Modelling Labour Supply for NATA Accreditation Assessment Services

Centre for Business and Social Innovation

UTS CRICOS 00099F

The UTS Research Team

Research Investigators

Associate Professor Christopher Bajada

Associate Professor Renu Agarwal

Research Associates

Dr Mikhail Shashnov

Dr Katrina Skellern

For further information about this research study, please contact Associate Professor Christopher Bajada at:
Email address: chris.bajada@uts.edu.au

Acknowledgements

UTS would like to thank the following for their contributions: Mr Eric Lo, Ms Kirsty Outhred and the NATA team for providing their time and knowledge to enable the UTS research team to compile a thorough understanding of the Assessment Effort model.

December 2019

Contents

Executive Summary

In December 2017, NATA developed a new fee methodology - the Assessment Efforts (AE) Model - to replace the Chargeable Unit system. The purpose of the AE Model is to provide a basis for recovering the cost of operations from NATA members in an accurate, fair, equitable and efficient manner. The AE model facilitates resources planning for each financial year, across the different assessment cycles, and enables the projection of NATA resources required whilst comparing the quantum of actual resources with existing and budgeted staff resources.

NATA commissioned the Centre for Business & Social Innovation (CBSI), UTS Business at the University of Technology Sydney (UTS) to evaluate the efficacy and efficiency of the AE model. Hence, this report aims to:

a	Provide an economic assessment on the current processes and assumptions used to allocate labour for accreditation assessment services by reviewing the AE Model,
b	Evaluate the efficiency and predictive value of the AE Model by reviewing the validity of key assumptions of the model as applied to a sample of the clients' historical time records of assessments; and
c	Make recommendations, to improve the robustness, accuracy, and sensitivity of the AE Model.

Overall the findings illustrate that on average the AE model gets it right!

When applying the AE model assumptions to general organisational assessment hours, it provides an accurate picture of assessment effort as a whole and therefore, is a reliable model in the use of fee generation. However, some refinements to the AE model are needed to improve the accuracy of forecasts at the activity and service level. Here, system generated hours differ to actual assessment hours (at times under-estimating or over-estimating actual assessment hours). This report suggests ways that the AE model can be improved to better estimate assessment hours at the activity and service level and providing improvements to the labour assessment hours estimates will allow NATA to better allocate scarce resources amongst their various clients knowing that this primarily a volunteer based scheme. A summary of the key results are presented in the recommendations section (section 3) at the end of this report.

1 Introduction

The National Association of Testing Authorities, (NATA) is Australia's national authority for the accreditation of laboratories, producers of reference materials, and a peak body for the accreditation of inspection bodies and proficiency testing scheme providers. Established in 1947, NATA is a not-for-profit organisation owned by its members.[1] The work of NATA contributes to Australia across several fields:

- Safeguarding the community from non-conforming products and services,
- Redressing information asymmetry between sellers and buyers of products and services,
- Reducing transaction costs, and
- Facilitating the reduction of technical and regulatory trade barriers.

NATA provides accreditation assessment services to its members and clients through the expertise of over 3000 technical volunteers on a time and material basis. Assessments cover a wide range of services and facilities in areas such as pathology, diagnostic imaging, environmental analysis, food, water, pharmaceuticals, concrete, asbestos, toxicology, electrical equipment, IT, biotechnology, and many more. The assessment activities undertaken by NATA vary with the 'scope of accreditation'; the larger the scope and/or more complex the test activities, the more time and effort will be expended by NATA in its assessment of the competence of member organisations.

Members pay an annual subscription fee to NATA and in return receive, among other services, onsite assessments and quality management system reviews of each accredited facility, associated activities and services. These reviews are conducted at approximately three year intervals[2]. On average, over a three year period, NATA will complete an accreditation assessment cycle of a) an onsite assessment, b) a desktop review at around 18 months after the onsite assessment, and finally c) an onsite assessment.

Until December 2017, NATA used a 'Chargeable Unit' (Unit) for estimating the annual subscription fee payable by each member. For example, a member organisation whose scope attracts three Chargeable Units worth of services paid three times as much as an organisation that attracts one Chargeable Unit. Due to the imprecise nature of the Unit calculation and the inconsistent application of the Unit method over the years, NATA believed the accuracy and validity of outputs had not advanced and in some cases had deteriorated. Therefore, in 2017,

[1] https://www.nata.com.au/about-nata/our-role
[2] Review assessment cycle timing varies from two to four years based on different service types.

NATA developed a new fee methodology to replace the 'Chargeable Unit' - the Assessment Efforts (AE) Model.

The AE model uses historical records of person hours and functional relationships with selected activity drivers to provide a more accurate fee estimate, as well as introducing a new method of charging fees to their prospective members.

The AE Model:

1	Provides a basis for recovering the cost of operations from members in an accurate, fair, equitable and efficient manner,
2	Facilitates resource planning for each financial year, across the assessment cycle, and
3	Enables the projection of NATA resources required by comparing the quantum of actual resources with existing and budgeted staff resources (all in equivalent person hours).

There is a potential risk that an inaccuracy in forecasting could have implications for resource planning which could result in either over-or under- estimation of the required labour resources needed for assessment, and therefore affecting planning decisions.

1.1 Project Objectives

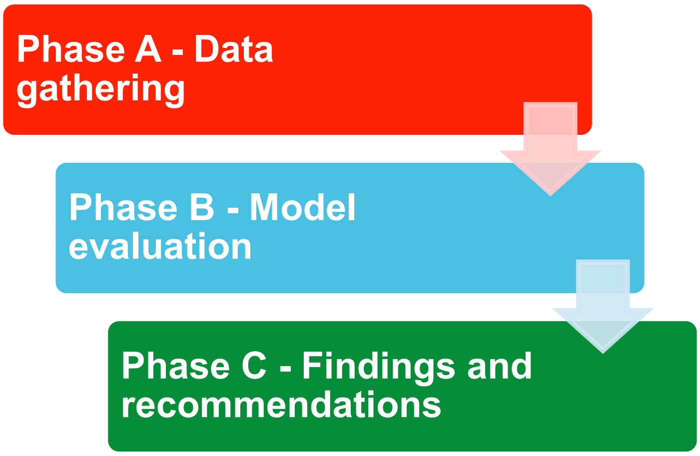
NATA has commissioned the Centre for Business & Social Innovation (CBSI), Business School at the University of Technology Sydney (UTS) to evaluate the efficacy and efficiency of the AE model and to assist NATA facilitate the applications addressed in points 1, 2 and 3 in Section 1 above. Hence, this report provides an evaluation of the current AE model used by NATA and provides improvements to the labour assessment hour estimates which will allow NATA to better allocate scarce resources amongst their various clients knowing that this is primarily a volunteer based scheme... The purpose of this project is to:

a	Provide an economic assessment on the current processes and assumptions used to allocate labour for accreditation assessment services by reviewing the AE Model,
b	Evaluate the efficiency and predictive value of the AE Model by reviewing the validity of key assumptions of the model as applied to a sample of the clients' historical time records of assessments; and
c	Make recommendations, to improve the robustness, accuracy, and sensitivity of the AE Model.

1.2 Research design and methodology

In order to ensure robust analysis of the AE Model, a structured research design and methodology was organised across three phases as illustrated in Figure 1.1.

Figure 1.1: Research design and methodology



Phase A involved several meetings and ongoing dialogue between the UTS research team and NATA staff representatives to ensure a sound understanding of the AE model. This step subsequently ensured the use of the appropriate data needed for the analysis and enabled the research team to replicate the model calculations using existing assumptions as deployed by NATA.

Phase B involved applying statistical techniques to evaluate the AE model by closely examining the underlying assumptions and forecast accuracy of the AE model.

Phase C enabled the UTS research team to articulate the findings on the evaluation of the AE model and recommend changes to improve the AE model forecasts.

2 Evaluation of Assessment Effort (AE) Model

Following consultation with NATA staff representatives and access to all relevant data, the UTS research team commenced their evaluation of the AE model. Section 2.1 outlines the current scope of accreditation as the context for this model evaluation, while Section 2.2 describes the construct and assumptions of the AE model. Section 2.3 discusses the evaluation and sample selection methods, Section 2.4 details further stages of the AE model evaluation, Section 2.5 outlines alternative AE model specifications and Section 2.6 provides an evaluation of the AE model assumptions.

2.1 Background

NATA's Accreditation Procedures Manual (APM) outlines the policy and procedures for the use of the AE Model under Section 6.1.1 'Assessment Effort (Sampling Plan for the Scope of Accreditation). In short, it contains the sampling activities covered by the scope of accreditation in order for NATA to establish and confirm ongoing competence of accredited facilities during assessments.

The principle of sampling activities is dependent upon defining and grouping tasks where the same (or like) competencies are applied. The Scheme is used to define the job types and assessment intervals for the assessment cycle of a particular Standard and is also used to define fee schedules. Fee codes are applied at the 'Standard' level for a site's scope.

The AE model accounts for the time necessary for NATA to service a client for one complete assessment cycle and in so doing provides the means to:

- Determine staff resource needs for the assessment cycle, and
- Inform how fees should be calculated and structured for the assessment.

The base AE model required to assess the quality management system (QMS) and the technical effort (TE)[3] is pre-set for each Activity and automatically applied to each Service within that Activity. These pre-set values enable fine tuning of effort required and sampling complexity to be applied at a service level.

[3] Technical complexity of the effort

Upon creating a scope for a site/department, each scheme within the Site's Scope of Accreditation has its QMS effort pre-set and each Service belonging to the Activity has its TE pre-set.

A complexity factor is also pre-set for each Service within an Activity that belongs to a particular Standard. This complexity factor describes the additional effort (in hours) required to assess a Service based on the number of determinants, products, or techniques the facility's service includes. The effort required to follow up minor non-conformances during an on-site assessment activity is factored into the base QMS effort rather than being separately specified. Appendix 1 provides an illustrative flow diagram of the current AE Model.

2.1.1 Job Types

Time expended on assessment activities can be broadly separated into time spent on scheduled jobs and non-scheduled jobs. The former includes reassessments (RES – see Appendix 2), (Formulas – see Appendix 3), surveillance visits (SRV – see Appendix 4) and online surveillance (OLN for medical testing only – see Appendix 5) which are non-chargeable and covered by annual fees.

The efficiency of the AE model can be evaluated on the basis of three criteria, namely:

(i) How accurately the AE model is able to predict actual hours of assessment,

(ii) The reliability of the assumptions used in the AE model, and

(iii) The sensitivity of these assumptions on the predictions of the actual assessment hours.

2.2 Model Description

The following section describes the AE Model in further detail for the purposes of replicating the model and determining the assessment hours required for accreditation.

2.2.1 Assessment Cycle

The surveillance cycle's duration varies depending upon the scheme used to assess a facility's services against a particular standard. APM 6.0 describes these surveillance cycles together with the various job types. In summary:

- Every 2 year cycle - GLP: RES
- Every 3 year cycle - All ISO/IEC 17025 (testing and calibration); ISO/IEC 17020 (Inspection); ISO/IEC 17043 (PTSP); ISO 17034 and Guide 34 (RMP): one assessment activity every 18 months sequentially ordered to conduct one RES and then one SRV.

- Every 4 year cycle - ISO 15189 (Medical Testing): yearly assessment activity sequentially ordered to conduct one RES, then one OLN then, one SRV and then one OLN,
    - RANZCR Stds (Medical Imaging): Assessment activity every 2 years sequentially ordered to conduct one RES and then one SRV,
    - ASA Stds (Sleep Disorders Services): Assessment activity every 2 years sequentially ordered to conduct one RES and then one OLN,
    - Note: RANZCR and ASA programs are both fee for service so the effort model is utilised as a guide to the duration of anticipated scheduled activities, i.e. effort required.

