



## How is it going? Performance assessment in major projects

Maylor, Harvey ; Johnson, Mark; Turner, Neil ; Geraldi, Joana

*Published in:*

Proceedings of the 77th annual meeting of the academy of management

*Publication date:*

2017

*Document Version*

Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*

Maylor, H., Johnson, M., Turner, N., & Geraldi, J. (2017). How is it going? Performance assessment in major projects. In *Proceedings of the 77th annual meeting of the academy of management* Design Society.

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# So, how's it going?

## Performance assessment in major projects

Harvey Maylor, Mark Johnson; Neil Turner and Joana Geraldi

### **Abstract**

Determining the performance of a major project is a challenge for both practitioners and scholars. In the context of operational change projects the challenge is exacerbated by the service-intensive nature of the transformation, temporal disconnects between contracting and delivery and lack of appropriate metrics. This paper considers performance from a service delivery perspective. Current measures of performance were noted to be inadequate in practice. The service operations literature provided frameworks which were investigated for their utility and supplemented by qualitative data to generate an enhanced service performance model. This was then tested using a survey and a structural equation model derived. Development of this yielded new classifications but most importantly, provided a more meaningful method for measuring the performance of operational transformation projects. Specifically, it is shown that expectations and perceptions are measured on different scales, and that quality performance is inseparable from other performance aspects. The contributions of this paper are to address an important practical problem but also to contribute to the development of the discussion of performance management in major projects from an OM perspective.

### **Keywords:**

Transformation projects, performance, service, SERVQUAL, temporal disconnects, satisfaction.

## Introduction

The point of departure for this study was a practitioner problem in the delivery of an operational transformation project. A major change to the operations of the client organisation was being undertaken by an outsourced service provider, involving the implementation of a new generation of ERP system. In the discussion of ‘how it’s going’ the research team noted that a number of managers complained, “*despite all of our key performance indicators showing green, the client still isn’t happy.*” ‘Green’ in this case indicated that performance in the defined indicators was at a pre-agreed (acceptable) level, but this wasn’t sufficient to imbue a state of satisfaction in the client. Any deviation between required levels of progress in projects was noted as a change in colour of a KPI from green to amber or red. When investigated further, it was noted that managers were ‘managing the gap’ between expectations and perceptions, as proposed by Maister (1993) – attention would be focused on any indicator that was not green. Preliminary investigation of the indicators, showed that they used a very limited conception of performance. Specifically, performance was assessed using product rather than service-based measures, as these had formed part of the contract between the client organisation and the outsourcing company. This paper reports the subsequent investigation into the conceptualisation of performance in the context of major operational change, in this case as carried out by an outsourced operation.

Outsourcing of IT is extensively employed by large organisations from all sectors; it is usually long-term and relational; and involves the service provider being embedded within the client organisation. In OM terms, outsourced IT provision is a complex B2B service offering, often comprising both repetitive operations (referred to in the industry as “*man-*

*aging the mess for less*” and non repetitive operations (projects) to develop new or enhanced capabilities. This second group of activities are referred to elsewhere as “*IT-enabled change*” (Benjamin and Levinson, 1993) and is the concern for this discussion.

The prevailing approaches to managing performance in a wide range of organisations involved in outsourced IT provision are seen to be much less mature than in other service operations, for example in repetitive B2C contexts (Chase and Apte, 2007). In the examples used above, they had a focus on product-based control and conformance to contract. Such manufacturing-centric definitions were noted to be inconsistent with the nature of the services offering (Haywood-Farmer, 1988; Parasuraman et al. 1985). In order to progress the development of the approaches to managing performance, a more appropriate conceptualisation is required.

This paper begins by considering the basis for a suitable service-based approach to performance measurement, from the service OM literature, and the nature of the context in which it is to be applied – that of operational change. The research design is then derived using a two-stage process – the first to explore the nature of performance in this context, the second to investigate whether this provides a suitable model for performance and how it can be enhanced. The findings are examined and areas for further development analysed.

### **Service-based performance measurement**

For the purposes of our discussion we will focus on what is known in projects as ‘the iron triangle’ of performance – time, cost and quality. Measures for performance of both time and cost attributes are relatively well-established (see e.g. Maylor, 2010), and so our initial

concern is with the quality performance. However, as will be shown empirically, this distinction is not sustained in practice with implications for further studies.

