

Mind the Gap:

Towards Performance Measurement beyond a Plan-Execute Logic

Abstract

Project performance measurement aims to identify deviations from intended goals and reduce 'the gap' between actual and expected performance. However, despite extensive measurement and control efforts, the gap is hard to close and, intriguingly, not necessarily related to the project's perceived performance, which is what will ultimately influence stakeholder's satisfaction. Based on service quality research, this study explores the differences between perception and expectations of performance. Our mixed method study involving eighteen interviews and 85 survey responses in an IT-enabled change context shows that expectations and perceptions are fundamentally different concepts. As they are different, managing the gap between expectations and perceptions may be a nugatory task. The paper expands the literature on project performance measurement by questioning its foundations and offering a first step towards developing a more dynamic and subjective understanding of project performance that is consistent with a project's evolving nature.

Keywords: Project success, transformation projects, performance, SERVQUAL, project performance measurement, subjective performance.

Introduction

Despite all of our key performance indicators showing green, the client still isn't happy. (Interview data)

Every project starts with an intent to succeed, yet, its prospects of success tend to be grim (Flyvbjerg, 2016), not least because project success involves hitting a moving and sometimes ambiguous target (Ika and Pinto, 2022). The ultimate goal of project performance measurement during project execution is to assess where the project stands in relation to its planned objectives. As a result, the project management team may adjust the course of the project (Browning, 2019), to keep the project 'relevant' (Kreiner, 1995) and 'on course' and, so it is suggested, enhance the chances of successful project delivery (Pesämaa et al, 2018).

However, counterintuitively, the research on project performance measurement is not in sync with developments in the concept of project success. While the last decade has experienced a move towards more holistic measurements of project performance beyond the iron triangle (Toor and Ogunlana, 2010), most of the literature maintained its basal assumption of measuring performance against plan unchecked (e.g. Pesämaa et al., 2018; Denicol et al., 2020; Browning, 2019). This is a problem, because the conceptualization of performance measurement clashes with our current understanding of project success, as a dynamic and subjective concept (Ika and Pinto, 2022; Kreiner, 2014). If success is dynamic and subjective, then the use of objective performance measures, and measuring against the plan will not necessarily be a good guide to achieving higher levels of success.

As performance measurement exists to enhance project success, the concepts should have walked hand and hand, but they didn't. Accordingly, as the call for papers for this Special Issue highlights, we need to move from “...*overly simplistic measures of project success to a broader and more nuanced consideration of project performance*” (Pesämaa et al., 2020), which embeds dynamism and subjectivity.

Our study joins an emerging line of inquiry in project performance that takes the *dynamic and subjective lived experience* of core project stakeholders seriously and as part of project success (Kreiner, 2014, 2020; Ika and Pinto, 2022), and hence, as something that should be part of project performance measurement (Abdahllah et al., 2022). Specifically, service performance measurement establishes a customer's *subjective* perceptions of service delivery and therefore are, in principle, in line with our overall aim. We found some reference to service-based models in the PM and New Product Development literatures, but little by way of measurement instrument development. In contrast, the service operations literature provided several instruments, of which SERVQUAL is dominant. As such, the model is robust and had been tested across multiple empirical settings, so providing a good starting point for our study.

Moreover, current research in performance measurement area has mostly explored project contexts that have a strong focus on material outcomes, like large infrastructure (Abdahllah et al., 2022), engineering (Pesämaa et al., 2020) and construction (Kreiner, 2014) projects. Yet, subjective perceptions are particularly salient where projects have a more dominant service component (e.g., Sampson 2010; Parasuraman et al. 1985; Johnston 1999), that is intangible and co-created at the interface between customer and service provider (e.g., Sampson 2010; Parasuraman et al. 1985; Bitner et al. 1997). We therefore conducted our study

in a context that has a dominant service component, namely IT-enabled change projects. Our empirical context involved strategic changes being implemented in the operating IT infrastructure of a client, by an outsourced provider, over months or years.

Based on these reflections, we ask: what are the *subjective* performance attributes of a transformation project *during* its execution? The paper contributes to the literature on project performance measurement with a model that addresses subjectivity and dynamism of project context in a novel way, based on a service-oriented logic. Our approach questions the basal assumption of project performance measurement.

The paper is organized in four sections. The first section explores the literature on project performance measurement and establishes the need for a service-oriented logic. It then describes our theoretical foundations in service-oriented performance measurements in general and SERVQUAL in particular. The second section is dedicated to our qualitative study. We explain our methodology and describe core findings. The third is focused on our quantitative study, and represents the bulk of our work. The paper concludes with discussions and implications for research and practice.

Literature Review and conceptual development

The intriguing relationship between project performance and project success

Project performance measurement refers to performance assessments as the project

progresses through its life cycle, from concept to completion. In this paper, we are concerned with ‘in-process performance’, that is measures *during* the project. We distinguish this from ‘project success measurement’ as referring to any measures related to performance *after* project completion, following the logic of Ika and Pinto (2022).

Our rationale is that these two concepts are different yet intrinsically connected, as the measurements during the project exist to adjust the project to perform well at outturn. Therefore, performance and success measurements should walk hand in hand both in practice and in theory, yet we contend that they do not. Project performance measurement has been subjected to extensive research but has retained its basal assumption of predominantly measuring objective performance against a plan. For all but the shortest and simplest of projects, this is a poor conceptualization as crucially, it ignores both dynamism and subjective dimensions. These two dimensions are now explored in the context of the project performance measurement literature in the following section, and become the focus of our subsequent empirical work.

[Subjectivism: Nuancing satisfaction in project performance measurement](#)

Our conceptualization of subjectivism comes from current developments in the project performance literature itself. Project performance is traditionally concerned with conformance to the plan, mainly through Key Performance Indicators (KPIs) and related tools and concepts (Brown and Eisenhardt, 1995; Clark and Fujimoto, 1991). Conformance to plan is mostly goal-oriented (Kagioglou et al., 2001) and expressed in techniques including Earned Value Management (Fleming and Koppelman, 1996; Lipke et al., 2009; Anbari, 2003), combining both cost and schedule performance into indices for assessment. It is common in practice for projects to have dashboards with key performance indicators/indices (e.g. Yuan et al., 2009; Geraldi and Arlt, 2015), a balanced scorecard (Kaplan and Norton, 2004) or even a maieutic machine (Quattrone, 2017). Such visual developments are designed to draw managers' attention to potential deviations from a plan and therefore reinforce the concern with conformance to plan (Geraldi and Lechler, 2012; Stingl and Geraldi, In Review).