A client's total assessment effort for one complete Surveillance Cycle is equal to *Assessment effort =Total Reassessment effort* + *Total Surveillance effort (where relevant)* + *Total On-line surveillance effort (where relevant).* This includes an estimate for Pre, UA and Post time, i.e. actual assessment hours recorded by the technical assessors. However, it does assume one Lead Assessor and one Technical Assessor are conducting the jobs, where relevant.

The next section describes the steps undertaken by the research team to evaluate the efficacy and efficiency of the AE Model.

2.3 Steps in evaluating the AE Model

To undertake this analysis, the UTS research team used a sample of 100 sites supplied by NATA. The process of evaluation compared the 'system generated hours' from the AE model with the actual hours recorded by the technical assessors at the aggregate, as well as the activity and service levels. The actual assessment hours recorded by the technical assessors were categorised as PRE, UA, POST and NCCA hours. The system generated hours do not account for time spent on other activities such as travel time to a site, therefore the NCCA hours (which capture such information) are not directly comparable to the technical hours generated by the AE assessment model.

2.3.1 Sample Selection

The UTS research team worked with NATA to select the appropriate sites used in the analysis. The choice of sites was based on the assessment of a single service (not including QMS). This is because actual hours recorded by the technical assessors do not distinguish between different services. The assessors only record the total hours spent at each site irrespective of how many services are being evaluated. Therefore, knowing exactly how many hours are spent on each service is impossible to determine. Given this constraint, and to allow for a direct comparison of system generated and actual assessment hours, only those sites with a single service (plus QMS) were considered, given that QMS is treated as a constant. A total of 100 sites were identified as having a single service (plus QMS). However from these 100 sites, only 73 sites had complete actual hours data, thus allowing the analysis using 73 observations (when system generated and actual hours were both required).

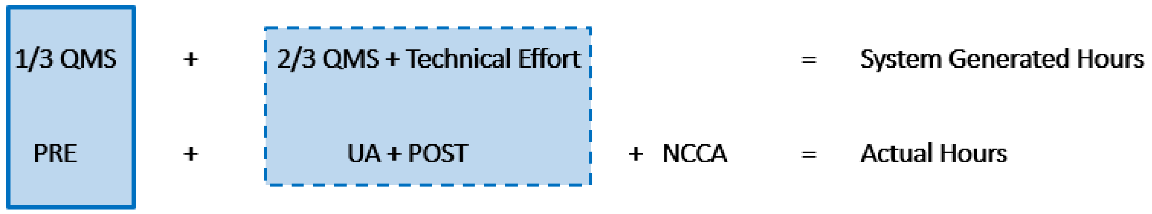
From these 73 selected site samples (offering one service only including QMS) there were a total of three activities and four services that spanned across these sites, thus limiting our analysis. The activities and services for these 73 sites are as follows:

Activities	Services
• Activity 1 - Human Pathology • Activity2 - Infrastructure and Asset Integrity • Activity 3 - Manufactured goods	• Service 1: Diesel engines for use in hazardous atmospheres - Overhaul and maintenance inspection, • Service 2: Integrity evaluation of air control equipment, • Service 3: Immunopathology - procedures related to the collection, processing, storage and issue of human haemopoietin progenitor cells, and • Service 4: Mechanical performance evaluation of components for building envelopes, framing and interior lining.

2.4 Evaluation of the AE Model

In Figure 2.1, the UTS research team compared the system generated hours with actual assessment hours. The actual PRE hours are approximated by a portion (1/3) of system QMS hours while the actual UA and POST assessment hours are approximated by the system generated total technical effort and a portion (2/3) of QMS hours. These approximations between variables are shown in Figure 2.1 as the solid box (PRE hours) and the dashed box (UA and POST) respectively. NCCA hours are not factored into the AE system generated hours but when combined with PRE, UA and POST comprise total actual hours spent by the technical assessor(s).

Figure 2.1 – Comparison of system generated and actual assessment hours



System generated hours were compiled across the 73 selected sample sites (including QMS) and were compared to total assessment hours (PRE, UA and POST) (excluding NCCA). The mean hours across all sites are presented in Table 2.1. The results in Table 2.1 suggest that the AE model provides on average a good approximation of the average assessment hours spent across the sample of 73 sites. The mean difference between the actual hours and system generated hours was not found to be statistically significant, suggesting that the AE model does a good job as a whole in forecasting actual assessment hours.

Table 2.1 – Comparing system generated and actual assessment hours

Variable	Mean hours	p-value	Sample size	Significantly different
Total Actual Hours (PRE, UA and POST)	32.04	0.515	73	No
TotalCycleEffort (System generated)	33.36		73	

Note: *TotalCycleEffort = total technical effort plus QMS effort (=system generated hours).*

To examine whether the performance of the AE model holds true at the activity and service level, the UTS research team repeated the process of comparing mean hours at each of these two levels. Tables 2.2 (a-d) provide the comparison of the actual vs the system generated hours at the service level and Table 2.3 (a-c) provide the comparison at the activity level.

Comparison of the actual vs the system generated hours at the service level

The results in Tables 2.2 (a-d) provide mixed results. For some services (Services 1 and 3) there is a statistically significant difference between the mean system generated and actual hours (excluding NCCA hours from total actual hours). For example, in Table 2.2a, the AE model estimated a total of 37.02 hours on average for Service 1 while the actual reported hours for the same sites in the sample was an average of 51.78 hours. Likewise, for Service 3 (see Table 2.2c) the AE model estimated a total of 20.3 hours on average for Services while the actual reported hours for the same sites in the sample was an average of 36.17 hours.

For the remaining two services (Services 2 and 4) there was no statistical significant difference identified between the mean hours generated by the system (AE Model) and actual hours reported. For example, in Table 2.2b, the AE model estimated a total of 30.43 hours on average for Services 2 while the actual reported hours for the same sites in the sample were an average of 31.91 hours. A similar result was found for Services 4 (see Table 2.2d).

Table 2.2a – Comparing system generated and actual assessment hours (service 1)

Variable	Mean	p-value	Sample size	Significantly different
Total Actual Hours (PRE, UA and POST)	51.78	0.0041	15	Yes
TotalCycleEffort (System generated)	37.02		15	

Note: TotalCycleEffort = total technical effort plus QMS effort (=system generated hours).

Table 2.2b – Comparing system generated and actual assessment hours (service 2)

Variable	Mean	p-value	Sample size	Significantly different
Total Actual Hours (PRE, UA and POST)	30.43	0.6198	22	No
TotalCycleEffort (System generated)	31.91		22	

Table 2.2c – Comparing system generated and actual assessment hours (service 3)

Variable	Mean	p-value	Sample size	Significantly different
Total Actual Hours (PRE, UA and POST)	20.30	0.0000	23	Yes
TotalCycleEffort (System generated)	36.17		23	

Table 2.2d – Comparing system generated and actual assessment hours (service 4)

Variable	Mean	p-value	Sample size	Significantly different
Total Actual Hours (PRE, UA and POST)	32.75	0.1021	13	No
TotalCycleEffort (System generated)	26.62		13	

These varying results may be due to the sampling size, but it may also be due to variances in the actual reported hours as a result of different assessors being involved. In the remainder of this section, we provide further detail of this comparison.

Comparison of the actual vs the system generated hours at the activity level

Next, the UTS research team compared the mean difference in assessment hours across the three activities (see Tables 2.3 a-c). Again the results are mixed. For Activity 2 (see Table 2.3b) there is no statistical significant difference between the mean system generated (including QMS) and actual hours (excluding NCCA hours) while for the other two activities (Activities 1 and 3 in Tables 2.3a and Table 2.3c respectively) a statistically significant difference in the mean hours was found. For example, the AE model over-estimated assessment hours for Activity 1 (36.17 hours) when compared with actual recorded hours by assessors (20.3 hours). On the other hand, the AE model under-estimated assessment hours for Activity 3 (32.19 hours) when compared with actual recorded assessment hours (42.95 hours).

Table 2.3a – Comparing system generated and actual assessment hours (Activity 1)

Variable	Mean	p-value	Sample size	Significantly different
Total Actual Hours (PRE, UA and POST)	20.30	0.0000	23	Yes
TotalCycleEffort (System generated)	36.17		23	

Table 2.3b – Comparing system generated and actual assessment hours (Activity 2)

Variable	Mean	p-value	Sample size	Significantly different
Total Actual Hours (PRE, UA and POST)	30.43	0.6198	22	No
TotalCycleEffort (System generated)	31.91		22	

Table 2.3c – Comparing system generated and actual assessment hours (Activity 3)

Variable	Mean	p-value	Sample size	Significantly different
Total Actual Hours (PRE, UA and POST)	42.95	0.0009	28	Yes
TotalCycleEffort (System generated)	32.19		28	

2.4.1 Summary of Results

In this section we compare the results of both the service and activity levels as presented in Tables 2.2 and 2.3 earlier. In Table 2.4 the symbol 'x' indicates that there is a statistical significance between actual reported and system generated hours (from AE model) and the symbol '✓' indicates that there is no statistical difference between actual hours and system generated hours (from AE model). For robustness, a comparison of actual total hours including NCCA was also added to this table. The results in Table 2.4 illustrate relatively consistent results with and without the inclusion of NCCA hours. These particular results can be attributed to mapping between these specific services and activities. Note also the following relationships are also influencing these results:

- All Service 2 are in Activity 2
- All Service 3 are in Activity 1
- All Service 1 + Service 4 are in Activity 3

Table 2.4 – Summary of results

		Service 1	Service 2	Service 3	Service 4
Comparing Total Actual and system generated hours	With NCCA	X	✓	X	X
	Without NCCA	X	✓	X	✓
		Activity 1	**Activity 2**	**Activity 3**	
Comparing Total Actual and system generated hours	With NCCA	X	✓	X	
	Without NCCA	X	✓	X	

Notes:

x – denotes instances where means of system generated hours are statistically significantly different from the mean actual hours (with and without NCCA hours).

✓ - denotes instances where means of system generated hours are not statistically significantly different from the mean actual hours (with and without NCCA hours).

Understanding the assumptions of the AE model

Understanding the assumptions of the AE model will provide insight and shed light on the reasons for this variation in prediction as illustrated in the earlier section. The research team explore these differences in predictions in two different ways:

1. Spread of assessment hours across service and activities – system generated versus actual reported hours

2. Assessing Activity and services by technique (simple/complex) by closely analysing the variation in total actual hours using visual box-whisker plots. The height of the rectangular boxes represents the inter-quartile range on the given dataset. The lines extending from the boxes are known as “whiskers”, used to indicate variability outside the upper and lower quartiles. Outliers are plotted as dots outside the lines. The horizontal line within the rectangular box represents the median. Together all this information provides a visual on the spread of the data.

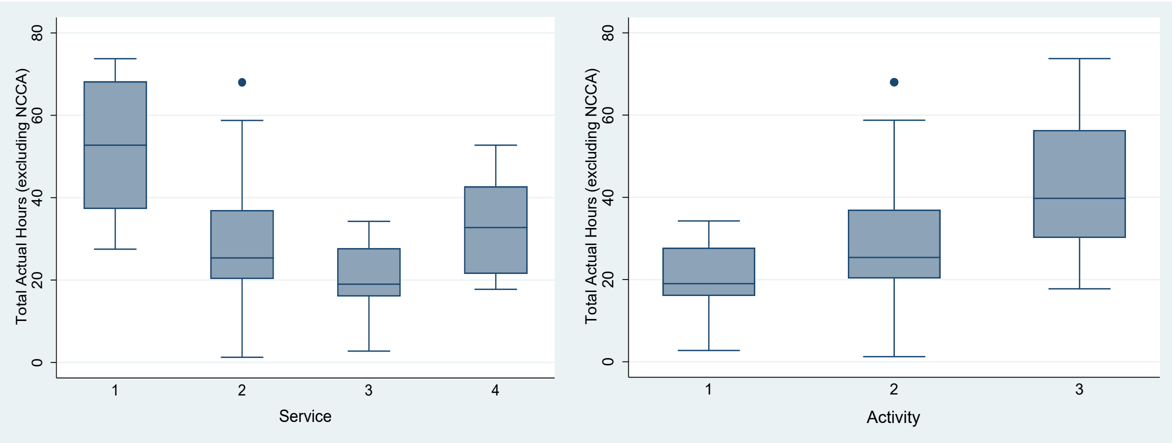
2.4.2 Spread of assessment hours across service and activities – system generated versus actual reported hours

In this section, the box-whisker plots are drawn by service and activity for total actual hours (excluding NCCA). This activity is also broken down by RES and SRV to examine whether any variation in hours may be explained during the reassessment or surveillance activities.

