

# The landscape of validity and reliability practices from applied performance analysts: Establishing a best practice framework

International Journal of Sports Science  
& Coaching  
1–23

© The Author(s) 2025



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/17479541251317043

journals.sagepub.com/home/spo



John William Francis<sup>1</sup> , Jamie Lee Kyte<sup>2</sup>, Michael Bateman<sup>1</sup>,  
and Scott Benjamin Nicholls<sup>3</sup> 

## Abstract

Whilst Theoretical Performance Analysts benefit from established validity and reliability frameworks, their practical application faces several challenges for Applied Performance Analysts (APAs). We explored the current practices and perceptions regarding validity and reliability by APAs to formulate best-practice recommendations. An open and closed-answer survey was completed by 175 APAs from a range of sports and countries, with responses analysed based on experience (early-career APAs vs experienced APAs). Findings reveal a desire for robust validity and reliability methods, particularly among experienced APAs. Collaborative approaches amongst APAs and other stakeholders were identified when determining performance indicators, yet decisions for live data collection were primarily made by APAs. Familiarisation and reliability processes varied, specifically the frequency of checks, with experienced APAs dedicating more time to learning new systems and checking data more frequently. Despite the existing theoretical frameworks, time constraints and personnel changes pose challenges for practical implementation, with APAs calling for an applied framework. These insights inform the development of a best-practice recommendations framework for collecting, analysing, and presenting accurate and meaningful performance analysis data in applied settings.

## Keywords

Career, collaborative learning, decision-making, expertise, technology, video

## Introduction

Validity and reliability are crucial in sports performance analysis, serving as the foundation for accurate and actionable insights for influencing coaches, players, and other stakeholders' decisions.<sup>1–3</sup> Theoretical Performance Analysts have developed frameworks for establishing key performance indicators (KPIs) and collecting data, for example, Thomson et al.,<sup>4</sup> focusing on content, construct and criterion validity.<sup>5,6</sup> However, the practices of Applied Performance Analysts (APA) are less understood. While Theoretical Performance Analysts focus on deriving general laws and reductionist approaches,<sup>7</sup> APAs navigate unique challenges, such as time constraints and personnel changes,<sup>8</sup> which may impact their ability to maintain rigorous data standards. Despite these challenges, maintaining high standards allows APAs to enhance performance insights and foster trust and collaboration.<sup>9,10</sup>

Theoretical Performance Analysts have created frameworks for establishing KPIs, defining operational terms,

and creating data collection systems. These frames often utilise expert reviewers to ensure accuracy, regression analysis to identify trends between variables, and inferential statistics to draw conclusions from the data.<sup>4,11,12</sup> However, these methods often apply league/tournament-wide or based on general sport KPIs, potentially overlooking between-team variation in play styles or venue impact.<sup>13,14</sup> Additionally, the involvement of experts with limited sport-specific

Reviewer: Kirsten Spencer (Auckland University of Technology, New Zealand)

<sup>1</sup>University of Worcester, Worcester, UK

<sup>2</sup>University of Birmingham, Birmingham, UK

<sup>3</sup>University of Derby, Derby, UK

### Corresponding author:

John William Francis, University of Worcester, Henwick Grove, Worcester WR2 6AJ, UK.

Emails: drjwfrancis@gmail.com; j.francis@worc.ac.uk

knowledge may lead to findings that compromise content and construct validity, as their lack of familiarity with the nuances and specific demands of the sport could result in the selection of inappropriate KPIs and misinterpretation of data.<sup>15</sup> This oversight potentially results in findings that invalidate the context, construct, and criterion validity principles, thereby affecting the meaningfulness of the objective data used in coaching and performance evaluation.<sup>16,17</sup>

Data collection systems, operational definitions, player identification, and possession state can vary significantly between different sports, teams and analysts. This variation makes it essential to thoroughly learn and understand these systems to ensure accuracy.<sup>18,19</sup> Yet, learning alone does not guarantee good agreement between analysts or consistency with established standards.<sup>12</sup> The time Theoretical Performance Analysts dedicate to learning new systems varies widely from no time<sup>20</sup> to several days.<sup>21,22</sup> This period aims to develop a routine that reduces errors and increases agreement, especially on more subjective variables.<sup>7</sup> Despite frameworks,<sup>18,23</sup> APAs, constrained by schedules and other workplace demands,<sup>24</sup> may not have the time for extensive training, impacting data effectiveness and accuracy.

Theoretical Performance Analysts highlight the importance of reliability assessment, recommending intra- and inter-observer checks,<sup>25,26</sup> with the second intra-observer test occurring after a period of four to six weeks to reduce recall.<sup>18,21</sup> For APAs, especially in sports with frequent fixtures, adhering to this schedule can be challenging, with multiple performances analysed before the recommended testing period elapses. Various methods to assess data reliability have been used, including the percentage difference,<sup>27,28</sup> Cohen's<sup>29</sup> Kappa or weight kappa,<sup>30,31</sup> and Bland and Altman's<sup>32</sup> 95% limits of agreement<sup>33</sup>; however, there is continued debate as to which method is most appropriate.<sup>19,34,35</sup> Reliability issues are also evident in media-presented statistics and secondary data (e.g. BBC, ESPN, Eurosport, InStat, Sky, WhoScored.com and WyScout) for several simplistic performance measures, e.g. corners, fouls, and offsides.<sup>35,36</sup> Establishing suitably reliable data is important for Theoretical Performance Analysts but perhaps more importantly for APAs, where the information is utilised to inform key coaching decisions.<sup>17,26,37</sup>

To overcome these challenges, FIFA's new Football Language and performance metrics aim to provide new insights and align technical expertise.<sup>38</sup> For example, by standardising terms and metrics used to describe player movements and game phases, coaches and APAs from different teams and countries can more easily share and compare data, leading to more consistent and accurate performance evaluations. However, collecting and applying these metrics in club settings with high demands, tight schedules, and limited personnel may hinder achieving the same levels of data accuracy.<sup>39</sup> Consequently, APAs face a dilemma between producing timely insights, which may

not always be reliable and taking the necessary time to ensure the accuracy and meaningfulness of the data. Furthermore, the importance of reliable data collection is frequently implied rather than explicitly stated in APA job descriptions.<sup>40</sup> Despite increasing emphasis on validity and reliability in academic programmes,<sup>41,42</sup> discrepancies between academic emphasis and real-world application still exist. This highlights the need for a greater understanding of the validity and reliability practices and challenges faced by APAs.

By exploring the current practices and perceptions regarding validity and reliability by APAs, this study seeks to develop a set of 'best-practice' recommendations for improving the collection, analysis, and presentation of accurate (reliable) and meaningful (valid) sports performance analysis data and information. The findings from the study are intended to be disseminated back to APAs, heads of sports science and medicine, coaches and other stakeholders involved in sports to assist in the knowledge and understanding of current processes and practices. It is hoped that the knowledge gained may also inform educational practices to better align with applied needs.

## Materials and methods

### Participants

A total of 175 APAs, currently employed full-time, part-time or in consultancy roles by sporting organisations and possessing a minimum of one year's experience, voluntarily participated in an online survey. Recruitment was facilitated through advertisements on social media platforms, namely LinkedIn and X (Twitter), as well as directly approaching contacts through the research team's industry network. Participants were encouraged to share the survey within their networks, using 'snowballing sampling'.<sup>43</sup> The survey was accessible for 11 weeks (17<sup>th</sup> July 2023–1<sup>st</sup> October 2023), with periodical social media posts every three weeks at different times to promote the survey.

The survey was not limited to a single sport and allowed multiple APAs from the same organisation or team. Participants were all over 18 years of age. Before completing the survey, all participants were presented with the participant information sheet on the first page and were advised by completing the survey, they were providing informed consent. The procedure was ethically approved by the lead researcher's University Ethics and Research Governance Committee.

### Survey design and distribution

The survey questions were developed by the first author, drawing on their 11 years of theoretical experience and 14 years of APA experience, along with relevant literature.<sup>4,12,44</sup> The initial list of questions was refined through

two focus groups to ensure content validity. These focus groups included the remaining authors, who collectively had 18 years of theoretical experience and 26 years of applied experience in sports performance analysis. Participants in the focus groups were selected based on their expertise and background in various sports. The digital survey platform, JISC OnlineSurveys V2, was used to design the structure of the survey. The survey was pilot-tested by an applied academic and two APAs, all with over 10 years of experience. These individuals were external to the research team, meaning they were not involved in the development of the survey. They were defined as experts based on their extensive experience and recognised contributions to the field. They did not complete the final survey to avoid any potential bias. Following pilot testing, minor adjustments were made to refine the wording and organisation of certain questions to aid readability and optimise the flow of the survey. The revised survey was collectively approved by the research team and three external experts. This process follows similar online surveys within sports performance analysis.<sup>39,45</sup>