In the last decade, researchers have observed a shift towards other measurements beyond planned project attributes. For example, Hoegl and Gemuenden (2001) defined performance not only as task-related project outcomes (i.e. cost, time, quality) but also as people-related outcomes (i.e. work satisfaction, learning of new skills), which are desirable but not precisely planned. The trend was confirmed in Toor and Ogunlana (2010)'s systematic literature review, that suggests, "other performance indicators such as safety, efficient use of resources, effectiveness, satisfaction of stakeholders, and reduced conflicts and disputes are increasingly becoming important." (p. 228). Most dominantly was the introduction of stakeholder satisfaction, e.g. user satisfaction (Nguyen and Ogunlana, 2004; Bryde and Brown, 2004) and customer satisfaction (Shenhar et al., 2001; Chan and Chan, 2004).

The developments slowly but surely opened a line of inquiry into more subjective performance measurements. A notable study in this regard is Pesämaa et al. (2018) who connect performance explicitly with learning. Unlike most work on performance measurement, their work was not interested in traditional project-related measures, but rather the role of feedback given to project employees on project process performance. They argued that *"different actors have different priorities, so feedback loops are required to establish processes (and outcomes) that all the key actors can find meaningful (Lin and Gerber 2014). As noted by Chang and Chiu (2005), poor process performance in projects may relate to the lack of joint understanding of these processes "* (Pesämaa et al., 2018, p. 3). Based on an impressive sample of 2,175 respondents, the authors demonstrated the relevance of performance feedback to a smooth project process, and for better collaboration and learning in projects. Yet, the definition

of in-process performance was still embedded in an “efficient delivery against established outcomes.”

Another relevant study in this regard is Abdallah et al. (2022). The authors took the potential misalignments of stakeholders interests and perceptions about the project to the core of their study and explored project performance measurement as patterns of convergence in project representations and stakeholder volition. Their conceptual model and qualitative study on megaprojects provide a solid foundation for the understanding of such patterns, and has inspired our study both in terms of dynamism and subjectivity.

Yet, despite developments in the project performance literature, we contend that the actual measurement of ‘stakeholder satisfaction’ is still problematic (e.g., Gatian, 1994), not least because the client is not a passive recipient of the service being provided. Indeed, their own actions and interactions with the project and the service provider will have an impact on the achievement or otherwise of project success (e.g., Bresnen and Marshall, 2001). Our challenge was to identify an instrument that could unpack the concept of ‘satisfaction’ while acknowledging the subjective lived experience of the client/customer as an agent in the transformation. We found such an alternative approach in the service quality literature, which we will introduce later in this section.

[Dynamism: Towards a loosely-coupled view on project performance measurement](#)

Our reflections on dynamism in project performance measurements arise from the engagement with the concept of project success. There is ample evidence that the assessment of the success of a project changes over time. Such a change is commonly expressed in the dichotomy between project output (focused on the short-term delivery measured against its initial plans) and project outcome (focused on the long-term outcomes and benefits generated through the

project) (e.g. Pinto and Slevin, 1988; Cooke-Davies, 2002; Munns and Bjeirmi, 1996; Jugdev and Muller, 2005). The challenge is that project output and project outcome are not necessarily related, and in some cases, might even be contradicting (Ika and Pinto, 2022). Consequently, project output and project outcome must be two *independent concepts*, that are *not causally related*.

Could we say the same about project performance and project success? Are project performance and success independent concepts that are not causally related? The classic example of the Sydney Opera House suggests we can. As detailed in Murray (2003), its project management was disastrous, leading to escalating costs, duration and conflicts, which led to the firing of the leading architect, Jørn Utzon. This means that its performance, as the project progressed, was disastrous. However, the project (outcome) is undoubtedly a success - the building is an architectural masterpiece, that attracts 10.9 million visitors a year and became the symbol of a continent. Utzon was finally recognized for his work with the Pritzker Prize in 2003, 30 years after the Opera House was inaugurated. (Murray, 2003). The building is cultural and iconic, with its value is estimated at 6.2 billion Australian dollars, dwarfing the magnitude of any cost overruns in the original project

(<https://www.sydneyoperahouse.com/general/corporate-information.html>). We, therefore, expect that project performance and project success are independent concepts.

Yet, uncoupling the concepts would suggest that project management is inconsequential, that is, how projects are being managed would be irrelevant to achieving success (Kreiner, 2020), which cannot be true as incompetence has an impact in projects, as, for example, the Berlin Brandenburg Airport illustrates (Geraldi and Stingl, 2016). Moreover, this

line of thinking would also render project performance measurement irrelevant, which is unlikely and unhelpful, as there is ample evidence that performance measurements influence behaviour (Maylor et al, 2015).

An alternative approach is to reframe the causal relationship with a loosely coupled one. As such, performance measurements can indicate success, but neither guarantees nor causes it. Accordingly, it is part of the management task to keep the concepts connected so they can evolve together. Such a process or temporal approach is '*through time*' (Hernes, 2022). That is, as the project progresses, project stakeholders work to understand how the project is evolving including through performance measurement. At each moment in time, performance measurement can guide them to judge their current perceptions of the project, compare it with their expectations and ambitions for the project. They then adjust not only the project but also the expectations for project success accordingly. Note here that we compare the lived experience with the expectations at each moment in time, not the plan. Hence, such an approach connects the past, present and future *in the moment*, and thereby encourages project stakeholders to 'mind the gap' between expectations and perceptions, connecting temporalities *through time*.

Such a view on project performance measurements accommodates the dynamics discussed in the project success literature and provides a novel approach to bridging its temporal disconnects between past hopes, future ambitions and present experience, while not falling back to the problematic 'plan vs. actual' mantra that dominates performance measurement in practice.

Our challenge then becomes to identify a performance measurement instrument that can accommodate such an approach, and we found it in a contextualized version of SERVQUAL, which we will explain later in the next section.

[Service quality: an approach to capture dynamic and subjective project performance](#)

We attempted to enrich current performance measurements with subjectivity and dynamism by evoking ideas and measurements from service quality. Service quality attempts to capture the 'lived experience' of those receiving services, in our case, those in the organisations being transformed. As such, they are seen as the customers of the transformation. We therefore frame the organizational transformation program as service.

There are three generalized and well-documented differences between the production of goods and services: intangibility (services are often performances rather than objects, putting hence more emphasis on subjectivity), heterogeneity (performance varies and is reliant upon consistent behaviour from personnel, which again highlights subjective characteristic of performance in projects) and inseparability (production and consumption are often simultaneous, which places the clients as co-producer of the transformation, which is the case in our context) (Parasuraman et al., 1985). The differences push the service literature to take a markedly different approach to performance assessment.

