COMPARATIVE ANALYSIS OF PROJECT PERFORMANCE BETWEEN DIFFERENT PROJECT DELIVERY SYSTEMS

Michael W. Ibrahim[1], and Awad S. Hanna[2]

ABSTRACT

Project Delivery System (PDS) defines the relationship and timing of involvement between different contracting parties. The main PDSs referred to in cited literature are: Design-Bid-Build (DBB), Construction Management at Risk (CMR), Design-Build (DB), and Integrated Project Delivery (IPD). By applying statistical tests such as Analysis of Variance (ANOVA) F-test and Kruskal-Wallis H-test to a dataset of 109 projects, this paper compares the performance of the four PDSs. As a result, statistically significant performance differences among the examined PDSs were identified in five performance areas: cost, schedule, quality, communication, and change management. Furthermore, performing pairwise comparisons using post-hoc statistical tests to each pair of PDSs shows that DBB performs markedly worse than the other examined PDSs, especially IPD. The findings presented in this paper should encourage industry professionals to move away from the DBB model, and towards IPD and other synergic PDSs.

KEYWORDS

Integrated Project Deliver (IPD), Collaboration, Relational, Project Performance, Quantitative Analysis.

INTRODUCTION

Despite being a large contributor to the global economy, the construction industry is fraught with waste and inefficiencies, leading to declining productivity (Teicholz 2013; Ibrahim 2018). An often-cited root cause of the poor productivity of the construction industry is systemic fragmentation which promotes confrontational culture (Yates and Battersby 2003). The result of this fragmentation has been an increase in the use of transactional contracts rather than a 'project-first' attitude (Thomsen et al. 2010). This has led to a trend of stakeholders increasing the pad to their estimates in an effort to protect themselves from a higher level of perceived risk (Iwanski 2013). These padded estimates have caused an increased reliance on the arbitration process. Thus, even more productivity is lost as project stakeholders devote time and resources to both issuing and facing claims, rather than the project.

---

[1]   California State University, Los Angeles, United States of America
[2]   University of Wisconsin-Madison, United States of America

One of the primary fronts on which the industry is attempting to combat the systemic fragmentation within itself is in the evolution of the PDSs. The general trend of PDSs chronologically is: DBB to CMR to DB to IPD. This paper studies the performance of the four aforementioned PDSs using metrics in five areas: cost, schedule, quality, communication, and change management. This is done identifying statistically significant performance differences among the four PDSs and investigating how each pair of PDSs performs differently using data collected from 109 projects.

LITERATURE REVIEW

Extensive research has been conducted to evaluate the performance of traditional PDSs. To varying degrees, much of this research concluded that more collaborative delivery systems achieved superior performance as compared to less collaborative ones. Most notably, it was shown that DB outperformed CMR, which in turn outperformed DBB in terms of unit cost, construction speed and delivery speed (Konchar and Sanvido 1998). This trend of performance improvement has suggested that even more collaborative PDS, such as IPD, will further enhance project delivery performance. With the introduction of IPD as a new PDS, many case studies were conducted to assess its performance. Almost all of these case studies showed that IPD projects finished on time, under budget, and with positive relations within the project team (Hanna 2016). By statistically analyzing 35 projects, it was shown that IPD outperformed non-IPD in 12 metrics spanning six performance areas: quality, communication, change management, business, recycling, and schedule (El Asmar et al. 2013). Also, from a subcontractor's perspective, it was shown that IPD outperformed non-IPD in four performance areas: quality, schedule, communication, and change management (Iwanski 2013). More recently, analyzing 32 projects presented statistical evidences for the superior performance of IPD/'IPD-ish' projects as compared to non-IPD projects in two metrics spanning two performance areas: communication and change management.

Although most of the cited studies claim that more collaborative PDSs outperformed less collaborative ones, none of them provided statistical analysis of the performance differences between the four main PDSs. Instead, they either studied the performance differences among the three traditional PDSs or compared the performance of IPD to the three traditional ones collectively. This paper aims to fill this gap by statistically analyzing the performance differences between the four PDSs.

