# CHAPTER VI

# USING QUANTITATIVE APPROACHES TO ENHANCE CONSTRUCTION PERFORMANCE THROUGH DATA CAPTURED FROM MOBILE DEVICES

**Aziz, Z., Barker, C. and Tezel, B.A.**

**University of Salford**

**UK**

## ABSTRACT

*Lack of primary empirical data within the Architecture, Engineering and Construction industry has in the past hindered developments of academic research in the field of performance measurement in construction. This paper presents quantitative analysis of construction production data captured through mobile information systems and provides empirical evidence that data captured through mobile devices has the potential to deliver new and enhanced methods for performance measurement and enhancement. Relational data gathered through mobile devices is used to generate metrics against which construction issue resolution performance is measured. Paper also discusses various methods for early identification and visualisation of performance deviations. Research approach and findings can be used for the development of academic performance measurement frameworks and also as an evidence base for further development by industry in the field of performance enhancement. The paper contributes insight regarding innovative ways to interrogate construction production data and provide stimulus to others to develop the methods and approaches taken.*

## BACKGROUND

Recent reviews of construction productivity performance indicate that the industry fell short in comparison to manufacturing and services based industry sectors. Some of the key factors hampering construction productivity include issues with quality, use of project controls and adequate levels of supervision (Merrow, et al., 2009). Similar observations have been made in UK Construction Industry Performance Report (2014), indicating that majority of construction projects continues to fail timely completion. This is coupled with falling profitability and client dissatisfaction with regards to product quality, service and value for money. While quantitative data presented in performance reports are subject to interpretation, it is obvious that there is tremendous potential of productivity growth within Construction Sector.

Review of academic literature and construction industry reports highlight inadequacies in terms of forecasting project costs, duration and other issues and inability of construction contractors to deliver quality products and services within a resource constrained environment. Need for process innovation to help industry deliver greater productivity and quality has long been identified. Also, literature identifies a need for more accurate data relating to on-site construction activities, and to develop process intelligence through better use of project data, to help improve the accuracy of project planning. The literature provides a considerable amount of existing research identifying weaknesses in current performance measurement methods including lack of availability of empirical data.

Recent developments in Information Communication Technologies make available a large and detailed digital sample of production performance data to test academic/theoretical hypotheses. After a review of academic research in the field of Total Quality Management (TQM) and its evolution from Deming's Management Method, (Rungtusanatham, et al., 2003) make reference to "The power of Primary Data". One of the issues they identify is the restrictions on further academic progress in the field of TQM without first validating Deming's principles against a sample of good empirical data. They suggest that until they can move beyond secondary analysis of potentially inaccurate and "weak" data, it will continue to stifle developments in this field.

With the developing Information Technologies (IT), the spectrum of automatic quantitative data capturing systems has been widening. IT based data capture systems minimise human intervention, bias and error in the data collection process. One of the important automated data collection implications in construction is the use of sensors. Sensors convert a physical parameter (e.g., temperature, distance, humidity, displacement, flow etc.) into an electrically measurable signal. Specific technologies such as Radio Frequency Identification (RFID), Near Field Communication (NFC) and Bluetooth beacons (nodes) are used for the automatic identification (AutoID) and monitoring of construction materials, equipment, plant and personnel. A Global Positioning System (GPS) and a Geographic Information System (GIS) can present data related to geography and location to the researcher. By providing an accurate point cloud at a reasonable speed, 3D Laser scanning applications (i.e., the LiDAR and LADAR technologies) automatically capture data on the shape/features of a surface, space or topography. The technical capabilities, IT architecture, cost and the researcher's experience in using those data collection systems should be well defined.

Literature review identified range of issues relating to the quality and accuracy of industry performance data, and consequently the capability of any existing performance measurement methods reliant upon it. It identified that data relating to production issues and performance is disparate, inconsistent, often subjective, and highly retrospective. There is also evidence to suggest a lack of much needed empirical primary data relating to construction performance. The main findings from literature review are highlighted in Table 1.

