display and relate the parameters in a coherent way (Knight and Fayek 2000).

Parameter quantification

We then developed a measurement scheme for all the identified parameters using appropriate quantitative and qualitative measurement scales. For example, the scales of measure for quantitative parameters (temperature) and qualitative parameters (complexity of task) are shown in Table 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Scale of measure (unit)</th>
<th>Cycle</th>
<th>Data source</th>
<th>Sample data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>Real number (°C)</td>
<td>Daily</td>
<td>Researcher</td>
<td>-5</td>
</tr>
<tr>
<td>Complexity of task</td>
<td>Rating (1–5 predetermined scale)</td>
<td>Daily</td>
<td>Crew members</td>
<td>2</td>
</tr>
</tbody>
</table>

Quantitative parameters have well defined numerical measures (e.g., crew size in terms of number of workers). Qualitative parameters (e.g., complexity of task) lack a well defined measurement scheme, hence, for each of them, a pre-determined 1–5 scale has been developed based on sub-factors referred to as variables. These variables are based on explicit concepts associated with the parameter. For example, the complexity of task parameter’s sub-factors were based on the number of alternatives, known means, and number of sub-tasks (Campbell 1988). The sub-factors, or variables, enabled the development an explicit qualitative measurement scale, as shown in Table 2.

In applying the quantification process to formulate appropriate measurement scales for all quantitative and qualitative parameters (factors and practices), we established a total of 314 input variables. Of this number, 36 are to be collected on daily basis, 63 weekly, 52 monthly, 156 initially, and 7 initially with a change in crew. Each parameter’s level within the project hierarchy determined its data source in terms of the appropriate project member to provide accurate values. The assumed variability of each parameter dictated its data collection cycle.
Table 2. Qualitative measurement scale: Complexity of task

<table>
<thead>
<tr>
<th>Scale</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>Many</em> alternatives, <em>well known</em> means, <em>very low</em> number of subtasks</td>
</tr>
<tr>
<td>2</td>
<td><em>Some</em> alternatives, <em>well known</em> means, <em>low</em> number of subtasks</td>
</tr>
<tr>
<td>3</td>
<td><em>Few</em> alternatives, <em>known</em> means, <em>high</em> number of subtasks</td>
</tr>
<tr>
<td>4</td>
<td><em>Few</em> alternatives, <em>unknown</em> means, <em>very high</em> number of subtasks</td>
</tr>
<tr>
<td>5</td>
<td><em>Very few</em> alternatives, <em>unknown</em> means, <em>very high</em> number of subtasks</td>
</tr>
</tbody>
</table>

Output parameter quantification

Determination of the output parameter (labor productivity) was based on the ratio of output (installed quantity) to input (total manhours). We collected installed daily quantities using a measurement of units, as in the project under study, the relevant activities could be completed in less than a shift and so counting the activity units completed was easily and accurately done. However, in cases without these conditions, a milestone method would be applied.

Critical parameter identification: Expert based approach

Researchers tend to consider construction labor productivity a micro (activity) level subject; however, parameters affecting labor productivity are multilevel, ranging from micro (activity), to meso (project), and to macro (organizational, provincial, national, and global) levels. In determining these parameters, it is important to consider construction practitioners’ expertise specific to a given project setting. To this end, we developed two surveys based on the practitioners’ level of involvement with a given project: 1) a project management (PM) survey (project manager/supervisor) and 2) a trade survey (craftsperson/foreman).

The surveys each have two sections: a background section to collect responses related to demographic information (i.e., age, gender, and education level), position of respondents, and overall satisfaction of the project productivity, and a parameter section including questions designed to assess 1) agreement or frequency and 2) impact for each parameter (Dai et al. 2009). The parameter section features agreement questions for parameters whose existence rather than frequency is of importance (e.g., the site has a very good material order tracking system) and frequency questions for parameters that occurred with varying frequencies (e.g., power equipment breakdown). Only activity level parameters related to materials and equipment required frequency questions. The survey design resembles that of the Voice of the Worker survey (CII 2006; Dai et al. 2009); bipolar seven-point Likert scales structured into positively and negatively worded statements collect ratings on agreement/frequency and impact of parameters, and the two scales then enable the analysis and ranking of the parameters. The parameters are presented in both positively and negatively worded statements in order to improve the accuracy of responses by requiring respondents to remain alert. The PM survey addresses a total of 141 parameters, all designed as agreement type statements, and refers to some micro (activity), some meso (project), and all macro (organizational, provincial, national, and global) level parameters. The trade survey addresses a total of 89 parameters derived from all micro (activity) and some meso (project) level
parameters. Only nine parameters in the trade survey were designed as frequency type statements (all negatively worded), while the rest were designed as agreement type statements.