RESEARCH METHODOLOGY

DATA COLLECTION AND DOCUMENTATION

To perform a comprehensive assessment of performance differences between the four main PDSs, an extensive project-based survey was used to collect data (El Asmar et al. 2013; Ibrahim 2016). This survey was designed to evaluate project delivery performance across key performance areas. As a result, this paper studies project delivery performance using eight performance metrics spanning five performance areas. These specific performance areas and metrics were selected to be consistent with cited literature and based on data

availability. Following are the five investigated performance along with their corresponding performance metrics and units of measurement:

- Cost performance area: construction cost growth (% of total cost).
- Schedule performance area: schedule growth (% of total duration).
- Quality performance area: enumeration of punch-list items (number/$1M) and overall systems quality (scale of very low to very high).
- Communication performance area: enumeration of Request for Information (RFI) forms (number/$1M) and RFI processing time (weeks).
- Change management performance area: overall project changes (% of total cost) and change order processing time (weeks).

Using this survey, substantial data was collected from 109 projects. This number of projects provides a considerable improvement, in terms of sample size, relative to related studies that collected data from 32 projects (Hanna 2016), 35 projects (El Asmar et al. 2013), and 49 projects (Cho et al. 2010). Also, the representativeness of the collected dataset was validated by making sure that its statistical distribution was consistent with cited literature (Ibrahim 2016).

DATA CHARACTERISTICS

Of the 109 studied projects, 28% were DBB, 32% were CMR, 23% were DB and 17% were IPD. From a geographical perspective, data was solicited with no specific geographic preference, thus the dataset of this study included projects from 31 states across the U.S., as well as projects from Canada, Colombia, and Ireland. Regarding the types of the studied construction projects, the dataset consisted mainly of institutional (39% of the data), industrial (29% of the data), and commercial projects (18% of the data). The rest of the dataset was infrastructure and large multi-story residential projects. The total dollar amount of construction work for the studied projects combined was around $16.2 billion, with an average project cost of each examined PDSs being around $150 million.

The distribution of project type within each PDS was akin to the distribution of all projects in the study. Of the DBB projects, 46% were institutional, 35% were industrial, 14% were commercial, and 4% were infrastructure. Of the CMR projects, 51% were institutional, 23% were industrial, 14% were commercial, 6% were infrastructure, and 6% were residential. Of the DB projects, 28% were institutional, 40% were industrial, and 32% were commercial. Of the IPD projects, 44% were institutional, 28% were industrial, 22% were commercial, and 6% were infrastructure. Statistical tests revealed that the project type was not a statistically significant factor impacting project performance, regardless of PDS.

In addition, distributions of further project factors were statistically examined within each PDS to ensure that the four examined PDSs had similar project characteristics. As a result, reported statistically significant performance differences between the examined PDSs should be directly linked to the type of PDS and not to confounding variables.

STATISTICAL FRAMEWORK

Using this data, statistical modelling techniques, including ANOVA and the Kruskal-Wallis H Test, were performed to determine if significant differences existed across the

performance of the examined PDSs. In this analysis, the independent variable was the PDS type, and the dependent variable was each of the eight-performance metrics. For each dependent variable, the null hypothesis was that the performance across the four PDSs was the same, whereas the alternative hypothesis was that the performance across the four PDSs was dissimilar. Normality and homoscedasticity assumptions were thoroughly examined through applying Shapiro-Wilk test and Levene's test when determining which statistical tests were most suitable for each performance metric.

If it was statistically evident that PDSs performed differently in a given performance metric, corresponding post-hoc statistical tests, including Tukey-Kramer and Conover-Iman with Ŝidák corrections, were applied to the set of pairwise comparisons. This additional analysis was performed to offer more statistical insight into how each pair of PDSs performs differently. The statistically proven findings of this paper should provide project stakeholders with the means to assess the performance of different PDSs, thus making well-informed decisions when choosing a delivery system for their projects.

RESEARCH FINDINGS

OVERALL FINDINGS

Table 1 lists the studied performance metrics, followed by the corresponding p-values of examining the null hypothesis of constant performance across the examined PDSs. The smaller the p-value, the stronger the evidence of statistically significant performance difference across the examined PDSs in the corresponding performance metric. To define statistical significance, this paper employed the common p-value threshold of 0.05, below which the performance differences between PDSs are to be considered significantly different with 95% confidence level.