First, whilst products can 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. The idea is that the difference between perception and expectation forms customer's subjective impression of the service performance. Harvey (1998:98) notes that when delivery is a key part of the service, then "*...perception is reality.*" As such, if what Grönroos (1984) terms the 'expressive performance' (what someone perceives) is

unsatisfactory, then the customer or client will still feel unsatisfied, irrespective of the 'instrumental' or 'objective' performance. Further, satisfaction is unlikely to be a binary and stable concept. As van Dyke et al (1999) recognised, satisfaction is more granular than this, with stakeholders prepared to accept lower than expected performance in some areas before being dissatisfied. They advocated for a conceptualization of 'satisfaction' which did not rely on direct matching of expected and perceived performance, but with 'zones of tolerance' (also Kettinger and Lee, 2005).

Service quality then measures the 'in the moment' subjective assessment of what success would look like and what is currently being experienced, without assuming that the image of success would be constant over time. In this regard, the assessment acts as an artefact that connects present and future, and thereby assists project actors *through time*. Therefore, service quality fits our conceptual development well.

Second, more than asking for "customer satisfaction", as it is often the case in project performance assessments (e.g. Nguyen and Ogunlana, 2004; Bryde and Brown, 2004; Shenhar et al., 2001; Chan and Chan, 2004), service quality models go a long way in attempting to operationalize perception and expectations based on the customer's or client's view of a "bundle" of service dimensions. Different models provide alternative service dimensions. For instance, Lehtinen and Lehtinen (1991) consider the categories to be: Physical Quality, Interactive Quality, Corporate Quality, Process Quality and Output Quality. Rust et al (1994) consider Service Product (technical quality of the outcome), Service Delivery (functional or process quality) and Service Environment. Philip and Hazlett (1997) classify criteria as being Pivotal, Core or Peripheral Attributes. Haywood-Farmer (1988) considers that Professional

Judgement, Physical Facilities and Processes, and People's Behaviour are the criteria for performance assessment.

From the different options, we opted for SERVQUAL (based on Parasuraman, Zeithaml, & Berry, 1985). The original model includes 22 attributes bundled around five dimensions:

- **Tangibles:** Physical facilities, equipment, and appearance of personnel.
- **Reliability:** Ability to perform the promised service dependably and accurately.
- **Responsiveness:** Willingness to help customers and provide prompt service.
- **Assurance:** Knowledge and courtesy of employees and their ability to inspire trust and confidence.
- **Empathy:** Caring, individualized attention the firm provides its customers.

SERVQUAL was chosen because it is a widely accepted instrument for evaluating service performance, that is robust and has been applied for different contexts, including IS (e.g. Kettinger and Lee, 1994; Pitt et al., 1997) and construction contexts (Winch et al, 1998).

The SERVQUAL instrument has been used to assess the quality of service being provided by an engineering design studio (Love et al, 2000), construction companies (Hoxley, 2000) and NPD (Allan Burström et al 2014). For example, Wikström et al (2009) focused on the business logic differences between a product and service-based project firm. Winch et al (1998) explored five gaps in understanding of quality. Moreover, while SERVQUAL has not yet been applied to IT-enabled organizational change, it has been applied to IS contexts. One of the studies suggests that SERVQUAL does not clash with IS contexts, nor could they *"discern any unique features from the IS domain that have been excluded from SERVQUAL."* (Pitt et al 1997, p. 212). This appears to justify SERVQUAL as an appropriate provisional framework and that it is useful for

our context. It is an accepted starting point to develop measurements for stakeholders' subjective experience of project performance.

Yet, the measurements needed to be adapted to address the particular features of the context of IT-enabled change. Hence, we conducted a mixed-method study. The first phase was a qualitative study with eighteen interviews, that helped us characterize the context and adapt service measurement instrument accordingly. In the second phase, we tested the model quantitatively based on a survey with 88 respondents from the same IT-enabled change context as the interviewees.

An immediate contextual difference occurs in the consideration of time. The SERVQUAL model was developed for services delivered over a relatively short period of time: minutes, hours or days. Accordingly, expectation and perception were not only measured at the same time but were also formed in close proximity in time. In our context, the service is delivered over months and years, yet we kept the simultaneous measurements of expectation (what would success look like) and perception (what is been experienced) 'in the present moment' rather than reverting to pre-project baseline expectations. At the outset, it was not clear what would be the effect of differences in time horizons on expectations, and how these could influence perceptions. We therefore empirically tested the validity of 'the gap'. *Mind the gap*, a phrase much heard by passengers on London's metro system, is our title, because our results suggested that expectations were not just different from perceptions, they were based on fundamentally different concepts. This finding reinforced our approach of seeing success and performance as independent, loosely coupled concepts changing *through time*.

Methodology

The empirical research presented in this paper was undertaken using a two-stage process. Stage 1 used a qualitative approach to explore the performance dimensions and attributes in our context. Stage 2 used a quantitative approach, intending to uncover the underlying relationships between the parameters.

Stage 1 Interview study - Tailoring SERVQUAL to context

This stage was designed to contextualise SERVQUAL for the assessment of the performance of IT-enabled transformation projects during their execution. We carried out a study comprising eighteen exploratory interviews with project and programme managers, and senior executives within a large IT provider specialising in business process transformation. The sample was purposive and was limited to this one sector.

The sample suited our purpose here because (1) all interviewees had significant experience with transformation projects, and (2) they were both providers and recipients of products and services. They provided products and services as part of the transformation to their customers, but themselves had a delivery supply chain for parts of those products and services (e.g. hardware, interfaces, hosting, software modules, training and administration services), for which they provided an integration function.

All interviews were conducted in English; took 30-90 minutes; and were recorded and transcribed. In addition, where commercial considerations allowed, examples of performance measures were collected. 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 and challenges related to measuring project performance and the nature of the delivery environment.

- **Defining performance** delved into the interviewees' understanding of performance and the respective attributes and measures by comparing projects they considered high performing and low performing. This was the main part of the interview. In this line of questioning, we deliberately distinguished between formal performance measures and how interviewees assessed performance and believed their performance was assessed. This meant that interviewees were not constrained by the limits of the formalized performance assessment system, but could take into account insights from the experience of working closely with clients / customers with whom they were embedded.

We did not mention any *a priori* constructs or attributes in the interviews to ensure that our own biases did not lead the interviewees. Once recorded and transcribed, the analysis was carried out using a simple manual coding method as used in Grounded Theory (e.g., Glaser and Strauss, 2017). We grouped the extracted codes by concepts, and concepts grouped into categories.