## Quality performance

Both scholars and practitioners have found that producing a definition of quality in general is far from straightforward. For instance, Garvin (1992:126) stated that “*Quality is an unusually slippery concept, easy to visualise, and yet exasperatingly difficult to define.*” The field of quality can be split into major two schools of thought. Firstly, the ‘product quality’ school developed in manufacturing, and has formed a key part of OM theory and practice. This has been supplemented by the second – the ‘service quality’ school, developed in both marketing and OM, with its focus on the B2C market. Developments in the school of product quality, following Shewhart (1931), focused on controlling manufacturing processes, minimising variation and developing tools and techniques to support these goals. Key figures and their major contributions are summarised in Table 1.

*Table 1 – ref Maylor (2000:254)*

<i>Instigator</i>	<i>Major Influence</i>
Deming	Initially developed statistical methods for control of repetitive processes, and their usage. Took the tools to Japan post world-war 2, and was seen as part of the Japanese quality revolution (though largely ignored in his native US until the last years of his life). Recommended 14 points for management, and the use of the Plan-do-check-act cycle (also know as the Deming Cycle).
Crosby	Championed the notion of ‘zero defects’, and ‘quality is free.’
Feigenbaum	The holistic approach to quality – company-wide or total quality control.
Taguchi	The proponents of Taguchi methods claim great results for the design of experiments, though good examples are few and far between.
Ishikawa	Quality circles and brainstorming tools including the fishbone or Ishikawa diagram.
Ohno	Architect of the Toyota Production System which took quality to new levels in manufacturing, through teamwork, training and education, ongoing continuous improvement and a focus on the absolute elimination of waste.
Juran	The engineer’s quality guru – established and compiled the requirements of systems and procedures for sampling and control with tangible products. Many of the routines are equally applicable to the tangible elements of service products.

The strategic nature of quality was highlighted by Garvin (1984), and showed how product quality could be considered as a competitive tool. It has led to the adoption of practices by organisations, such as TQM (Boaden, 1997; Dotchin and Oakland, 1992; Spencer, 1994), techniques such as Six-Sigma, and Quality Systems such as ISO 9000. This can improve operational performance (Ahire et al., 1996; Sousa and Voss, 2002), although the ‘softer’ aspects such as leadership, HR management and customer focus have been shown to be important in the implementation (Powell, 1995; Samson and Terziovski, 1999).

## Services

In OM terms, there are three generalised and well-documented differences between the production of goods and services: intangibility, heterogeneity and inseparability (Parasuraman et al., 1985). Consequently, the service literature takes a markedly different approach to the concept of quality from manufacturing. Whilst products can generally be measured in terms of attributes and evaluated by objective criteria, service quality is conceived as the difference between what the consumers perceive against what they expect from the service. It is therefore a subjective approach. The discussion of service quality has been dominated by two schools of thought. The “Nordic/European” school (characterised by Grönroos, 1984), and the “American” school, after Parasuraman et al. (1985). Both are based on evaluation of consumers, and this needs to be considered when applying the models to complex service offerings such as major projects. However, the frameworks are beneficial in offering accepted methods of analysis that provide a starting point for our consideration.

Grönroos’ (1984) study of service quality differentiates between technical quality (the “what” aspect, or “instrumental performance”) and functional quality (the “how” aspect,

or “expressive performance”) in a service delivery environment. The former is an evaluation based on what the consumer receives, the latter an evaluation of how the service is delivered. Adequate instrumental performance (i.e. the output) is a prerequisite for customer satisfaction in the contexts researched by Grönroos, and its role in major projects needs to be tested. This has some useful managerial implications, not least that if the expressive performance (the service delivery aspect) of the product is not considered satisfactory, then the consumer will still feel unsatisfied, irrespective of the degree of satisfaction caused by the instrumental performance. Harvey (1998:98) notes that when delivery is a key part of the service, then “*perception is reality.*” The perceived service is the consumers’ view of a “bundle” of service dimensions. When this perceived service is compared to the expected service, we get the perceived service quality. However Grönroos also considers a third quality dimension, that of corporate image, and asserts that the expectations of consumers are influenced by their view of the company. Kang and James (2004) identify that this acts as a form of filter in terms of the consumer’s perception of quality. Other similar models are also relevant in the context of service quality, as shown in Table 2.

