

공공공사 시공평가 항목의 객관성 확보를 위한 주요 개선 항목 도출에 관한 연구

서세덕* · 김옥규** · 박형근***

Seo, Se Deok*, Kim, Ok Kyue**, Park, Hyung Keun***

A Study on Construction Evaluation Criteria for Securing the Objectivity in Public Construction

ABSTRACT

The government introduced the comprehensive evaluation bidding system with the goal of pursuing the best value and the global standard in 2016. However, for the evaluation criteria on the construction evaluation reflected to the comprehensive evaluation bidding system, the problems of the objectivity insufficiency, the inclusion of multiple subjective evaluation items, and the irrationality of the weight for each evaluation item are continue to be presented. The central office group, the local government, the relevant industry, and the expert group share recognition, but the solution is not derived. Hence, the major evaluation items to be improved were derived with the characteristics analyzed to secure the objectivity of the construction evaluation. For the analysis method, the standard deviation and the Fleiss Kappa analysis method were used by utilizing the characteristics that the construction evaluation criteria consist of all 4-point measures (good, average, insufficient, and poor). According to the result, the 10 evaluation items of the total 25 construction evaluation items were derived as the evaluation items to be improved. It was found in the analysis on the major characteristics of the derived evaluation items that the qualitative evaluation criteria such as 'Very Suitable' and 'Suitable' were commonly included in the detailed evaluation guidelines. Hence, as far as the future construction evaluation standards are concerned, the qualitative evaluation standards are sublated, and the improvement should be made mainly for the quantitative evaluation criteria enabling the objectivity assurance.

Key words : Construction evaluation, Improvement item, Fleiss kappa, Comprehensive evaluation bidding system

초 록

정부는 최고 가치와 글로벌 스탠다드를 추구한다는 목표 아래 2016년 종합심사낙찰제도를 도입하였다. 하지만 종합심사낙찰제도에 반영되는 시공평가의 평가기준은 객관성 부족, 주관적 평가 항목 다수 포함, 평가 항목별 가중치 비합리성 등의 문제가 지속적으로 제기 되고 있으며 중앙부처, 지자체, 관련 업계, 전문가 집단 모두 그 인식을 같이하고 있으나 해결 방안을 도출하지 못하고 있는 실정이다. 따라서 본 연구는 시공평가의 객관성 확보를 위해 개선이 필요한 주요 평가항목을 도출하고 그 특성을 분석하였다. 분석 방법은 시공평가 평가기준이 모두 4점 척도(우수, 보통, 미흡, 불량)로 구성된 특성을 활용하여 표준 편차, Fleiss Kappa 분석 방법을 활용하였으며, 그 결과 총 25개로 구성된 시공평가의 평가항목 중 10개의 평가항목이 개선이 필요한 평가항목으로 도출되었다. 도출된 평가 항목의 주요 특성을 분석한 결과 세부 평가지침에 '매우적정', '적정'과 같은 정성적 평가 기준이 공통으로 포함된 것을 확인할 수 있었다. 따라서, 향후 시공평가의 평가기준은 정성적 평가기준은 지양하고 객관성 확보가 가능한 정량적 평가기준 중심으로 개선되어야 할 것으로 판단된다.

검색어 : 시공평가, 개선항목, Fleiss Kappa, 종합심사낙찰제

* 정희원 · 충북대학교 토목공학과 박사수료, 공학석사 (Chungbuk National University · seosd@chungbuk.ac.kr)

** 충북대학교 건축공학과 교수, 공학박사 (Chungbuk National University · okkim@chungbuk.ac.kr)

*** 중신회원 · 교신저자 · 충북대학교 토목공학부 교수, 공학박사 (Corresponding Author · Chungbuk National University · parkhk@chungbuk.ac.kr)

Received September 11, 2019/ revised October 21, 2019/ accepted October 22, 2019

1. 서론

1.1 연구의 배경 및 목적

그동안 300억 원 이상의 정부 발주공사에 획일적으로 적용되던 최저가낙찰제의 폐해가 지속적으로 대두되면서, 정부는 최고 가치와 글로벌 스탠다드를 추구한다는 목표 아래 2016년 새로운 입찰제 도입 ‘종합심사낙찰제’를 도입하였다.

시공부분의 ‘종합심사낙찰제’ 반영은 건설기술 진흥법 제50조(건설기술용역 및 시공 평가 등)와 ‘동법 시행령 제82조(건설기술용역 평가 및 시공평가의 대상)-제84조(종합평가의 기준 및 절차)’, ‘동법 시행규칙 제44조(건설기술용역 및 시공평가)’ 및 시행규칙 제44조 제7항에 의거한 ‘국토교통부 고시 제2017-531호 건설기술용역 및 시공 평가지침’(이하 건설기술용역 및 시공 평가 지침)에 따라 평가되며 그 결과가 종합심사낙찰제에 반영되고 있다.

하지만 건설기술용역 및 시공평가 지침의 평가기준은 객관성 부족, 주관적 평가항목 다수 존재, 평가 항목별 가중치 비합리성, 절대적인 평가 시간 부족 등의 문제가 지속적으로 제기되고 있다. 이와 같은 문제를 해결하기 위해 국토해양부(현 국토교통부)는 평가자에 따라 점수가 상이하고 평가기관(발주청)은 일률적으로 높은 점수를 주는 등 평가결과에 대한 신뢰성 및 변별력이 없다고 판단하여 2009년부터 시공평가항목을 100% 정량화 하는 노력을 지속적으로 하고 있다(MLTM, 2009).