The final survey took an average of 18 min to complete by the respondents, with responses remaining as no identifiable information was requested. The survey (Supplementary Document: Survey) began with participant information and consent, a series of demographic questions and a glossary of terms regarding validity<sup>a</sup> and reliability<sup>b</sup>. Demographic information was collected using pre-populated responses to uphold confidentiality. The main survey included a series of open and closed-ended questions focusing on four key areas: (1) establishing KPI processes, (2) operational definitions processes, (3) familiarisation of systems and (4) reliability processes. Questions featured Likert-scale responses on a 6-point scale relating to frequency, multiple-choice responses, and free-text response fields. The free-text fields allowed respondents to provide additional context to their closed-ended responses where necessary.<sup>46</sup> The language and style of the questions were tailored to the language, expressions and phrases used by APAs.<sup>47</sup>

### *Data reduction and analysis*

Survey responses were exported to Microsoft Excel and analysed in RStudio (version 2024.04.1 + 748, RStudio, PBC, Boston, MA). Frequency analysis was conducted for multiple-choice and categorical questions to compare the responses of APAs with varying levels of experience. APAs with less than five years of experience are referred to as “Early-Career APAs”, while those with over five years of experience are referred to as “Experienced APAs”. This categorisation was based on combining the closed-answer response data from less than 1 year, 1–2 years, and 3–4 years of experience into the ‘Early-career’ group and 5–6 years and 7+ years of experience into the ‘Experienced’ group. The rationale for the five-year

threshold is that APAs with less than five years of experience are likely to have less applied experience but potentially more current theoretical underpinning. Frequency counts and percentage distributions were calculated to explore the distribution of responses among these groups for the demographic questions and non-Likert scale questions. A 2-sample test for equality of proportions with continuity correction and, chi-square test and Cramér’s V were used to examine any statistically significant differences between these groups.

Likert scale responses were quantified by assigning integers, facilitating quantitative analysis. These numeric values were then interpreted using qualitative terms, and the median response reported.<sup>48</sup> The differences in responses between APAs’ experience levels were reported as differences in the median responses, accompanied by interquartile ranges. These differences were further evaluated using the Mann-Whitney U test to assess for statistical differences due to the data being non-parametric. Additionally, to aid interpretation, any substantial difference between groups was displayed by indicating whether a meaningful change of 1 point in the Likert scale was observed.<sup>49</sup>

Open-ended responses were manually analysed using inductive thematic analysis, following the approach outlined by Braun and Clarke.<sup>50</sup> Each question’s open-ended response data was analysed separately by the second author, who assigned a code to each response. To ensure trustworthiness, the developed codes were reviewed for relevance and re-coded by the research team, where necessary, until consensus was reached. Themes were consolidated, and data saturation was achieved for the emerging themes.

### *Creation of ‘best-practice’ PRECISE recommendations*

Following the analysis of both quantitative and qualitative data, the research team met on two occasions to understand participants’ responses concerning validity, familiarisation and reliability practices. The focus was to begin identifying areas of best practice and areas of development. The lead author and the fourth author then reviewed relevant research papers and theories from sports performance analysis (e.g. measurement issues in sports performance analysis<sup>12</sup>), learning (e.g. variation theory<sup>51</sup>) and psychology (e.g. deliberate practice<sup>52</sup>) to devise a set of ‘best-practice’ recommendations. The wider research team then met to finalise the recommendations and develop a framework for their presentation. Following this, the recommendations framework was shared with seven experts whose combined expertise in sports performance analysis (both applied and theoretical) amounted to over 80 years of experience across various sports. These experts provided suggestions and revisions to the acronym and terminology to ensure a multi-sport framework was developed.

## Results

### Respondent demographics

Among the 175 survey respondents, 72 (41.1%) were early-career APAs, while 103 (58.9%) were experienced APAs. Early-career APAs were predominately under 29 years of age ( $n = 66$ ; 91.7%), whereas experienced APAs were typically between 25 and 34 years of age ( $n = 72$ ; 71.9%). Soccer emerged as the prominent sport ( $n = 106$ , 59.4%; early-career = 52, 72.2%; experienced = 54, 52.4%), followed by Rugby ( $n = 39$ , 22.3%; early-career = 12, 6.9%; experienced = 27, 15.4%). Full-time employment was prominent ( $n = 131$ , 74.9%; early-career = 42, 58.3%; experienced = 89, 86.4%), while part-time employment was more common among those with fewer years of APA experience ( $n = 38$ , 21.7%; early-career = 27, 37.5%; experienced = 11, 10.7%). Most respondents worked with senior or first-team performers ( $n = 117$ , 66.9%; early-career = 43, 59.7%; experienced = 74, 71.8%) rather than with academy or development-level performers. Respondents generally held at least a post-graduate qualification ( $n = 130$ , 74.3%; early-career = 48, 66.6%; experienced = 82, 79.6%). Early-career APAs often lacked or held only a Level 1 coaching qualification ( $n = 94$ , 53.7%; early-career = 49, 68.1%; experienced = 45, 43.6%), while experienced APAs commonly held a Level 2 or higher coaching qualification ( $n = 81$ , 46.3%; early-career = 23, 31.5%; experienced = 58, 56.3%). However, 29.1% ( $n = 30$ ) of experienced APAs did not hold any coaching qualification. Geographically, responses were diverse, spanning African, American, Asian, European and Oceania countries; however, the UK notably represented the majority of respondents ( $n = 125$ , 71.4%; early-career: 53, 73.6%; experienced: 72, 69.9%). Further demographical insights can be seen in Table 1.

### Perceptions of validity and reliability

All participants perceived validity processes as highly important, with 52.8% ( $n = 38$ ) of early-career APAs and 54.4% ( $n = 56$ ) of experienced APAs rating it as essential. The median rating for early-career APAs was 5 (IQR = 1), indicating 'Essential', while for experienced APAs, the median was also 5 (IQR = 1), indicating 'Essential' ( $p = 0.426$ ) (Figure 1). For reliability processes, 34.7% ( $n = 25$ ) of early-career APAs viewed it as very important, and 45.8% ( $n = 33$ ) viewed it as essential. Among experienced APAs, 29.1% ( $n = 21$ ) rated it as very important and 60.2% ( $n = 62$ ) as essential (early-career: 4 (IQR = 1), experienced: 5 (IQR = 1)) ( $p = 0.036$ ).

Over 90% of all APAs utilised primary<sup>c</sup> video and data for pre-analysis<sup>d</sup> and post-analysis<sup>e</sup> techniques. Specifically, 93.1% ( $n = 163$ ) of APAs engaged in primary pre-analysis, with 90.3% ( $n = 65$ ) of early-career APAs and 95.1% ( $n =$

98) of experienced APAs participating ( $p = 0.342$ ). For primary post-analysis, 96.8% ( $n = 169$ ) of APAs were involved, including 95.8% ( $n = 69$ ) of early-career APAs and 97.1% ( $n = 100$ ) of experienced APAs ( $p = 0.979$ ). In terms of secondary<sup>f</sup> video and data analysis, over 70% of APAs employed these techniques for both pre-analysis and post-analysis. For secondary pre-analysis, 73.7% ( $n = 129$ ) of APAs were involved, with 68.1% ( $n = 49$ ) of early-career APAs and 77.6% ( $n = 80$ ) of experienced APAs ( $p = 0.212$ ). For secondary post-analysis, 72.6% ( $n = 127$ ) of APAs participated, including 65.3% ( $n = 47$ ) of early-career APAs and 77.7% ( $n = 80$ ) of experienced APAs ( $p = 0.102$ ). Experienced APAs used primary live<sup>g</sup> analysis more frequently ( $n = 132$ , 75.4%) compared to early-career APAs ( $n = 47$ , 65.3%;  $p = 0.015$ ). However, there was no statistically significant difference in the use of secondary live analysis between early-career APAs ( $n = 24$ , 33.3%) and experienced APAs ( $n = 44$ , 42.7%;  $p = 0.273$ ).

A total of 121 (69.1%) APAs used Hudl SportsCode for generating and curating their primary data (early-career = 41, 56.9%; experienced = 80, 77.7%;  $p = 0.025$ ), while the remaining 54 APAs used various other software providers. Among the 130 (74.3%) APAs using data providers, 18 different providers were identified with Wyscout (early-career = 31, 43.1%; experienced = 19, 18.5%), StatsBomb (early-career = 4, 5.5%; experienced = 17, 16.5%) and StatsPerform/Opta (early-career = 8, 11.1%; experienced = 25, 24.3%) being the most frequently used.

### Validation processes

The coaching team and APA/analysis department typically worked collaboratively when deciding what KPIs to use for primary data and video in pre-analysis ( $n = 163$ , 93.1%; early-career = 52, 80.0%; experienced = 80, 81.6%) and post-analysis ( $n = 124$ , 70.9%; early-career = 54, 79.0%; experienced = 79, 79.0%). In contrast, decisions regarding primary video and data for live analysis ( $n = 132$ , 75.4%; early-career = 26, 55.3%; experienced = 46, 54.1%) and secondary video and data for pre-analysis ( $n = 129$ , 73.7%; early-career = 29, 59.2%; experienced = 54, 67.5%), live-analysis ( $n = 68$ , 38.9%; early-career = 14, 58.3%; experienced = 30, 68.2%) and post-analysis ( $n = 126$ , 72.0%; early-career = 24, 51.1%; experienced = 43, 53.8%) was more frequently undertaken by the APA/analysis department, regardless of experience.