Appendix 8 provides additional results including NCCA. The results with and without NCCA are quite similar in their variance and distribution.

In Figure 2.2, the box-whisker plots are provided for total actual hours (excluding NCCA) by service and activity. Looking at the services, there is considerable variation in the median (line within the whisker plots) and spread in the inter-quartile range (size of box). In some cases the inter-quartile range for one service is entirely above the inter-quartile range of another service (for example, compare Service 1 and 3).

Figure 2.2 – Box-Whisker Plots by Service and Activity



Note 1: *Service descriptors include:*

- *Service 1: Diesel engines for use in hazardous atmospheres - Overhaul and maintenance inspection*
- *Service 2: Integrity Evaluation of air control equipment*
- *Service 3: Immunopathology - Procedures related to the collection, processing, storage and issue of human haemopoietin progenitor cells*
- *Service 4: Mechanical performance evaluation of components for building envelopes, framing & interior lining*

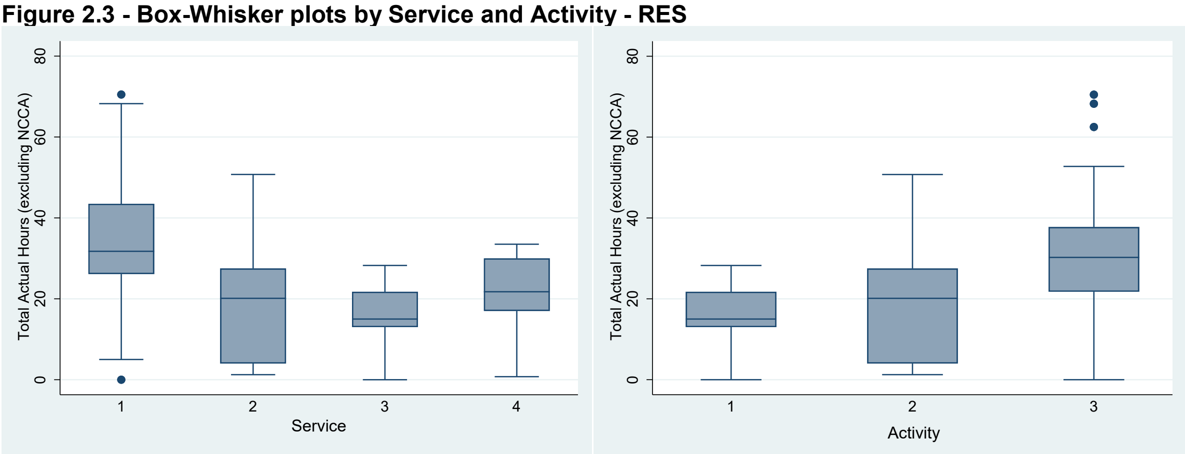
Note 2: *Activity descriptors include:*

- *Activity 1 - Human Pathology*
- *Activity 2 - Infrastructure and Asset Integrity*
- *Activity 3 - Manufactured Goods*

A similar result is demonstrated in the box-whisker plots for activities. Both the median and the size of the inter-quartile range are substantially different across the three types of activities. This might suggest that a model that assumes homogeneity across services and activities would be problematic.

In Figures 2.3 and 2.4, the UTS research team provides the same data in Figure 2.2 by its two components: RES and SRV (which includes OLN for medical testing). Comparing the results in Figure 2.3 with the results in Figure 2.2, the findings illustrate that results are quite similar with some notable variations in Service 2 and Activity 2.

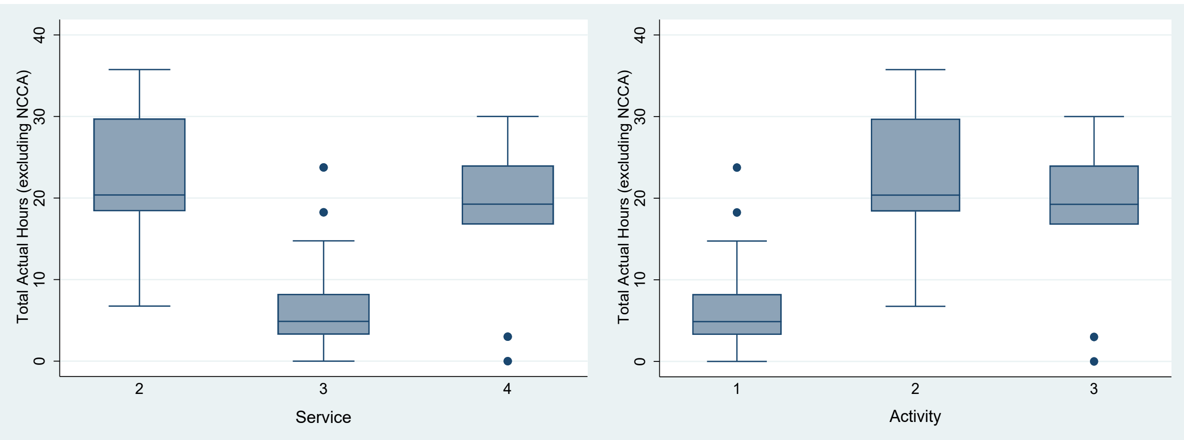
Figure 2.3 - Box-Whisker plots by Service and Activity - RES



Note: *see Figure 2.2 for definitions of activity and service*

In Figure 2.4, we plot the box-whisker plots for SRV. Comparing the results in this figure with those in Figure 2.3, we note that there are quite different variance and inter-quartile range across SRV and RES for activities and services. The findings in Figures 2.3 and 2.4 suggest different assumptions are required for RES and SRV and that the overall variation in total hours (for both service and activity) need to account for difference in assessment hours in RES and SRV. These have important implications for the assumptions used in the model.

Figure 2.4 - Box-Whisker plots by Service and Activity – SRV



Note 1: *including OLN for medical testing. Service 1 is not present in the figure because of missing observations.*
Note 2: *for service and activity definitions see Figure 2.2.*

2.4.3 Assessing Activity and services by technique (simple/complex)

The difficulty and the time consuming nature involved in an assessment is an important factor explaining the variation in assessment hours observed in the box-whisker plots. This difficulty or time consuming nature is represented / accounted for in the AE model as technique type, which is classified into three categories: Classic (or simple); Mixed and Instrument (or complex). Each of these three types are given weightings in the model, namely, 0.5, 0.75 and 1 respectively. Table 2.5 illustrates the average hours by technique type across all services and activities in this sample of sites.

Table 2.5 - Average Assessment Hours by Technique Type – Full Sample (all activities and services)

Technique type	Average Assessment Hours	Sample size
Classic	40.28	32
Mixed	28.99	21
Instrument	22.06	20

Note: average assessment hours excludes NCCA hours.

Table 2.5 indicates that the average actual hours (without NCCA hours) is higher for the classic technique classification compared to the mixed and instrument technique classifications. This finding appears to be inconsistent with the assumptions of the model which assumes that instrument type (being complex) should take on average more time to assess than the simpler (Classic) assessment technique type, hence the reason for being allocated a greater weight.

To understand this result a little better, the research team examined the mean hours at the level of technique type by service and activity. These results are presented in Tables 2.6 and Table 2.7.

Table 2.6 - Average Assessment Hours by Technique Type – by Service

Technique type	Service			
	1	2	3	4
Classic				
Average Assessment Hours	51.78	20.19	27.25	33.69
Sample size	15	4	1	12
Mixed				
Average Assessment Hours		30.63	22.00	
Sample size		17	4	
Instrument				
Average Assessment Hours		68.00	19.54	21.50
Sample size		1	18	1

Note 1: excluding NCCA hours

Note 2: service descriptors include:

- Service 1: Diesel engines for use in hazardous atmospheres - Overhaul and maintenance inspection
- Service 2: Integrity Evaluation of air control equipment
- Service 3: Immunopathology - Procedures related to the collection, processing, storage and issue of human haemopoietin progenitor cells
- Service 4: Mechanical performance evaluation of components for building envelopes, framing and interior lining

For some/or for most services and activities (such as Services 3 and 4, and Activities 1 and 3) the average value of actual hours is higher for the classic technique type than for Instrument (complex) technique type. While for others (Service 2 and Activity 2), the average value of assessment hours is higher for instrument technique type than for classic technique type as assumed in the AE model based on the technique type weights. Therefore, a review of the simple/complex nature of the assessment tasks would be useful in helping improve the accuracy of the AE model.

Table 2.7 – Average Assessment Hours by Technique Type – by Activity

Technique type	Activity			
	1	2	3	
Classic				
Average Assessment Hours	27.25	20.19	43.74	
Sample size	1	4	27	
Mixed				
Average Assessment Hours	22.00	30.63		
Sample size	4	17		
Instrument				
Average Assessment Hours	19.54	68.00	21.50	
Sample size	18	1	1	

Note 1: excluding NCCA hours

Note 2: Activity descriptors include

- Activity 1 - Human Pathology
- Activity 2 -Infrastructure and Asset Integrity
- Activity 3 - Manufactured Goods

Appendix 9 provides the results including NCCA which are consistent with those reported here.

2.5 Alternative AE Model Specification

In this section and Section 2.6, the research team test the validity of the AE model by examining the various assumptions on which the AE model is based. Specifically, the assumptions examined are:

(i) the approach taken on factor rounding,

(ii) (ii) the technique type weights, and

(iii) (iii) the pre-determined constant values for each field as presented in Table 1A in Appendix 2; and the determination of QMS.

To replicate the system generated results from the AE model using the sample dataset, we express the AE model algebraically from which we derive a simplified expression that can be used to replicate the system generated results. The AE model can be represented algebraically as follows:

$$SysGenHours = [A + B.F_1]P + [E + G.F_2]P \quad (1)$$

where

SysGenHours = system generated assessment hours as estimated using the AE model.

A = STE_RES values from Table 1A

B = CFE_RES values from Table 1A

C = CFQ_RES values from Table 1A

D = number of unique determinations

E = STE_SRV from Table 1A

F = CFQ_SRV from Table 1A

G = CFE_SRV from Table 1A

P = technique type weighting

$F_1 = \frac{1}{C} \times D$ (adjusted using NATA rounding)

$F_2 = \frac{1}{F} \times D$ (adjusted using NATA rounding)

Using the implied linear relationships given in Table 1A of Appendix 3 between A, B, C, E, F and G and then simplifying equation (1), results in the following equation:

$$SysGenHours = P.A\left[\frac{4}{3} + \frac{1}{2}F_1 + \frac{1}{6}F_2\right] \quad (2)$$

Equation (2) was used to produce the system generated assessment hours consistent with those generated using the AE model. Using this specification (equation 2), observations illustrate that the system generated hours are influenced by the technique type weighting (variable: P), the assumed values for STE_RES (variable: A), the factor scores which in turn depend on CFQ_RES, and the rounding method used to adjust the factor scores. With this in mind, the research team considered how sensitive the results are to changes in these assumptions and how valid the forecasts are of assessment hours benchmarked against actual assessment hours as reported by the assessors. The research team use this simplified representation of the AE model to undertake the analysis that follows.