*Table 2 – Alternative Service Quality Models*

<b>Authors</b>	<b>Dimensions of Service Quality Model</b>
Lehtinen and Lehtinen (1991)	Physical Quality, Interactive Quality, Corporate Quality, Process Quality and Output quality.
Rust and Oliver (1994)	Service Product (technical quality of the outcome), Service Delivery (functional or process quality) and Service Environment.
Philip and Hazlett (1997)	Pivotal, Core and Peripheral Attributes.
Haywood-Farmer (1988)	Professional Judgement, Physical Facilities and Processes, People’s behaviour.

## The context: major projects

In OM terms, the context of projects in general has received relatively little attention (Walker et al, 2015). This is despite the ubiquity and importance of the context. This is beginning to be addressed including through sessions at AoM and special issues of OM journals, but the field remains wide open for OM scholars. We use the definition of a major project as *a set of interrelated operations uniquely and temporarily established to achieve a significant purpose and on a significant scale* (author ref). Major projects include infrastructural development (buildings, roads, dams) which in budgetary terms tend to steal the headlines. However, just as important is the operational transformations that take place in the effort by organisations to deploy strategy. These are also of a significant scale, less so in budgetary terms, but in terms of the extent of their reach organisationally. Mergers and acquisitions, re(dis)organisations, the implementation of a new IT system and are all examples that have significant impact on the operations of any organisation. Indeed, the distinction between infrastructure and transformation projects is only partially helpful as most infrastructure projects involve some transformation (e.g. a new hospital requiring new operating methods for staff) and vice versa (a major reorganisation requiring new facilities for staff).

In analysing the process aspects present in our scenario, there are two operations of interest – the one doing the transforming, and the one being transformed. We are interested in the performance of the former, as assessed by the latter.

The context of a major transformational project needs some further description before we can consider how the performance of the process can be assessed. The following are the key criteria of interest:



1. We are considering a business process with a high degree of uniqueness. By definition a project has a high variety of process and low volume of throughput. Due to its scale and impact, it is a relatively infrequent event.
2. It contains inherent conflict in purpose for the two operations. It is often the case that the operation to be transformed has the objective to ‘keep the shop open’ – sustaining operations during the change, whilst the transforming operation exists solely to execute the change.
3. It is a lengthy engagement, lasting months or potentially years.
4. The transforming operation is embedded into the one being transformed and as a result is a highly relational process.

Applying the above criteria to the assessment of performance in major projects is clearly a challenge for those managers involved as either clients or service providers – as evidenced by the scenario outlined at the start of the paper. With a high degree of uniqueness in a process, the ability to set appropriate performance measures will have limited experience on which to draw. These will also need to be cognisant of the inherent conflict between operations and change. The length of the engagement does also present a challenge as performance criteria established at the start of the project may lose their relevance as the project progresses. Lastly, we noted that our initial scenario used performance measures that were product-focused (as assessed against a specification for the product), progress-related (achieving milestones for delivery as set out in the plans) and resource-related (number of suitably qualified and experienced people that the contractor had allocated to the project). The only ‘service’ measure was a ‘Service Level Agreement’ concerning how

quickly system problems would receive attention and be rectified. However, as the scenario demonstrated, despite measuring these at a high level of detail, client satisfaction was proving elusive, and that these do not recognise the relational nature of the transformation process.

We posit that the instrumental or product performance measures did not reflect the client requirements well or the nature of the process and that as a result, the measures themselves were incomplete. These may be considered ‘pre-requisites’, potentially even ‘hygiene factors’ but in and of themselves are insufficient, no matter how many indicators may be included on a project dashboard. From the service operations literature, we notice that for instance good expressive or service performance may also be required, and potentially too, Grönroos’ third dimension of corporate image.