APAs collected a diverse range of information across key moments (goals/shots/points awarded/cards), play-by-play or ball-by-ball, team/opposition spatial interactions and player actions/technical information. Pre-analysis and post-analysis procedures, whether using primary or secondary video or data, were considered equally important across all types of information, regardless of experience. However, when APAs were collecting primary video and data for live analysis, there was a preference to focus on key moments

**Table 1.** Demographic information of Applied Performance Analysts (APAs) based on experience level expressed as raw frequency count and percentage distribution across experience levels.

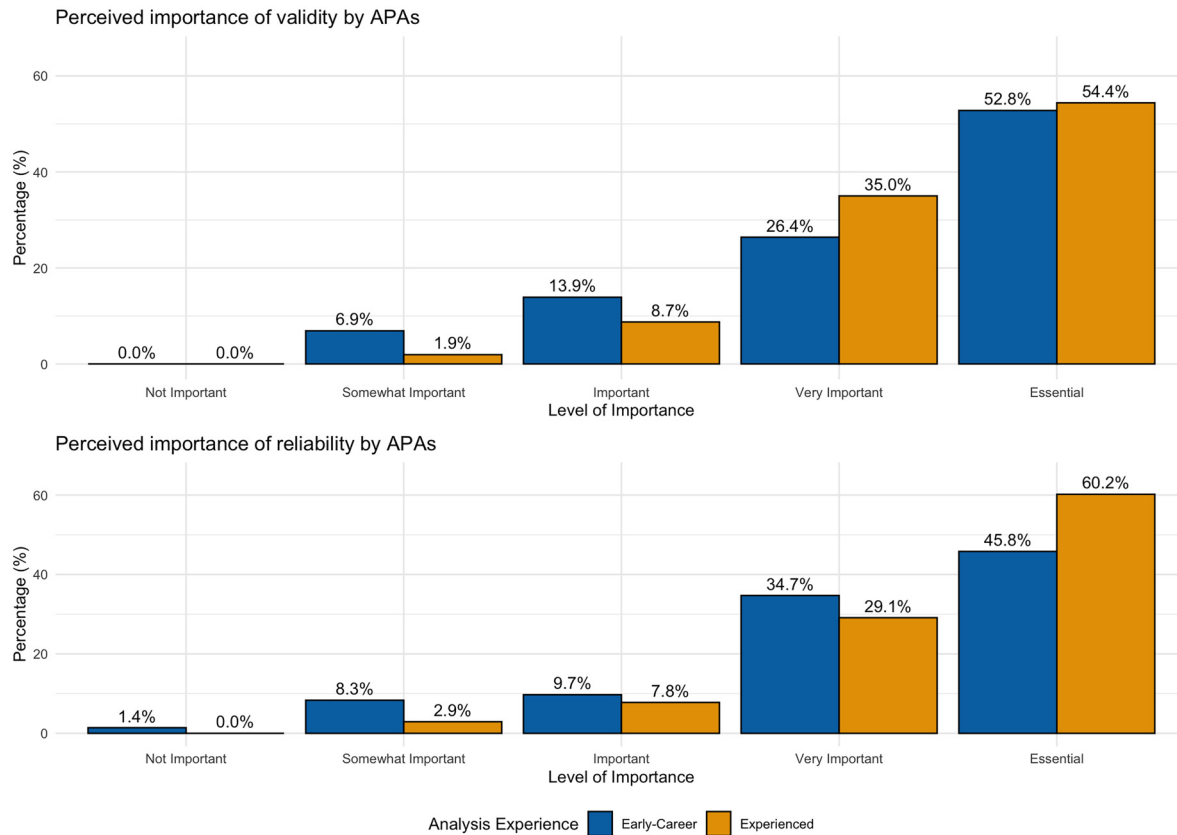
	Experience level		P value ( $\varphi_c$ )
	Early-career APA (n = 72; 41.1%)	Experienced APA (n = 103; 58.9%)	
Age			
18–24 years	35 (48.6%)	8 (7.8%)	<.001 (0.552)
25–29 years	31 (43.1%)	42 (40.8%)	
30–34 years	3 (4.2%)	32 (31.1%)	
35–39 years	2 (2.8%)	5 (4.9%)	
40+ years	1 (0.6%)	16 (16.5%)	
Sport			
AFL	0	3 (2.9%)	0.035 (0.308)
Basketball	0	2 (1.9%)	
Cricket	4 (5.6%)	1 (1.0%)	
Soccer	52 (72.2%)	54 (52.4%)	
Netball	2 (2.8%)	4 (3.9%)	
Olympic	2 (2.8%)	5 (4.9%)	
Other Sports	0	5 (4.9%)	
Racquet Sports	0	2 (1.9%)	
Rugby	12 (6.9%)	27 (15.4%)	
Employment			
Full-Time	42 (58.3%)	89 (86.4%)	<.001 (0.327)
Part-Time	27 (37.5%)	11 (10.7%)	
Consultant	3 (4.2%)	3 (2.9%)	
Role Level			
Academy or Development	29 (40.3%)	29 (28.3%)	0.130 (0.127)
Senior or First Team	43 (59.7%)	74 (71.8%)	
Highest Academic Qualification			
Doctoral Degree	1 (1.4%)	12 (11.7%)	0.032 (0.225)
Masters Degree	47 (65.3%)	70 (68.0%)	
Undergraduate Degree	23 (31.9%)	20 (19.4%)	
No Degree	1 (1.4%)	1 (1.0%)	
Highest Coaching Qualification			
No Level	17 (23.6%)	30 (29.1%)	0.001 (0.347)
Level 1	32 (44.4%)	15 (14.6%)	
Level 2	13 (18.1%)	35 (34.0%)	
Level 3	8 (11.1%)	15 (14.6%)	
Level 4	2 (2.8%)	6 (5.8%)	
Level 5	0	2 (1.9%)	
Country or Continent			
UK	53 (73.6%)	72 (69.9%)	0.914 (0.109)
Ireland	1 (1.4%)	3 (2.9%)	
Other European Country	9 (12.5%)	11 (10.7%)	
African Country	2 (2.8%)	3 (2.9%)	
American Country	2 (2.8%)	2 (1.9%)	
Asian Country	2 (2.8%)	3 (2.9%)	
Oceania Country	3 (4.2%)	9 (8.7%)	

(early-career = 45, 37.5%; experienced = 79, 35.4%) over the other types of performance information, irrespective of APA experience.

Early-career APAs ‘Very Frequently’ aligned the collected video and data to the team or individual’s playing style for primary video and data used in pre-analysis, live analysis and post-analysis (Table 2). In contrast, experienced APAs consistently rated this practice as ‘Always’,

showing a clear 1-point difference. For secondary video and data, both early-career and experienced APAs ‘Very Frequently’ aligned these for pre-analysis, live analysis, and post-analysis.

Operational definitions for KPIs collected during primary video and data processes were established by 63.1% to 79.0% of APAs, regardless of experience (pre-analysis = 162, 92.6%; early-career = 41, 63.1%; experienced = 76, 77.6%; live



**Figure 1.** Perception of the perceived importance of validity and reliability, presented as proportion of responses for early-career and experienced APAs.

analysis = 132, 75.4%; early-career = 34, 72.3%; experienced = 64, 75.3%; post-analysis = 169, 96.6%; early-career = 47, 68.1%; experienced = 79, 79.9%). In secondary data and video processes, experienced APAs consistently had a higher percentage of acknowledging the definitions from the data providers across each stage of the sports performance analysis process compared to their early-career counterparts (pre-analysis = 130, 74.3%; early-career = 29, 58.0%; experienced = 62, 77.5%; live analysis = 68, 38.9%; early-career = 14, 58.3%; experienced = 34, 77.3%; post-analysis = 127, 72.6%; early-career = 27, 57.5%; experienced = 66, 82.5%).

Regarding the alignment of video and data definitions with the playing philosophy, early-career APAs 'Very Frequently' do so for primary pre-analysis and live analysis, while experienced APAs rate this alignment as 'Always', showing a significant 1-point difference (Table 2). Whilst for primary post-analysis, all APAs, regardless of experience, align their definitions with their playing philosophy 'Very Frequently'. In contrast, when working with secondary video and data, both early-career and experienced APAs align 'Very Frequently' their definitions with their playing philosophy across all analysis stages.

Operational definitions for primary data and video were typically reviewed at the start of the season (pre-analysis:

early-career = 23, 34.9%; experienced = 25, 25.5%; live analysis: early-career = 23, 35.4%; experienced = 24, 24.2%; post-analysis: early-career = 22, 33.3%; experienced = 26, 26.5%) or not at all (pre-analysis: early-career = 15, 22.7%; experienced = 24, 24.5%; live analysis: early-career = 14, 21.4%; experienced = 30, 30.3%; post-analysis: early-career = 12, 18.2%; experienced = 17, 17.4%) (Table 3). If disagreements with definitions from third-party/secondary data companies, APAs only 'Occasionally' would reanalyse the footage to align data with their own definitions, regardless of experience (Table 2). Operational definitions were stored on servers, online clouds, or openly shared documents/files with APAs and coaches, ensuring accessibility (n = 102, 58.3%; early-career = 38, 62.3%; experienced = 64, 67.4%). Storage formats included Word/PDF documents (n = 31, 17.7%; early-career = 12, 19.7%; experienced = 19, 20.0%), PowerPoint/Keynote presentations (n = 12, 6.9%; early-career = 3, 4.9%; experienced = 9, 9.5%) and Excel/Numbers sheets (n = 7, 4.0%; early-career = 4, 6.6%; experienced = 3, 3.2%). Some APAs integrated their operational definitions within their data collection tools, such as embedding them within their code window in Hudl SportsCode. Video clips were also used to support operational definitions by 59.7% of early-career APAs (n = 43) and 71.2% of experienced APAs (n = 74).