*Table 1: Performance measurement issues identified from literature review*

| Issue Identified | Issue Description | Supporting References |
|---|---|---|
| Consistency and Standardisation | • A lack of standard relational data for performance measurement and organisational learning; | (Cheng & Wu, 2012; Dissanayake & Fayek, 2008) |

| | | |
|---|---|---|
| | • Inability to identify relationships and patterns between performance outcomes and on site decision making | |
| Complexity and Data Quality | • The subjective nature of performance measurement data. Not enough detail, visibility or science surrounding performance measures.<br>• A lack of detail, accuracy and scale for performance metrics | (Cheng & Wu, 2012, Akhavian & Behzadan, 2012) |
| Context and Communication | • Poor context for causes relating to project performance failures;<br>• A lack of detailed industry metrics to measure performance against benchmarks | (Son, et al., 2012). |
| Lagging Metrics and Leading Metrics | • Retrospective measurement of performance outcomes that incur extensive lag time<br>• A lack of forecasting and lead performance measures | (Barber, 2004) |
| Visualisation | • A need for improved analytical reasoning through visualisation, for project performance issues | Russell et al (2009) |

***Research Approach***

This chapter presents a quantitative analysis of production data captured through mobile devices on a wide range of projects, undertaken by a multi-national infrastructure and services company. Before explaining the details of the research approach, the particular research philosophy that contains important assumptions about the researchers' view of the relationship between knowledge and the process by which it is developed should be clarified. The adopted research philosophy will indicate four important underlying assumptions (Saunders et al., 2009); (i) the ontological stance as the researchers' view of reality or nature of being, (ii) the epistemological stance as the researchers' view on what constitutes an acceptable knowledge, (iii) the axiological stance as the researchers' view of the role of values in research and (iv) the data collection techniques as the means to obtain data to generate knowledge.

The adopted research philosophy in this particular research is positivism. Positivists believe that reality is stable and can be observed and described from an objective viewpoint (Punch, 2005), i.e. without interfering with the phenomena being studied. Therefore, ontologically, reality is external, objective and independent of social actors. To positivists, reality must be investigated through the rigorous process of scientific inquiry. The positivist philosophy calls for focusing on facts and locating causality between variables (Easterby-Smith et al., 2012). Thus, epistemologically, only observable phenomena can provide credible data. As reality is external and out of researchers' control, axiologically, research is undertaken in a value-free way, independent of researchers values (objective stance). Data collection and analysis methods are generally quantitative from large samples with highly structured data collection approaches.

Along with the philosophical stance of a research, the nature of data accessible to researchers is also a defining parameter in research approaches. Generally, when numerical or quantifiable data or "hard data" are more readily-accessible as in this particular research, the quantitative research methodology is employed (Neuman, 2007). Quantitative research is "explaining phenomena by collecting numerical data that are analysed using mathematically based methods (in particular statistics)" (Aliaga and Gunderson, 2005). Quantitative research tends to be explanatory and generally provide 'snapshots' or instantaneous results and, are used to address questions such as what, how much, how many? (Fellows and Liu, 2015). Quantitative research allows for a broader study (wider breadth), involving a greater number of subjects, and enhancing the generalisability of the results with the capabilities of replicability and comparison with similar studies (Kruger, 2003). However, it doesn't generally yield an in-depth analysis of the studied phenomenon as the results are limited, providing numerical descriptions rather than detailed narratives with less elaborate accounts of human perception (Bryman, 2012). The quantitative research approach is well suited for the deductive reasoning in which hypothesis are tested with experiments, statistical methods, observations etc. for generalisable confirmations or rejections (top down approach), as opposed to the inductive reasoning in which observations lead to theories (bottom up approach) (Blaikie, 2009). Quantitative researchers design studies that enable testing hypotheses, which are tentative explanations that account for a set of facts open to further investigation. Both the quantitate and qualitative research methodologies and data collection methods can be simultaneously employed in the same research to take advantage of their particular strengths (multi-method research). The study outlined in this paper however, adopts a mono-method approach that exploits solely the quantitative research methodology and data collection methods as they permit the inference from and gathering of a greater number of quantitative data elements in a relatively shorter time and at a relatively lower cost (Balnaves and Caputi, 2001).

The research in question aims to develop an understanding on how emerging structured and detailed quantitative data, created and made possible using a plethora of mobile ICT devices, could be applied to meet performance measurement requirements of construction operations. Key objectives included: a) assessment of whether mobile data can provide the detail, accuracy and scale for performance measurements and metrics; b) assessment of whether mobile data could reduce lag times associated with the identification of poor performance and c) assessment of whether mobile data could provide improved visualisation of project performance. Those points identified in the key objectives constitute the starting point or the hypothetical stance of the quantitative deductive approach with some explorative motives.