또한, 부산광역시시는 현행 ‘건설기술용역 및 시공 평가지침’의 문제점을 개선하기 위하여 자체적으로 개정한 ‘부산광역시 건설기술용역 및 시공 평가지침 전부개정(안)’을 마련하였다. 시공평가와 관련된 주요 개정안으로는 총공사비가 500억 원 이상인 건설공사는 시공평가를 총 2회 실시하도록 하는 등의 내실 있고 객관성있는 평가로 유도하려는 노력이 계속하고 있다(Busan Metropolitan City, 2017).

이처럼, 유사환경 내 동일 수준의 공사를 수행했음에도 불구하고 평가위원 및 발주청에 따라 상이한 시공평가 결과가 종합심사낙찰제에 반영된다면 객관성 부족 및 형평성 문제는 지속적으로 야기 될 수밖에 없다.

하지만, 시공평가 결과의 객관성 부족 등의 문제 인식을 중앙부처, 지자체, 관련 업계, 전문가 집단 등이 모두 같이하고 있으나 해결 방안을 도출해내지 못하고 있는 실정이다.

따라서 본 연구의 목적은 시공평가의 평가위원별 평가 결과를 활용하여 평가항목 별 평가결과 차이를 분석함으로써 시공평가 지침의 객관성 확보를 위한 평가항목 도출에 그 목적이 있다.

1.2 연구의 범위 및 방법

본 연구의 범위는 국토교통부 고시 제2015-505호 이후(MOLIT, 2015) 개정된 ‘건설기술용역 및 건설공사 시공 평가지침’의 평가지침으로 평가된 365개 공공공사의 평가위원별 평가결과를 대상으로

하였다. 시공평가의 구성은 2개의 대분류, 9개의 중분류, 25개의 세분류로 구분되며 이중 세분류를 기준으로 연구를 수행하였으며, 가점(세분류: 공사 특성 및 난이도 등에 따른 보정, 시공자 제안으로 인한 공사비 절감비율)과 감점(세분류: 평가위원에게 급품·항을 제공) 항목은 출현 빈도가 낮아 본 연구범위에서는 제외하였다. 본 연구의 방법은 Fig. 1과 같다.

첫째, 자료수집 및 보정 단계이다. 통상적으로 시공평가에 참여하는 평가위원은 5인으로 구성되나 평가위원이 중도 교체 등의 내·외부적 요인으로 5인 미만 및 6인 이상이 참여하여 평가한 시공평가 결과는 제외하였으며, 2015년 이전 고시된 건설기술용역 및 시공 평가 지침을 활용한 평가 결과는 현재의 평가 기준과는 다소 차이가 있어 분석에서 제외하였다.

둘째, 표준편차 계산을 통한 평가항목 도출 단계이다. 평가위원별 평가점수 결과를 활용하여 세부평가항목별로 표준편차 발생 비율 및 누적 표준편차 계산을 통해 주요 평가항목을 도출하였다.

셋째, Fleiss Kappa 분석을 통한 평가항목 도출 단계이다. 본 단계에서는 평가항목별로 도출된 Fleiss Kappa값의 해석을 위해 기준을 정의하고 그 결과를 바탕으로 시공평가 항목의 주요 평가항목을 도출하였다.

마지막으로 표준편차를 활용하여 도출된 주요 평가항목과 Fleiss Kappa분석을 통해 도출된 주요 평가항목의 세부 평가지침 특성을 도출하였다.

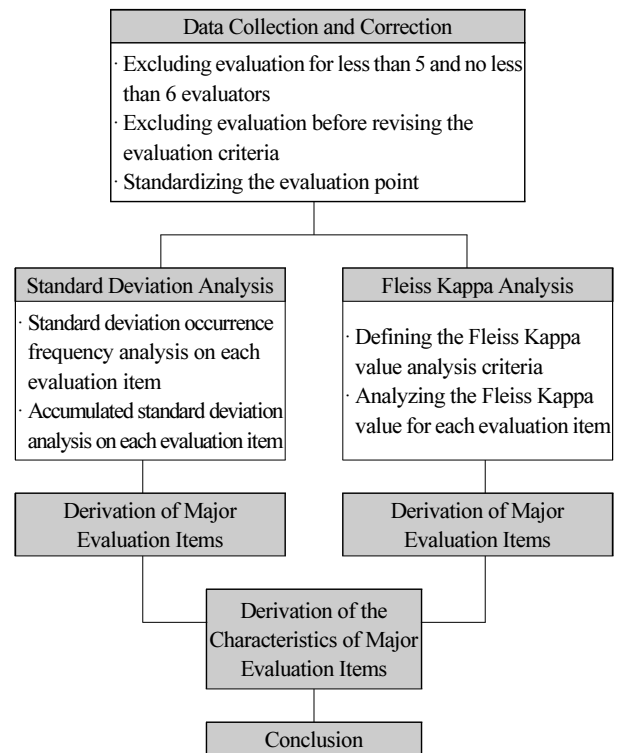


Fig. 1. Research Methodology

2. 이론적 고찰

2.1 종합심사낙찰제

종합심사낙찰제는 2016년부터 국가 또는 공공기관이 발주하는 추정가격 300억 원 이상의 시설공사에 적용되는 낙찰방식을 말한다. 입찰평가는 크게 공사수행능력, 입찰가격, 사회적 책임, 계약실행도의 4개 분야로 이루어져 있으며, 각 분야에서 받은 점수의 총합이 가장 큰 입찰자가 낙찰자로 선정되는 제도이다(Park, 2016).