When we explored with participants how the (largely unmeasured, at least formally) service delivery process was assessed, the concepts that emerged from the coding process were consistent with the 22 measures in 5 categories from the original SERVQUAL framework but with some additional concepts. These five additional concepts fitted well within the existing categories (see Table 1).

Table 1: 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 included 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’ 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. The ability to move from this disturbance to a restored sense of purpose with the work being carried out, was important. These modified dimensions and attributes were used for the quantitative study described below.

Stage 2: Survey study – Quantitative assessment of the contextualized SERVQUAL

Following a well-described scale development procedures (mirroring Parasuraman *et al.* 1986)

and our modified list of dimensions and attributes, we then produced an instrument comprising 2 banks of questions, ‘expectations’ and ‘perceptions’, using 7-point Likert scales. Each bank contained the full set of questions (SERVQUAL plus the five additional criteria), which allowed comparison between the contextualised model and the original (by excluding the new five from the analysis). We determined to explore whether there was validity in a measure of service performance that could supplement the current instrumental measures. There was already evidence from the contextual interviews, and from contrasting with the context in which SERVQUAL had been originally developed, that the approach to assessment might need further amending to account for a project’s dynamism and subjectivity.

To deal with this, we asked respondents to consider expectations in terms of performance dimensions of what they would consider to be a high-performing project at this moment in time, across all 27 of the contextualized SERVQUAL dimensions. For instance such a high-performing project might *“ensure the availability of appropriate processes and tools to meet the contract requirements.”* If these are not immediately available, they will make them available by improving, developing or acquiring adequate processes and tools. Immediately after asking for their expectations, we asked for their perception: “In your current work, [project leaders] are making the appropriate processes and tools available.” It is this ‘in-the-moment’ assessment that we are using here, with the intention to better understand both perceptions and expectations (Harvey, 1988), in line with our research objectives. Moreover, the in-the-moment perception and expectation help us unlock dynamism in the project *in time*. Longitudinal studies would also

be valuable to connect 'in-the-moment' assessments and thereby capture patterns through time (Hernes, 2022). We recommend such a study in our areas for further research.

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, readability and to determine the likely time requirement for participants.

The instrument was then used to test our enhanced list of service attributes of performance in a cluster sample of the customers of 40 business units based in several countries in Europe with the same IT outsourced service provider. Each of the business units operated as an independent financial entity, but within a common owning firm, and with many operational similarities and the potential to draw on common resources. 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 of the customers ($n=200$) 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 2 details the geography, role and experience of the respondents.

Table 2: Sample Frame

Country		Experience in current role (years)	
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
Current Role			
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

Findings

We started the analysis with a qualitative description of IT-enabled change context and their existing project performance measurement instruments. We noted, as expected, that the context was particularly marked by dynamism and subjectivity, and its projects and programs had a salient service component. We then move to the quantitative data, which 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, as the model fit for expectations and perceptions was seen to be different. Finally, we constructed and tested the new model using confirmatory factor analysis (CFA) to determine whether the new constructs fit the data better than the existing SERVQUAL constructs. We found that the new model was indeed better at representing variances in our sample than both original

and contextualized SERVQUAL and required a reduced set of attributes (12 attributes for each for expectation and perception rather than 27 were required).

Stage 1 - characterizing the context: Dynamic and subjective nature of IT-enabled change performance was not considered in performance measurements in place at the time

Our interviews confirmed our assumption that IT-enabled transformation have a highly salient service component. We observed a high degree of dependency and interaction between client and provider, and high levels of relationship between staff from different organisations in the supply chain. At the same time, there was an inherent conflict in purpose for the two organisations. Respondents noted that the organisation being transformed frequently has the objective 'to keep the shop open' (i.e., sustain operations during the change), while the transforming organisation exists solely 'to execute the change';

The transformations were organised as programmes or projects, lasting months or years, and included the provision of a range of products (both hardware and software), systems design and architecture, consultancy, target operating model design, installation and training. The scope had both products and services, yet they were orchestrated to bring about change, and such changes could only be assessed long after the project was completed. As such, the subjective perceptions of the actual progress of the work were important to guide the work, as well as to create and maintain confidence in the change project. As would be expected, there was a high level of uncertainty at the outset as to the precise requirements. As a result, the supplier was typically embedded in the client organisation and frequently on-site, engaged in co-creating the end solution (the IT-supported target operating model). This also meant that the initial plans were not representatives of 'in the moment' ambitions, and expectations would 'mature' as the

project progressed. Thus, the context was indeed characterized by salient dynamism and subjectivity.

In contrast to the context, the classic 'conformance to plan' performance assessments were being used but were not functioning to the needs of either client or supplier. Indeed, many of the interviewees reported problems with the approaches to measuring the performance of the transformation project. Typically, it was assessed through pre-agreed Key Performance Indicators (KPIs). The KPIs were product-focused (assessed against a specification for the product), progress-related (achieving milestones for delivery as set out in the plans) and resource-related (the number of suitably qualified and experienced people that the contractor had allocated to the project). All of these are 'instrumental' in Grönroos' terms. It was common practice that the only 'service' measure was a 'Service Level Agreement' concerning how quickly system problems would receive attention and be rectified.

Despite measuring performance in such detail, client satisfaction was proving elusive. It was commonly found that, as one manager described: "Despite all of our KPIs being green, the customer still isn't happy." This echoes Grönroos (1984)'s findings that instrumental measurements focused on what is delivered, such as the KPIs, are not enough to express a client's (dis)satisfaction (also Kang and James, 2004). We speculated that this may reflect the belief on the part of those who set up the measures that transformation projects are hard systems, but this is inconsistent with the lived experience of our interviewees. Suppliers and the project teams were involved in transforming how people worked, and all appeared to understand very well that this required more than just changing the IT. However, the performance assessments then were geared towards technical delivery, rather than achieving the desired organisational change.

Our reflection on this performance measurement challenge concluded that these instrumental measures may be '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. One programme had 142 KPIs.

Further, interviewees noted that formal performance assessment had little credibility with either the client or the provider. This state was maintained except where there was a breakdown of relationships, and each side started scrutinising the terms of their contracts. Many noted that expectations were indeed highly dynamic. For instance, where a major problem had been encountered, expectations were reduced. Similarly, early successes in a project had raised expectations. Neither of these is particularly surprising but does provide an indication that 'expectations' as a measure is indeed unstable over the timescales of months and years.

Thus, an alternative project performance measurement was required, one that would break away from conformance to plan and embrace the dynamism and subjectivity of IT-enabled transformations.