2.6 Evaluating AE Model Assumptions

As a complement to the data used in this research project, the research team make use of interview data from a previous project conducted by UTS with NATA. The interview data from the 2017 NATA project titled 'Economic Evaluation of NATA in Australia'[4], highlighted several challenges for clients with the accreditation process that may explain some of the results found in this study[5]. Concerns raised by clients were themed into the following categories:

(i) Organisational size impacts - in particular how the technical assessment process differs between small and larger firms,

(ii) Limitations in the assessment process,

(iii) Simplification of the assessment process, and

(iv) Assessment completion timeframe.

Knowing that the AE model assumes a degree of homogeneity across assessors as well as a standardised assessment process, it may not be surprising to see variations between system generated assessment hours and actual reported hours by the assessors.

For example, the interviewees provided feedback that indicated firm size influenced different sets of challenges. On the one hand, smaller firms were concerned with cost, time and limited resources to take part in ongoing assessment tasks. On the other hand, larger firms felt a certain amount of inequity due to the number of assessments that needed to be conducted across multiple laboratories distributed in different parts of Australia, thus causing the assessment process to be drawn out and conducted in a piecemeal manner. Some examples of these challenges and concerns are illustrated in Table 2.8.

[4] Agarwal, R., Bajada, C., Green, R. and Lo, E (2017). Economic Value of NATA Accreditation to Australia, University of Technology Sydney

[5] Although this data was useful but not used directly in the 2017 NATA report, it has significant merit and value in this report as the focus is on assessors and assessments.

Table 2.8. Client views on NATA assessment processes

Small Firms	Relevance to AE Model Evaluation
*'…you know enormous big companies can accommodate certain things and smaller companies can't…. everyone's still got to comply (*Calibration sector).	Although, this firm is agreeing with the need to comply, this firm expressed frustration with NATA's assumption that smaller firms can provide resources for assessments similar to larger firms.
*'we're the furthest south, we go up to Cairns, out to Tamworth and all up the eastern seaboard.......some of our labs are quite small, we find that's really quite an impost, from a financial point of view, because instead of buying a piece of glassware that might cost you $100 to get it certified and it might cost you $500. Like our pipettes, piston pipettes, again they said they had to be done by an external provider when we were still doing them every quarter ourselves, doing calibration checks (*Life Sciences sector*).*	Again, there is an assumption by NATA that smaller firms (even though geographically distributed) can resource rigorous assessment processes and costs that may increase with remoteness
Large Firms	
*' We have guys from head office they just came here for a couple of years… and their question is … why so many accreditation labs? But then they travel around Australia – to understand the needs …and they kind of understand why we have to do it that way. (*Infrastructure sector).	A larger firm expressing the geographical size of Australian assessment processes and the challenges this causes firms who have distributed premises.
*'…we need sometimes flexibility and understanding from NATA because we have projects – we have shutdowns sometimes can be for a month or a few weeks you know so we sometimes literally struggle to provide feedback in time (*Infrastructure sector).	Large companies in certain sectors need to shut down for maintenance periods and this firm is frustrated that NATA expect unrealistic deadlines to be met during operational periods.

The challenges associated with the technical assessment process consider whether accreditation actually captures the technical capability levels of the organisation, illustrated below by some member interviews from the previous NATA project. In the context of this project, the feedback suggests that the AE model does not capture the complexity of the task nor differentiates across firm capabilities:

> *I don't think audits quite capture technically the capabilities of a laboratory – they capture more the compliance.*

> *You have to dig deeper to tell one type of lab from a different type of lab. So, I wish NATA would just set the bar a little bit higher so that some labs couldn't make it*

> *Universities that have got no relevant experience in testing, they sit behind a HP calculator and say yep, that's the way it should be tested. But when it comes time for something to be physically installed, then, 99% of the time, the installation is outside the design test parameters of the relevant Australian Standard, so therefore it goes out the window. So it's no different to a car tyre was tested and says you can get 75,000km driving on roads, now all of a sudden you only get 10,000km of out of it. Why? Because hey, it was tested in a perfect test laboratory.*

The amount of resources, time and effort that are invested in the internal accreditation process is considered unnecessarily bureaucratic at times. The administrative type assessment process tends to slow down company productivity and profitability levels as highlighted by member responses below. In the context of this project, this feedback suggests the environment is such that it may be difficult for assessors to be efficient in their tasks and thus results in increased assessment time.

> *Right now we are putting internal training, then we actually we wait for the next level of NATA certification … audit and then ask NATA to verify competence of our signatories……..It's quite a long process to get NATA and organise for us a special signatory assessment which is also costly and … is finding it very costly for us.*

> *There is quite a lot of resource required to maintain a NATA level, and I would probably say that some NATA requirements seem unnecessarily onerous and some seem to change without any particular reason. So we, we would invest a lot of time to, to change to be compliant and then a few years later, a few short years later we would have to change again because somehow NATA's view on a particular thing has changed, it doesn't really make any material difference but we have to rejig our systems.*

2.6.1 Uniformity of assessment

Achieving uniformity across the assessment process is one of the challenges that needs to be taken into account when reviewing actual data. The AE model assumes that there are no differences between assessors undertaking assessment. However the reality can be vastly different from this assumption. Interviewees across the NATA member sectors highlighted the difficulties in managing different assessors' agendas and the fact that there can be a change in the audit / assessment process depending on which assessor is visiting on the day. This may explain the variability in actual assessment hours when more than one assessors is involved. The feedback below indicates that at times there is a lack of uniformity in the assessments undertaken.

> *When they come on site they look at different things every time. And you need to look at the important things or something that may have come up as an issue in a previous one, so it's trying to get the sampling plan right.*

> *I wish NATA would be stricter with everybody. I wish the entry baseline…there's not a specific difference can be made between a high performing lab and a lab that just squeaks in to meet NATA's requirements.*

> *So, it's nice to have peer review from actual peers................one can say you've got to do A and then the next one comes in 18 months later – 3 years later now and … say, no, you shouldn't be doing A, you should be doing B and it's the opposite of what the other one said.*

> *I don't think it's important that they pay for assessors but I do think it's important that they focus on getting them to do the same type of assessment.*

> *There are differences in how NATA interpret some of the requirements and there's variation between different assessors, and things like that. So that's a challenge*

> *And to a degree it varies on which NATA person you get and certainly one of the issues we've had is that you know, one NATA person will interpret something differently to another Nata person, and there's some inconsistency there.*

2.6.2 Suggestions to improve the AE Model

The auditing and assessment process could be streamlined by ensuring adoption of a common standard by assessors for specific assessments to minimise heterogeneity between assessors. This would also assist clients understand what is expected and what the role of the assessor is in the context of evaluating a standard, activity and service. This would not only address some of the feedback (as noted below) but also improve the reliability of the forecast produced by the AE model.

> *I think uniformity in technical assessments, in other words, the type of information gathered, the method of gathering...So if NATA could simplify the method of auditing…..*

> *If they had better assessment tools, like a software…to capture this type of information. So whether that's loadable to your phone, or your iPad, or your tablet and you go along and you check the boxes…...will help the labs also know a consistency of how the audits going to be.*

> *So every 18 months we technically have three audits – so three packs of auditors do come to us and we have to have three different audits for that just for one lab I am talking about just one lab, and what would be good in the future is maybe somehow to combine it you know what I mean?*

The research team suggests that:

1	A process for ensuring consistency amongst assessors be implemented so as to improve the accuracy of the AE model.

2.6.3 Factor Rounding

The AE Model uses two adjustment factor scores to calculate the system generated assessment hours. The factor scores are adjusted using a specific rounding approach (referred to here as NATA rounding). This is different to the standard mathematical rounding where increments of 0.5 or more adjust up a score to the nearest whole number. The research team examined the sensitivity of the system generated assessment hours to using different rounding techniques for the factor score, namely a comparison between the use of NATA rounding, mathematical rounding and the option of no rounding at all. Table 2.8 illustrates the different mean values for system generated assessment hours for technical effort (excluding QMS) under the different rounding regimes.

Table 2.8 – Sensibility to different factor rounding

Variable	Mean Assessment Hours	Sample Size
Total_effort_1	11.90	100
Total_effort_2	14.01	100
Total_effort_3	13.54	100

Note 1: *Technical effort excludes QMS*

Note 2: *variable description:*

Total_effort_1 – system generated total technical effort (NATA rounding / simplified)

Total_effort_2 - system generated total technical effort (no rounding)

Total_effort_3 - system generated total technical effort (mathematical rounding)

The findings in Table 2.8 illustrate that the system generated hours for technical effort using NATA rounding is much smaller (11.90 hours) compared to no rounding (14.01 hours) and using mathematical rounding (13.54 hours). Table 2.9 shows statistical comparisons of the results from using the different rounding methods.

Table 2.9 – Comparing different rounding techniques (t-tests)

t-tests	p-value	Significantly different?
Total_effort_1 vs Total_effort_2	0.0000	Yes
Total_effort_1 vs Total_effort_3	0.0000	Yes
Total_effort_2 vs Total_effort_3	0.0000	Yes

Note: *variable description:*

Total_effort_1 - total technical effort (NATA rounding / simplified)

Total_effort_2 - total technical effort (no rounding)

Total_effort_3 - total technical effort (mathematical rounding)

The t-tests comparing the mean assessment hours using NATA rounding with no rounding is found to be statistically different. A similar result is found when comparing the mean assessment hours using NATA rounding and standard mathematical rounding. The two rounding methods are found to be statistically different. There is also a statistical difference between the use of no rounding with mathematical rounding. This suggests that the use of factors in the AE model produce results that are sensitive to the rounding approaches and so caution is recommended when using any rounding approach that may significantly produce different results.

2.6.4 Suggestions to improve AE Model

Alternative approaches to modify or improve the current AE model is to consider the rounding techniques for factor values. As part of the team's analysis, comparisons between the use of the NATA rounding methodology and the alternatives to mathematical rounding or no rounding were investigated. The results of such comparisons of different methods of rounding are presented in Table 2.10.

Table 2.10 Difference between system generated and actual hours for a full sample

	Mean	Sample size
NATA round dif1	-9.188356	73
NO Round dif2	-7.186558	73
Math Round dif3	-7.664384	73

Note: *NATA Round dif1 – difference between system generated and actual hours calculated by using the simplified expression of the AE model in Section 2.5. No Round dif2 - difference between system generated and actual hours calculated by using no rounding. Math Round dif3 - difference between system generated and actual hours calculated by using a proper mathematical rounding.*

The research team repeated the use of rounding alternatives at the field level. These are presented in Table 2.11.

Table 2.11. Difference between system generated and actual hours, by field

	Mean			
Field	**NATA Round dif1**	**No Round dif2**	**Math Round dif3**	**Sample size**
Inspection	-7.73	-3.69	-3.18	15
Medical	-10.13	-8.46	-8.26	23
Mechanical	-9.19	-7.85	-9.19	35

***Note1**: NATA Round dif1 - difference between system generated and actual hours calculated by using simplified NATA formula. NO Round dif2 - difference between system generated and actual hours calculated by using no rounding. Math Round dif3 - difference between system generated and actual hours calculated by using a proper mathematical rounding*

As Tables 2.10 and 2.11 illustrate, applying no rounding and mathematical rounding for calculating factor values (while keeping the variables the same) reduces the difference between system generated and actual hours for the full sample and for almost every field. Hence, the forecasting capacity of the AE model can be further improved by using the traditional mathematical rounding technique. The research team revisit the effects of rounding again in the next section to explore the impact of changes in other assumptions combined with these different methods of factor rounding.