As the quantitative research methodology generally follows a linear research path (hypothesis-data collection-analysis), speaks the language of statistics with variables and hypotheses, and emphasises precisely measuring quantitative or "hard" data and testing hypotheses that are linked to general causal explanations, one needs employ a relevant quantitative data collection method or data collection means through which quantifiable data are obtained for further analysis and manipulation (Gray, 2004). Observations, tests/experiments, surveys (questionnaires), and archive, document and secondary data (i.e. databases, company records etc.) studies are the frequently used quantitative data collection methods (Creswell, 2013). In this particular case, existing (secondary) construction production data were statistically analysed to determine their effectiveness to enhance the existing performance measurement methods. Secondary data provide researchers with a relatively quick and cost-effective data source that may otherwise not be acquired from first hand. They also enable both cross-sectional (one specific point in time) and longitudinal (extending over a period of time) analysis (Vartanian, 2010). However, because secondary data were collected by another entity at some point in the past for another purpose before the research effort, secondary analysts have no opportunity to influence the initial data collection method in terms of data quality (completeness and consistency in the data set), bias and compatibility with their research aims (Smith and Smith, 2008). This research demonstrates a longitudinal study of a secondary data source with its specific limitations (e.g. data completeness) underlined in the discussion section.

The secondary data source, production data sample, used in this study for quantitative analysis contained a large number of tables and data fields. A data sample is the representative subset of the whole population, in this case the whole construction productivity data. The number of records relating to key data variables used for query and analysis as part of this research has been outlined In Table 2. A variable is a particular characteristics of the studied phenomenon that vary or have different values. Variables are used to statistically test hypotheses or assumptions. It is obvious from Table 2 that there is a deficit in the number of "date due" and "date closed" records when compared to the total number of production issues, immediately highlighting a shortfall in the number of resolution periods available to contribute towards averages.

*Table 2: Count of data variables used for query and analysis*

| **Project Variables** | |
|---|---|
| **Total Number of Projects** | 34 |
| **Total Number Project Type / Construction Sector** | 10 |
| **Geographical Variables** | |
| **Total Number of Regions** | 7 |
| **Company and User Variables** | |
| **Total Number Companies Reporting Issues** | 716 |
| **Total Number of Creator Roles / Professions** | 8 |
| **Total Number Users Creating Data** | 259 |
| **Total Number of Trade Classifications** | 46 |
| **Issue Variables** | |
| **Number of General Issue Types** | 5 |
| **Total Number of Production Issues** | 149,733 |
| **Total Number of Date Due Records** | 136,628 |
| **Total Number of Date Closed Records** | 140,571 |

After obtaining the raw data sample, which requires relatively less effort with readily available secondary data sources, the data analysis process necessitates management and preparation of the data sample in terms of data coding, data entry and data consistency for further descriptive and inferential statistical analysis. The data preparation and analysis procedures for the study were largely automated to minimise human bias, error and intervention. The data sample was collated into a structured cloud database with the aid of a proprietary application that utilises mobile touch screen tablets as the primary means for data capture (Figure 1). The database also has a web interface that allows the mobile data to be updated, amended and reviewed via a desktop computer. An export of the cloud data has been provided as a Microsoft Access database file. The database file is intended to be used for grouping and running calculations against large numbers of records. The detailed findings were exported to Microsoft Excel for more detailed analysis and visualisation. The approach requires the data sample to be updated via SQL query to generate elapsed periods between the creation and resolution of issues. The resolution periods were then used to determine performance averages, standard deviations and z scores. Averages were established by grouping issue records according to the content of geographical, trade and role variables.