Table 1: Statistical analysis of performance differences among the examined PDSs

Performance Area	Performance Metric	p-value
Communication	Enumeration of RFIs	0.00
Change management	Change order processing time	0.00
Schedule	Schedule growth	0.00
Communication	RFI processing time	0.00
Cost	Construction cost growth	0.00
Change management	Overall project change	0.00
Quality	Overall systems quality	0.00
Quality	Enumeration of punch-list items	0.02

Figure 1 presents comparative box and whisker plots demonstrating the performance differences between the examined PDSs.
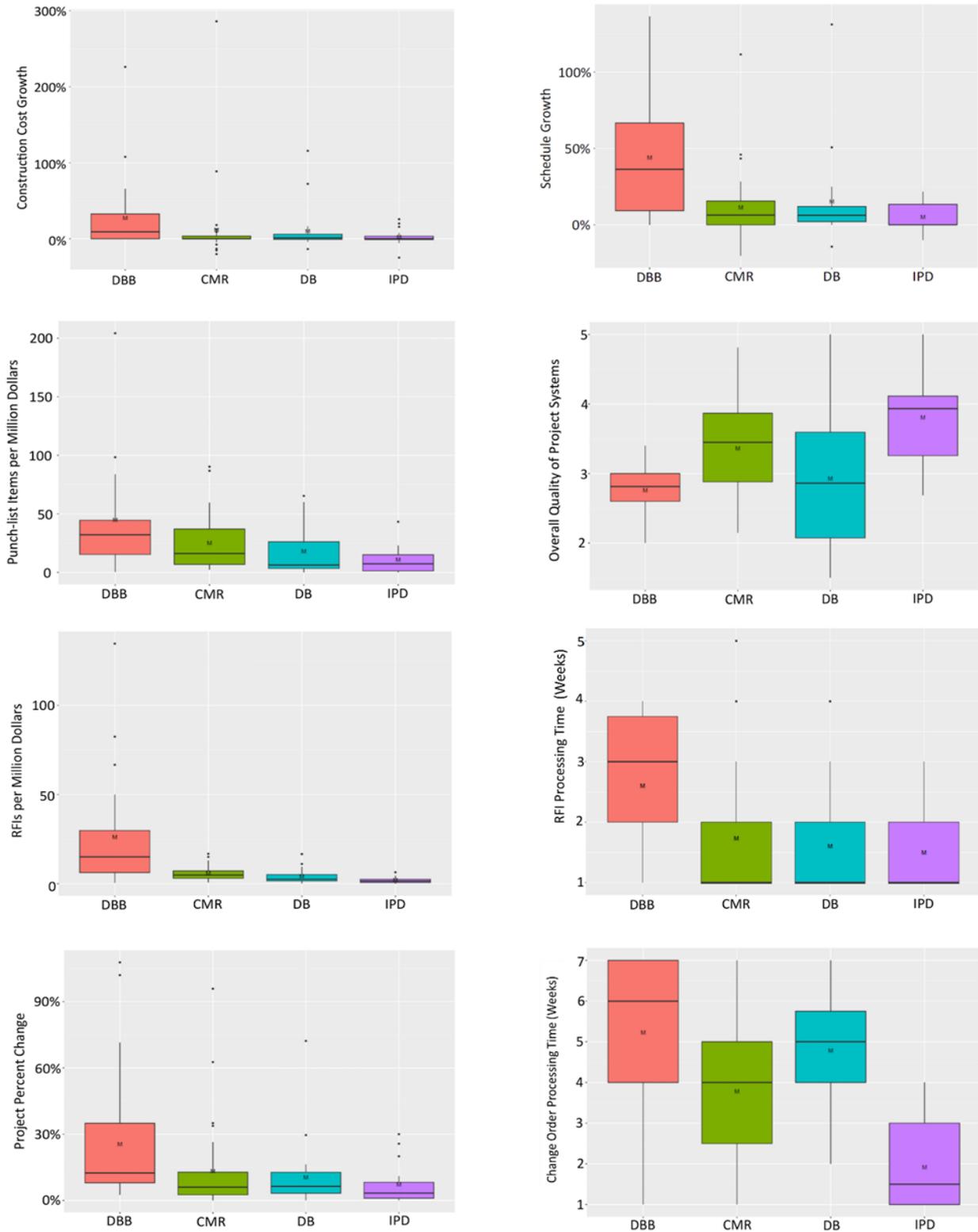
Figure 1: Comparative box-and-whisker plots for the performance of examined PDSs