For now, it is this second set of expressive or service performance elements that are of interest. However, the diverse and generalised requirements described in Table 2, appeared to be either inappropriate or incomplete for the context we are considering; no clear method was apparent that could adequately be used to determine the nature of quality in the context of transformation projects. In the past researchers have used SERVQUAL (Parasuraman *et al.* 1988) to achieve an appropriate (cf. manufacturing) conceptualisation of quality. Given its ubiquity and apparent flexibility, we chose SERVQUAL as the starting point for this context.

## Research Design

The first stage of this study was designed to answer the question, *what are the quality performance attributes of operational change?* We assessed this contextualisation through a pilot comprising eighteen exploratory interviews with project and programme managers, quality managers, and senior executives within a large IT provider. This sample was purposive and was limited to this one sector.

The interviews were conducted in English; took 30-90 minutes; and were taped and transcribed. The interviews were structured around two areas:

- **Context** explored the background, job and responsibilities of the interviewee in general as well as the current practices in management and measurement of quality.
- **Defining Quality** delved into the interviewees' understanding of quality and the respective key attributes by comparing projects they considered of high quality and low quality. This was the main part of the interview. We did not mention SERVQUAL or its constructs and attributes in the interviews.

We analysed the data by a process of coding, and then contrasting the concepts extracted with the attributes in SERVQUAL. Five additional concepts were found over and above those in the standard SERVQUAL framework and these fitted well within the existing categories (see Table 3).

*Table 3: additional SERVQUAL items*

<i>Additional SERVQUAL item</i>	<i>Category</i>
Adapt styles	Empathy
Zero defects	Reliability
Smooth execution	Reliability

Change management	Responsiveness
Back on track	Responsiveness

‘Adapt styles’ concerned the ability of the delivery team to match their style to the culture, state and level of people that they were working with at that time. This would include the ability to handle change in an appropriate manner, recognising that this can be politically sensitive. ‘Zero defects’ might appear to be a product characteristic, but in this context the reliability of the service was described in this manner. The zero defects applied to the level of errors that the client perceived had been made in the process. ‘Smooth execution’ relates to the general experience of the client in the delivery of the work. ‘Change management’ (as distinct from ‘the management of change’) is a standard feature of most projects, and so it is not surprising to see this concept emerge strongly here. Specifically, this concerned how responsive the supplier would be to both formal and informal change requests. ‘Back on track’ reflected the reality that in such work there would be disturbances, for instance caused by a change in the requirements or an issue that has arisen from the work carried out. For the client, the ability to move from this disturbance to a restored sense of purpose with the work being carried out, was important.

Following well described scale development procedures (Parasuraman *et al.* 1986), we then produced an instrument comprising 44 questions, using 7-point Likert scales, 22 which assess the expectations and 22 the perceptions of quality. Before deploying the instrument to the sample, the revised framework was piloted (and subsequently went through minor amendments) with two academics and three practitioners not involved in the research project to check for clarity, understanding and readability.

The instrument was then used to test the attributes of quality in a cluster sample of 40 business units based in several countries in Europe within a global IT outsource service provider. Each of the business units operated as an independent company, but still within the same business and with certain similarities. Through these means we reduced the number of other possible contextual variables (e.g. corporate culture, industry specific factors) that might have influenced the data, but guaranteed a still heterogeneous set. The instrument was deployed via a web-based survey where a link was sent via email to five managers of each business unit ( $n= 200$ ) at different levels and with different responsibilities. Forty-one responses were received in the first week, and forty-four in the second week after a reminder. From the survey, 85 usable assessments were obtained, a return rate of 42.5%. Table 4 details the geography, role and experience of the respondents.

*Table 4: Sample Frame*

<b>Country</b>		<b>Experience in current role (years)</b>	
UK	47	1 to 5	42
Italy	21	6 to 10	26
Other European Countries	17	11 to 20	14
		Over 21	2
		No Answer	1
<b>Current Role</b>			
PMO and support Staff	8	Project Manager	25
Line manager and staff	7	Account Manager	17
Team Manager	5	Programme Manager	12
No Answer	1	Technical Staff	10

## **Analysis, Results and Discussion**

The data were analysed in three steps. In the first step, we evaluated the validity and reliability of the original and contextualised SERVQUAL constructs. We then conducted a cluster analysis to determine whether there were different configurations of the scales and constructs possible within the data. We then constructed and tested a measurement model using confirmatory factor analysis (CFA) to determine whether the new constructs fit the data better than the existing SERVQUAL constructs.