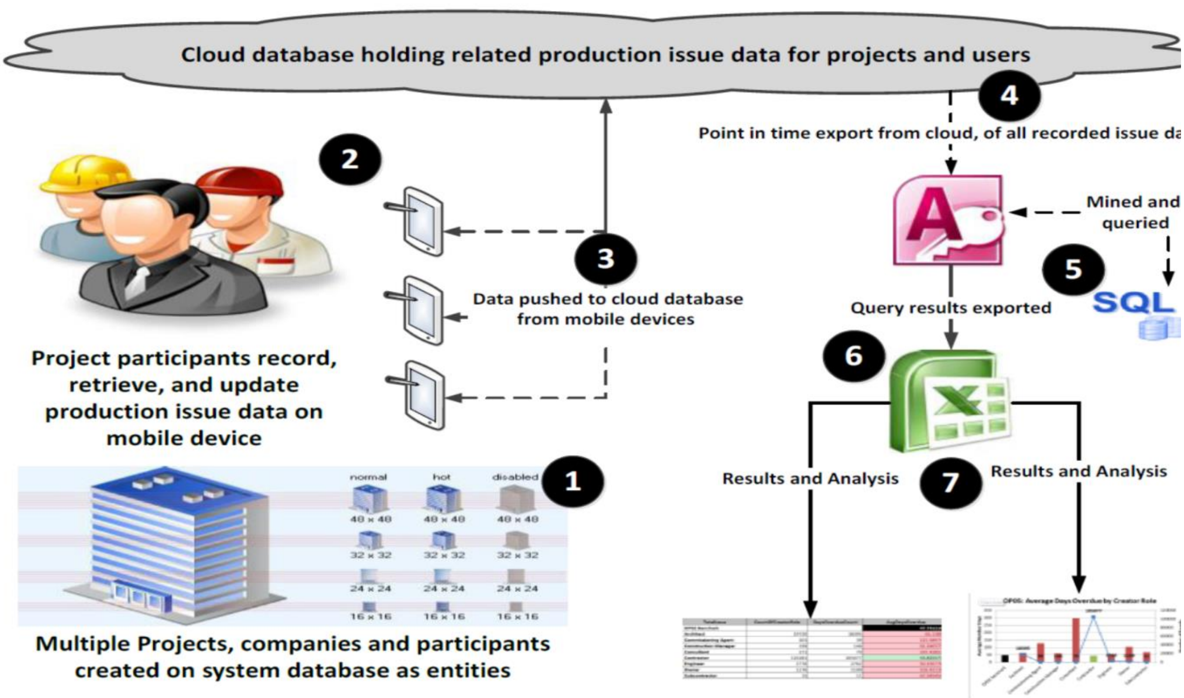


*Figure 1: Data collection and analysis strategy*

The production issues contained in the sample have been recorded across multiple projects, and deliberately structured to hold a number of standard data variables. The variables deemed to be of interest for this research included Project type, Issue classification, location, Trade, Dates relating to issue identification, required resolution, and actual resolution and roles and professions of the individuals capturing the issues. These variables were used as the basis for SQL queries, designed to separate records into groups, for further analysis and performance calculation.

Approach to quantitative analysis included generating a query and running it against the date entries in the database, to calculate and create resolution periods, against each production issue according to issue type, project type, creator role and trade variables. Time was employed as a standard unit (days), and used as a metric for performance measurement. Once defined, the averages were compared against their hierarchical benchmarks to establish any variance to performance averages. Further, statistical analysis were conducted to determine standard deviations for resolution periods, and from this, z scores were used to identify proximity to average resolution period. A z-score or a standard score indicates how many standard deviations an element is from the mean. It is a standardised value that lets researchers compare raw data values between different data sets allowing the comparison of "apples" and "oranges" by converting raw scores to standardised scores relative to population mean(s) (i.e. comparing one project's productivity values to another's productivity values) (Field, 2009). This was identified as a proposed alternative and more robust statistical measurement for performance against a benchmark value. The following statistical analysis methods were used to undertake quantitative analysis.

**Average Resolution Period:**
Determination of the average resolution period was the first step in quantitative analysis, to help evolve issue data into a standard measurement of performance. This was based on the premise that project performance is reliant on the timely resolution of production related issues. Two averages were established for timely resolution of production issues.

1. ***"Total Resolution Period"*** i.e. the average time (in days) to resolve a production issue;
2. ***"Overdue Resolution Period"*** i.e. the average time (in days) in relation to the required resolution period.

$$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{n}}$$

Standard Deviation Formula

$$Z = \frac{x - \mu}{s}$$

Z-Scores Formula

| | |
|---|---|
| S | Total |
| s | Standard Deviation (Square Root of Variance) |
| µ | Average resolution period |
| n | Number of records |
| x | Individual Resolution period |

*Figure 2: Formulas used and legend to support analysis of average resolution, standard deviation and z-score calculations*

**Standard deviations for performance records**

In this research, standard deviation values were used as an indication of the size of any variability, spread and distribution of resolution periods contained within the sample data. The larger the standard deviation, the more disparate and varied the sample data is assumed to be.