**Table 2.** Summary of APAs' responses to Likert Scale questions on validity processes (median and IQR, including median difference, p-values and overall IQR).

	Early-career APA	Experienced APA	Clear 1-Point Difference (Median Diff; P Value; IQR)
<b>Is The Performance Information Collected Based on A Playing Philosophy For...Analysis:</b>			
Primary Video/Data Used for Pre-Analysis	Very Frequently 5 (IQR = 1) (N = 65)	Always 6 (IQR = 1) (N = 98)	Yes (1; p = 0.064; 1)
Primary Video/Data Used for Live-Analysis	Very Frequently 5 (IQR = 1) (N = 47)	Always 6 (IQR = 1) (N = 85)	Yes (1; p = 0.093; 1)
Primary Video/Data Used for Post-Analysis	Very Frequently 5 (IQR = 1) (N = 69)	Always 6 (IQR = 1) (N = 100)	Yes (1; p = 0.183; 1)
Secondary Video/Data Used for Pre-Analysis	Very Frequently 5 (IQR = 2) (N = 49)	Very Frequently 5 (IQR = 2) (N = 80)	No (0; p = 0.523; 2)
Secondary Video/Data Used for Live-Analysis	Very Frequently 5 (IQR = 2.25) (N = 24)	Very Frequently 5 (IQR = 1) (N = 44)	No (0; p = 0.051; 2)
Secondary Video/Data Used for Post-Analysis	Very Frequently 5 (IQR = 2) (N = 47)	Very Frequently 5 (IQR = 2) (N = 80)	No (0; p = 0.828; 2)
<b>Are the definitions based on a playing philosophy for...analysis?</b>			
Primary Video/Data Used for Pre-Analysis	Very Frequently 5 (IQR = 2) (N = 61)	Always 6 (IQR = 1) (N = 89)	Yes (1; p = 0.005; 2)
Primary Video/Data Used for Live-Analysis	Very Frequently 5 (IQR = 2.5) (N = 59)	Always 6 (IQR = 1) (N = 87)	Yes (1; p = 0.013; 2)
Primary Video/Data Used for Post-Analysis	Very Frequently 5 (IQR = 2.5) (N = 63)	Very Frequently 5 (IQR = 1) (N = 93)	No (0; p = 0.034; 2)
Secondary Video/Data Used for Pre-Analysis	Very Frequently 5 (IQR = 3) (N = 53)	Very Frequently 5 (IQR = 4.75) (N = 78)	No (0; p = 0.450; 3.5)
Secondary Video/Data Used for Live-Analysis	Very Frequently 5 (IQR = 3) (N = 43)	Always 5.5 (IQR = 3.75) (N = 58)	No (0; p = 0.220; 3)
Secondary Video/Data Used for Post-Analysis	Very Frequently 5 (IQR = 2.5) (N = 47)	Very Frequently 5 (IQR = 5) (N = 79)	No (0; p = 0.483; 3)
If you disagree with the definitions provided by third-party/secondary data companies, do you re-analyse the data for...analysis:			

(continued)

Table 2. (continued)

	Early-career APA	Experienced APA	Clear 1-Point Difference (Median Diff; P Value; IQR)
Secondary Video/Data Used for Pre-Analysis	Occasionally 4 (IQR = 3.25) (N = 48)	Occasionally 4 (IQR = 2) (N = 75)	No (0; p = 0.564; 3)
Secondary Video/Data Used for Live-Analysis	Occasionally 4 (IQR = 3.25) (N = 41)	Occasionally 4 (IQR = 2) (N = 61)	No (0; p = 0.966; 3)
Secondary Video/Data Used for Post-Analysis	Occasionally 4 (IQR = 3.25) (N = 43)	Occasionally 4 (IQR = 2) (N = 77)	No (0; p = 0.574; 3)

### Familiarisation

When changes were implemented or new systems introduced, experienced APAs typically dedicated time 'Very Frequently' to familiarise themselves with the primary data collection system, whilst early-career APAs 'Occasionally' dedicated time (Table 4). Notably, between 16.7% and 18.1% of early-career APAs (early-career: pre-analysis = 13, 18.1%; live analysis = 12, 16.7%; post-analysis = 13, 18.1%) and 30.1% (n = 31) of experienced APAs 'Always' allocated time to learn the new primary data collection system before using them for performance evaluation. For secondary data collection systems, APAs, regardless of experience, generally set aside time 'Occasionally' to become familiar with the new data and video.

While 9.1% (n = 16) APAs (early-career = 11, 15.3%; experienced = 5, 4.9%) did not engage in any form of familiarisation, those who did vary greatly in the time dedicated, from a few minutes to a few hours or weeks. An early-career APA in Rugby captured the timing ambiguity, stating, "This depends largely on the magnitude of the change in variable/definition. It is hard to put a specific time frame on it as it's more just when I am confident that I'm getting it right". APAs provided insights into their familiarisation processes, specifically, a high proportion highlighted a lack of formalised processes (Informal Practice or Recording = 66, 41.8%; early-career = 32, 54.2%; experienced = 34, 34.4%), (Figure A1). Experienced APAs adopt a more formal process, specifying dedicating time to learning the new system, in comparison to early-career APAs (Formal Practice or Recording = 43, 27.2%; early-career = 12, 20.3%; experienced = 31, 31.3%). However, uncertainty emerged regarding what system familiarisation entailed, with some APAs showing their work to coaches and other APAs, engaging in 'Seeking Confirmation' (n = 37, 23.3%; early-career = 13, 22.0%; experienced = 24, 24.2%), whilst other APAs undertook 'Reliability Testing Processes' (n = 7, 4.4%; early-career = 1, 1.7%; experienced = 6, 6.0%).

### Reliability process

APAs conducted reliability checks "Occasionally" (Table 5), with no discernible differences based on experience levels or between primary and secondary video and data. Early-career APAs typically checked primary video and data after 'Each Performance' (15, 22.7%) or 'In Performance Blocks' (14, 21.2%). Experienced APAs tended to check their primary video and data 'In Performance Blocks' (31, 31.3%) or at the 'Start of the Season' (20, 20.2%). Notably, 11.1% of experienced and 13.6% of early-career APAs did not check the quality of their primary video and data at all. For secondary video and data, APAs frequently checked 'In Performance Blocks' (early-career = 11, 20.4%; experienced = 25, 28.4%) or 'Never' (early-career = 10, 18.5%; experienced = 17, 19.3%).



**Table 3.** Frequency and percentage distribution of when APAs review their operational definitions.

	Each Performance	At the end of season	Start of the season	In Performance blocks	Mid-Season	Weekly	Monthly	Never
<b>Primary Video/Data used for Pre-Analysis</b>								
Early-career APA	3 (4.5%)	8 (12.1%)	23 (34.9%)	10 (15.2%)	1 (1.5%)	3 (4.6%)	4 (4.6%)	15 (22.7%)
Experienced APA	4 (4.1%)	22 (22.5%)	25 (25.5%)	16 (16.3%)	2 (2.0%)	4 (4.1%)	1 (1.0%)	24 (24.5%)
<b>Primary Video/Data used for Live-Analysis</b>								
Early-career APA	4 (6.2%)	9 (13.9%)	23 (35.4%)	13 (20.0%)	0	0	2 (3.1%)	14 (21.5%)
Experienced APA	2 (2.0%)	18 (18.2%)	24 (24.2%)	16 (16.2%)	3 (3.0%)	4 (4.0%)	2 (2.0%)	30 (30.3%)
<b>Primary Video/Data used for Post-Analysis</b>								
Early-career APA	3 (4.6%)	8 (12.1%)	22 (33.3%)	13 (19.7%)	1 (1.5%)	2 (3.0%)	5 (7.6%)	12 (18.2%)
Experienced APA	5 (5.1%)	24 (24.5%)	26 (26.5%)	19 (19.4%)	2 (2.0%)	3 (3.1%)	2 (2.0%)	17 (17.4%)
<b>Secondary Video/Data used for Pre-Analysis</b>								
Early-career APA	3 (5.9%)	4 (7.8%)	15 (29.4%)	7 (13.7%)	1 (2.0%)	2 (3.9%)	1 (2.0%)	18 (35.3%)
Experienced APA	2 (2.6%)	12 (15.8%)	14 (18.4%)	13 (17.1%)	1 (1.3%)	3 (4.0%)	1 (1.3%)	30 (39.5%)
<b>Secondary Video/Data used for Live-Analysis</b>								
Early-career APA	4 (15.4%)	2 (7.7%)	5 (19.2%)	5 (19.2%)	0	0	1 (3.9%)	9 (34.6%)
Experienced APA	0	6 (14.3%)	8 (19.1%)	10 (23.8%)	1 (2.4%)	1 (2.4%)	0	16 (38.1%)
<b>Secondary Video/Data used for Post-Analysis</b>								
Early-career APA	3 (6.7%)	1 (2.2%)	14 (31.1%)	8 (17.8%)	1 (2.2%)	1 (2.2%)	1 (2.2%)	16 (35.6%)
Experienced APA	1 (1.4%)	12 (16.9%)	15 (21.1%)	17 (23.9%)	1 (1.4%)	3 (4.2%)	0	22 (31.0%)

**Table 4.** Summary of APAs responses to Likert Scale questions on familiarisation processes (median and IQR, including median difference, p-values and overall IQR).