Table 5 shows the results for the reliability and validity tests and the measurement model.

*Table 5: Cronbach's Alpha, AVE, Path Loading, for Original and Contextualised ServQual*

	Expectations		Perceptions	
	Original	Context	Original	Context
<b>Assurance</b>	$\alpha=0.811$ / AVE= 0.629		$\alpha=0.811$ / AVE= 0.640	
Confidence		0.75		0.75
Safety		0.67		0.67
Courteous staff		0.60		0.60
Having necessary knowledge		0.79		0.79
<b>Empathy</b>	$\alpha=0.775$ AVE= 0.590	$\alpha=0.652$ AVE= 0.584	$\alpha=0.695$ AVE= 0.600	$\alpha=0.790$ AVE= 0.597
Accessible	0.58	0.59	0.58	0.62
Focus on client's interest	0.42	0.53	0.42	0.79
Understand needs of clients	0.74	0.82	0.74	0.64
Adapt styles	-	0.63	-	0.83
<b>Visual appeal</b>	$\alpha=0.806$ / AVE= 0.596		$\alpha=0.599$ / AVE= 0.422	
Updated equipment		0.54		0.54
Professional appearance of people		0.61		0.61
Appearance of physical facilities		0.59		0.59
Professional appearance of materials		0.92		0.92
<b>Reliability</b>	$\alpha=0.863$ AVE= 0.628	$\alpha=0.845$ AVE= 0.760	$\alpha=0.827$ AVE= 0.736	$\alpha=0.860$ AVE= 0.623
Keep promises on time	0.84	0.85	0.84	0.81
Interest in solving client's problems	0.84	0.84	0.84	0.74
Error-free records	0.77	0.75	0.77	0.61
Zero Defect	-	0.60	-	0.75
Smooth execution	-	0.70	-	0.76
<b>Responsiveness</b>	$\alpha=0.703$ AVE= 0.622	$\alpha=0.830$ AVE= 0.597	$\alpha=0.612$ AVE = 0.545	$\alpha=0.782$ AVE= 0.530
Inform progress	0.78	0.80	0.78	0.49
Prompt service	0.81	0.77	0.81	0.70
Commitment	0.37	0.43	0.37	0.65
Change management	-	0.73	-	0.76
Back on track	-	0.74	-	0.61
NFI=	0.739	0.689	0.786	0.710
CFI=	0.813	0.787	0.888	0.839
$\chi^2/df=$	2.520	2.316	1.643	1.795
$p=$	0.000	0.000	0.000	0.000

We analysed the internal consistency of the original constructs through Cronbach's Alpha using SPSS 22.0. The majority of the values were between 0.7 and .8, which are considered adequate and good for exploratory purposes (Nunally, 1978), although some values fall below this limit which indicates a lack of internal consistency within the constructs.



The contextualised SERVQUAL constructs had better results than the original SERVQUAL constructs, with only one value under 0.7, and an average of .795.

To assess the convergent validity of constructs (extent to which indicators are related to their respective construct), we firstly assessed the discriminant validity (degree of overlap between indicators within the same construct) through AVE (average variance extracted) as proposed by Fornell and Larcker (1981). The values were greater than .50, and thus justify the use of the construct (Hair *et al.* 2009). We then conducted a CFA using AMOS 16 of the original SERVQUAL and the extended SERVQUAL models with full information and maximum likelihood estimation. One of the indicators of each construct was fixed to a value of 1.0, so that the constructs were scale-variant (Jöreskog and Sörbom 1984). The overall fit of both models was not adequate ( $<.9$ ), especially if expectations and perceptions are considered together. This clearly shows that the model can be improved. The results also indicated that perception and expectation had different reliabilities, and if tested in two different models would generate better overall fit. Due to the inadequate fit indices we ran a hierarchical cluster analysis using Ward's method with centroid weighting and standardised z-scores (e.g. Hair *et al.* 2009) in SPSS 17.0 to determine whether the observed variables clustered differently to the SERVQUAL scheme.