Table 1과 같이 종합심사낙찰제 평가기준 중 공사수행능력 점수는 40~50점이며 이 중 공공공사 시공평가 점수의 가중치는 30~50%를 차지해 가장 크다(MOEF, 2016).

2.2 시공평가

시공평가는 ‘건설기술용역 및 시공 평가지침’에 의거하여 실시되며, 이 지침은 「건설기술진흥법」 제50조 제6항, 같은 법 시행령 제84조 제4항 및 같은 법 시행규칙 제44조 제7항에 따라 세부 평가기준 및 방법 등이 정해져 있다(MOLIT, 2017).

시공평가표의 모든 세분류(25개) 평가 등급은 매우우수, 우수,

보통, 미흡 4단계로 구성되어 있으며 Table 2와 같이 평가항목 별로 세부 평가 기준은 다르게 규정하고 있으며 평가항목별로 배점도 상이하다

2.3 Fleiss Kappa

동일한 측정 대상들에 대해 평가자(rater)들이 평가한 결과의 일치 정도를 나타내는 지표로 일치도(agreement)를 들 수 있다. 일치도 분석 방법 중 카파통계량(Kappa Statistic)은 평가치가 범주형 자료일 때 일치도의 척도로 자주 쓰이며 2인 평가자의 일치도는 Cohen’s Kappa 통계량, 3인 이상의 평가자 일치도는 Fleiss Kappa 통계량을 활용할 수 있다. 본 연구에서 활용하는 시공평가의 평가자는 3인 이상이므로 Fleiss Kappa통계량을 활용하였다(Fleiss, 1971; Berry and Mielke, 1988).

Fleiss의 방법은 Scott의 π (Scott, 1955)의 개념을 확장시킨 형태로 평가 결과의 범주는 명목형이며 평가자는 서로 독립임을 가정하며 Fleiss 방법의 k 식은 다음 Eq. (1)과 같다

$$k = \frac{P_a - P_{e,k}}{1 - P_{e,k}} \quad (1)$$

Table 1. Evaluation Standards of the Comprehensive Evaluation Bidding System

Review Area	Review Item		Weight (%)	
			General Construction Work	Most Challenging Construction Work
Construction Work Performance Capability (40~50 Points)	Specialty	Construction Work Result (Construction Manpower)	20~30	20~30
		Sales Amount Weight	0~20	0~20
		Assigned Engineer	20~30	20~30
	Capability	Public Work Construction Evaluation Point	30~50	30~50
		Construction Capability for Each Size	0~20	0~20
		Consortium Composition	1~5	1~5
Subtotal		100	100	
Bidding Amount (50~60 Points)	Amount		100	100
	Suitability of Price Calculation	Unit Price	Point Subtraction	-
		Subcontracting Plan	Point Subtraction	Point Subtraction
		Quantity	-	Point Subtraction
		Construction Plan	-	Point Subtraction

Table 2. Example of the Composition of the Construction Evaluation Table

Evaluation Criteria	Point Distribution	Evaluation Grade			
		Good (×1.0)	Average (×0.8)	Insufficient (×0.6)	Poor (×0.4)
2.2 Contract Construction Period Observance	4	Construction Period Reduction	Scheduled Construction Period Observance	No More than 1 % Delay	More than 1 % Delay
3.1 Suitability of the Site Manpower Assignment	3	Very Suitable	Suitable	Unsuitable	Very Unsuitable
⋮	⋮	⋮	⋮	⋮	⋮

Eq. (1)에서 P_a 는 같은 범주로 분류한 평가대상자들의 비율이며 Eq. (2)와 같이 나타낼 수 있다.

$$P_a = \frac{1}{nr(r-1)}(\sum n_{ij}^2 - nr) \quad (2)$$

여기서, r 은 평가자 수이며, n 은 총 평가 개수이다. n_{ij} 은 i 번째 평가대상을 j 번째 범주로 평가한 평가자 수를 의미한다.

또한, 우연에 의해 기대되는 일치비율은 $P_{e.k}$ 이며 Eq. (3)과 같다.

$$P_{e.k} = \sum_{j=1}^q (\sum_{i=1}^n r_{ij}/r)^2 \quad (3)$$

Eq. (1)에 의해 도출된 k 값이 1에 가까울수록 평가자간의 평가 결과가 같음을 의미하며 0에 가까울수록 다름을 의미한다.

3. 주요 평가항목 도출 결과

3.1 표준편차를 활용한 도출 결과

3.1.1 평가항목별 표준편차 발생 빈도·발생률 결과

5인의 평가위원으로 구성된 365개 시공평가 결과를 바탕으로 평가항목별 평가위원 점수의 표준편차를 도출하였다. 각각의 현장과 평가위원은 모두 독립이므로 Table 3과 같이 현장별 동일 평가항목을 대상으로 계산 하였으며, 이때 표준편차=0일 경우 평가위원 5인의 평가 결과가 모두 같음을 의미하고, 표준편차≠0일 경우 평가위원 1인 이상이 평가 결과가 다르다고 해석할 수 있다.