	Early-career APA	Experienced APA	Clear 1-Point Difference (Median Diff; P Value; IQR)
When changes are made to variables or definitions, do you undertake a period of familiarisation before using the template to collect video and data:			
Primary Video/Data used for Pre-Analysis	Occasionally 4 (IQR = 2) (n = 57)	Very Frequently 5 (IQR = 2) (n = 91)	Yes (1; p = 0.044; 3)
Primary Video/Data used for Live-Analysis	Occasionally 4 (IQR = 2) (n = 47)	Very Frequently 5 (IQR = 2) (n = 92)	Yes (1; p = 0.190; 2)
Primary Video/Data used for Post-Analysis	Occasionally 4 (IQR = 2) (n = 59)	Very Frequently 5 (IQR = 2) (n = 95)	Yes (1; p = 0.154; 2)
Secondary Video/Data used for Pre-Analysis	Occasionally 4 (IQR = 2) (n = 51)	Occasionally 4 (IQR = 2) (n = 74)	No (0; p = 0.824; 2)
Secondary Video/Data used for Live-Analysis	Occasionally 4 (IQR = 3) (n = 40)	Occasionally 4 (IQR = 2) (n = 58)	No (0; p = 0.509; 2)
Secondary Video/Data used for Post-Analysis	Occasionally 4 (IQR = 3) (n = 46)	Occasionally 4 (IQR = 2) (n = 76)	No (0; p = 0.897; 2.75)
Do you test the reliability or accuracy of your video and data:			
Primary Video/Data used for Analysis	Occasionally 4 (IQR = 2.75) (n = 70)	Occasionally 4 (IQR = 1.75) (n = 102)	No (0; p = 0.559; 2)
Secondary Video/Data used for Analysis	Occasionally 4 (IQR = 2) (n = 67)	Occasionally 4 (IQR = 2) (n = 92)	No (0; p = 0.645; 2)

**Table 5.** Summary of APAs responses to Likert Scale questions on reliability processes (median and IQR, including median difference, p-values and overall IQR). Along with the frequency of counts and percentage distribution of video/data checking used for analysis.

	Early-career APA		Experienced APA		Clear I-Point Difference (Median Diff; P Value; IQR)	
<b>Do you test the reliability or accuracy of your video and data:</b>						
Primary Video/Data used for Analysis	Occasionally 4 (IQR = 2.75) (n = 70)		Occasionally 4 (IQR = 1.75) (n = 102)		No (0; p = 0.559; 2)	
Secondary Video/Data used for Analysis	Occasionally 4 (IQR = 2) (n = 67)		Occasionally 4 (IQR = 2) (n = 92)		No (0; p = 0.645; 2)	
<b>How often are the reliability processes completed?</b>						
	Each Performance	At the end of season	In Performance blocks	Weekly	Monthly	Never
Primary Video/Data used for Analysis						
Early-career APA	15 (22.7%)	7 (10.6%)	14 (21.2%)	2 (3.0%)	5 (7.6%)	4 (6.0%)
Experienced APA	16 (16.1%)	8 (8.1%)	31 (31.3%)	6 (6.0%)	1 (1.0%)	6 (6.0%)
Secondary Video/Data used for Analysis						
Early-career APA	9 (16.7%)	5 (9.3%)	11 (20.4%)	3 (5.6%)	3 (5.6%)	6 (11.1%)
Experienced APA	9 (10.2%)	11 (12.5%)	25 (28.4%)	4 (4.6%)	4 (4.6%)	6 (6.8%)

Out of the 130 (74.3%) APAs who shared insights into their checking procedures, the majority, regardless of experience (Figure 2), relied on visual checks, discussions with coaches or other APAs (early-career = 19, 40.0%; experienced = 31, 37.4%) or reviewing frequency tables or matrices (early-career = 12, 25.5%; experienced = 17, 20.5%). Only experienced APAs (10, 8.3%) reported having automated systems to check for errors during the video to data input or extraction. Some APAs had

developed output windows that facilitated the visual assessment of data quality (early-career = 3, 6.4%; experienced = 5, 6.0%). The lowest checking procedure was the undertaking of conducted bivariate statistical analysis tests (early-career = 2, 4.3%; experience = 11, 13.3%) using various methods, including percentage error, Kappa, T-Test, Altman plots or Coefficient of Variation. An experienced Rugby APA provided insight into the limited use of statistical tests: “No statistical tests, the process is more about

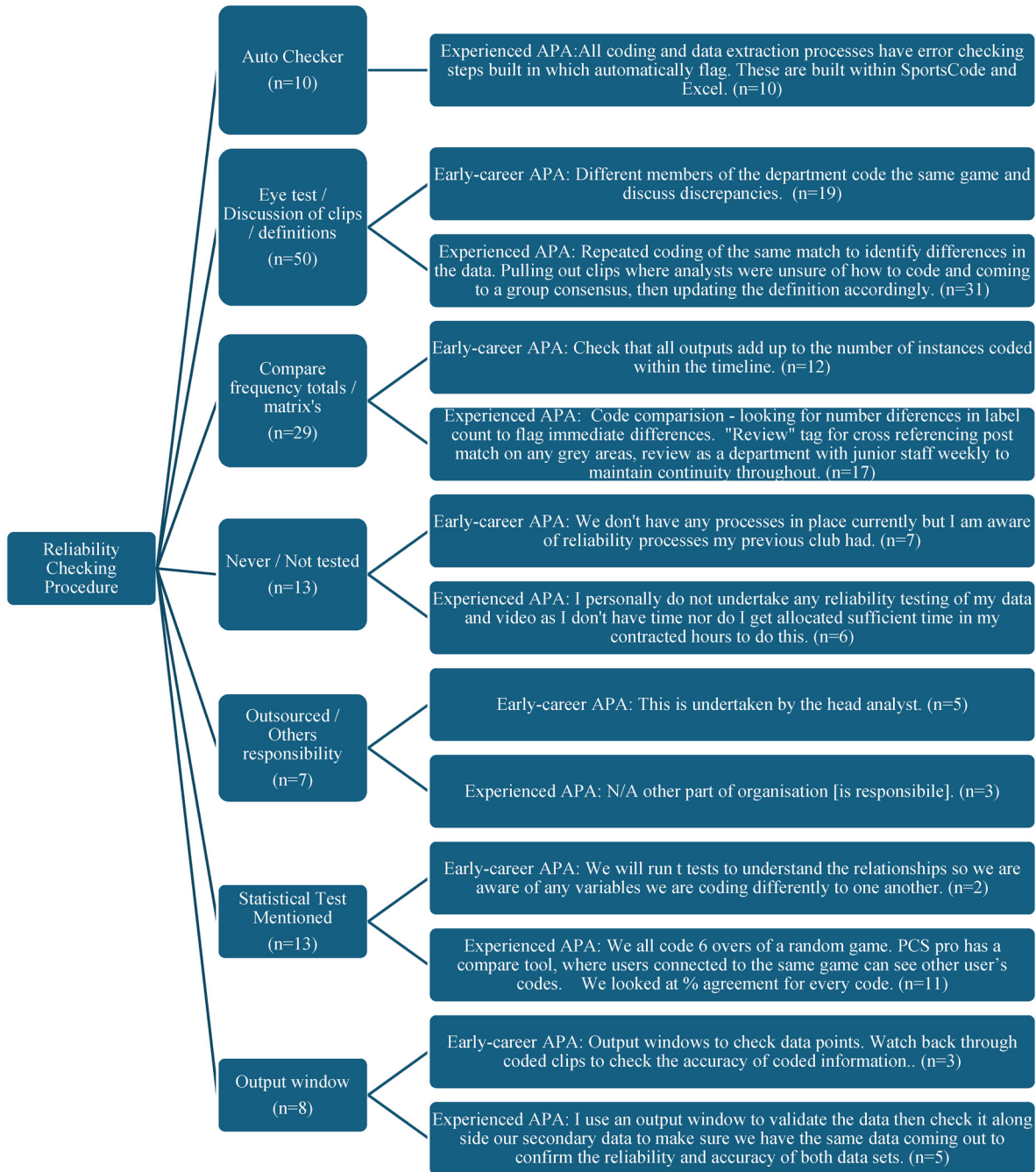


Figure 2. Thematic analysis of open-ended responses regarding reliability processes, specifically regarding checking procedures.

getting our definitions correct...If we are coding appropriately to the definitions and our data capture is consistent, the reporting will take care of itself'. This highlights that experienced APAs place greater emphasis on adherence to defining variables and consistent data collection procedures over extensive statistical testing.