2.6.5 Technique Type Weight Assumptions

The AE model makes several assumptions on different technique types. Three weights are used to reflect the 'level of difficulty' associated with a given assessment. The weights used are 0.5, 0.75 and 1 for Classic (simple), Mixed and Instrument (Complex) respectively. Given the relatively small sample size, the research team was limited to validating only the weights across these three fields (Mechanical, Medical and Inspection). The available actual assessment hour's data was then used to compare the relative weightings of Classic and Mixed assessment hours relative to Instrument assessment hours from actual assessment hours data. To do this, the research team constructed ratios of different technique types for different hours (system generated and actual hours). Instrument technique type value is used as the base value and given a value equal to 1. The results are reported in Table 2.12.

Table 2.12 - Weights assigned to different technique type

Technique Type	Classic	Mixed	Instrument
Sys_gen	1.44	1.45	1.00
AHrs_3	1.50	1.17	1.00
AHrs_4	1.55	1.29	1.00

***Note:** variable description: Sys_gen - system generated hours: total technical effort only plus 2/3 QMS is included, AHrs_3 - actual hours with NCCA but without PRE hours. AHrs_4 - actual hours without NCCA and without PRE hours.*

The results in Table 2.12 illustrate that ratio values for various technique types (system generated less QMS and actual hours less PRE hours (with and without NCCA hours) are different from the 0.5, 0.75 and 1 assumed in the AE model. Classic technique type has the highest ratio while instrument technique type has the lowest for system generated and actual hours.

2.6.6 Suggestions to improve AE Model

This finding suggests that the weights for technique type should be reviewed. The types of activities undertaken by the assessors classified as either classic, mixed or instrumental should be reviewed to reflect the actual hours / complexity being reported by the assessors. Without knowing the specific details of the work involved across each of these services, it is very difficult for the research team to make any inference on the relative weighting between classic, mixed and instrumental technique types.

The UTS research team suggests the following:

2	NATA should review the definition of classic, mixed and instrumental to ensure alignment with the current activities and services
3	Technique type weights be re-evaluated in accordance to the type of assessment activities being undertaken and a judgement made based on actual assessment hours being reported.
4	NATA should consider activity or field based weight for technique type (classic, mixed and instrumental)

2.6.7 Assessment hours by field

In Table A1 of Appendix 3, each Field is assumed a given number of hours for RES and SRV (for STE, CFE and CFQ). The research team used actual reported assessment hours to determine whether the actual data reflects the constant hours assumed in the AE model as presented in Table A1 of Appendix 3. Limited to a sample dataset covering the fields of

Inspection, Mechanical and Medical, the analysis is therefore constrained only to these three areas. It may not be possible to extrapolate these findings to the other fields as there are at times, significant variations in the actual hours of assessment reported just in these fields alone. This may be due in some part by the uniformity of assessment issues described earlier.

The simplified AE model presented in Section 2.5 suggest that this task can be simplified by only having to focus on that value for STE_RES. The research team produced an estimate of STE_RES (STE_RES_est) based on actual data (PRE hours were subtracted from total actual hours excluding NCCA reported hours). The results are presented in Table 2.13.

Table 2.13 – Estimating STE_RES – Full sample

Variable	Sample size	Mean	Ratio
STE_RES	73	12.45	1.00
STE_RES_est	73	22.20	1.78

Note: *variable description, STE_RES - value from NATA field codes (Appendix 3). STE_RES_est - estimated value calculated based on actual hours (without NCCA and without PRE hours) and factor values. Full sample (STE_RES value is used as a base and equal = 1).*

The first row of Table 2.13 is the average value of STE_RES using the NATA Field Codes Table A1 (Appendix 3) based on the full sample. The research team used sites/observations that represent three different fields (Mechanical, Medical and Inspection) and each field has different STE_RES value from Table A1. STE_RES value of 12.45 is the weighted average value (for full sample) based on corresponding STE_RES value of each field and number of observations from this field. This is shown in the first row of Table 2.13. The estimated assessment hours or STE_RES is reported in the second row of Table 2.13. The results illustrate that the average estimated value of STE_RES is greater than the STE_RES value derived from Table A1 (Appendix 3). In fact, the estimated value is 1.78 times more than based on Table A1.

Next, the analysis was repeated across three fields using actual assessment hours and data from Table A1 (Appendix 3). The results of this analysis are reported in Tables 2.14 (a-c). For each of these fields, the estimated value of STE_RES is greater than that reported in Table A1 of Appendix 3. However the scale factor is different across each of these fields. In Table 2.14a, the estimated values of STE_RES for Mechanical is 1.79 times higher than the assumed number of hours in the AE model, while for Medical it is 3.53 times higher than the assumed value of hours in the AE model (see Table 2.14b). For Inspection (in Table 2.14c) the estimated value if 1.39 times the assumed number of hours. The t-tests (see Appendix 7) demonstrate that there is significant difference between average values of STE_RES and STE_RES_est for different fields. The results suggest that of the 3 fields, Medical has the largest margin of error. A comparison of STE_RES_est with STE_RES by field is given in the following three tables (Tables 2.14a, 2.14b and 2.14c).

Table 2.14a - Estimating STE_RES – by Mechanical field

Variable	Sample size	Mean	Ratio
STE_RES	35	15.00	1.00
STE_RES_est	35	26.80	1.79

Note 1: Field 13 = Mechanical (STE_RES value is used as a base and equal = 1).

Note 2: variable description:

STE_RES - value from NATA table 1

STE_RES_est - estimated value calculated based on actual hours (without NCCA and without PRE hours) and factor values

Table 2.14b – Estimating STE_RES – by Medical field

Variable	Sample size	Mean	Ratio
STE_RES	23	3.00	1.00
STE_RES_est	23	10.58	3.53

Note 1: Field 10 = Medical (STE_RES value is used as a base and equal = 1).

Note 2: variable description:

STE_RES - value from NATA table 1

STE_RES_est - estimated value calculated based on actual hours (without NCCA and without PRE hours) and factor values

Table 2.14c - Estimating STE_RES – by Inspection field

Variable	Sample size	Mean	Ratio
STE_RES	15	21.00	1.00
STE_RES_est	15	29.28	1.39

Note 1: Field INSPECTION = Inspection (STE_RES value is used as a base and equal = 1).

Note 2: variable description:

STE_RES - value from NATA table 1

STE_RES_est - estimated value calculated based on actual hours (without NCCA and without PRE hours) and factor values

2.6.8 Suggestions to improve AE Model

For this reason the research team do not make any specific suggestions on the weightings between these different technique types, rather changes are considered to the constant values in Table 1A (Appendix 3) as a way to improve the assessment hours forecast of the

AE model (while maintaining the factor weight assumptions). In estimating the revised values in Table 1A, the rounding techniques are taken into account and discussed in Section 2.6.2. Given that equation (2) is used to generate the forecast of assessment hours, the research focus is on the changing value of STE_RES and the factor scores[6]. By attempting to minimise difference between system generated hours (total technical effort excluding QMS effort) and actual hours (total excluding PRE and NCCA hours) the research team produced an estimate for values in Table 1A for each of the three fields that ensure the AE model produces the best forecast of assessment hours.

Once the optimal STE_RES values are estimated for each field, the research team re-estimated the model using the full sample set of observations and calculated the difference between actual and the new system generated assessment hours. Based on the three fields available in the sample set, the relevant rows in NATA Table 1A were revised.

The steps involved in producing the revised values in Table 1A required calculating the STE_RES value for each of the three fields. In Tables 2.15, 2.16 and 2.17, the optimum STE_RES value is shown for Mechanical, Medical and Inspection fields respectively. As Table 2.15 illustrates, the minimum difference between system generated and actual assessment hours (=0.34 hours) occurs when the STE_RES value is equal 26 hours. Currently the value for STE_RES for Mechanical is 15.

Table 2.15. Field_13 - Mechanical

	STE_RES=15	STE_RES=20	STE_RES=24	STE_RES=25	STE_RES=26	STE_RES=27	STE_RES=30	Sample size
Difference between new system generated and actual hours (mean value)	-9.19	-4.86	-1.39	-0.53	0.34	1.21	3.81	35

Table 2.16 illustrates, the minimum difference between system generated and actual assessment hours occurs when the STE_RES value is equal 10. Currently the value for STE_RES for Medical is 3.

[6] We do recalculate the factor scores, instead we change the score only by changing the rounding methodology.

Table 2.16. Field_10 - Medical

Stats	STE_RES=1	STE_RES=2	STE_RES=3	STE_RES=5	STE_RES=7	STE_RES=10	STE_RES=15	Sample size
difference (mean)	-13.06	-11.59	-10.13	-7.20	-4.28	0.12	7.43	23

Finally, in Table 2.17 the minimum difference between system generated and actual assessment hours occurs when STE_RES value is between 29 and 30 hours. Currently the value for STE_RES for Inspection is 21.

Table 2.17. Field: Inspection

Stats	STE_RES=21	STE_RES=25	STE_RES=26	STE_RES=28	STE_RES=29	STE_RES=30	STE_RES=35	Sample size
difference (mean)	-7.73	-4.11	-3.21	-1.39	-0.49	0.42	4.94	15

A summary of the results in the previous three tables (Tables 2.15, 2.16 and 2.17) is shown in Table 2.18. The results in the table show difference between system generated and actual assessment hours using both the current values and the revised values for the three fields in Table 1A. The forecast error significantly reduces with these revised values for Table 1A.

Table 2.18. Summary of differences by field

	Mean		
Field	Actual dif_1	Updated dif1_est	Sample size
Inspection	-7.73	-0.49	15
Medical	-10.13	0.12	23
Mechanical	-9.19	0.34	35

Note. Actual dif1 – difference between system generated and actual hours calculated by using simplified NATA formula. Updated dif1_est - difference between system generated and actual hours calculated by using STE_RES values that minimise difference for each field

Using these estimated values for STE_RES it is possible to update the values in Table 1A (Appendix 3) for each of the three fields – inspection, medical and mechanical. The revised values are shown in Table 2.19.

Table 2.19. NATA Table 1 with new estimates of STE_RES (3 fields only)

Field	STE - RES	CFE –RES	STE - SRV	CFE - SRV	CFQ RES	CFQ SRV	CFS
INSP	29	14.5	9.7	4.8	6.9	13.8	Prod
MED	10	5.0	3.3	1.7	3.3	6.7	Tec
MECH	26	13.0	8.7	4.3	17.3	34.7	Prod

Using the results of Table 2.18, the forecast error of the AE model is calculated before and after the change. The results are shown in Table 2.20.

Table 2.20. Difference between system generated and actual hours for a full sample

Variable	Sample size	Mean
Actual dif1 (using current Table 1A in Appendix 3)	73	-9.19
Updated dif1_est (using revised Table 1A in Appendix 3)	73	0.10

***Note**. Actual dif1 – difference between system generated and actual hours calculated by using simplified NATA formula. Updated dif1_est - difference between system generated and actual hours calculated by using STE_RES values that minimise difference for each field*

As Table 2.20 highlights, these revised values significantly reduce the difference between system generated and actual assessment for the full sample from -9.19 to 0.1 hours. However, it is important to note that these estimations are based on the sample used in this study. Increasing the sample size to include other clients may produce different estimated values of STE_RES and therefore the values in Table 1A in Appendix 3. To obtain the right approximations of STE_RES (and therefore the values in NATA field codes of Table 1A), it is desirable to increase the sample size to the population database.

2.6.9 Evaluation of QMS effort

The next step in the analysis is to include an estimate of QMS effort. From Figure 2.1 PRE hours is approximately 1/3 QMS hours.

There is a large variance in values across different fields, services and activities. For example, average QMS effort value for the inspection field is 2 times more than average for a full sample while average QMS effort value for medical field (field 10) is almost 2 times less than average for a full sample (see first table in Appendix 6 for details).