평가항목별 표준편차 발생 빈도 결과는 Fig. 2 및 Table 4와 같으며, 평가항목 ‘2.1 공정관리계획 적정성 및 적기제출’이 전체 365개 현장 중 242개 현장에서 평가위원별로 평가 결과가 일치하지 않아 가장 큰 표준편차 발생률(66 %)을 보였다. 또한, 평가항목 ‘8.2 중대건설현장 사고 등의 발생 여부’가 365개 현장 중 단 3개 현장에서만 평가위원별로 평가 결과가 다른 것으로 나타났다.

평가항목 별 표준편차 발생률은 Fig. 2와 같이 표준편차 발생률 10 %를 기준으로 그 차이가 명확히 구분되었다. 표준편차 발생률이 10 % 이하로 계산된 평가항목은 전체 25개 평가항목 중 14개 항목이었으며, 10 %를 초과한 평가항목은 전체 25개 항목 중 11개 항목으로 나타났다.

표준편차 발생률 10 % 이하 평가항목의 평균 표준편차 발생률은 5 %였으며, 10 % 초과 평가항목의 평균 표준편차 발생률은 50 %로 나타나 두 그룹간의 표준편차 발생률은 차이는 45 %로 분석되어 상당한 차이가 있음을 알 수 있다.

3.1.2 평가항목별 누적 표준편차 결과

3.1.1.절에서 분석한 표준편차 발생빈도(비율)와 표준편차 누적 값이 가지는 의미는 다르다. 그 예로 5인의 평가위원 중 ‘4인은 우수등급, 1인은 보통등급’으로 평가한 평가결과와 ‘4인은 우수등급, 1인은 미흡등급’으로 평가한 평가결과의 표준편차 발생빈도(비율)에 미치는 영향은 같지만 여기서 발생한 표준편차 값의 차이는 반영이 되지 않는다. 즉, 평가항목별 누적 표준편차 분석을 통해 도출할 수 있는 결과는 누적 표준편차의 크기가 클수록 해당 평가항목을 평가하는 평가위원들 간의 평가 등급 결과가 크게 다르다고 추론 할 수 있다.

Table 3. Example of Standard Deviation Calculation for Each Evaluation Item

Site	Evaluation Criteria	Rater					Standard Deviation
		A	B	C	D	E	
1. A site	2.1 Suitability and timely submission of the process control plan	1.6	1.6	1.6	1.6	1.6	0
2. B site	2.1 Suitability and timely submission of the process control plan	2.0	2.0	2.0	2.0	2.0	0
3. C site	2.1 Suitability and timely submission of the process control plan	1.6	1.6	1.6	2.0	2.0	0.220
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
365. n site	2.1 Suitability and timely submission of the process control plan	2.0	2.0	2.0	2.0	2.0	0
1. A site	3.1 Suitability of the site manpower assignment	1.8	1.8	1.8	1.8	1.8	0
2. B site	3.1 Suitability of the site manpower assignment	2.4	3.0	2.4	2.4	3.0	0.329
3. C site	3.1 Suitability of the site manpower assignment	3.0	3.0	3.0	3.0	3.0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
365. n site	3.1 Suitability of the site manpower assignment	2.4	2.4	2.4	3.0	2.4	0.268
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

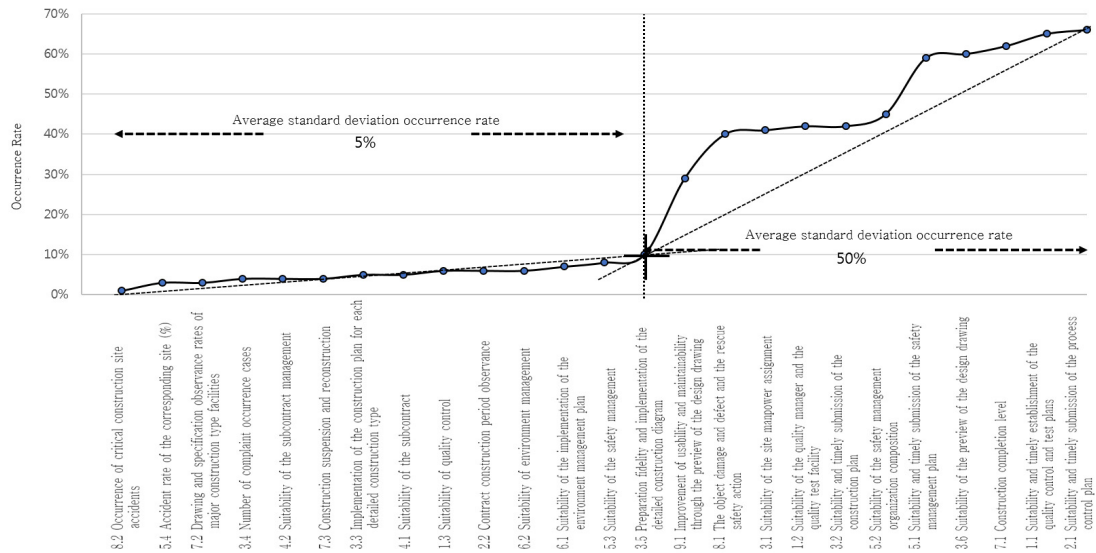


Fig. 2. Rate of Standard Deviation by Evaluation Criteria

Table 4. The Standard Deviation Occurrence Frequency and the Occurrence Rate for Each Evaluation Item