Stage 2 - testing the original and contextualized SERVQUAL

Table 3 shows the results for the reliability and validity tests and the measurement model.

Table 3: Data internal consistency: Cronbach's Alpha, AVE and Path Loading, for original and contextualised SERVQUAL instruments

	Expectations		Perceptions	
	Original	Context	Original	Context
Assurance	$\alpha = 0.811$ / 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
	$\alpha = 0.775$	$\alpha = 0.652$	$\alpha = 0.695$	$\alpha = 0.790$
Empathy	AVE= 0.590	AVE= 0.584	AVE= 0.600	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
Tangibles	$\alpha = 0.806$ / 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
	$\alpha = 0.863$	$\alpha = 0.845$	$\alpha = 0.827$	$\alpha = 0.860$
Reliability	AVE= 0.628	AVE= 0.760	AVE= 0.736	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
	$\alpha = 0.703$	$\alpha = 0.830$	$\alpha = 0.612$	$\alpha = 0.782$
Responsiveness	AVE= 0.622	AVE= 0.597	AVE= 0.545	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
χ^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 0.8, which are considered adequate and good for exploratory purposes (Nunnally, 1978), although some values fall below this limit which indicates a lack of internal consistency within the constructs. The contextualised SERVQUAL constructs overall had somewhat better results than the original SERVQUAL constructs, with only one value (empathy expectations) under 0.7, and an average of 0.795.

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, 2009). We then conducted a CFA using AMOS 17 of the original SERVQUAL and the contextualized 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, 2006). The overall fit of both models was not adequate ($<.9$), 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 a better overall fit. This in itself is interesting: it showed that expectations and perceptions, even 'in the moment,' are different in structure, in contrast to the assumptions underlying the original Parasuraman et al study.

[Reconsidering expectations and performance: Cluster analysis](#)

Due to the inadequate fit indices and to explore these differences in structure, we ran a hierarchical cluster analysis using Ward's method with centroid weighting and standardised z-

scores (e.g. Hair, 2009) to determine whether the observed variables clustered differently to the SERVQUAL scheme.

Table 4 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 to tangibles (updated equipment, professional appearance of materials and commitment). However, the observed variables were clustered around different constructs. Thus, items in expectations were not seen in perception and vice versa. Therefore, the new clustering has reduced the number of items in the instrument to just 12 for both expectations and perceptions. This is most helpful as we found the contextualised SERVQUAL to be unwieldy in application.

We carried out the same set of tests as for contextualised and original SERVQUAL, and the results improved significantly for the contextualised model. All AVE values were greater than 0.5, Cronbach's Alpha were above 0.8, with the exception of Visual Appeal and Personal Empathy. This indicated acceptable convergent validity. The new clusters also showed a significant improvement 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 4: New Clusters and proposed instrument for assessing service performance in IT-enabled transformation projects

Contextualized SERVQUAL dimension	Proposed model of Expectations	Path loading	Proposed model of Perceptions	Path loading
Empathy: Providing caring, individualized attention to the customers	Personal empathy: Going the extra mile in the relationship with the client ($\alpha=0.781$ / AVE= 0.674)		Professional empathy: Knowing and understanding what the client wants ($\alpha=0.801$ / AVE= 0.693)	
<i>Adapt styles</i>	<i>Adapt styles</i>	0.67	<i>Adapt styles</i>	0.83
<i>Accessible</i>	<i>Accessible</i>	0.84		
<i>Focus on client's interest</i>			<i>Focus on client's interest</i>	0.53
<i>Understand needs of clients</i>	<i>Courteous</i>	0.76	<i>Having necessary knowledge</i>	0.73
Reliability: Performing promised service dependably and accurately	Problem-orientation: Solving problems for and with the client ($\alpha=0.903$ / AVE= 0.770)		Problem-solving: Solving problems reliably ($\alpha=0.851$ / AVE= 0.669)	
<i>Keep promises on time</i>	<i>Keep promises on time</i>	0.87	<i>Keep promises on time</i>	0.82
<i>Interest in solving client's problems</i>	<i>Interest in solving client's problems</i>	0.86	<i>Interest in solving client's problems</i>	0.86
<i>Error-free records</i>			<i>Error free records</i>	0.66
<i>Zero Defect</i>	<i>Understand client's needs</i>	0.86	<i>Prompt services</i>	0.75
<i>Smooth execution</i>	<i>Inform progress</i>	0.82		
Responsiveness: Helping customers and providing prompt service.	Resilience: Dealing with change ($\alpha=0.825$ / AVE= 0.742)		Dependability: Delivering reliably along with changes ($\alpha=0.863$ / AVE= 0.646)	
	<i>Smooth execution</i>	0.73	<i>Smooth execution</i>	0.77
<i>Back on track</i>	<i>Back on track</i>	0.81	<i>Back on track</i>	0.66
<i>Responsive to change requests</i>	<i>Responsive to change requests</i>	0.83		
<i>Inform progress</i>			<i>Zero defect</i>	0.75
<i>Prompt service</i>			<i>Confidence</i>	0.79
<i>Commitment</i>			<i>Safety</i>	0.78
Tangibles: Offering physical facilities, equipment, and appearance of personnel.	Visual appeal: Looking professional ($\alpha=0.773$ / AVE= 0.642)			
<i>Professional appearance of people</i>	<i>Professional appearance of people</i>	0.61		
<i>Appearance of physical facilities</i>	<i>Appearance of physical facilities</i>	0.90		
<i>Updated equipment</i>				
<i>Professional appearance of materials</i>				
Model fit indices	NFI=	0.880		0.879
	CFI=	0.940		0.954
	χ^2/df =	1.794		1.494
	p =	0.001		0.013

Finally, we ran a bootstrap comparison (Linhart and Zucchini, 1986; Arbuckle, 2006) of the six models: perception and expectation of each of the three instruments - SERVQUAL original, contextualized SERVQUAL and the new model. This final test determines whether the new, clustered multi-item scales provide a better fit to the data. Table 5 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 5: Model Fit Measurements

Model		<i>Failures</i>	<i>Mean Discrepancy</i>	<i>AIC</i>	<i>BCC</i>	<i>CAIC</i>	<i>RMSEA</i>
Models of interest	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
Bootstrap comparisons	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 Table 5, the new models fit the data better for both expectations and perceptions than the existing and contextualised SERVQUAL scales and the saturated and independence models. This can be seen by the mean discrepancy, AIC (Akaike Information Criteria), BCC (Browne-Cudeck) and CAIC (Consistent AIC) being lower for the new expectations and perception 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, 1992). 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 performance within project delivery.