**Z Scores reflecting performance in relation to standard deviations**

The z score was used to determine how many standard deviations from the sample norm (average) any specific resolution periods are. If successful, this was used to evidence the capability to measure performance beyond simple mean calculations. It also confirms whether results sit above or below average resolution periods due to their positive or negative score. The z scores were then converted to a percentage, to show where the records sit in respect to their performance against other data in the samples i.e. the percentage of records with quicker or slower response and resolution periods.

To better structure and organise the analysis process, the data sample was divided into smaller related data tables using Standard Query Language (SQL). SQL queries were named according to query type and its intended use and the level at which it groups any data.

**Data Distribution:** The sample was queried to count the number of records by project type and region and to establish whether data was evenly distributed or clustered within geographical areas. The results indicate that records were clustered and not uniform, with heavier concentrations in the “NTX” and “WDC” regions, among projects types “Education”, “Healthcare” and “Other Ancillary Facilities”. The numerical distribution and 3D column chart is shown in Figure 3.

| Project Type | Count of Issues | FLA | GEO | HOU | HSC | MHF | NTX | WDC |
|---|---|---|---|---|---|---|---|---|
| EDUCATION PREK-12 | 38580 | 14277 | 1405 | | | | 22898 | |
| HEALTHCARE | 34782 | | | | | | 8245 | 26537 |
| HIGHER EDUCATION_TRAINING | 20983 | 20333 | | | | | 650 | |
| OTHER_ANCILLARY FACILITIES | 20975 | | | 16 | | | | 20959 |
| MULTI FAMILY | 15687 | | | | 9538 | 6149 | | |
| OFFICE_CORPORATE | 10882 | 9608 | | | 974 | | 12 | 288 |
| AIRPORT | 5983 | | | | | | 5983 | |
| PUBLIC USE_CIVIC | 1051 | 695 | | | 182 | | 174 | |
| POWER_ENERGY | 590 | | | | | | | 590 |
| HOSPITALITY | 220 | | | | 220 | | | |

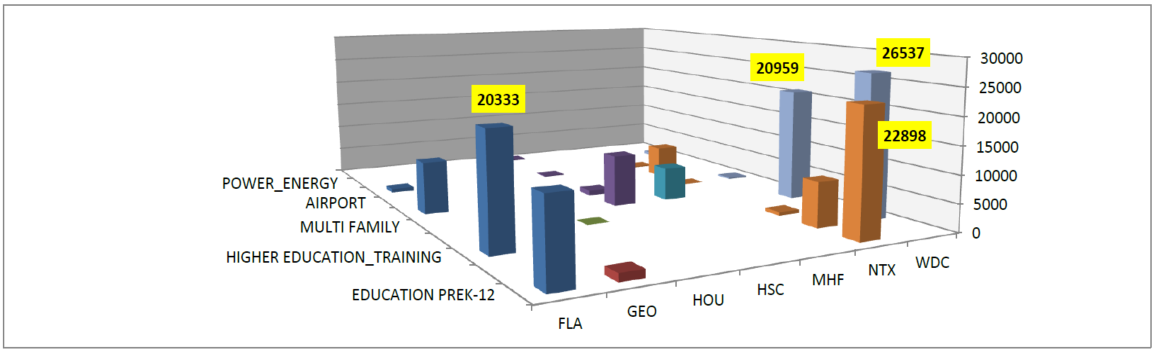


*Figure 3: 3D Column Chart showing distribution of issue records by project type and region*

Figure 4 below identifies a dominant record set captured by the “Contractor” having provided 83.79% of issues used to generate total resolution periods (117,787 of 140,571), and another majority related to general issue of type “Punch List”, with 85.8% of issues (120,665 of 140,571).