Table 6 shows the resulting clusters for perception and expectation, the rationale explaining their grouping, reliabilities, path loadings and model fit indices. Only three observed variables were not considered in the clusters, two of them related with visual appeal (updated equipment, professional appearance of materials and commitment). However the observed variables were clustered around different constructs, and there was a significant

difference between constructs explaining perceptions and expectations. The constructs related to expectations are similar to the contextualised SERVQUAL. Assurance did not appear as a group, empathy and visual appeal were akin to what we termed, respectively, relationship and aesthetics. Responsiveness was focused on to the ability to deal with change, and reliability was refocused into a professional relationship with the client, in other words the expectations clusters emphasised the ability to deal with change and the strength of the relationship with the client. However, the constructs related to perceptions were very different. The original set of SERVQUAL constructs were reclassified into three clusters: reliability, client centricity and understanding. The clusters suggest an increase in the levels of tangibility of the measures. We suggest that this is due to greater clarity around what and how the programme or project is being delivered.

We carried out the same set of tests as for contextualised and original SERVQUAL, and the results improved significantly. All AVE values were greater than 0.5, Cronbach's Alpha were above 0.8, with the exception of Aesthetics and Empathy. This indicated acceptable convergent validity. The new clusters also showed a significant improve in overall fit. Although the values of NFI are still below 0.9, the values are close to it, especially in comparison to the previous models (Table 6).

Table 6: New Clusters

Construct	Rationale	Reliabilities / Items	Path loading
<b>Expectations</b>			
<b>Professionalism</b>	"Providing the solution the client wants" $\alpha = 0.903 / AVE = 0.770$	Keep promises	0.87
		Interest solve client's problems	0.86
		Understand client	0.86
		Inform progress	0.82
<b>Resilience</b>	"The ability to deal with change" $\alpha = 0.825 / AVE = 0.742$	Back on track	0.81
		Change management	0.83
		Smooth execution	0.73
<b>Aesthetics</b>	"Looking professional" $\alpha = 0.773 / AVE = 0.642$	Professional appearance of people	0.61
		Professional appearance of physical facilities	0.90
<b>Relationship</b>	"Going the extra mile in the relationship with the client" $\alpha = 0.781 / AVE = 0.674$	Accessible	0.84
		Adapt styles	0.67
		Courteous	0.76
<b>Perception</b>			
<b>Reliability</b>	"Deliver what you are expected to deliver even with changes" $\alpha = 0.863 / AVE = 0.646$	Smooth execution	0.77
		Back on track	0.66
		Zero defect	0.75
		Confidence	0.79
		Safety	0.78
<b>Client Centricity</b>	"Delivery of solution on time and with minimum error throughout the process" $\alpha = 0.851 / AVE = 0.669$	Keep promises	0.82
		Interest solve client's problems	0.86
		Error free record	0.66
		Prompt service	0.75
<b>Understanding</b>	"Know and understand what the client wants" $\alpha = 0.801 / AVE = 0.693$	Adapt styles	0.83
		Focus client	0.53
		Have knowledge	0.73
<b>Model fit indices</b>			
		<b>Expectation</b>	<b>Perception</b>
	NFI=	0.880	0.879
	CFI=	0.940	0.954
	$\chi^2/df$ =	1.794	1.494
	p=	0.001	0.013

Finally, we ran a bootstrap comparison of the six models (Linhart and Zucchini, 1986; Arbuckle, 2006) to determine whether the new, clustered multi-item scales provide a better

fit to the data. Table7 shows the model fit measurements. The saturated models were tested against the new clusters, meaning that the observed variables in each model were the same.