Evaluation Criteria	Number of Standard Deviation Occurrence	Occurrence Rate
1.1 Suitability and timely establishment of the quality control and test plans	236	65 %
1.2 Suitability of the quality manager and the quality test facility	152	42 %
1.3 Suitability of quality control	23	6 %
2.1 Suitability and timely submission of the process control plan	242	66 %
2.2 Contract construction period observance	21	6 %
3.1 Suitability of the site manpower assignment	150	41 %
3.2 Suitability and timely submission of the construction plan	154	42 %
3.3 Implementation of the construction plan for each detailed construction type	19	5 %
3.4 Number of complaint occurrence cases	16	4 %
3.5 Preparation fidelity and implementation of the detailed construction diagram	35	10 %
3.6 Suitability of the preview of the design drawing	220	60 %
4.1 Suitability of the subcontract	17	5 %
4.2 Suitability of the subcontract management	16	4 %
5.1 Suitability and timely submission of the safety management plan	214	59 %
5.2 Suitability of the safety management organization composition	166	45 %
5.3 Suitability of the safety management	31	8 %
5.4 Accident rate of the corresponding site (%)	11	3 %
6.1 Suitability of the implementation of the environment management plan	26	7 %
6.2 Suitability of environment management	21	6 %
7.1 Construction completion level	225	62 %
7.2 Drawing and specification observance rates of major construction type facilities	11	3 %
7.3 Construction suspension and reconstruction	13	4 %
8.1 The object damage and defect and the rescue safety action	147	40 %
8.2 Occurrence of critical construction site accidents	3	1 %
9.1 Improvement of usability and maintainability through the preview of the design drawing	106	29 %

따라서 본 절에서는 평가항목별 평가점수를 모두 10점 만점으로 환산하여 표준편차를 계산하였으며 평가항목별로 도출된 표준편차 누적 값의 대푯값을 활용하여 Fig. 3과 같이 상자그림으로 표현하였다. 누적 표준편차의 크기가 가장 큰(Maximum) 평가항목으로는 '2.1 공정관리계획 적정성 및 적기제출' 이었으며, 그 다음으로는 '3.6 설계도서 사전 검토의 적정성' 평가항목 순이었다. 중앙값(Median) 기준으로 가장 큰 누적 표준편차를 보인 평가항목으로는 '7.1 공사완성도', '3.6 설계도서 사전 검토의 적정성' 순이었으며, 3분위(Q3, 75번째 백분위수)값 기준으로는 '7.1 공사 완성도', '2.1

공정관리계획 적정성 및 적기제출' 순으로 도출되었다(Table 5).

3.1.1절 표준편차 발생 빈도 기준 상위권으로 분류된 평가항목과 3.1.2절 누적 표준편차 기준으로 도출된 주요 평가항목은 대표 해석값 따라 순서의 차이는 있으나 상위 11개의 평가항목은 모두 동일하게 도출되었다.

3.2 Fleiss Kappa Value(k)를 활용한 도출 결과

3.2.1 Fleiss Kappa Value(k)의 해석 기준

Kappa value(k)의 해석 기준에 대해서는 연구자마다 견해의

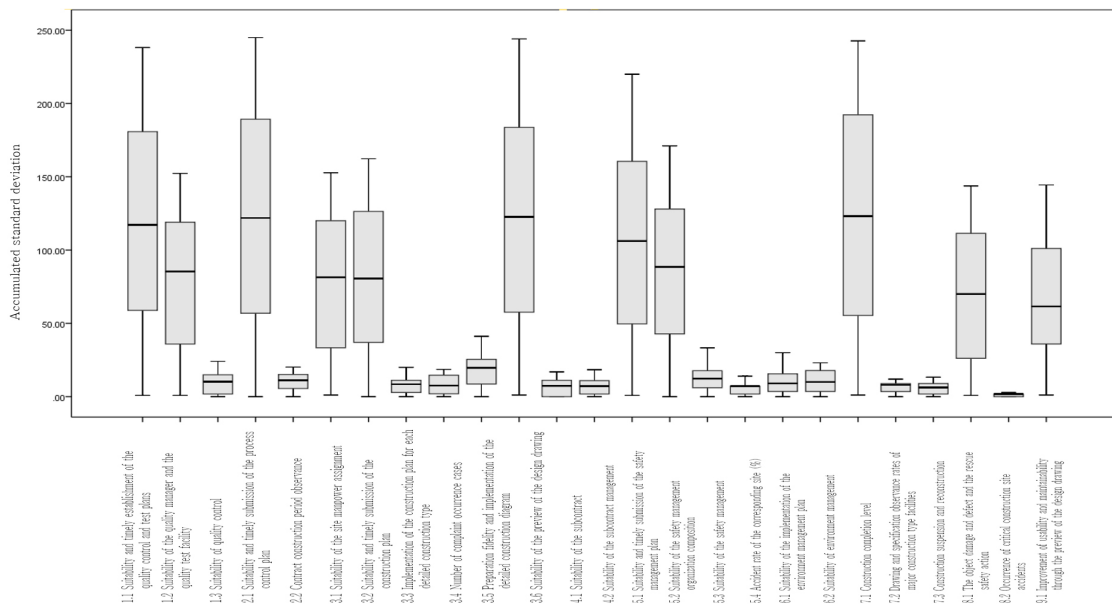


Fig. 3. Cumulative Standard Deviation by Evaluation Criteria

Table 5. Accumulated Representative Standard Deviation Value for Each Evaluation Item