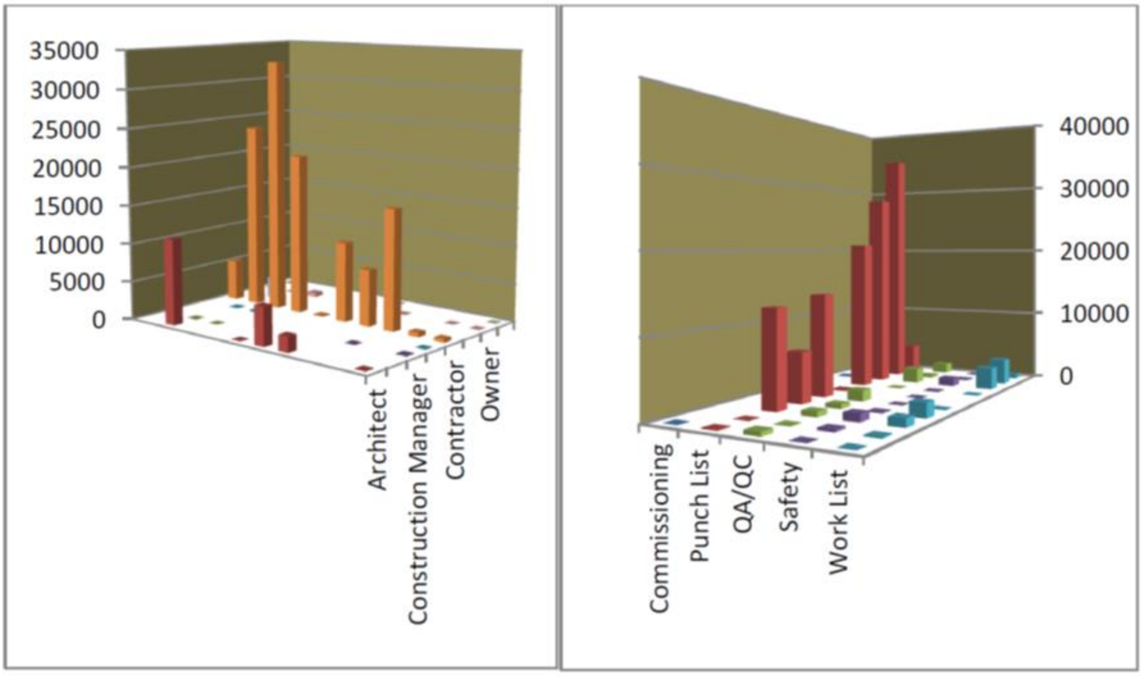


*Figure 4: Record distribution by creator role and general issue type used to calculate total resolution periods.*

Figure 5 represent percentage of data entries that are missing for every data field used to conduct SQL queries for analysis. This shows a generally good level of information has been provided with only 1.99% of data fields (35,796 of a total 1,796,796) containing blank records. Most of the missing entries related to date issue was due or closed.