The QMS effort values are required in the next stage of the analysis in order to calculate new system generated hours based on new STE_RES values/estimates and QMS effort values. Table 2.21 compares the current system generated hours (including QMS effort) with estimated values of system generated hours (based on different rounding approaches) and actual hours for the full sample. The analysis is replicated by field, service and activity in the subsequent tables.

Table 2.21. System generated and actual hours for a full sample.

Variable	NATA Sys_gen	NATA Round Sys_gen_1	No Round Sys_gen_2	Math Round Sys_gen_3	Actual Hours (excluding NCCA)
Mean	33.74	32.14	31.75	31.85	32.04
Sample size	73	73	73	73	73

Note. *Variable description: Sys_gen – current system generated hours (including QMS effort). Sys_gen_1 – calculated system generated hours by using NATA formula and including QMS effort. Sys_gen_2 – calculated system generated hours by using no rounding and including QMS effort. Sys_gen_3 – calculated system generated hours by using a proper mathematical rounding and including QMS effort. Actual – actual hours (without NCCA).*

The current system generated hours (variable: Sys_gen) and calculated system generated hours (variable: Sys_gen_1) are close to actual hours for a full sample. The analysis was also replicated by field, service and activity and the results are provided in Tables 2.22 – 2.24.

Table 2.22. System generated and actual hours (average values), by field

Field	NATA Sys_gen	NATA Round Sys_gen_1	No Round Sys_gen_2	Math Round Sys_gen_3	Actual Hours (excluding NCCA)	Sample size
Inspection	37.02	51.29	51.39	51.97	51.78	15
Medical	36.17	20.42	19.92	20.39	20.30	23
Mechanical	30.74	31.63	31.09	30.77	31.29	35

Table 2.23. System generated and actual hours (average values), by services

Service	NATA Sys_gen	NATA Round Sys_gen_1	Actual Hours (excluding NCCA)	Sample size
1	37.02	51.29	51.78	15
2	31.91	34.14	30.43	22
3	36.17	20.42	20.30	23
4	28.77	27.40	32.75	13

Table 2.24. System generated and actual hours (average values), by activity

Activity	NATA Sys_gen	NATA Round Sys_gen_1	Actual Hours (excluding NCCA)	Sample size
1	36.17	20.42	20.30	23
2	31.91	34.14	30.43	22
3	33.19	40.20	42.95	28

***Note**. Variable description: Sys_gen – current system generated hours (including QMS effort). Sys_gen_1 – calculated system generated hours by using NATA formula and including QMS effort. Sys_gen_2 – calculated system generated hours by using no rounding and including QMS effort. Sys_gen_3 – calculated system generated hours by using a proper mathematical rounding and including QMS effort. Actual – actual hours (without NCCA).*

The results highlight that there is a variance in system generated hours across different fields, services and activities. The preferred approximation of actual hours (and the minimum difference between system generated and actual hours) is achieved by variable Sys_gen_1 and taking into account QMS effort.

2.6.10 Suggestions for improvement

The UTS research team suggests the following:

5	The constant values in Table 1A should be re-calibrated to achieve better forecasts of assessment hours required across different fields.
6	QMS values should be re-evaluated by field using all available data on PRE assessment hours.
7	NATA continue using its rounding methodology when the optimal values for Table 1A (as in point i above) is adopted in the AE model.

3 Recommendations

As has been recognised throughout the report findings presented in Section 2, the current AE Model performs adequately for the purpose of forecasting assessment hours for accreditation purposes. However, based on the analysis outlined in Section 2, the UTS research team has proposed a number of recommendations for enhancing the forecast capacity of the AE model.

It is recommended that:

1	**A process for ensuring consistency amongst assessors be implemented so as to improve the accuracy of the AE model.**

Feedback from NATA client interviews during the NATA project Economic Value of NATA Accreditation in Australia suggested there are variations in activities between assessors in undertaking similar assessment tasks. Without a set process or procedural document, such difference between assessor’s approaches are natural and expected. Ensuring consistency would not only assist clients understand what is expected in an assessment but also the role of the assessor is in the assessment process. This would not only address some of the feedback received from clients on this issue but also improve the reliability of the forecast of assessment hours produced by the AE model.

2	**NATA should review the definition of classic, mixed and instrumental technique type to ensure alignment with the current activities and services.**

Such a review requires an examination of each activity based on complexity (expertise required by the assessor) and the actual hours necessary to undertake the assessment. For example, an assessment may be complex in nature but the assessor is assisted with technology to undertake the task, hence less hours are needed in the assessment. Classic, mixed and instrumental technique types should be defined not by scientific complexity alone, but by the intensity of the assessment activity involved. NATA should review the definition of technique types: classic, mixed and instrumental, as they currently apply to activities and services.

3	**Technique type weights be re-evaluated in accordance to the type of assessment activities being undertaken and a judgement decision made based on actual assessment hours being reported.**

The technique types should be reviewed along with a sample of assessors to determine appropriate classification of classic, mixed and instrumental technique type. This review should reflect the experiences of the assessors noting the reported hours / complexity of the tasks.

4	**NATA should consider activity or field based weight for technique type (classic, mixed and instrumental).**

It is possible that there are similarities in the technique adjustment factors across various types of activities, meaning that activities can be grouped and the same technique adjustment factors can be used. However, without the entire data population set, this is not possible to determine at this stage.

5	**The constant values in Table 1A should be re-calibrated to achieve better forecasts of assessment hours required across different fields**

Any constant values in the AE model should be re-evaluated on a time cycle or better still be constructed internally by the model through a process of adaptive estimation method that recalibrates the parameters of the model based on the regular input of data coming from the assessors.

6	**QMS values should be re-evaluated by field using all available data on PRE assessment hours.**

The AE model currently uses global values for QMS, but the report analysis suggests QMS should be treated differently across fields. It is recommended QMS is treated differently across the fields.

7 NATA continue using its rounding methodology when the optimal values for Table 1A (as in point 1 above) is adopted in the AE model.

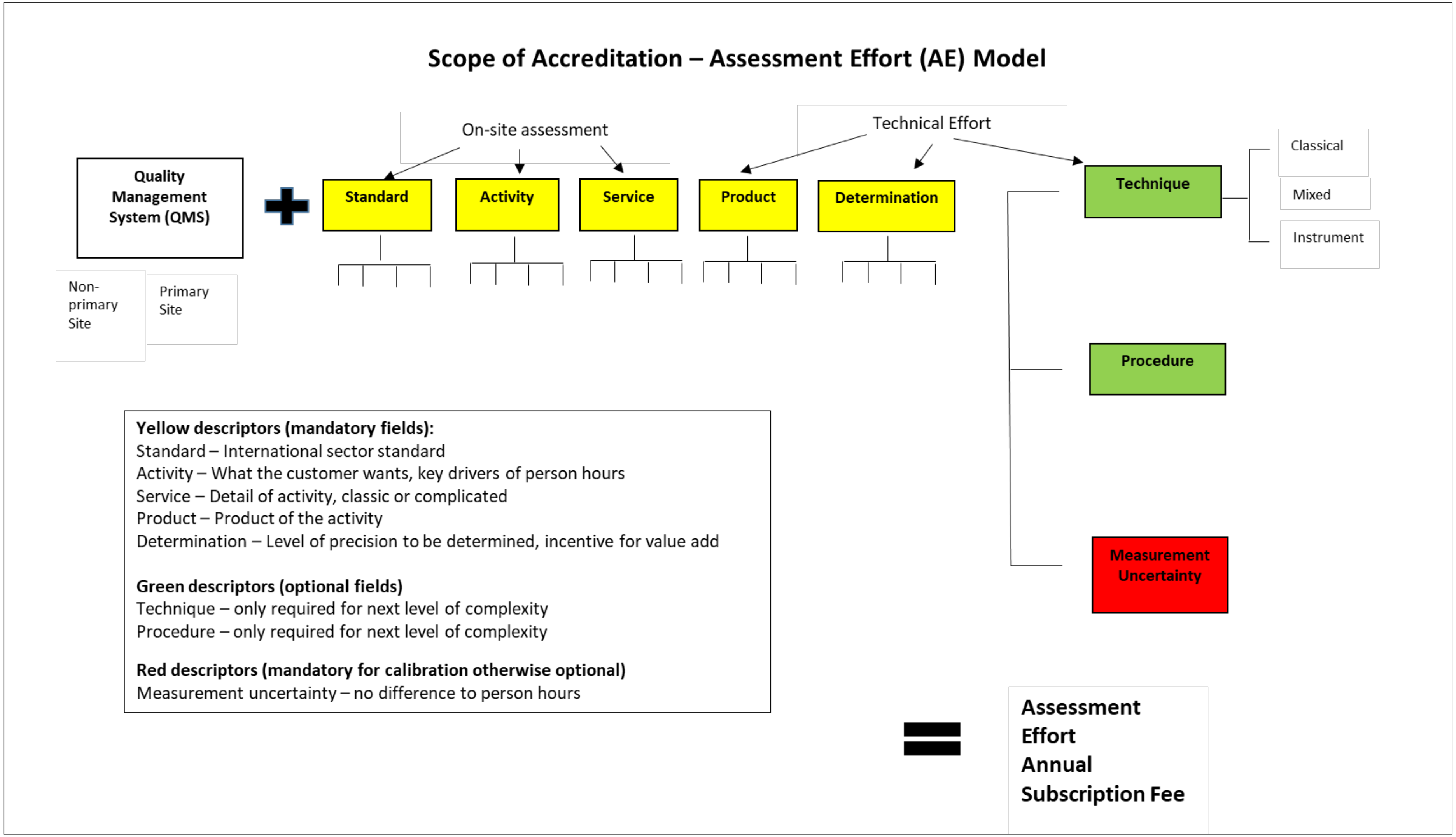
The AE model uses a table of constant values by a Field classification that serve as scalars in the AE model’s forecasts of assessment hours. The results in Section 2 suggest that there is a need to make some adjustments to these constant values. Given that the sample set used in this analysis only has three fields, the research team could only make several suggestions for these fields. These adjustments were derived from actual assessment hours (excluding PRE and NCCA hours reported by assessors). Using the balance of the actual assessment hour’s dataset, the revised constant values for the field classifications can also be calculated.

4 Conclusion

Overall, the report findings illustrate the AE model is suitable to forecast assessment hours for the purposes of allocating labour resources for accreditation purposes and providing. improvements to the labour assessment hours estimate to enable NATA to better allocate scarce resources amongst their various clients knowing that the assessment process is primarily a volunteer based scheme However, forecast analysis specifically conducted by the NATA team at the service and activity levels can be improved by updating assumptions and these suggestions are contained in the recommendations of this report.

Appendices

Appendix 1 – Current Scope of Accreditation



Appendix 2 – Reassessment Effort (RES)

Reassessment effort = QMS effort res + [{(STE res) + (CFT)} * PD]

QMS effort res = **the effort to review QMS records and to follow up on minor non-conformances from the previous visit.**

The value is applied uniformly within the scheme and at this time is set as 6 hours for a primary site, i.e. Primary QMS for a RES.

For corporate accreditations that are not the primary site at this stage this time is set as 3 hours, i.e. Discounted QMS for a RES.

STE res = the effort to review at least one Determination for each competency. This value was determined for our old field/program descriptors and thus has been applied to different Standard/Activity combinations.

See Table 1

CFT = the additional effort to review more than one determination/product or technique where the size of the scope requires additional sampling.

CFT = [CFE res x ($^{1}/_{\text{CFQ res}}$ x Count of CFQ within the service)]

If ($^{1}/_{\text{CFQ res}}$ x Count of CFQ within the service) is ≤1 then no additional factor is required.