Having established that the new clusters score higher in model fit indices and require less variables, we turn to understanding the qualitative differences between the constructs. To display the differences at a variable level, Table 4 also contains the items of the contextualized SERVQUAL and items shared between SERVQUAL and the new clusters are marked in italics. We note two significant differences.

First, not all SERVQUAL constructs were clustered together. ‘Assurance’ did not appear as a group. ‘Tangibles’, which referred to material aspects of the service like the appearance of the physical facilities was only relevant to explain variations in expectations, but did not appear as a construct of perception, strengthening the intangible nature of IT enabled transformation context. When they did appear, only one to three variables of each contextualized SERVQUAL construct were shared across the new clusters.

Second, and most surprisingly, there was a significant difference between constructs explaining perceptions and expectations. Performance was formed of four constructs while expectations was formed of three. Perception and expectations shared only one or two items in each construct. As the constructs were so different, we attempted to extract each’s rationale inductively from the items associated to each construct and renamed them accordingly.

The SERVQUAL ‘empathy’ originally focused on caring and individualized attention to customers. The new cluster for expectations followed a similar ethos, with a strong focus on personal empathy, adapting styles, and being accessible and courteous. Perceptions, in contrast,

had a stronger focus on what we termed ‘professional empathy’, which was more about knowing and understanding what the customers actually want.

The SERVQUAL original ‘reliability’ moved from a focus on the delivery of the promised service dependably and accurately to a focus on problems. The new clusters showed a stronger understanding of the dynamism of projects, expressed in the items ‘inform progress’ and ‘prompt services’ in expectations and perceptions, respectively. While expectations were more focused on adjustments in relation to customer’s needs and with the customer, expectations had a stronger focus on quick and error-free delivery. We have therefore termed them ‘problem-orientation’ and ‘problem-solving’, respectively.

Responsiveness maintained a focus on dynamism, but also on the ‘smooth execution’ under turbulence, particularly. While expectations had a stronger focus on resilience, perceptions focuses on dependability, and included items such as ‘safety’ and ‘zero defect.’ Finally, tangibles were not relevant to explain differences in perceptions, and in expectations, they were mostly confined to professional appearance of people and buildings.

Thus, the qualitative analysis of the differences between the new clusters in performance and expectations point to a divide between idealistic expectations of working collaboratively (e.g. ‘inform progress’, ‘understand client’s needs’) and developing personal relationships (e.g. ‘being courteous’) surrendered to a perception focused on getting problems sorted (e.g. ‘having necessary knowledge’, ‘prompt services’) and faultless delivery (e.g. ‘error free’, ‘zero defect’). Even the expected ‘visual appeal’ did not materialize as relevant in the perception measures. The indicative differences also suggest a tendency to surrender any idealistic view of collaborative

processes to a wish for project managers to simply understand what is needed and get it done, reliably.

Discussion and conclusions

In this paper, we start with a paradox: project performance measurement exists to improve project success, yet the concepts of performance and success have not developed hand in hand, so that performance measurement is focused on an inadequate understanding of project success. We set out to explore an alternative view of performance measurement, that is more in line with the current understanding of project success. We, therefore, venture into the development of a more dynamic and subjective understanding of performance. Our research question became, what are the *subjective* performance attributes of a transformation project *during* its execution?

Our journey brought us to the context of IT-enabled transformations, where projects are often co-created with the customers, and are more dynamic and more intangible than projects such as engineering and construction. We also bring the literature on service quality to assist us in the development of an understanding of performance as a subjective lived experience, that takes place “in the moment” of service delivery. Our findings, based on a combination of eighteen qualitative interviews and a survey with 85 respondents, quantitatively testing our performance measurement model.

Our interviews explored performance measurement in the context of the delivery of IT-enabled transformation projects. The tools currently in use demonstrated the prevalence of Grönroos’ instrumental performance assessment (specifically progress according to pre-defined

plans) and limited use of non-instrumental or expressive assessment. Given the nature of the transformation projects, this was surprising to us. For the managers in our study, this led to an ongoing source of dissonance; despite instrumental performance being good, customers were not satisfied. Service-based business logic has long been established in repetitive operations, and an instrument for the measurement of its performance, SERVQUAL, provided a starting point for assessing performance.

Our qualitative study yielded a list of concepts for assessing expressive aspects of the project, supplementing SERVQUAL to produce a 'contextualised model'. It did also raise questions about the use of the traditional gap model for performance assessment. Specifically, both the passing of time and a customer's experience of expressive aspects did affect their expectations. Noting that expectation was itself not an objective function, we progressed with assessing expectations beyond the notion of 'what you originally expected' to considering perceptions 'in the moment.' We concluded with a proposed 27-item instrument for performance assessment.

The analysis of quantitative data collected using original and modified SERVQUAL instruments, yielded a revised instrument (Table 4) which demonstrated improved validity over the original SERVQUAL measures. We saw that expectations and perceptions were different in structure and that the performance assessment could be reduced to 12-item scales for both perception and expectation assessment. Further, this validated the use of 'in the moment' assessment of expectations, rather than a reliance on historic expectations.

Overall, we derive two implications about the differences between original, contextualized and new clusters for measurements of project performance. First, we extend

current literature on performance measurements with a model that integrates subjectivity and dynamism in project performance. When compared to the original and contextualized SERVQUAL, the new clusters shared a focus on variables related to *dynamism*, such as ‘prompt service’, ‘adapt styles’ and ‘responsiveness to change’. Moreover, the new clusters emphasised *subjective* aspects related to the relationship with the client over e.g. tangible aspects of the service delivery. This was visible in items such as ‘an interest in solving client’s problems’ and ‘understanding of client’s needs’, which are indeed subjective and dependent on the relationship between the client and service provider. In other words, the new clusters relied more explicitly on variables measuring dynamism and subjectivity to explain differences in how project performance was perceived in the moment, and how expectations of performance were formed. Our data, therefore, reinforce our initial observations, that performance measurements gain from a stronger relationship with *conceptual* developments of project success (Ika and Pinto, 2022), as performance was indeed enriched by subjectivity and dynamism.

In the same vein, the paper contributes to the literature on project performance measurement with a model that provides more nuanced measurements of project performance, such as stakeholder satisfaction, than has been established as a relevant performance measure in the last decade (Hoegl and Gemuenden, 2001; Toor and Ogunlana, 2010; Nguyen and Ogunlana, 2004; Bryde and Brown, 2004; Shenhar et al., 2001; Chan and Chan, 2004), but have not been precise enough to break down such performance into specific constructs and variables.