Evaluation Criteria	Q1	median	Q3	Maximum
2.1 Suitability and timely submission of the process control plan	56.923	121.895	189.259	244.976
3.6 Suitability of the preview of the design drawing	57.646	122.622	183.674	244.013
7.1 Construction completion level	55.402	123.112	192.267	242.708
1.1 Suitability and timely establishment of the quality control and test plans	58.867	117.176	180.762	238.177
5.1 Suitability and timely submission of the safety management plan	49.656	106.177	160.507	219.972
5.2 Suitability of the safety management organization composition	42.792	88.559	127.982	171.066
3.2 Suitability and timely submission of the construction plan	37.004	80.580	126.347	162.284
3.1 Suitability of the site manpower assignment	33.335	81.411	120.023	152.755
1.2 Suitability of the quality manager and the quality test facility	35.908	85.343	118.970	152.305
9.1 Improvement of usability and maintainability through the preview of the design drawing	35.924	61.544	101.064	144.445
8.1 The object damage and defect and the rescue safety action	26.160	70.027	111.392	143.723
⋮	⋮	⋮	⋮	⋮

차이가 있지만 본 연구에서는 일반적으로 가장 많이 사용하는 Landis and Koch(1977)의 기준을 적용하여 해석하였다. 이 분류가 절대적이지는 않지만 유용하다고 판단되어 널리 활용되고 있다 (Donner and Eloasziw, 1987)

Table 6. Analysis of Kappa Value(k)

k	Analysis
<0	poor agreement
0.000-0.200	slight agreement
0.201-0.400	fair agreement
0.401-0.600	moderate agreement
0.601-0.800	substantial agreement
0.801-1.000	almost perfect agreement

Table 7. Number of Items for Each Fleiss Kappa Value(k) Range

k	Number of Items
<0	0
0.000-0.200	1
0.201-0.400	1
0.401-0.600	4
0.601-0.800	8
0.801-1.000	8

Table 8. Fleiss Kappa Value(k) of Each Evaluation Criteria

Evaluation Criteria	k
1.1 Suitability and timely establishment of the quality control and test plans	0.488
1.2 Suitability of the quality manager and the quality test facility	0.692
1.3 Suitability of quality control	0.896
2.1 Suitability and timely submission of the process control plan	0.446
2.2 Contract construction period observance	0.926
3.1 Suitability of the site manpower assignment	0.687
3.2 Suitability and timely submission of the construction plan	0.690
3.3 Implementation of the construction plan for each detailed construction type	0.738
3.4 Number of complaint occurrence cases	0.938
3.5 Preparation fidelity and implementation of the detailed construction diagram	0.927
3.6 Suitability of the preview of the design drawing	0.526
4.1 Suitability of the subcontract	0.934
4.2 Suitability of the subcontract management	0.776
5.1 Suitability and timely submission of the safety management plan	0.536
5.2 Suitability of the safety management organization composition	0.602
5.3 Suitability of the safety management	0.920
5.4 Accident rate of the corresponding site (%)	0.969
6.1 Suitability of the implementation of the environment management plan	0.953
6.2 Suitability of environment management	0.805
7.1 Construction completion level	0.315
7.2 Drawing and specification observance rates of major construction type facilities	0.676
7.3 Construction suspension and reconstruction	0.874
8.1 The object damage and defect and the rescue safety action	0.260
8.2 Occurrence of critical construction site accidents	0.814
9.1 Improvement of usability and maintainability through the preview of the design drawing	0.758

Landis와 Koch가 제시한 해석기준은 kappa value(k) 기준 0.000~0.200은 약간의 일치도, 0.201~0.400은 어느 정도 일치도, 0.401~0.600은 적당한 일치도, 0.601~0.800은 상당한 일치도, 0.801~1.000은 완벽한 일치도로 판단하고 있다(Table 6).

3.2.2 Fleiss Kappa Value(k)의 해석 결과

Fleiss Kappa value(k) 분석 결과 약간의 일치도(0.000~0.200) 범위에 포함된 평가항목은 1개였으며 해당항목은 ‘8.1 목적물 손상 및 결함, 구조안전 조치 여부’이다. 어느 정도 일치도(0.201~0.400) 범위에 포함된 평가항목도 1개 도출되었으며 해당 항목은 ‘7.1 공사 완성도’ 항목이다. 적당한 일치도 기준인 0.401~0.600에 포함된 평가항목은 4개 항목으로 ‘2.1 공정관리계획 적정성 및 적기제출’, ‘1.1 품질관리계획 및 품질시험계획의 적기수립 및 적정성’ 등 순이었으며, 상당한 일치도 기준인 0.601~0.800에 포함된 평가항목은 8개 항목으로 ‘5.2 안전관리조직 구성의 적정 여부’, ‘7.2 주요 공중 시설물의 도면 시방서 준수비율’ 등의 순이다. 완벽한 일치도 기준인 0.801~1.000에 포함된 평가항목은 총 8개 항목이었다(Table 7). 이중 Fleiss Kappa Value(k)가 가장 높은 항목은 ‘5.4 당해 현장의 재해율(%)’ 항목이 도출 되었다(Table 8).