If ($^{1}/_{\text{CFQ res}}$ x Count of CFQ within the service) is >1 and ≤2 then the value 1 is applied to the above equation)

If ($^{1}/_{\text{CFQ res}}$ x Count of CFQ within the service) is 2 and ≤3 then the value 2 is applied to the above equation) etc.

Note: The Count of CFQ is based on the item defined in the CFS res field for the service, i.e. determinant, product, technique.

PD = the reduction value applicable for each service based on the composition of technique types.

The value applied will be 1, 0.75 or 0.5 as listed on page 3

Appendix 3 – Formulas for determining assessment effort (PRE, UA and POST Time)

The assessment effort should relate back to competencies and be determined per job type per assessment cycle per site.

In each service the following parameters are defined for each scheduled job type:

STE = base rate to which various modifiers can be applied.

CFE = Complexity factor hour value used when scaling the effort up. (At this time this has been designated as ½ of STE for all areas).

CFS = what item the complexity factor count is based on (determinant, product or technique).

CFQ = the rate at which complexity factor scaling is applied, e.g. every 5 determinations.

In each scheme the following parameters are also defined:

Primary QMS = value(s) defined for each scheduled job type (used for stand-alone sites and primary sites in a corporate accreditation).

Discounted QMS = value(s) defined for each scheduled job type (used for non-primary sites in a corporate accreditation).

Percentage discount PD =

A percentage discount can be applied uniformly within each service in a scheme to recognise that the total technical effort for a service is affected by the composition of technique types recorded within each service, i.e. simple (Classical) versus complex (Instrumental).

0% reduction is applied when all techniques in the scope for that service are complex, i.e. of the type Instrumental. The value for PD is 1.

25% reduction is applied when there is a combination of simple and complex, i.e. of the type Classical and Instrumental. The value for PD is 0.75.

50% reduction is applied when all techniques in the scope for that service are simple, i.e. of the type Classical. The value for PD is 0.5.

Table A1 – Field codes

Field	STE - RES	CFE -RES	STE - SRV	CFE - SRV	CFQ RES	CFQ SRV	CFS
INSP	21	10.5	7	3.5	5	10	Prod
R&D	30	15	10	5	1	2	Det
IT	30	15	10	5	2	4	Det
PAT	18	9	6	3	3	6	Prod
MED	3	1.5	1	0.5	1	2	Tec
FOR	24	12	8	4	15	30	Det
MECH	15	7.5	5	2.5	10	20	Prod
PTSP	30	15	10	5	3	6	Prod
CAL	9	4.5	3	1.5	10	20	Prod
RMP	21	10.5	7	3.5	2	4	Prod
GLP	30	15			3		Prod
CMT	3	1.5	1	0.5	2	4	Det
BIO	12	6	4	2	3	6	Tec
NDT	15	7.5	5	2.5	4	8	Det
VET	9	4.5	3	1.5	5	10	Tec
CHEM	9	4.5	3	1.5	10	20	Tec
Sleep	6	3	2	1	5	10	Det
RANZCR	6	3	2	1	5	10	Det
DIAS	6	3	2	1	5	10	Det

Appendix 4 – Surveillance Effort (SRV)

Surveillance effort = QMS effort srv + [{(STE srv) + (CFT)} * PD]

QMS effort srv = **the effort to review the QMS in full and to follow up on minor non-conformances from the previous visit.**

The value is applied uniformly within the scheme and at this time is set as 4 hours on site (12 hours Pre, UA and Post) for a primary site, i.e. Primary QMS for a SRV.

For corporate accreditations that are not the primary site at this stage this time is set as 2 hours on site (6 hours Pre, UA and Post), i.e. Discounted QMS for a SRV.

STE srv = the effort to conduct a vertical audit for each competency. STE srv is typically half the STE res effort.

Refer to Table 1 in Appendix 2 for values.

CFT = the additional effort to conduct more than 1 vertical audit for each competency where the size of the scope requires additional sampling.

CFT = [CFE srv x (${}^{1}/_{\text{CFQ srv}}$ x Count of CFQ within the service)]

If (${}^{1}/_{\text{CFQ srv}}$ x Count of CFQ within the service) is ≤1 then no additional factor is required.

If (${}^{1}/_{\text{CFQ srv}}$ x Count of CFQ within the service) is >1 and ≤2 then the value 1 is applied to the above equation)

If (${}^{1}/_{\text{CFQ srv}}$ x Count of CFQ within the service) is 2 and ≤3 then the value 2 is applied to the above equation) etc.

Note: The Count of CFQ is based on the item defined in the CFS srv field for the service, i.e. determinant, product, technique.

Appendix 5 – Online Effort (OLN)

On-line assessment effort = QMS effort oln + STE oln

QMS effort oln = **the effort to review QMS activities and major C's and required technical elements (i.e. new staff, equipment changes, method changes, QAP).**

The value is applied uniformly within the scheme and at this time is set as 8 hours for a primary site, i.e. Primary QMS for an OLN.

Note: OLNs for clients are only done under the auspices of the primary site and thus there is no need to utilise the discounted OLN QMS value.

STE oln = the effort to review QMS activities and major C's and required technical elements (i.e. new staff, equipment changes, method changes, QAP) for all sites.

The effort required for an OLN is affected by the number of sites/departments covered in the job (taken out against the primary site). As such 4 services have been created to add value to the OLN's as required.

STE oln values:

Corporate OLN class 1	<5 sites/hospital departments	0
Corporate OLN class 2	5 - 10 sites/hospital departments	8
Corporate OLN class 3	11 - 20 sites/hospital departments	16
Corporate OLN class 4	>20 sites/hospital departments	24

Appendix 6 – QMS Effort

The following tables demonstrate distribution of QMS effort by field, service and activity.

QMS effort, by field

Field	Mean	Sample size
INSPECTION	25.03	15
10	5.78	23
13	9.10	35
Total	11.33	73

QMS effort, by service

Service	Mean	Sample size
1	25.03	15
2	9.32	22
3	5.78	23
4	8.73	13

QMS effort, by activity

Activity	Mean	Sample size
1	5.78	23
2	9.32	22
3	17.46	28

Appendix 7 – RES & SRV Box Whisker Plots

Table 7A T-Tests

Field 13 - Mechanical - STE_RES vs STE_Res_est

	Mean	p-value	Sample size	Significantly different
STE_RES	15.00	0.0000	35	Yes
STE_Res_est	26.80		35	

Field 10 - Medical - STE_RES vs STE_Res_est

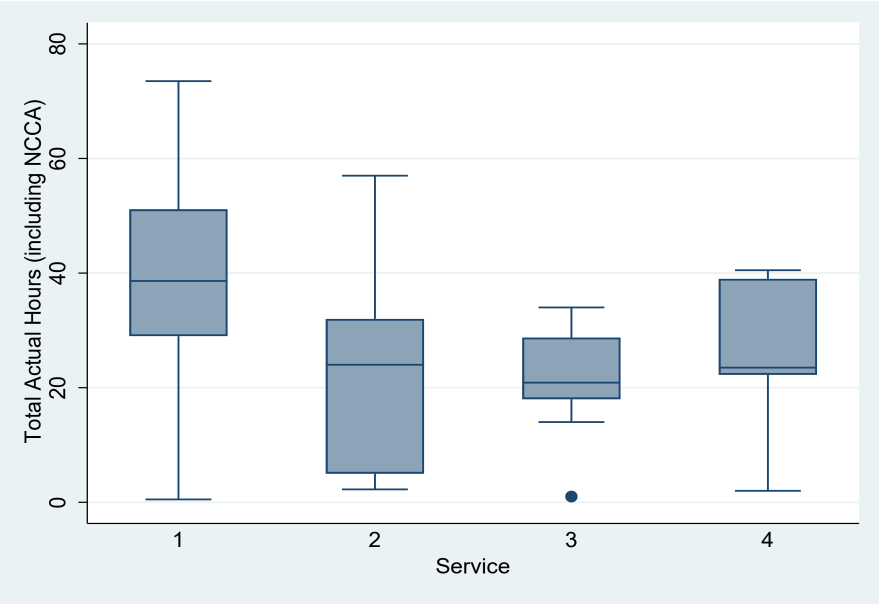
	Mean	p-value	Sample size	Significantly different
STE_RES	3.00	0.0000	23	Yes
STE_Res_est	10.58		23	

Field - Inspection - STE_RES vs STE_Res_est

	Mean	p-value	Sample size	Significantly different
STE_RES	21.00	0.0031	15	Yes
STE_Res_est	29.28		15	

Boxplots. RES component.

By service



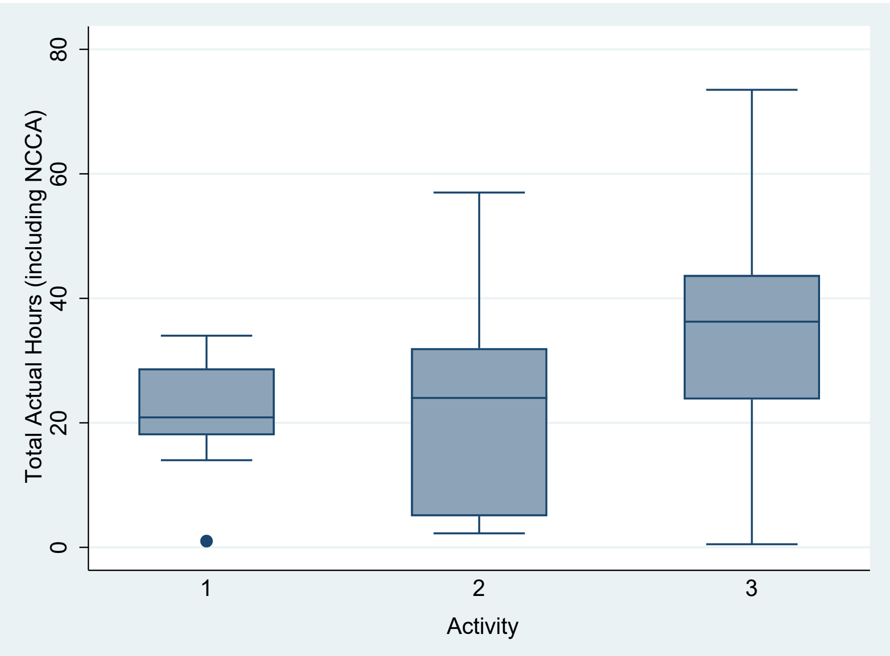
Service 1: Diesel engines for use in hazardous atmospheres - Overhaul and maintenance inspection

Service 2: Integrity Evaluation of air control equipment

Service 3: Immunopathology - Procedures related to the collection, processing, storage and issue of human haemopoietin progenitor cells

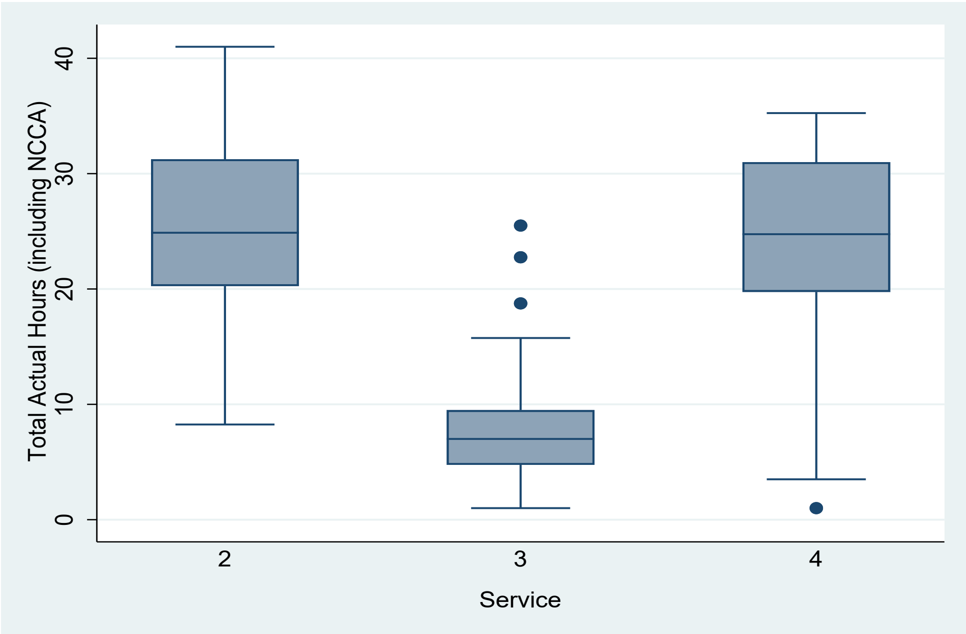
Service 4: Mechanical performance evaluation of components for building envelopes, framing and interior lining