When addressing errors, APAs either took a quick approach, like re-analysis (early-career = 18, 37.5%; experienced = 25, 27.8%) or engaged in discussions with staff (early-career = 8, 16.7%; experienced = 24, 26.7%) (Figure A2). Or adopted a more comprehensive approach, identifying the source of the problem (process identification: early-career = 7, 14.6%; experienced = 8, 8.9%), adjusting definitions (early-career = 4, 8.3%; experienced = 8, 8.9%), or providing additional training (early-career = 7, 14.6%; experienced = 11, 12.2%). An experienced Rugby APA noted, "Where major discrepancies exist, definitions are reviewed, and the process repeated until results are consistent. This is done pre-season due to time constraints in season". For issues with secondary video footage (n = 15), APAs either replaced the incorrect video or data or communicated with the provider to find a resolution (early-career = 3, 6.3%; experienced = 12, 13.3%).

## Overview

APAs recognised the importance of validity and reliability and were generally 'Satisfied' (4; IQR = 1;  $p = 0.216$ ) with their overall process. However, all APAs emphasised the need for more robust reliability-checking methods (early-career = 21, 36.8%; experienced = 25, 33.3%) and refining operational definitions (early-career = 14, 24.6%; experienced = 11, 14.7%) (Figure 3). Some APAs also suggested incorporating video clips for subjective variables. A key constraint contributing to current satisfaction levels with processes was time, particularly for those experienced APAs (early-career = 2, 3.5%; experienced = 13, 17.3%).

APAs proposed developing an applied framework of practice to enhance education and training (early-career = 8, 14.0%; experienced = 11, 14.7%) or automating workflows (experienced = 5, 6.7%). An early-career Rugby APA highlighted the need for a time-saving process, stating,

*"Definitely to test for both validity and reliability more than we do now, however, we just need to find a way to make this a less time-consuming process for in-season testing, especially when there are games every week and less time to do this sort of thing".*

It became evident that APAs were not solely responsible for ensuring accurate and meaningful data collection. Educating other stakeholders on the importance of a thorough validity and reliability process is essential (early-career = 3, 5.2%; experienced = 3, 4.0%). Improved

communication was needed for those using secondary data or relying on National Governing Bodies to better understand the data being collected, while others suggested reducing dependency on secondary data sources altogether (early-career = 2, 3.5%; experienced = 2, 2.7%).

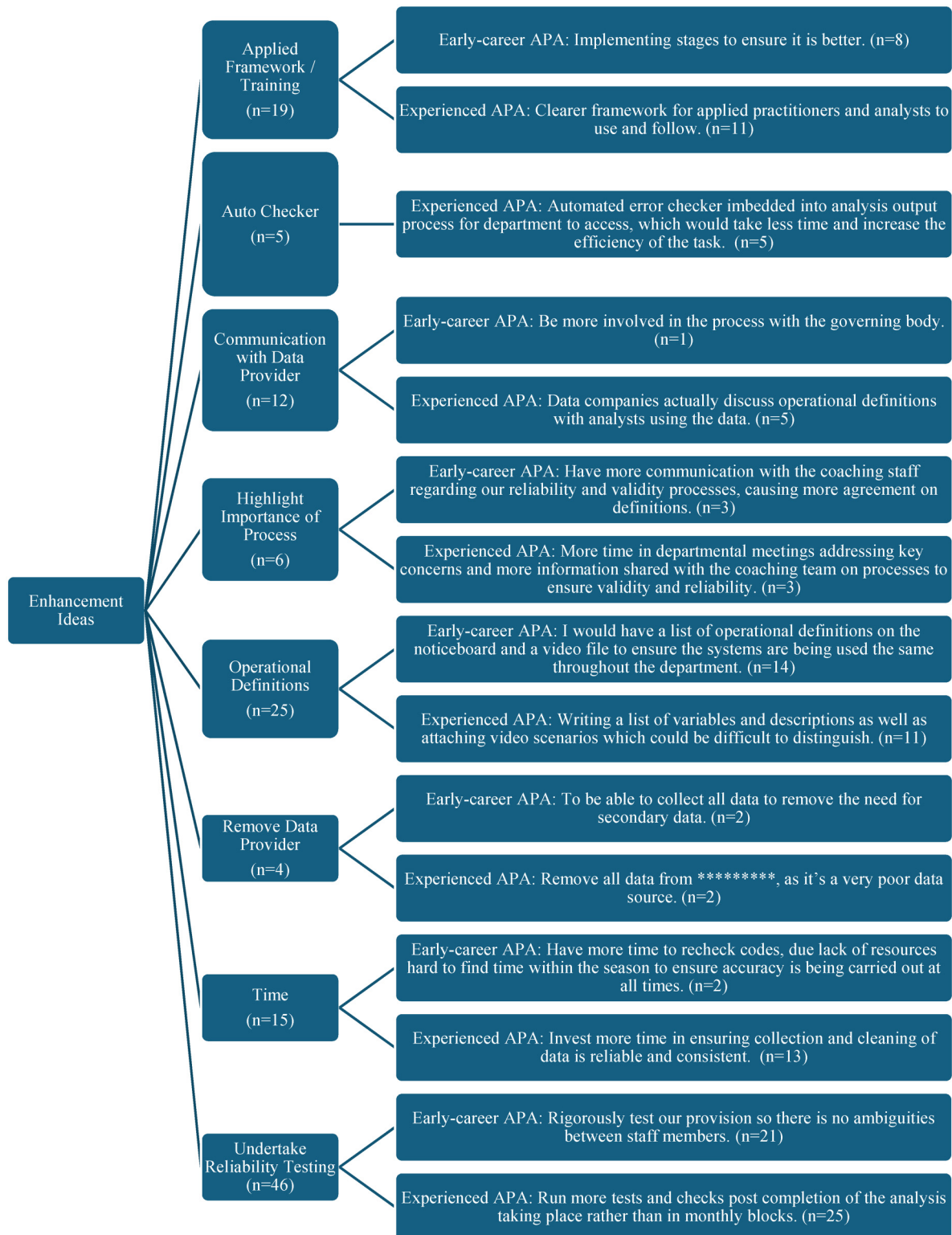
## Discussion and PRECISE framework presentation

The purpose of this study was to explore the current practices and perceptions regarding validity and reliability by APAs, formulating a set of 'best-practice' recommendations for improving the collection, analysis, and presentation of accurate (reliable) and meaningful (valid) sports performance analysis data and information. The findings highlighted notable differences across experience levels, particularly in collaborative decision-making in the alignment of KPIs and operational definitions with team/organisational goals/philosophies, familiarisation with new and existing data collection systems, and assessing the accuracy of the collected data.

Experienced APAs generally indicated greater satisfaction with their processes, attributed to a deeper understanding of aligning KPIs and operational definitions with organisation goals/philosophies, better navigation of data collection systems, and higher importance of data accuracy. However, both early-career and experienced APAs acknowledged the need for refining operational definitions and implementing more robust reliability procedures. Notably, time constraints emerged as a critical barrier to these enhancements, highlighting a need for efficiency in processes to ensure meaningful data collection, ultimately facilitating coach, player and stakeholder decisions.

The study also highlighted that APAs recognised the importance of enhancing validity, familiarisation and reliability processes within their work. Ensuring that data collection processes lead to valid and reliable insights is essential for the co-creation of knowledge, which, in turn, facilitates the development of learning opportunities that add value to coaches, players and stakeholders alike.<sup>24</sup> Validity and reliability processes are fundamental responsibilities for APAs,<sup>3,40</sup> as they ensure the accuracy, objectivity and meaningfulness of the information presented. While some APAs have established robust processes and express satisfaction with their practices, others continue to struggle with these aspects, indicating a need for ongoing improvement.

In response to these findings, a best practice recommendations framework for validity (Figure 4), familiarisation (Figure 5), and reliability (Figure 6) centred around the acronym "PRECISE" was developed. This framework, inspired by practices from the APAs who completed the survey and insights from theoretical performance analysts and researchers in other disciplines,



**Figure 3.** Thematic analysis of open-ended responses regarding ideas for future enhancement of APAs processes.

emphasises the meticulous and systematic approach required to maintain valid and reliable data throughout all stages of the data process. The following discussion

elaborates on our results and how they have informed the development of the specific constructs within the framework of the recommendations.