Moreover, the developments of a measurement 'in-the-moment' and the focus on items related to change reinforce the dynamic nature of the projects, and the need to consider performance as an ongoing pursuit (Kreiner, 2020). Future research could explore how performance in the moment changes in the course of the project, longitudinally. As we observed changes between perceptions and expectations, can one also observe changes in perception at each phase of the project? Following the line of reasoning of Abdallah et al (2022), are there any patterns? For example, how do project managers work and adapt to a dynamic set of expectations, for instance where improving performance increases future expectations? What is the nature of any zone of tolerance (van Dyke et al 1999; Kettinger and Lee, 2005) and how does this change with time?

Second, despite this overall shared focus on dynamism and subjectivity, perceptions and expectations have different constructs and variables. The difference is important as it reinforces our initial challenge that the basal assumption of performance measurements needed to be reconsidered, not only because environments drift (Kreiner, 1995) and expectations for future performance mature (Pitsis et al, 2003), but much more fundamentally, the way we frame what would be expected is inherently different from the way we actually experience the service. Therefore, determining project success by focusing on the gap between what was the intention and what is the actual performance is inherently wrong, as these are different concepts, measured by different constructs, at least when it concerns the service dimension of projects. We need further reflections to understand whether the enduring notion of assessment expressed as a gap between desired and achieved performance, is indeed still appropriate for complex projects, or whether an alternative conceptualisation of performance is now required.

Our findings therefore contribute to the call to advance our understanding of project performance measurements by proposing a novel approach for measuring performance in-the-moment, and thereby, freeing project performance measurements from the shadow of plan-execute models. This is a radical step, and would require more significant empirical research to confirm and dwell into the differences between perception and expectations, both quantitatively and qualitatively, and crucially, over time.

References

- Abdallah, S. B., El-Boukri, S., Floricel, S., Hudon, P. A., Brunet, M., Petit, M. C., & Aubry, M. (2022). A process-oriented framework to measure development performance and success of megaprojects. *International Journal of Project Management*, 40(6), 685-702.
- Allan Burström, T., Jacobsson, T., M., & L. Wilson, T. (2013). Integrating service practice into project management: a matter of “do or die”? *International Journal of Managing Projects in Business*, 7(1), 5-22.
- Anbari, F.T., 2003. Earned value project management method and extensions. *Project Management Journal*, 34(4), pp.12-23.
- Arbuckle, J. L. (2006). 17.0 user's guide. In Crawfordville, FL. Amos Development Corporation.
- Bitner, M. J., Faranda, W. T., Hubbert, A. R., & Zeithaml, V. A. (1997). Customer contributions and roles in service delivery. *International journal of service industry management*.
- Bresnen, M., & Marshall, N. (2001). Understanding the diffusion and application of new management ideas in construction. *Engineering, construction and architectural management*.
- Brown, S. L., & Eisenhardt, K. M. (1995). Product development: Past research, present findings, and future directions. *Academy of management review*, 20(2), 343-378.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological methods & research*, 21(2), 230-258.
- Browning, T. R. (2019). Planning, tracking, and reducing a complex project's value at risk. *Project Management Journal*, 50(1), 71-85.

- Bryde, D. J., & Brown, D. (2004). The influence of a project performance measurement system on the success of a contract for maintaining motorways and trunk roads. *Project Management Journal*, 35(4), 57-65.
- Chan, A. P., & Chan, A. P. (2004). Key performance indicators for measuring construction success. *Benchmarking: an international journal*.
- Clark, K.B. & Fujimoto, T. (1991). Product Development Performance: Strategy, Organization and Management in the World Auto Industry. HBS Press.
- Cooke-Davies, T., 2002. The “real” success factors on projects. *International Journal of Project Management*, 20(3), pp.185-190.
- Denicol, J., Davies, A., & Krystallis, I. (2020). What are the causes and cures of poor megaproject performance? A systematic literature review and research agenda. *Project Management Journal*, 51(3), 328-345.
- van Dyke, T. P., Prybutok, V. R., & Kappelman, L. A. (1999). Cautions on the Use of the SERVQUAL Measure to Assess the Quality of Information Systems Services. *Decision Sciences*, 30(3), 877–891.
- Fleming, W. and J. Koppelman (1996). *Earned value project management*. Newton Square, Project Management Institute.
- Flyvbjerg, B. (2016). The fallacy of beneficial ignorance: A test of Hirschman’s hiding hand. *World Development*, 84, 176-189.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, 18(1), 39-50.

Gatian, A. W. (1994). Is user satisfaction a valid measure of system effectiveness?. *Information & management*, 26(3), 119-131.

Geraldi, J., and Arlt, M. (2015). *Visuals Matter! Designing and using effective visual representations to support project and portfolio decisions*. Project Management Institute.

Geraldi, J. and Lechler, T. (2012), "Gantt charts revisited: A critical analysis of its roots and implications to the management of projects today", *International Journal of Managing Projects in Business*, Vol. 5 No. 4, pp. 578-594.

Geraldi, J., & Stingl, V. (2016). *From Visions of Grandeur to Grand Failure: Alternative schools of descriptive decision theories to explain the Berlin Brandenburg*. In EURAM 2016 (p. 2298). Academy of Management.

Glaser, B. G., & Strauss, A. L. (2017). *The discovery of grounded theory: Strategies for qualitative research*. Routledge.

Grönroos C., 1984. A Service Quality Model and its Marketing Implications, *European Journal of Marketing*, 18 (4) 36-44.

Hair, J. F. (2009). *Multivariate data analysis: a global perspective*. 7th edition. Upper Saddle River: Prentice Hall.

Harvey, J. (1998). Service quality: a tutorial. *Journal of operations management*, 16(5), 583-597.

Haywood-Farmer, J. (1988). A conceptual model of service quality. *International journal of operations & production management*.

Hernes, T. (2022). *Organization and Time*. Oxford: Oxford University Press.