**Critical parameter identification: Data-driven approach**

As discussed in the literature review, researchers have previously identified critical parameters using factor surveys based on a limited number of parameters. This approach has not provided a better understanding of the parameters and their impact on labor productivity (Panas and Pantouvakis 2010). To develop more accurate labor productivity models, the methodology described in this paper utilizes a data-driven approach based on detailed qualitative and quantitative parameter data collection; to determine the effects of project factors and practices on labor productivity, the methodology applies appropriate measurement scales and data collection techniques in various project settings. The data-driven approach relies on field data collected on 314 input variables used to measure the 169 project parameters for the same project for which the two expert groups (PM and trade) have shared their knowledge regarding the critical parameters. The parameter and productivity data, collected over a minimum period of three months, is then used to identify the relevance of each parameter based its relation to productivity.

We investigated the use of a data-driven filter feature selection technique that produces a rank of relevant parameters in order to determine the critical parameters affecting labor productivity using field data already collected as part of this study. Filter feature selection is suitable for this purpose as it has the ability to preserve the original representation of the parameters while providing better understanding of the underlying process that generated the data (Guyon and Elisseeff 2003). Filter feature selection techniques are indispensable in supervised learning when the number of parameters is very large (Marono et al. 2007). Filter feature selection methods identify critical parameters via a ranking algorithm; Marono et al. (2007) recommend the RELIEF algorithm for its “low bias”, “ability to include interaction among parameters”, and its potential to “capture local dependencies that other methods miss”. The RELIEF algorithm estimates the statistical relevance of the parameters in the neighborhoods around target samples. For each target sample, the algorithm finds the “hit” sample—the nearest sample in variable space of the same category—and measures the distance between the target and hit samples. It also finds the nearest sample of the other category—the “miss” sample—and then performs the same procedure. RELIEF then uses the difference between those measured distances for the weight of the target parameter for ranking (Marono et al. 2007).

Thus, by applying the RELIEF algorithm to the collected parameters and productivity field data, we have established a critical parameter list. The established list ranks the critical parameters based on data analysis of their relationships to labor productivity, and provides an important source of knowledge to verify experts’ perspectives on the critical parameters. However, this approach is highly dependent on the size of the data set. Extensive data collection from a number of projects is required in order to acquire adequate data for a comprehensive investigation of the critical parameter list using filter feature selection techniques.
Data Analysis

The data used for analysis in this paper was collected from electrical works in a residential project in Edmonton, Alberta, Canada. Following the methodology previously described, we first collected a survey from 22 respondents directly involved with the project’s electrical work. Of these respondents, 5 were from the PM level and occupied the roles of project manager, project control, estimator, and safety officer while 17 were from the trade level and occupied the roles of foreman and craftsperson. We conducted the survey with all of the possible project respondents, resulting in full sampling.

The respondents at the PM level had an average of 6 years’ experience at their stated occupations. All of these respondents were above 30 years of age. The respondents at the trade level were all electricians with an average of 7 years’ experience; 53% were above 30 years of age. The agreement/frequency and impact scale values for each of the statements chosen by the respondents were collected and analyzed to determine the evaluation scores. The evaluation scores were then used to rank each parameter for its positive and negative effect on labor productivity. For example, we computed the positive evaluation scores for positively worded agreement parameter statements as follows: First, the weighted percentage of agreement (RA) with a given parameter statement by a number of respondents was computed using Equation (1), where the maximum possible weighted percentage of agreement is equal to 50, in the case where \( C = 100\% \). In Equation (1), \( A \) = percentage of respondents rating the positively worded parameter as 5 (Slightly Agree), \( B \) = percentage of respondents rating the positively worded parameter as 6 (Agree), and \( C \) = percentage of respondents rating the positively worded parameter as 7 (Strongly Agree).

\[
R_A = \left( \frac{A \times 1 + B \times 2 + C \times 3}{5} \right) \times 100
\]

(1)

The weighted percentage of positive impact (Ip) of a given agreement type parameter statement by a number of respondents was computed using Equation (2), where the maximum possible weighted percentage of impact is equal to 50, in the case where \( Z = 100\% \). In Equation (2), \( X \) = percentage of respondents rating the impact of the parameter as 5 (Slightly Positive), \( Y \) = percentage of respondents rating the impact of the parameter as 6 (Positive), and \( Z \) = percentage of respondents rating the impact of the parameter as 7 (Strongly Positive).

\[
I_p = \left( \frac{X \times 1 + Y \times 2 + Z \times 3}{5} \right) \times 100
\]

(2)

Next, for the positive effect of a positively worded parameter the evaluation index and evaluation score was computed using equations (3) and (4). First, the evaluation index based on the product of the agreement and impact scores was computed. Then, the evaluation score was computed by dividing the evaluation index of a given parameter by the maximum possible evaluation score. The maximum possible evaluation score is equal to 2500, which is the product of the maximum values of agreement (50) and impact (50). Similarly, we analyzed the frequency type
parameters using frequency and impact rating computations. Finally, we used the evaluation scores for positive and negative effect to rank each parameter in the respective surveys according to its positive or negative effect on labor productivity.