# VALIDITY

## Applied Performance Analysts

### Best Practice Guidelines



**Figure 4.** The “PRECISE” validity applied performance analysts’ best practice guidelines.

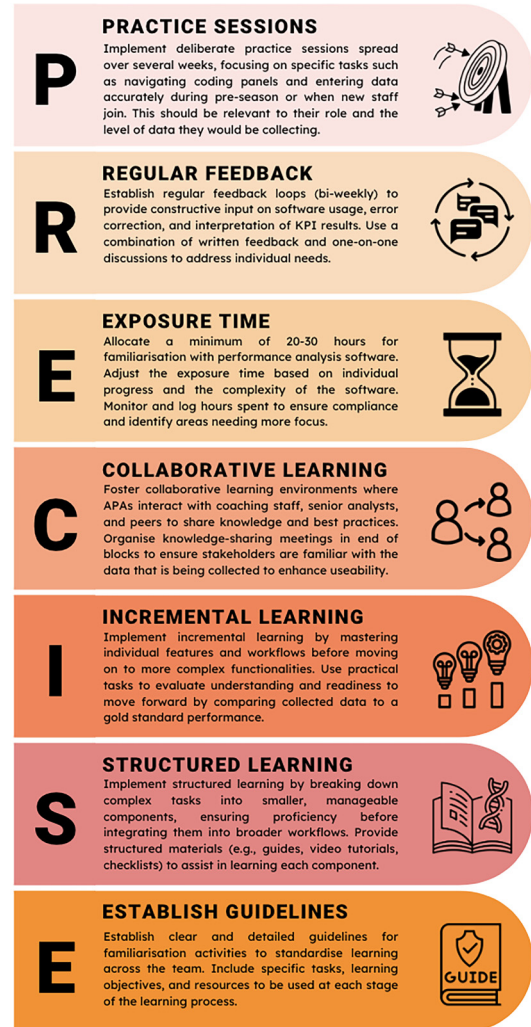
### Collaborative decision-making of KPIs/operational definitions

APAs often engage in collaborative processes with the coaching staff and sports performance analysis departments, especially in pre-analysis processes ( $n = 163$ , 93.1%; early-career = 52, 80.0%; experienced = 80, 81.6%). This approach enriches content validity by incorporating a range of perspectives<sup>53</sup> and mitigates individual biases.<sup>54–56</sup> Involving stakeholders in KPI selection ensures that the data collected is contextually relevant and aligned with the team’s specific goals and philosophies.<sup>53,57</sup> Thus, regular inter-disciplinary

# FAMILIARISATION

## Applied Performance Analysts

### Best Practice Guidelines



**Figure 5.** The “PRECISE” familiarisation applied performance analysts’ best practice guidelines.

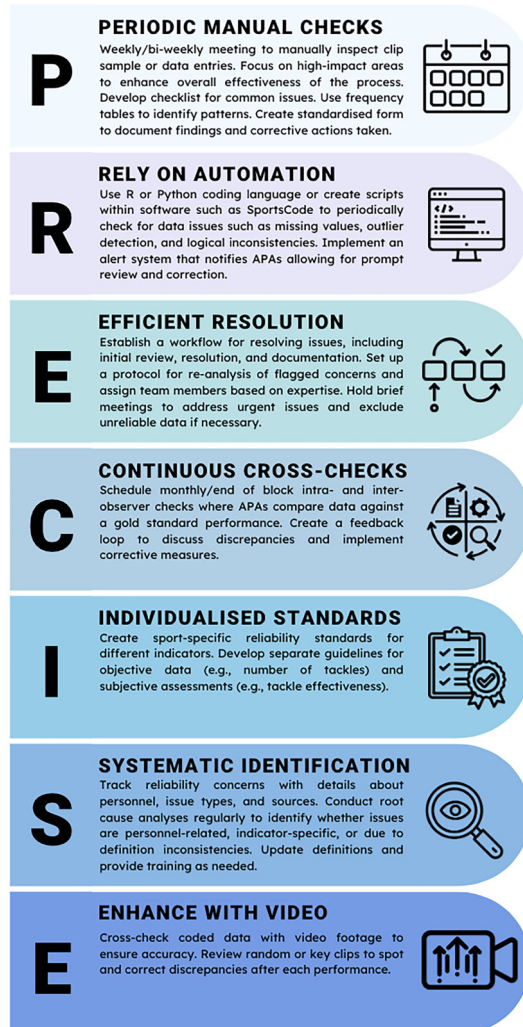
discussions can further foster this collaborative effort, ensuring everyone is on the same page.

While Theoretical Performance Analysts consult experts to select or validate KPIs, these tend to be broad-ranging rather than tailored to a specific playing philosophy, aiming to derive general statistical laws.<sup>6,7</sup> Aligning KPIs to reflect a particular philosophy can enhance feedback, decision-making and learning for stakeholders, deepening the understanding of why and how behaviours emerged and presenting actionable information.<sup>6,58</sup> Having the most relevant individuals involved in the reviewing and classifying of actions can help enhance meaningfulness

# RELIABILITY

## Applied Performance Analysts

### Best Practice Guidelines



**Figure 6.** The “PRECISE” reliability applied performance analysts’ best practice guidelines.

and reduce bias. However, challenges arise when APAs work with live or secondary video/data that may not fully align with a team’s playing style, often leading to more generalised rather than specific, actionable insights.<sup>8,59</sup>

The fluid environment of elite sports and personnel changes require APAs to adapt processes rapidly, impacting data collection.<sup>60</sup> Despite these challenges, aligning information with playing styles enhances ecological and content validity, providing meaningful insights for future learning and adding value to the APAs’ work.<sup>24</sup> Therefore, APAs should create flexible data collection systems that can adapt to changes in

coaching strategies or team dynamics, resulting in meaningful data collection.

Theoretical Performance Analysts increasingly include operational definitions for KPIs in published works,<sup>21,59</sup> which enhances construct validity by providing a clear framework for interpreting and analysing performance data.<sup>61</sup> However, creating clear operational definitions does not guarantee good agreement or construct validity, nor does the lack of learning these definitions guarantee poor agreement or lack of construct validity.<sup>12</sup> Tailoring operational definitions to specific philosophies ensures KPIs accurately reflect performer behaviours,<sup>17,62</sup> a practice supported by experienced APAs.

Recent trends among Theoretical Performance Analysts include using video clips to support agreed KPIs and their definitions, aiding in the accurate categorisation of observed behaviour.<sup>61</sup> They also support data collectors by providing scenarios that illustrate the concepts and definitions, seeking to produce a more accurate categorisation of specific observed behaviours.<sup>21</sup> Likewise, over 70% of the APAs now incorporate video clips, particularly for more subjective KPIs, which enhances understanding, external criterion validity and data accuracy. While some APAs and Theoretical Performance Analysts suggest a single clip is sufficient to showcase a KPI, others suggest using a minimum of five clips per KPI. Using more clips aligns closely with Marton’s<sup>51</sup> variation theory, which infers exposing learners to multiple examples allows them to understand the critical features and variations of an event. By presenting at least five clips, learners can discern the range of possible scenarios, leading to a deeper understanding of its application in different contexts. This approach not only enhances understanding but also saves time in learning the system in elite sports settings.<sup>53,61</sup> This approach supports KPI interpretation and consistency, enhancing the overall validity and robustness of the sports performance analysis system.

### *Familiarisation of data collection systems*

The findings indicate that experienced APAs engage in familiarisation with new sports performance analysis systems ‘Very Frequently’, whilst those early-career APAs ‘Occasionally’. This suggests a greater awareness among experienced APAs of the need to understand new sports performance analysis systems for capturing accurate performance information. Recent research highlights the importance of this familiarity, specifically when new sports performance analysis templates or definitions are created or changes are made to existing processes.<sup>18,21,63</sup> Familiarisation or piloting of the sports performance analysis system is likely to reduce errors and ensure users understand its functions, making it advisable to incorporate this process into an APA’s workflow.<sup>17,64</sup>



There is considerable variation in the time spent on this stage. For instance, Theoretical Performance Analysts have been shown to spend less than an hour before completing a reliability test, while other Theoretical Performance Analysts engaged in four weeks of intensive training.<sup>4,65</sup> Painczyk et al.<sup>18</sup> employed a seven-stage training protocol to understand the system, but no specific length of time was provided for this training period. Our findings reflect similar variation among APAs, influenced by factors such as the magnitude of KPIs, changes in KPIs or definitions and individual confidence levels. For example, APAs dealing with higher magnitude KPIs may adopt more rigorous familiarisation processes to ensure accuracy. While frequent changes in KPIs or their definitions can lead to a preference for informal methods to quickly integrate new information. Learning theories suggest sufficient time and structured practice are essential for mastering complex skills and systems.<sup>66</sup> Ericsson's<sup>52</sup> work on time-critical communication highlights the importance of adequate exposure time to ensure consistent low latency and high system usage and learning skills reliability. While he does not specify an exact duration, proficiency can only be achieved through sufficient practice and familiarisation, such as over 20 h, depending on the system's complexity. Therefore, establishing minimum duration guidelines for familiarisation, informed by theory, could improve data consistency and reliability.

Similarly, ambiguity exists regarding the nature of familiarisation activities. Experienced APAs tend to adopt formal processes and dedicate specific time to learning new systems or changes over early-career APAs. This preference for formal processes among experienced APAs may stem from their recognition of the complexities involved in data analysis and the need for structured learning to ensure accuracy and efficiency. However, informal processes were still preferred by all APAs, likely due to the flexibility and immediacy they offer in adapting to new information or tools. Uncertainty persists regarding the specific components and activities of familiarisation. We found some APAs engage in 'Seeking Confirmation' by showing their work to coaches and other APAs ( $n = 37$ , 23.3%), while others undertake 'Reliability Testing Processes' ( $n = 7$ , 4.4%). The preference for informal processes over formalised ones could be attributed to the collaborative nature of the work environment and the value placed on real-time problem-solving. Martin et al.<sup>24</sup> illustrate the practical challenges faced by APAs, with one stating, "Time is your enemy as an analyst" (p.841), which highlights the critical importance of finding sufficient time for thorough familiarisation. This diversity emphasises the varied approaches to familiarisation and highlights the necessity for clear guidelines outlining recommended activities. By defining such activities, organisations can ensure that APAs receive appropriate preparation, enhancing both the reliability of data collection and the consistency of analytical practices.