*Table 7: Model Fit Measurements*

	<i>Failures</i>	<i>Mean Discrepancy</i>	<i>AIC</i>	<i>BCC</i>	<i>CAIC</i>	<i>RMSEA</i>
Original Expectations	14	429.86(2.68)	396.50	429.77	513.97	0.134
Original Perceptions	14	323.13(1.41)	267.13	291.13	418.61	0.088
Cont Expectations	10	749.76(3.65)	568.60	609.32	754.51	0.123
Cont Perceptions	27	629.01(2.33)	465.06	505.78	650.96	0.096
New Expectations	0	142.63(.95)	146.07	157.06	249.35	0.096
New Perceptions	0	136.93(.81)	154.173	168.455	223.13	0.077
New Expectations Saturated	-	-	156.000	184.563	424.527	-
New Perceptions Saturated	-	-	180.000	212.958	-	-
New Expectations Independence	-	-	740.970	745.365	782.282	0.343
New Perceptions Independence	-	-	678.454	687.243	-	0.319

As indicated in Table7, the new models fit the data better than the existing and contextualised SERVQUAL scales and the saturated and independence models. This is due to the mean discrepancy, AIC (Akaike Information Criteria), BCC (Browne-Cudeck) and CAIC (Consistent AIC) being lower for the new models. Root Mean Square Error of Approximation (RMSEA) values of the new cluster are between 0.05 and 0.10 indicating an adequate model fit (Browne and Cudeck, 1993). This improvement in AIC, BCC, CAIC and RMSEA indicates that our new constructs may be a more appropriate set of scales to use when measuring the expectations and perceptions of service quality within programme and project delivery.

## Conclusions

This study began with a practitioner problem – how to measure performance in operational transformation projects. This was not well understood and attempts using product-based measures yielded high performance measures, but low customer satisfaction. The service management literature provided some indication of why this was the case. The context is characterised by high levels of uniqueness, inherent conflict, long duration and a highly embedded and relational processes.

The study was therefore carried out to determine the attributes of quality in delivery of operational change. We used SERVQUAL as a point of departure and introduced new observed variables based on a qualitative study to contextualise the framework. The reliability and model fit measurement of both original and contextualised models indicated that this was an improvement. This solution was further developed by carrying out a cluster analysis with all the observed variables, and the resulting constructs had better values for reliability and model fit. Constructs for expectations and perceptions were clustered differently – a key finding. Expectations approximated to contextualised SERVQUAL, while constructs of perceptions were different and emphasised tangible aspects of process and outcome of the change projects.

This suggests that some of what is expected is not relevant in the perceptions of performance once the project is ‘in flight’. At the beginning of a project, uncertainty is higher, and expectations underline the “soft” aspects of relationship, such as accessibility and even courteousness. As projects become a daily “reality” for the transformed operation, tangible

aspects such as zero defects, error free records and having knowledge become more relevant. Thus, given that operational change is often delivered over an extended period, one explanation for the difference is the time delay between the setting of expectations and the forming of perceptions concerning the services. Another possible explanation is that such variables are only observable as the transformation progresses.

This change in performance assessment over time has implications for managers and researchers. For managers working in this context, a different approach to managing performance from that of other operations areas may be appropriate, changing the rationale and focusing on and the different requirements set out by the initial expectations and final perceptions of service delivery. The objective can still be to ‘mind the gap’ but with the knowledge that expectations and perceptions are formed on different sets of attributes as the work progresses. Moving earlier from the focus on short-term contractual requirements to assessment of the longer-term perceptual measures is more likely to generate satisfaction.

The nature of the performance assessment changed too. We initially attempted a separation of time, cost and quality performance, with the intention to focus on the third of these. However, and possibly inextricable in the context of service operations, aspects of time performance were included in the quality assessment.

For OM researchers, there is considerable potential for researching major projects, as levels of customer satisfaction are often low. Performance management is well developed for OM generally but less well so for major projects, and particularly operational transformation. This gap in the knowledge base is clear and the potential impact of such work could

be significant. Specifically, as a result of this work, we can say that managerial and organisational solutions to both recognise and work with the features of the context would be beneficial. We have shown that the time between designing performance management and its enactment, causes misalignment which benefits neither client nor provider.

‘How’s it going’ seems such a simple question, but as might be expected in the context of major projects, the answer is far less so.