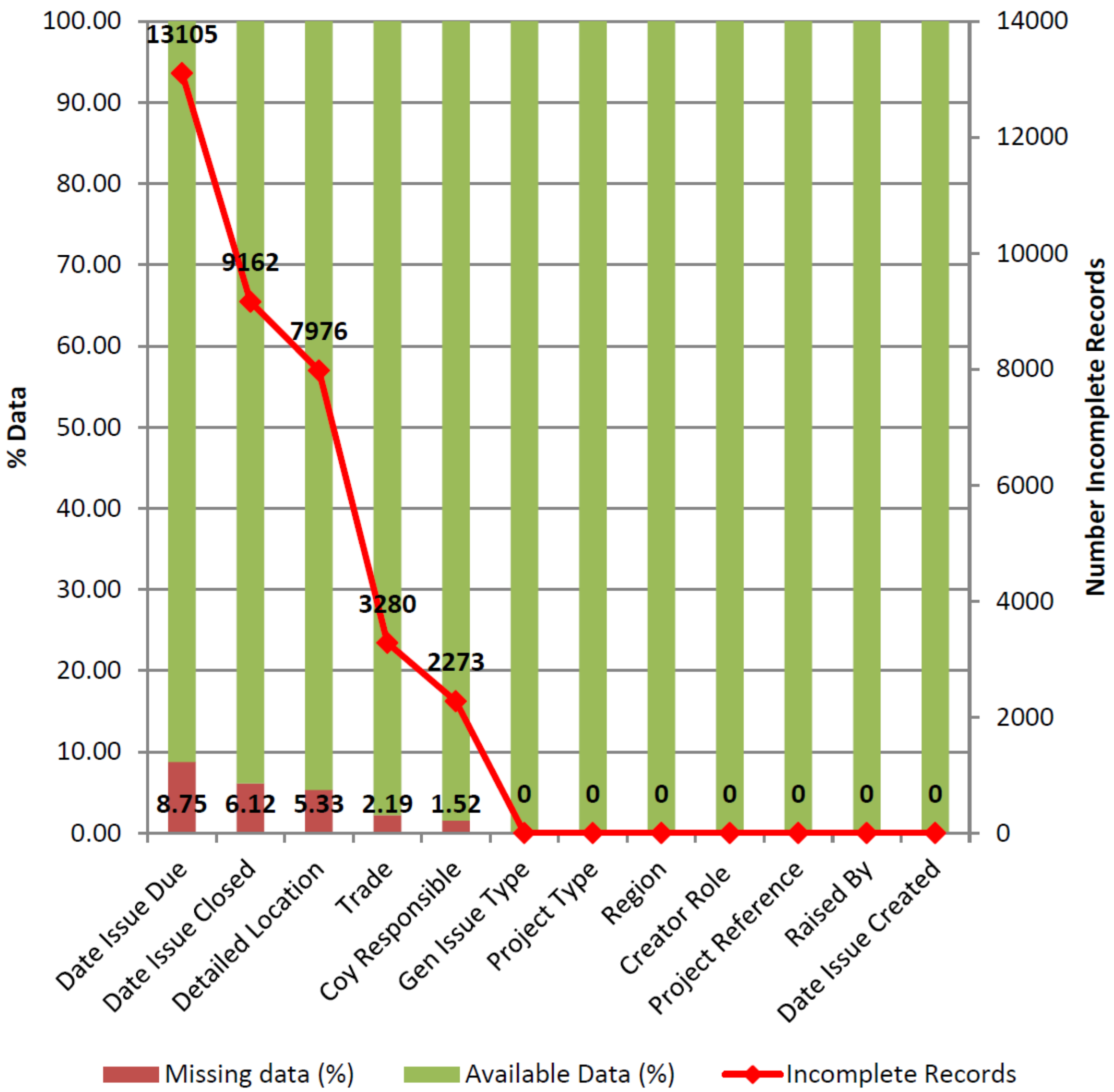


**Figure 5:** ***Quantity of missing entries for queried data field in the sample***

**Z-Score:** A database query was created to establish the global organisation average and standard deviation for overdue resolution periods and total resolution periods from the data sample, to generate a metric for measurement that was statistically viable as an organisational performance standard. A further query of resolution periods for all issues was made against 4 projects. The data records were grouped by project and the z scores then calculated for individual issue resolution periods, to establish the number of standard deviations from the data sample mean. The results were plotted into scatter graphs to help visualise general performance distribution. The results of the distribution of z scores showed that 2 projects (i.e. PR20 and PR06) performed relatively well in context to organisational delivery standards with the majority of resolution periods sitting within + / -1 standard deviation. The other 2 projects (PR15 and PR04) did not appear to perform very well, with large volumes of records creeping up to and beyond 2 standard deviations.

## Discussion

While the quantitative analysis approach relied on data sample of a considerable size, the content was inconsistent and the levels of available information varied depending on what part of the sample was being analysed. The data sample used was perfectly adequate for the research, with availability of key fields related to who, when, where and what, being available and available for analysis. However, while designing the research process, it was initially presumed to establish time based performance as standard units to measure performance. However, this approach was limited due to the number of missing date fields within the sample.

Quantitative analyses undertaken in this research has highlighted a number of development areas. It is apparent that the quality of the sample is largely dependent on the process used to prepare the system and collate the data. The sporadic and inconsistent volumes of data within the sample suggest that the organisation is using the platform in an ad hoc capacity rather than as a standard, systemic means to capture and communicate production issues. The user base is also significantly biased towards contracting staff. The system used to capture the data may well benefit from employing more controls to the data capture process, to ensure fields such as date required and trade are made mandatory. Further standardisation of automated functions would also benefit the data sample. The removal of generic defaults for data entry such as "other" would also help prevent large portions of intelligent data from becoming unwieldy as found with issue types.

The number of unsolved items in the sample, and the considerable lengths of time elapsed before the majority of issue are closed; again suggest process issues surrounding the implementation and management of the platform. It would appear that items are raised more quickly than they are closed, and that perhaps there is a lack of accountability for ensuring the system is updated.