4. 주요 평가항목 비교 및 특성 고찰

표준편차 및 Fleiss Kappa 분석을 통해 도출된 평가항목을 편차 발생률과 Kappa value(k) 기준, 상위 11개 평가항목은 Table 9와 같다.

표준편차 발생률 기준으로 '2.1 공정관리계획 적정성 및 적기제

출', '1.1 품질관리계획 및 품질시험계획의 적기수립 및 적정성' 등의 순으로 높은 표준편차 발생률을 보였으며, Kappa value(k) 기준으로는 '8.1 목적물 손상 및 결함, 구조안전 조치여부', '7.1 공사 완성도' 등의 순으로 평가위원별 평가결과의 불일치도를 보였다. 두 분석방법으로 도출된 평가항목별 순위는 약간의 차이를 보이

Table 9. Rank of the Major Derivation Items for Each Analysis Method

Rank	Major improvement items using the standard deviation		Major improvement items through the Fleiss Kappa Value(k)	
	Evaluation Item	Occurrence Rate	Evaluation Item	k
1	2.1 Suitability and timely submission of the process control plan	66 %	8.1 The object damage and defect and the rescue safety action	0.260
2	1.1 Suitability and timely establishment of the quality control and test plans	65 %	7.1 Construction completion level	0.315
3	7.1 Construction completion level	62 %	2.1 Suitability and timely submission of the process control plan	0.446
4	3.6 Suitability of the preview of the design drawing	60 %	1.1 Suitability and timely establishment of the quality control and test plans	0.488
5	5.1 Suitability and timely submission of the safety management plan	59 %	3.6 Suitability of the preview of the design drawing	0.526
6	5.2 Suitability of the safety management organization composition	45 %	5.1 Suitability and timely submission of the safety management plan	0.536
7	1.2 Suitability of the quality manager and the quality test facility	42 %	5.2 Suitability of the safety management organization composition	0.602
8	3.2 Suitability and timely submission of the construction plan	42 %	7.2 Drawing and specification observance rates of major construction type facilities	0.676
9	3.1 Suitability of the site manpower assignment	41 %	3.1 Suitability of the site manpower assignment	0.687
10	8.1 The object damage and defect and the rescue safety action	40 %	3.2 Suitability and timely submission of the construction plan	0.690
11	9.1 Improvement of usability and maintainability through the preview of the design drawing	29 %	1.2 Suitability of the quality manager and the quality test facility	0.692

Table 10. Evaluation Criteria for Major Improvement Items

Evaluation Criteria	Evaluation Grade			
	Good ($\times 1.0$)	Average ($\times 0.8$)	Insufficient ($\times 0.6$)	Poor ($\times 0.4$)
1.1 Suitability and timely establishment of the quality control and test plans	Submitted timely and very suitable	Submitted timely and suitable	Submitted late or one supplementation	Unsubmitted or 2 or more supplementations
1.2 Suitability of the quality manager and the quality test facility	Very suitable	Suitable	Unsuitable	
2.1 Suitability and timely submission of the process control plan	Submitted timely and very suitable	Submitted timely and suitable	Submitted late or 1st supplementation	Unsubmitted or 2 or more supplementations
3.1 Suitability of the site manpower assignment	Very suitable	Suitable	Unsuitable	Very unsuitable
3.2 Suitability and timely submission of the construction plan	Submitted timely and very suitable	Submitted timely and suitable	Submitted late or 1st supplementation	Unsubmitted or 2 or more supplementations
3.6 Suitability of the preview of the design drawing	Very suitable	Suitable	Insufficient	Unimplemented
5.1 Suitability and timely submission of the safety management plan	Submitted timely and very suitable	Submitted timely and suitable	Submitted late or 1st supplementation	Unsubmitted or 2 or more supplementations
5.2 Suitability of the safety management organization composition	Very suitable	Suitable	Unsuitable	Very unsuitable
7.1 Construction completion level	95 % or more satisfied	90 % or more satisfied	80 % or more satisfied	Less than 80 % satisfied
8.1 The object damage and defect and the rescue safety action	Few problems with the overall major subsidiary materials	Minor damages or defects of major subsidiary materials	Doing the precision safety check on the structural safety	Implementing strengthening for securing the structural safety

나 11개 항목 중 10개 항목(91 %)이 중복 도출되어 양 분석방법 모두 유사한 결과가 도출되었다.

중복 도출된 평가항목 및 평가기준은 Table 10과 같으며 평가기준의 공통된 특성을 도출 할 수 있었다. 10개의 평가항목 중 8개의 평가항목 세부 평가기준에 ‘매우적정’, ‘적정’과 같은 정성적 평가 기준이 포함되어 있었으며, 이는 시공평가 세분류 25개의 평가항목 중 ‘매우적정’, ‘적정’이 평가기준으로 정의된 8개의 평가항목 전체가 개선이 필요한 평가항목으로 도출된 것이다.

‘건설기술용역 및 시공 평가지침’에서 적정성 관련 평가 기준은 “관련 기준을 만족한 경우 적정”, “다수 항목이 기준 이상을 만족하는 경우 매우적정”으로 정의하고 있으며, 이처럼 모호한 정의가 평가위원의 평가 결과에 영향을 준 것으로 판단된다.