- Hoegl, M., & Gemuenden, H. G. (2001). Teamwork quality and the success of innovative projects: A theoretical concept and empirical evidence. *Organization Science*, 12(4), 435-449.
- Hoxley, M. (2000). Measuring UK construction professional service quality: the what, how, when and who. *International Journal of Quality & Reliability Management*.
- Ika, L. A., & Pinto, J. K. (2022). The “re-meaning” of project success: Updating and recalibrating for a modern project management. *International Journal of Project Management*, 40(7), 835-848.
- Johnston, R. (1999). Service operations management: return to roots. *International Journal of Operations & Production Management*, 19(2), 104-124.
- Jöreskog, K. G., & Sörbom, D. (2006). *LISREL 8.80*. Lincolnwood, IL: Scientific Software International Inc.
- Jugdev, K., & Müller, R. (2005). A retrospective look at our evolving understanding of project success. *Project management journal*, 36(4), 19-31.
- Kagioglou, M., Cooper, R., & Aouad, G. (2001). Performance management in construction: a conceptual framework. *Construction management and economics*, 19(1), 85-95.
- Kang, G. D., & James, J. (2004). Service quality dimensions: an examination of Grönroos's service quality model. *Managing Service Quality: An International Journal*.
- Kaplan, R. S., & Norton, D. P. (2004). Focusing your organisation on strategy-with the balanced scorecard. Cambridge: Harvard Business School Publishing.
- Kettinger, W. J., & Choong C. Lee, W. J. (1994). Perceived Service Quality and User Satisfaction with the Information Services Function. *Decision Sciences*, 25(5/6), 737–766.

- Kettinger, W. J., & Lee, C. C. (2005). Zones of Tolerance: Alternative Scales for Measuring Information Systems Service Quality. *MIS Quarterly*, 29(4), 607–623.
- Kreiner, K. (1995). In search of relevance: Project management in drifting environments. *Scandinavian Journal of Management*, 11(4), 335-346.
- Kreiner, K. (2014). Restoring project success as phenomenon. In *Advancing research on projects and temporary organizations* (pp. 19-38). Copenhagen Business School Press.
- Kreiner, K. (2020). Conflicting notions of a project: The battle between Albert O. Hirschman and Bent Flyvbjerg. *Project Management Journal*, 51(4), 400-410.
- Lehtinen, U., & Lehtinen, J. R. (1991). Two approaches to service quality dimensions. *Service Industries Journal*, 11(3), 287-303.
- Lim, C. S., & Mohamed, M. Z. (1999). Criteria of project success: an exploratory re-examination. *International journal of project management*, 17(4), 243-248.
- Lin, S. H. E., & Gerber, D. J. (2014). Designing-in performance: A framework for evolutionary energy performance feedback in early stage design. *Automation in Construction*, 38, 59-73.
- Linhart, H., & Zucchini, W. (1986). *Model selection*. John Wiley & Sons.
- Lipke, W., Zwikael, O., Henderson, K. and Anbari, F., 2009. Prediction of project outcome: The application of statistical methods to earned value management and earned schedule performance indexes. *International journal of project management*, 27(4), pp.400-407.
- Love, P. E., Smith, J., Treloar, G. J., & Li, H. (2000). Some empirical observations of service quality in construction. *Engineering, Construction and Architectural Management*.

- Maylor, H., Turner, N., & Murray-Webster, R. (2015). "It worked for manufacturing...!": operations strategy in project-based operations. *International journal of project management*, 33(1), 103-115.
- Murray, P. (2003). *The saga of Sydney Opera House: the dramatic story of the design and construction of the icon of modern Australia*. Routledge.
- Munns, A. K., & Bjeirmi, B. F. (1996). The role of project management in achieving project success. *International journal of project management*, 14(2), 81-87.
- Nguyen, L. D., & Ogunlana, S. O. (2004). A study on project success factors in large construction projects in Vietnam. *Engineering, Construction and Architectural Management*, 11(6), 404-413.
- Nunnally, J. C. (1978). An overview of psychological measurement. *Clinical diagnosis of mental disorders: A handbook*, 97-146.
- Parasuraman, A., Zeithaml, V.A. and Berry, L.L., 1985. A conceptual model of service quality and its implications for future research. *Journal of marketing*, 49(4), pp.41-50.
- Pesämaa, O., Bourne, M., Bosch-Rekvelde, M., Kirkham, R. and Forster, R., 2020. Call for papers: Performance measurement in project management. *International Journal of Project Management*, 38(8), pp.559-560.
- Pesämaa, O., Larsson, J., & Eriksson, P. E. (2018). Role of performance feedback on process performance in construction projects: Client and contractor perspectives. *Journal of Management in Engineering*, 34(4), 04018023.
- Philip, G., & Hazlett, S. A. (1997). The measurement of service quality: a new P-C-P attributes model. *International Journal of Quality & Reliability Management*.

- Pinto, J. K., & Slevin, D. P. (1988). Critical success factors across the project life cycle. Drexel Hill, PA: Project Management Institute.
- Pitsis, T. S., Clegg, S. R., Marosszeky, M., & Rura-Polley, T. (2003). Constructing the Olympic dream: A future perfect strategy of project management. *Organization science*, 14(5), 574-590.
- Pitt, L. F., Watson, R. T., & Kavan, C. B. (1997). Measuring information systems service quality: concerns for a complete canvas. *MIS quarterly*, 209-221.
- Quattrone, P. (2017). Embracing ambiguity in management controls and decision-making processes: On how to design data visualisations to prompt wise judgement. *Accounting and Business Research*, 47(5), 588-612.
- Rust, R. T., Verhoef, P. C. and Oliver. R.L. (1994). Service quality: Insights and managerial implications from the frontier. *Service Quality: New Dimensions in Theory and Practice*. Thousand Oaks, CA: Sage.
- Sampson, S. E. (2010). The unified service theory: A paradigm for service science. *Handbook of service science*, 107-131.
- Shenhar, A.J., Dvir, D., Levy, O. & Maltz, A.C. (2001), "Project Success: A Multidimensional Strategic Concept," *Long Range Planning*, Vol. 34, pp. 699-725.
- Thompson, J. (1967). D.(1967) *Organisations in action*. New York.
- Toor, S and Ogunlana, S. O. (2010). Beyond the 'iron triangle': Stakeholder perception of key performance indicators (KPIs) for large-scale public sector development projects. *International journal of project management*, 28(3), 228-236.

- Turner, R., & Zolin, R. (2012). Forecasting success on large projects: developing reliable scales to predict multiple perspectives by multiple stakeholders over multiple time frames. *Project management journal*, 43(5), 87-99.
- Wikström, K., Hellström, M., Artto, K., Kujala, J., & Kujala, S. (2009). Services in project-based firms—Four types of business logic. *International Journal of Project Management*, 27(2), 113-122.
- Winch, G., Usmani, A., & Edkins, A. (1998). Towards total project quality: a gap analysis approach. *Construction Management & Economics*, 16(2), 193-207.
- Yuan, J., Zeng, A. Y., Skibniewski, M. J., & Li, Q. (2009). Selection of performance objectives and key performance indicators in public–private partnership projects to achieve value for money. *Construction management and economics*, 27(3), 253-270.