\begin{equation}
\text{Evaluation Index}_{AP(+\text{or} -)} = R_A \times I_P
\end{equation}

\begin{equation}
\text{Evaluation Score}_{AP(+\text{or} -)} = \frac{\text{Evaluation Index}_{AP(+\text{or} -)}}{2500} \times 100
\end{equation}

For the data-driven analysis, we collected the 169 parameters and their associated detailed variables for a five-month period from the same project along with the productivity values of several electrical activities. We identified a total of 32 instances, made up of parameter and labor productivity (metres/manhour) data points for the wire pulling activity, as suitable for illustrating the establishment of critical parameters using the data-driven approach. The instances were loaded on the WEKA data mining tool, and using the inbuilt RELIEF algorithm (ReliefFAttributeEval), the respective parameters were ranked. The rankings of the 169 parameters also included the weighted distance measure to identify the statistical relevance of each parameter to labor productivity. To facilitate comparison with expert-based scores, the weighted distance measures of the ranked parameters were normalized against the maximum weight of the top ranked parameter.

Table 3 shows the top 10 parameters and their respective evaluation score values, based on the data-driven approach, in Column (a). These top 10 parameters provided a basis for comparison and verification of the experts’ perspectives on critical parameters. In columns (b) and (c), respectively, Table 3 also summarizes the evaluation scores for the top 10 parameters collected from both the PM survey and trade survey results. The comparison of the PM and trade perspectives was carried out using the difference between the absolute evaluation scores and the data-driven values (Dai et al. 2009). The results shown in Table 3 indicate that there is a major gap between the three rankings of the critical parameters affecting electrical works in the project.

The data-driven results indicate that the weather (humidity and wind speed) and work conditions (noise and site access) parameters strongly relate to the actual productivity and its variability. The evaluation score differences shown in Table 3 indicate that the greatest difference is observed for noise in work conditions. The least difference between the data-driven approach and the PM perspective related to craftperson motivation; similarly, between the data-driven approach and trade perspective the least difference related to cooperation between craftsmen. The overall comparison between the PM and trade perspectives in terms of sum of score differences indicated that the trade perspective is closest to the data-driven one.

As previously indicated, a limitation of this study is the reliance on the size of the data set. The data set used for analysis was small, with only 32 instances of the wire pulling activity. The results presented are an illustration of the proposed approach. Additional data on electrical activities are being collected from residential and commercial projects in order to acquire adequate data for verification of the results and further investigation of the data-driven ranking approach.
Table 3. Perspectives on critical parameters: Expert and data-driven results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Evaluation score</th>
<th>Evaluation score difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data-Driven (a)</td>
<td>PM (b)</td>
</tr>
<tr>
<td>Work conditions (noise)</td>
<td>100.00</td>
<td>19.05</td>
</tr>
<tr>
<td>Cooperation between craftsmen</td>
<td>76.05</td>
<td>12.07</td>
</tr>
<tr>
<td>craftsmen flexibility (task changes)</td>
<td>59.79</td>
<td>12.07</td>
</tr>
<tr>
<td>Weather (humidity)</td>
<td>55.19</td>
<td>37.28</td>
</tr>
<tr>
<td>Site access</td>
<td>46.31</td>
<td>21.30</td>
</tr>
<tr>
<td>Misplacement of tools</td>
<td>42.88</td>
<td>0.00</td>
</tr>
<tr>
<td>Natural gas price</td>
<td>42.10</td>
<td>20.63</td>
</tr>
<tr>
<td>Weather (wind speed)</td>
<td>41.87</td>
<td>37.28</td>
</tr>
<tr>
<td>Unmet labor requirement</td>
<td>38.60</td>
<td>24.85</td>
</tr>
<tr>
<td>Craftsperson motivation</td>
<td>36.59</td>
<td>39.68</td>
</tr>
<tr>
<td>Sum</td>
<td>321.35</td>
<td>259.91</td>
</tr>
</tbody>
</table>

CONCLUSIONS AND FUTURE RESEARCH

In this paper, three perspectives for ranking and comparing the critical parameters affecting labor productivity were established and compared to identify gaps or differences. The first two perspectives were based on expert opinions at the project and trade levels, while the third was based on actual data analysis using filter feature selection. The comparison between the perspectives indicated that there is a major gap between the three perspectives. The preliminary results illustrated in this paper suggest that a data-driven approach facilitated by field parameter data collection be employed to further explore knowledge gaps and improve the understanding of the critical parameters and their relation to labor productivity. The results reveal the necessity of verifying expert-based critical parameters in productivity studies. Further research based on continued data collection from projects with similar electrical activities is being conducted. In the short term, we will continue to test and verify the methodology with several filter ranking algorithms and to identify knowledge gaps of experts. In the long term, we will develop construction labor productivity prediction models based on the identified critical parameters so as to improve labor productivity performance and increase the accuracy of productivity prediction in different project settings.

REFERENCES