나머지 2개(‘7.1 공사완성도’, ‘8.1 목적물 손상 및 결함, 구조안전조치 여부’)의 평가항목은 대부분 기준 ‘목적물의 품질 및 상층’에 포함되는 평가항목이다. 평가기준은 각 공종별 완성도를 우수, 보통, 미흡, 불량 등급으로 평가한 후, 이를 공종별 공사비 가중치를 반영하여 평가하도록 정의하고 있다. 하지만 평가 시점인 준공 후 60일까지는 하자가 거의 발생되지 않으며 발생하여도 귀책사유가 밝혀지기까지 시일이 소요되어 평가위원의 판단이 쉽지 않아 주관적인 시일이 반영된 것으로 판단된다.

반면, 최소 표준편차 발생빈도 및 최대 Fleiss Kappa value(k)로 분석된 ‘5.4 당해 현장의 재해율(%)’의 경우, 한국산업안전보건공단의 평균환산 재해율을 기준으로 평가되어 그 기준이 명확하고 정량적 세부기준으로 평가기준이 설정되어 본 연구에서 도출된 개선필요 항목과 쉽게 대비 되었다.

5. 결론

2014년 최고 가치와 글로벌 스탠다드를 추구한다는 목표 아래 도입된 ‘종합심사낙찰제’가 운영 중에 있다. 하지만, 공사수행능력 심사분야 중 최대 50 %를 차지하는 공공공사 시공평가는 객관성 부족, 주관적 평가항목 다수 존재 등의 문제가 지속적으로 제기되고 있으며 이를 해결하기 위해 중앙부처, 지자체 등의 노력이 지속되고 있으나 그 효과는 미비한 실정이다.

따라서 본 연구에서는 365개 공공공사의 평가위원별 시공평가 점수를 활용하여 표준편차 및 Fleiss Kappa 분석을 통해 객관성 확보가 필요한 주요 평가항목을 도출하고 그 특성을 도출하였다.

그 결과 표준편차 발생률 기준 상위 11개 평가항목과 Fleiss Kappa value(k) 기준 상위 11개 평가항목을 개선이 필요한 평가항목으로 분석 되었으며, 두 방법으로 도출된 평가항목의 순서에는 차이가 있으나 10개의 평가 항목이 중복 도출되어 두 분석방법 모두 유사한 결과를 보였다.

또한, 중복 도출된 평가 항목의 특성을 파악한 결과 세부 평가지

침에 ‘매우적정’, ‘적정’과 같은 정성적 평가 기준이 공통으로 포함 된 것을 확인할 수 있었다.

따라서, 향후 본 연구에서 도출된 평가항목 중심으로 평가기준 개선이 필요하며, 개선 시에는 모호한 평가기준과 정성적 평가기준은 지양하고 객관성을 확보할 수 있는 정량적 평가기준 중심으로 시공평가 기준이 개선되어야 한다. 객관성이 확보된 평가기준이 수립이 이루어져야 비로소 종합심사낙찰제도 도입 목적을 달성할 수 있다고 판단된다.

본 연구에 활용된 자료는 서로 상이한 발주청의 자료를 모두 활용한 연구로써 특정 발주청의 평가결과만을 대상으로 연구를 수행하였을 경우 우선개선 항목으로 분류된 평가항목이 다소 상이할 수 있다. 또한 대표 공종별로 분류하여 본 연구를 수행할 경우 또한 그 결과를 다소 상이할 수 있다.

References

- Berry, K. J. and Mielke, P. W. (1988). “A generalization of Cohen’s kappa agreement measure to interval measurement and multiple raters.” *Educational and Psychological Measurement*, Vol. 48, No. 4, pp. 921-933.
- Busan Metropolitan City (2017). *Full Revision (Draft) of ‘Busan metropolitan city construction technology service and construction evaluation guidelines* (in Korean).
- Donner, A. and Eliasziw, M. (1987). “Sample size requirements for reliability studies.” *Stat Med.*, Vol. 6, No. 4, pp. 441-449.
- Fleiss, J. L. (1971). “Measuring nominal scale agreement among many raters.” *Psychological Bulletin*, Vol. 76, No. 5, pp. 378-382.
- Landis, J. R. and Koch, G. G. (1977). “The measurement of observer agreement for categorical data.” *Biometrics*, Vol. 33, No. 1, pp. 159-174.
- Ministry of Economy and Finance (MOEF) (2016). *Review criteria for the comprehensive review successful bid system*, Established Rules No. 283 (in Korean).
- Ministry of Land, Infrastructure and Transport (MOLIT) (2015). *No. 2015-505 Construction technology service and construction work construction evaluation guidelines* (in Korean).
- Ministry of Land, Infrastructure and Transport (MOLIT) (2017). *No. 2017-531 Construction technology service and construction evaluation guidelines* (in Korean).
- Ministry of Land, Transport and Maritime Affairs (MLTM) (2009). *100 % quantification of the construction evaluation item -reflecting the construction cost reduction rate of the constructor-* (in Korean).
- Park, H. S. (2016). *A brief view on the comprehensive review successful bid system*, Korea Institute of Public Finance (in Korean).
- Scott, W. A. (1955). “Reliability of content analysis : the case of nominal scale coding.” *Public Opinion Quarterly*, Vol. 19, No. 3, pp. 321-325.