Figure 1 demonstrates the performance differences between the examined PDSs. In addition, the p-values presented in Table 1 show that the investigated data provided enough evidence to conclude that the examined PDSs significantly differ, at 95% confidence level, in the eight studied performance metrics. The following five subsections are separated by performance area to provide the detailed findings for these eight metrics.

COST PERFORMANCE AREA

Application of the Kruskal-Wallis H test to the percentage of cost growth returned a p-value of 0.00. This provided sufficient statistical evidence at 5% significance level to reject the null hypothesis, thus concluding that construction cost growth percentage significantly differs across the examined PDSs at a confidence level of 95%. Table 2 demonstrates the p-values resulting from applying pairwise comparisons between each pair of PDSs.

Table 2: Pairwise comparison about construction cost growth

Pair of PDSs	p-value	Significantly Better PDS at 95% Confidence Level
IPD and DBB	0.004	IPD
IPD and CMR	0.891	Not applicable
IPD and DB	0.621	Not applicable
DB and DBB	0.047	DB
DB and CMR	0.816	Not applicable
CMR and DBB	0.003	CMR

Applying post-hoc Conover-Iman tests provided statistical evidence at 95% confidence level to support the following findings: DBB projects have higher construction cost growth percentage compared to IPD projects; DBB projects have higher construction cost growth percentage compared to DB projects; and DBB projects have higher construction cost growth percentage compared to CMR projects. Combining these three findings shows that DBB has the poorest performance level regarding this metric.

SCHEDULE PERFORMANCE AREA

Application of the Kruskal-Wallis H test to the percentage of schedule growth returned a p-value of 0.00. This provided sufficient statistical evidence at 5% significance level to reject the null hypothesis, thus concluding that schedule growth percentage significantly differs across the examined PDSs at a confidence level of 95%. Table 3 demonstrates the p-values resulting from applying pairwise comparisons between each pair of PDSs.

Applying post-hoc Conover-Iman tests provided statistical evidence at 95% confidence level to support the following findings: DBB projects have higher construction schedule growth percentage compared to IPD projects; DBB projects have higher construction schedule growth percentage compared to DB projects; and DBB projects have higher construction schedule growth percentage compared to CMR projects. Combining these three conclusions, DBB was, again, proven to be the poorest performing PDS with regard to this performance metric.

Table 3: Pairwise comparison about schedule growth

Pair of PDSs	p-value	Significantly Better PDS at 95% Confidence Level
IPD and DBB	0.000	IPD
IPD and CMR	0.685	Not applicable
IPD and DB	0.582	Not applicable
DB and DBB	0.005	DB
DB and CMR	0.944	Not applicable
CMR and DBB	0.000	CMR

QUALITY PERFORMANCE AREA

Application of the Kruskal-Wallis H test to the number of punch-list items per million dollars resulted in a p-value of 0.02. Therefore, statistical evidence at 5% significance level was presented to reject the null hypothesis, thus concluding, at 95% confidence level, that the number of punch-list items per million dollars changes when the PDS changes. Table 4 demonstrates the p-values resulting from applying pairwise comparisons between each pair of PDSs.

Table 4: Pairwise comparison about number of punch-list items per million dollars

Pair of PDSs	p-value	Significantly Better PDS at 95% Confidence Level
IPD and DBB	0.014	IPD
IPD and CMR	0.155	Not applicable
IPD and DB	0.815	Not applicable
DB and DBB	0.041	DB
DB and CMR	0.406	Not applicable
CMR and DBB	0.447	Not applicable

Application of post-hoc Conover-Iman tests provided sufficient statistical evidence to support the following conclusions at 95% confidence level: IPD projects have fewer punch-list items per million dollars than DBB projects; and DB projects have fewer punch-list items per million dollars than DBB projects.