This research has set out to review the potential for mobile data capture to enhance performance measurement on construction projects. The research has identified potential areas and methods for further industry development where mobile data could be used to derive new methods for future performance benchmarking. There are new opportunities emerging from technology that allow organisations to gather and record production data quickly and on site. However, the research has shown that while these technologies can provide a great deal of insight and clarity into the nature and cause of certain performance issues, there are still considerable limitations to a purely statistical approach. Also, results highlight that getting people to use the equipment correctly is the most important and pertinent issue at present. While increased detail and improved methods to capture detail surrounding performance issues will be of value to many, there is still the requirement for context and perception. This is not just an issue for construction; this is an issue for most industries as technology pushes forward and allows us to measure things with greater degree of accuracy.

**References:**


Akhavian, R. & Behzadan, A., 2012. An integrated data collection and analysis framework for remote monitoring and planning of construction operations. *Advanced Engineering Informatics*, 26, 749-761.

Aliaga, M. & Gunderson, B., 2005. *Interactive statistics*. 3rd Ed., Pearson, New Jersey, USA.

Balnaves, M. & Caputi, P., 2001. *Introduction to quantitative research methods: an investigative approach*. Sage, London, UK.

Barber, E., 2004. Benchmarking the management of projects: a review of current thinking. *International Journal of Project Management*, 22, 301 - 307.

Blaikie, N., 2009. *Designing social research*. 2nd Ed., Polity, Cambridge, UK.

Bryman, A., 2012. *Social research methods*. 4th Ed., Oxford University Press, Oxford, UK.

Cheng, M.Y. & Wu, Y.W., 2012. Improved construction subcontractor evaluation performance using ESIM. *Applied Artificial Intelligence*, 26, 261--273.

Creswell, J. W., 2013. *Research design: qualitative, quantitative, and mixed methods approaches*. 4th Ed., Sage, London, UK.

Dissanayake, M. & Fayek, A., 2008. Soft computing approach to construction performance prediction and diagnosis. *Canadian Journal of Civil Engineering*, 35, 764-776.

Easterby-Smith, M., Thorpe, R. & Jackson, P. R., 2012. *Management research*. 4th Ed., Sage, London, UK.

Fellows, R. F. & Liu, A. M. M., 2015. *Research methods for construction*. 4th Ed., Wiley-Blackwell, Oxford, UK.

Field, A., 2009. *Discovering statistics using SPSS*. 3rd Ed., SAGE, London, UK.

Gray, D. E., 2004. *Doing research in the real world*. SAGE, London, UK.

Kruger, D. J., 2003. Integrating quantitative and qualitative methods in community research. *The Community Psychologist*, 36(2), 18-19.

Merrow, E. W. et al., 2009. *Productivity in the UK engineering construction industry*. S.l.: Department for Business, Innovation and Skills.

Neuman, W. L., 2007. *The basics of social research: Qualitative and quantitative approaches*. 2nd Ed., Pearson Education, Boston, USA.

Punch, K. F., 2005. *Introduction to social research: quantitative and qualitative approaches*. 2nd Ed., Sage, London, UK.

Rungtusanatham, M., Ogden, J. A., & Wu, B., 2003. Advancing theory development in total quality management: A “Deming management method” perspective. *International Journal of Operations & Production Management*, 23(8), 918-936

Russell, A. D., Chiu, C.-Y. & Korde, T., 2009. Visual representation of construction management data. *Automation in Construction*, 18, 1045 - 1062.

Saunders, M., Lewis, P. & Thornhill, A., 2009. *Research methods for business students*. 5th Ed., Pearson Education, Harlow, UK.

Smith, E. & Smith Jr, J., 2008. *Using secondary data in educational and social research*. McGraw-Hill Education, Berkshire, UK.

Son, H., Park, Y., Kim, C. & Chou, J.S., 2012. Toward an understanding of construction professionals acceptance of mobile computing devices in South Korea: An extension of the technology acceptance model. *Automation in Construction*, 28, 82-90.

UK Construction Industry Performance Report (2014) [Online] https://www.glenigan.com/market-analysis/reports/construction-kpi-report-2014

Vartanian, T. P., 2010. *Secondary data analysis*. Oxford University Press, New York, USA.