## References

- Ahire S., Waller M. and Golhar D. (1996) "Quality Management in TQM versus non-TQM Firms: An Empirical Investigation" *International Journal of Quality and Reliability Management* 13 (8) 8-27.
- Arbuckle, J.L. (2007). *Amos 16.0 Users Guide*. Chicago: SPSS Inc.
- Robert I. Benjamin and Eliot Levinson (1993), "A Framework for Managing IT-Enabled Change", *Sloan Management Review*, Vol. 34, No. 4, pp. 23-33.
- Boaden R. (1997) "What is Total Quality Management ... and Does it Matter?" *Total Quality Management* 8 (4) 153-171.
- Browne, M. W., and R. Cudeck. 1993. Alternative ways of assessing model fit. In: *Testing structural equation models*, K. A. Bollen and J. S. Long, eds. Newbury Park, CA: Sage Publications, 136-162.
- Chase, R. and Apte, U. (2007) "A History of Research in Service Operations: What's the Big Idea?" *Journal of Operations Management*, 25 (2) 375-386.
- Dotchin J. and Oakland J. (1992) "Theories and Concepts in Total Quality Management" *Total Quality Management* 3 (2) 133-145.
- Fornell, C. and Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*. 18 (1) 39-50.
- Garvin D. (1992) "*Operations Strategy, Text and Cases*", Prentice Hall, Englewood Cliffs, NJ.
- Garvin D. (1984) "What Does Product Quality Really Mean?" *Sloan Management Review* (26) 25-43.
- Grönroos C. "A Service Quality Model and its Marketing Implications" *European Journal of Marketing* 18 (4) 36-44.
- Hair, J. F. Jr., Black, W. C., Babin, B. J. and Anderson, R. E. (2009) *Multivariate Data Analysis: A Global Perspective*. (7<sup>th</sup> ed.) Pearson.
- Harvey J. (1998) "Service Quality: A Tutorial" *Journal of Operations Management* (16) 583-597.
- Haywood-Farmer J. (1988) "A Conceptual Model of Service Quality" *International Journal of Operations and Production Management* 8 (6) 19-29.
- Jöreskog, K. G., and D. Sörbom. 1984. *LISREL-VI user's guide*. 3rd ed. Mooresville, IN: Scientific Software.
- Kang G. and James J. (2005) "Service Quality Dimensions: An Examination of Grönroos's Service Quality Model" *Managing Service Quality* 14 (4) 266-277.
- Lehtinen U. and Lehtinen J. (1991) "Two Approaches to Service Quality Dimensions" *The Services Industries Journal* 11 (3) 287-303.
- Linhart, H., and Zucchini, W. (1986). *Model selection*. New York: John Wiley and Sons.
- Maylor, H (2000) "Strategic Quality Management," in Moutinho, L. (2000), *Strategic Management in Tourism*, Wallingford: CABI Press.
- Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Parasuraman A., Zeithaml V. and Berry L. (1988) "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality" *Journal of Retailing* 64 (1) 12-40.
- Philip G. and Hazlett S. (1997) "The Measurement of Service Quality: a New P-CP Attributes Model" *International Journal of Quality and Reliability Management* 14 (3) 260-286.
- Powell T. (1995) "Total Quality Management as Competitive Advantage: A Review and Empirical Study" *Strategic Management Journal* (16) 15-37.
- Rust R. and Oliver R. (1994) "Service Quality: Insights and Managerial Implications from the Frontier" in Rust R. and Oliver T. (eds) *Service Quality: New Directions in Theory and Practice*, Sage Publications, Thousand Oaks, CA p1-19.
- Samson D. and Terziovski M. (1999) "The Relationship Between Total Quality Management Practices and Operational Performance." *Journal of Operations Management* 17 (1999) 393-409.
- Shewhart W. (1931) "Economic Control of Manufactured Products" Van Nostrand Reinhold, New York.
- Sousa R. and Voss C. (2002) "Quality Management Revisited: A Reflective Review and Agenda for Future Research" *Journal of Operations Management* (20) 91-109.
- Spencer B. (1994) "Models of Organisation and Total Quality Management: A Comparison and Critical Evaluation." *Academy of Management Review* 19 (3) 446-471.
- Walker, H., Chicksand, D., Radnor, Z. and Watson, G. (2015), Theoretical perspectives in operations management: an analysis of the literature, *International Journal of Operations and Production Management*, 35 (8), 1182-1206.