To calculate the second metric (overall quality of project systems), respondents evaluated the quality of each of 11 major project systems as well as the quality of the entire project, using a scale of 1 to 5. The average of these ratings corresponds to the project's overall systems quality. These major project systems are foundation, structure, interior finishes, exterior enclosure, roofing, mechanical systems, electrical systems, site, process equipment, conveying systems, and specifications. Application of ANOVA to this metric returned p-values of less than 0.00. Therefore, sufficient statistical evidence existed to reject the null hypothesis at 5% significance level, thus concluding that the overall quality of project systems differs across the studied PDSs at a confidence level of 95%.

Table 5 demonstrates the p-values resulting from applying pairwise comparisons between each pair of PDSs.

Table 5: Pairwise comparison about overall quality of project systems

Pair of PDSs	p-value	Significantly Better PDS at 95% Confidence Level
IPD and DBB	0.002	IPD
IPD and CMR	0.221	Not applicable
IPD and DB	0.013	IPD
DB and DBB	0.799	Not applicable
DB and CMR	0.293	Not applicable
CMR and DBB	0.057	Not applicable

Application of post-hoc Tukey-Kramer tests found statistical evidence at 95% confidence level to support the following conclusions: IPD projects have higher overall systems quality than DBB projects; and IPD projects have higher overall systems quality than DB projects.

COMMUNICATION PERFORMANCE AREA

Application of the Kruskal-Wallis H test to the number of RFIs per million dollars returned a p-value of 0.00. This low p-value provided statistical evidence at a significance level of 5% to reject the null hypothesis, thus concluding that there is a statistical evidence at 95% confidence level that the number of RFIs per million dollars changes when PDS changes. Table 6 demonstrates the p-values resulting from applying pairwise comparisons between each pair of PDSs.

Applying post-hoc Conover-Iman tests showed statistical significance at a confidence level of 95% for the following conclusions: CMR projects have fewer RFIs per million dollars than DBB projects; DB projects have fewer RFIs per million dollars than DBB projects; and IPD projects have fewer RFIs per million dollars than DBB projects. Therefore, DBB was shown to be the PDS that has the poorest performance with regard to this performance metric. Additionally, it was found that, at 95% confidence level, IPD projects have fewer RFIs per million dollars than CMR projects.

When the Kruskal-Wallis H test was applied to RFI processing time (weeks), a p-value of 0.00 was returned. This low p-value provided statistical evidence at a significance level of 5% to reject the null hypothesis, from which it was concluded that there is statistical evidence at 95% confidence level that RFI processing time changes when PDS changes. Table 7 demonstrates the p-values resulting from applying pairwise comparisons between each pair of PDSs.

Table 6: Pairwise comparison about number of RFIs per million dollars

Pair of PDSs	p-value	Significantly Better PDS at 95% Confidence Level
IPD and DBB	0.000	IPD
IPD and CMR	0.000	IPD
IPD and DB	0.154	Not applicable
DB and DBB	0.000	DB
DB and CMR	0.104	Not applicable
CMR and DBB	0.001	CMR

Table 7: Pairwise comparison about RFI processing time

Pair of PDSs	p-value	Significantly Better PDS at 95% Confidence Level
IPD and DBB	0.002	IPD
IPD and CMR	0.854	Not applicable
IPD and DB	0.978	Not applicable
DB and DBB	0.000	DB
DB and CMR	0.855	Not applicable
CMR and DBB	0.002	CMR

Results from the application of the post-hoc Conover-Iman tests showed statistical significance at a confidence level of 95% for the following conclusions: IPD projects have shorter RFI processing times than DBB projects; DB projects have shorter RFI processing times than DBB projects; and CMR projects have shorter RFI processing times than DBB projects. These statistically significant findings demonstrate that DBB has the poorest performance with respect to this metric relative to the examined PDSs.

CHANGE MANAGEMENT PERFORMANCE AREA

Application of the Kruskal-Wallis H test to the project percent change returned a p-value of 0.00. This provided statistical evidence at 5% significance level to reject the null hypothesis, thus concluding that project percent change significantly differs across the examined PDSs at a confidence level of 95%. Table 8 demonstrates the p-values resulting from applying pairwise comparisons between each pair of PDSs.