By activity



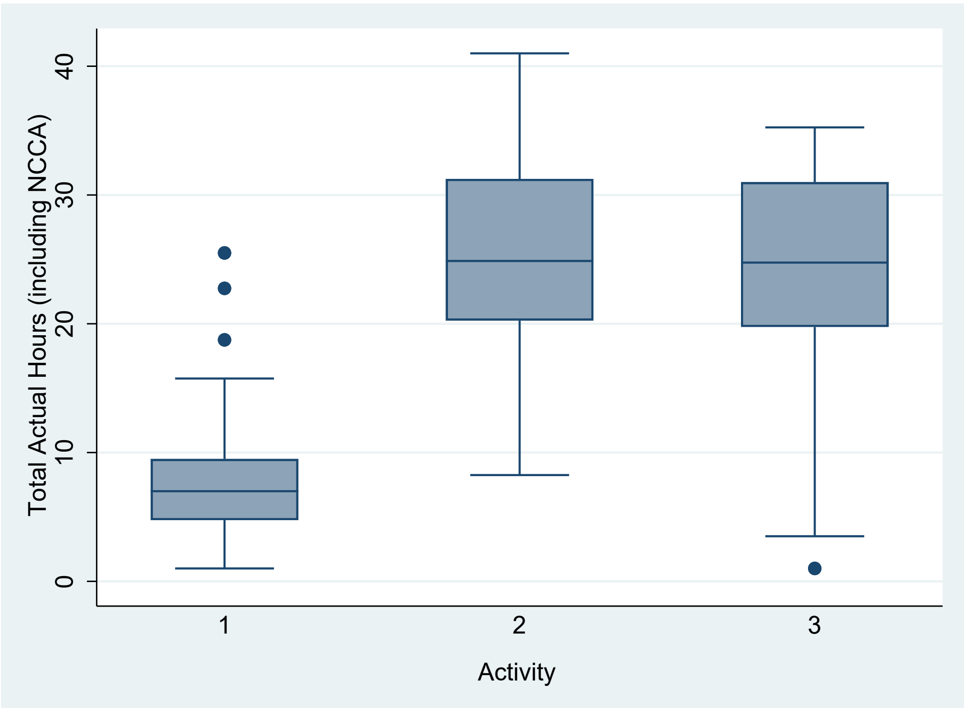
Boxplots. SRV component.

By service



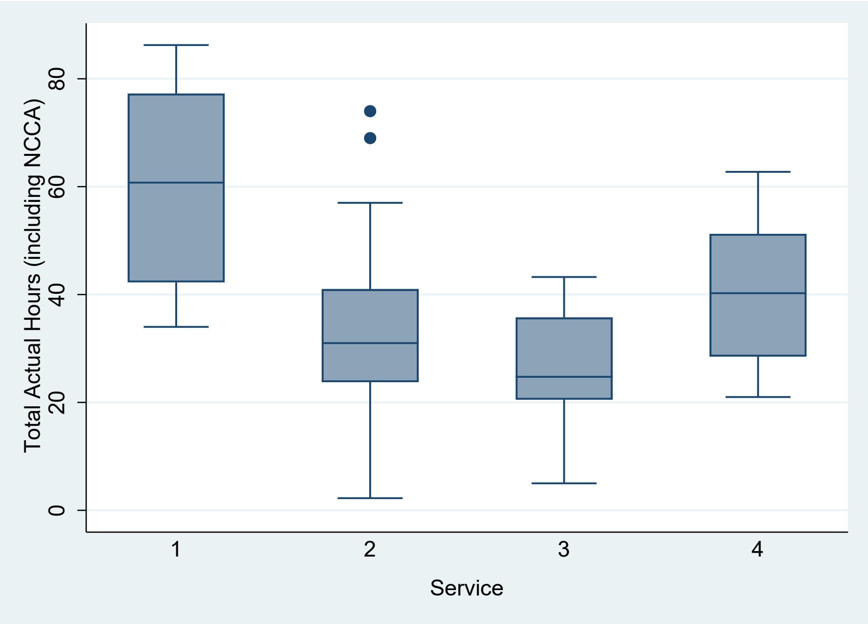
Service 2: Integrity Evaluation of air control equipment
Service 3: Immunopathology - Procedures related to the collection, processing, storage and issue of human haemopoietin progenitor cells
Service 4: Mechanical performance evaluation of components for building envelopes, framing and interior lining

By Activity



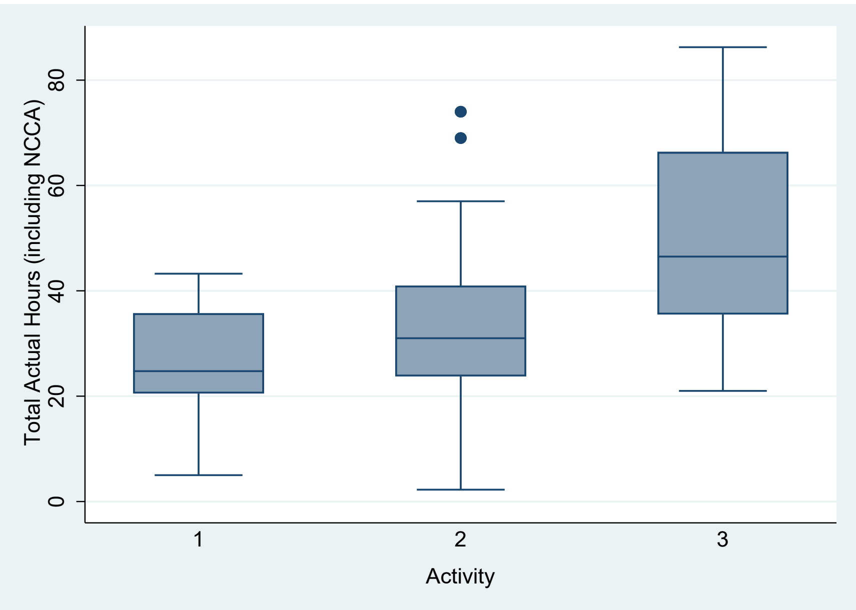
Boxplots. Total hours

By service



Service 1: Diesel engines for use in hazardous atmospheres - Overhaul and maintenance inspection
Service 2: Integrity Evaluation of air control equipment
Service 3: Immunopathology - Procedures related to the collection, processing, storage and issue of human haemopoietin progenitor cells
Service 4: Mechanical performance evaluation of components for building envelopes, framing and interior lining

By Activity



Appendix 8 – Other results with NCCA

Table 8A – Comparing System generated and Actual Assessment Hours (including NCCA)

Variable	Mean	p-value	Sample size	Significantly different
Total Actual Hours (PRE, UA, POST and NCCA)	38.25	0.0305	73	Yes
TotalCycleEffort (System generated)	33.36		73	

Note: variable description:

TotalCycleEffort = total technical effort plus QMS effort (=system generated hours)

Appendix 9 – Results with and without NCCA

Results excluding NCCA:

Technique type	Mean	Sample size
Classic	29.71	45
Mixed	20.76	18
Instrument	17.63	12

Notes: (i) RES component only; (ii) Actual hours without NCCA

Technique type	Service			
	1	2	3	4
Classic				
RES Component	33.95	1.50	20.00	20.82
Sample size	32	1	1	11
Mixed				
RES Component		20.13	25.88	
Sample size		16	2	
Instrument				
RES Component		48.50	14.82	
Sample size		1	11	

Notes: RES component only, by Service

Technique type	Activity		
	1	2	3
Classic			
Actual hours	20.00	1.50	30.59
Sample size	1	1	43
Mixed			
Actual hours	25.88	20.13	
Sample size	2	16	
Instrument			
Actual hours	14.82	48.50	
Sample size	11	1	

Notes: (i) Actual hours without NCCA, by Service; (ii) RES component only, by Activity; (iii) Actual hours without NCCA, by Activity

Technique type	Mean	Sample size
Classic	17.45	15
Mixed	18.89	20
Instrument	7.17	34

Notes: (i) SRV component only; (ii) Actual hours without NCCA

Technique type	Service		
	2	3	4
Classic			
Actual hours	19.81	7.25	17.53
Sample size	4	1	10
Mixed			
Actual hours	22.77	7.25	
Sample size	15	5	
Instrument			
Actual hours	19.50	6.34	21.50
Sample size	1	32	1

Notes: (i) SRV component only, by Service; (ii) Actual hours without NCCA, by Service

Technique type	Activity		
	1	2	3
Classic			
Actual hours	7.25	19.81	17.53
Sample size	1	4	10
Mixed			
Actual hours	7.25	22.77	
Sample size	5	15	
Instrument			
Actual hours	6.34	19.50	21.50
Sample size	32	1	1

Notes: (i) SRV component only, by Activity; (ii) Actual hours without NCCA, by Activity

Results including NCCA

Technique type	Mean	Sample size
Classic	35.11	45
Mixed	24.39	18
Instrument	22.29	12

Notes: (i) RES component only; (ii) Actual hours with NCCA

Technique type	Service			
	1	2	3	4
Classic				
Actual hours	39.78	3.75	28.75	24.93
Sample size	32	1	1	11
Mixed				
Actual hours		23.39	32.38	
Sample size		16	2	
Instrument				
Actual hours		54.00	19.41	
Sample size		1	11	

Notes: (i) RES component only, by Service; (ii) Actual hours with NCCA, by Service

Technique type	Activity		
	1	2	3
Classic			
Actual hours	28.75	3.75	35.98
Sample size	1	1	43
Mixed			
Actual hours	32.38	23.39	
Sample size	2	16	
Instrument			
Actual hours	19.41	54.00	
Sample size	11	1	

Notes: (i) RES component only, by Activity; (ii) Actual hours with NCCA, by Activity

Technique type	Mean	Sample size
Classic	21.72	15
Mixed	21.91	20
Instrument	8.92	34

Notes: (i) SRV component only; (ii) Actual hours with NCCA

Technique type	Service		
	2	3	4
Classic			
Actual hours	23.63	8.50	22.28
Sample size	4	1	10
Mixed			
Actual hours	26.15	9.20	
Sample size	15	5	
Instrument			
Actual hours	20.00	8.08	24.75
Sample size	1	32	1

Notes: (i) SRV component only, by Service; (ii) Actual hours with NCCA, by Service

Technique type	Activity		
	1	2	3
Classic			
Actual hours	8.50	23.63	22.28
Sample size	1	4	10
Mixed			
Actual hours	9.20	26.15	
Sample size	5	15	
Instrument			
Actual hours	8.08	20.00	24.75
Sample size	32	1	1

Notes: (i) SRV component only, by Activity; (ii) Actual hours with NCCA, by Activity

Contact UTS

For more information about
this report, contact:

Chris Bajada
Associate Professor
Centre for Business and Social Innovation
E: chris.bajada@uts.edu.au

cbsi.uts.edu.au

Centre for Business and Social Innovation
UTS Business School
PO Box 123
Broadway NSW 2007