Application of post-hoc Conover-Iman tests found statistical evidence to support the following conclusions at the 95% confidence level: DBB projects have higher project percent change compared to IPD projects; DBB projects have higher project percent change compared to DB projects; and DBB projects have higher construction project percent change compared to CMR project. Combining these three conclusions, DBB was proven to be the PDS that had the poorest performance regarding this metric.

Table 8: Pairwise comparison about overall project percent change

Pair of PDSs	p-value	Significantly Better PDS at 95% Confidence Level
IPD and DBB	0.001	IPD
IPD and CMR	0.526	Not applicable
IPD and DB	0.625	Not applicable
DB and DBB	0.019	DB
DB and CMR	0.976	Not applicable
CMR and DBB	0.013	CMR

Application of ANOVA to the change order processing time returned a p-value of 0.00, providing statistical evidence to reject the null hypothesis at 5% significance level, thus concluding that this metric significantly differs across the examined PDSs. Table 9 presents the p-values resulting from applying pairwise comparisons between each pair of PDSs.

Table 9: Pairwise comparison about change order processing time

Pair of PDSs	p-value	Significantly Better PDS at 95% Confidence Level
IPD and DBB	0.000	IPD
IPD and CMR	0.010	IPD
IPD and DB	0.000	IPD
DB and DBB	0.859	Not applicable
DB and CMR	0.227	Not applicable
CMR and DBB	0.061	Not applicable

Application of post-hoc Tukey-Kramer tests found statistical evidence at 95% confidence level to support the following conclusions: IPD projects have higher project systems quality than DBB projects; and IPD projects have higher project systems quality than DB projects.

CONCLUSIONS

The comparative statistical analysis presented in this paper shows that there are statistically significant differences across the four examined PDSs, at 95% confidence level, in eight metrics spanning five performance areas: communication, change management, schedule, cost and quality. For these eight metrics, post-hoc statistical tests were applied to investigate the performance differences between each pairing of PDSs. As a result, it was shown that IPD outperformed DBB in the eight metrics, IPD outperformed CMR in two metrics, and IPD outperformed DB in two metrics. Results also demonstrated that DB outperformed DBB in six metrics, and CMR outperformed DBB in five metrics. Overall, DBB was proven to be the lowest performing PDS in five metrics. Based on this paper's findings, industry practitioners should be encouraged to move away from DBB and towards IPD to create an environment that fosters collaboration and optimal project performance.

REFERENCES

Cho, S., Ballard, G., Azari, R., and Kim, Y. (2010). "Structuring Ideal Project Delivery System." Proc., IPPC4, 2010.

El Asmar, M., Hanna, A., and Loh, W. (2013). "Quantifying Performance for the Integrated Project Delivery System as Compared to Established Delivery Systems." J. Constr. Eng. Manage., 10.1061/(ASCE)CO.1943-7862.0000744, 04013012.

Hanna, A. (2016). "Benchmark Performance Metrics for Integrated Project Delivery." J. Constr. Eng. Manage. ,10.1061/(ASCE)CO.1943-7862.0001151 , 04016040.

Ibrahim, M.W. (2016). "Modeling, Benchmarking, and Maximizing Project Delivery Performance." Master's Thesis. University of Wisconsin–Madison.

Ibrahim, M. W. (2018). "Improving Project Performance by Mitigating Out-of-Sequence Work and Assessing Construction Readiness." Ph.D. Dissertation. The University of Wisconsin-Madison.

Iwanski, M. (2013). "Performance of Integrated Project Delivery (IPD) for Mechanical and Electrical Contractors." Master's Thesis. University of Wisconsin-Madison.

Konchar, M., and Sanvido, V. (1998). "Comparison of U.S. project delivery systems." J. Constr. Eng. Mgmt., 124, 435.

Teicholz, P. (2013). "Labor-Productivity Declines in the Construction Industry: Causes and Remedies" (another look). AECbytes Viewpoint, 67.

Thomsen, C., Darrington, J., Dunne, D., and Lichtig, W. (2010). "Managing Integrated Project Delivery." Construction Management Association of America, Mclean, VA.

Yates, J. K., and Battersby, L. C. (2003). "Master builder project delivery system and designer construction knowledge." Journal of Construction Engineering and Management, 129(6), 635-644.