APAs generally spend less time familiarising themselves with secondary data collection systems compared to primary data systems, reflecting the perceived importance of primary data in performance evaluation. This prioritisation suggests that APAs focus more on systems directly impacting their sports performance analysis tasks, possibly indicating greater trust in the reliability of established secondary data providers such as Opta, Wyscout, and StatsBomb. Companies like Opta are known to employ rigorous formal processes of user training,<sup>19,67</sup> highlighting the need for a formalised approach. However, the specifics of these processes are closely guarded, which further complicates understanding and learning from their familiarisation practices. The limited time available during the season, with games occurring weekly, exacerbates this issue. As captured earlier from one APA, this time constraint forces individuals to prioritise primary data systems, which are seen as more critical to immediate performance evaluation needs. Implementing formalised familiarisation practices with systematic instruction, practice opportunities and feedback loops<sup>66</sup> could enhance data accuracy and efficiency. By standardised familiarisation processes and providing systematic training across different experience levels, APAs can improve the accuracy and efficiency of data collection.<sup>12</sup> Establishing such practices would help mitigate the impact of time constraints, ensuring that both primary and secondary data systems are effectively utilised.

### Reliability

Historically, sports performance analysis has relied on manual data capture using software such as Hudl SportsCode, SBG Focus, and FulcrumTec Angles, among others. This method is susceptible to human error, which can significantly impact data reliability,<sup>12,26</sup> especially if data collectors have not received sufficient training. Even with automated collection techniques beginning to exist, algorithmic or optical limitations can still impact data collection.<sup>26,68</sup> Ensuring reliability through regular checks and rectifying issues is crucial to producing data representative of observed performances, aiding in meaningful objective feedback for coaching, player and stakeholder decisions.<sup>18</sup>

Whilst the APAs acknowledged the need for conducting reliability checks, they only occasionally undertook checking methods at the start of the season or in performance blocks (see Table 5), with no significant differences observed based on experience levels or the type of data (primary vs. secondary). However, experienced APAs tend to check data quality more frequently, indicating a potential correlation between experience and reliability practices. This suggests that more experienced APAs are potentially more aware of reliability issues by their experience of seeing the impact of unreliable data within applied practice provision.<sup>69</sup>

While some Theoretical Performance Analysts have assessed the reliability of secondary data providers (e.g. Opta, Wyscout, and StatsBomb),<sup>19,36</sup> this process has not been undertaken universally across the wide variety of data providers on offer. This lack of universal assessment means that the reliability of data from many providers remains uncertain, potentially impacting the accuracy of the data and insights. Similarly, the internal reliability results, that are presumed to be undertaken, are also unfortunately not shared publicly. This lack of transparency further complicates the ability of APAs to trust and effectively use secondary data sources. Therefore, APAs should review secondary data sources to understand where (if any) reliability issues are evident within the data. By conducting their own assessments, APAs can ensure that the data they rely on meets the necessary standards of accuracy and consistency. This practice not only enhances the quality of their analyses but also builds a more robust foundation for performance evaluation.

APAs commonly relied on quick manual checking methods such as visual inspection (i.e. 'the eye test'), discussions with colleagues, or reviewing frequency tables. These methods typically serve as a quick fix to identify observational errors (e.g. missing data entries) or definitional errors (e.g. unsure of how to categorise a behaviour).<sup>70,71</sup> However, these approaches often fail to address the root cause of errors, as they are prone to confirmation bias, lack structure, involve only partial examination, fail to spot persistent error trends, are employed under time constraints, and do not uncover complex, multifaceted issues,<sup>72</sup> often leading to their reoccurrence. Confirmation bias, for instance, can lead APAs to overlook errors that do not fit their expectations, while the lack of structure in manual checks means that not all data is scrutinised equally. This can result in persistent errors going unnoticed and uncorrected, ultimately compromising the quality of the analysis. A minority of experienced APAs have implemented automated systems using programming languages like R and Python for error detection or created scripts in sports performance analysis software to visually display data accuracy. This shift towards more efficient reliability checks should be encouraged because automated systems can systematically and consistently identify errors, reducing the likelihood of human bias and oversight. However, developing these programming languages and scripts requires significant time and expertise. Despite this, the investment in automated systems is justified by the long-term benefits of improved data accuracy and efficiency. By reducing the time spent on manual checks and minimising errors, APAs can focus more on in-depth analysis and interpretation, ultimately enhancing the quality and impact of their work.

In the current study, 13 APAs mentioned conducting statistical tests like Percentage Error, Kappa, T-Test, Altman plots, and Coefficient of Variation. However,

there was no consensus on the preferred or optimum methods. This lack of standardised data-checking procedures mirrors the absence of specific methods reported by Theoretical Performance Analysts in published research papers,<sup>21,73,74</sup> despite O'Donoghue<sup>26</sup> highlighting the importance of reliability procedures almost 25 years ago. This raises questions about progress in understanding and implementing reliability processes by both Theoretical Performance Analysts and APAs.

Early guidance recommended that APAs complete intra- and inter-observer reliability tests and assess the data through methods like percentage error,<sup>75</sup> confidence intervals,<sup>76</sup> or absolute agreement.<sup>33</sup> More recently, O'Donoghue and Hughes<sup>12</sup> recommend using Kappa or Weighted Kappa when data collection involves "neighbouring" values, such as pitch or court locations. Despite this, percentage error and Kappa remain favoured approaches, with Theoretical Performance Analysts often using 5–10% percentage error bands or 0.8–1.0 Kappa values to indicate acceptable reliability.<sup>21</sup> While these values may be appropriate for some situations, they may be too stringent in others. Therefore, it might be more appropriate for APAs to select a target value based on the sport, scenario or KPI being assessed.<sup>12</sup> However, this approach may not effectively guide APAs when faced with the decision of where to set their reliability targets, highlighting the need for more tailored guidelines that consider the specific context of their work.

## Conclusion

As a much-needed contribution to the field, this work brings APA validity, familiarisation and reliability practices to the fore, marking a crucial advancement from previous Theoretical Performance Analysts' recommendations that often overlook barriers faced by APAs. We observed notable differences based on experience levels, particularly in collaborative decision-making processes for aligning KPIs and operational definitions with playing philosophy and in the implementation of robust reliability processes. Experienced APAs expressed greater satisfaction with their processes, highlighting their deeper understanding and effectiveness in navigating these systems. However, both experienced and early-career APAs acknowledged the ongoing need to refine operational definitions and strengthen reliability procedures amidst time constraints, posing a notable challenge.

The insights from APAs have been essential in shaping a best-practice recommendations framework aimed at enhancing the accuracy and meaningfulness of sports performance analysis data. The generation of the PRECISE recommendation framework, which prioritises meticulous data collection and analysis practices, will assist APAs in achieving valid and reliable sports performance analysis data. These recommendations ensure that APAs are well-familiarised with the data collection system and that reliability is maintained

throughout the data collection, analysis, translation and feedback stages. These recommendations benefit not only APAs but also provide valuable insight for stakeholders, contributing to the overall development of the APA role and the broader field of sports performance analysis. By adopting these best practices recommendations, APAs can generate more accurate and reliable data, fostering a more informed and effective environment for learning, decision making and performance improvement. This, in turn, leads to the fostering of trusting relationships and enhancing the impact of the APA role.<sup>9,24,40</sup> Subsequently, future research should explore how APA practices have evolved, the effectiveness of training programmes for APAs and the influence of contextual factors on the implementation of validity and reliability procedures. These research areas will help further enhance data reliability practices and the overall effectiveness of sports performance analysis.

### Acknowledgements

The authors wish to thank the Applied Performance Analysts who completed the survey, without whom this study would not have been possible. The authors would also like to thank Tom Corden, Daniel Ferrar, Andy Filer, Connor Flanagan, Dan Hiscocks and Denise Martin for either reviewing the survey or the PRECISE framework. Their suggestions were invaluable in connecting the theoretical aspects of our work to practical applications.

### Data availability

The raw data file of the survey responses can be found here: <https://eprints.worc.ac.uk/id/eprint/14170>

### Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### ORCID iDs

John William Francis  <https://orcid.org/0000-0001-7457-5665>  
 Scott Benjamin Nicholls  <https://orcid.org/0000-0002-6200-1418>

### Supplemental material

Supplemental material for this article is available online.

### Notes

1. *Validity is generally referred to as the ability of a measurement tool to reflect what it is designed to measure. Therefore, do the variables and performance indicators with the performance analysis system/code window/mark up collect what we want them to collect in order to aid an individual's or team's learning and decision making.*
2. *Reliability refers to the reproducibility of values of a test or other measurement in repeated trials on the same individuals*

*and repeatability over different observers. Therefore, if we or others were to collect the same variables and performance indicators of the same performance, would the data be the same.*

3. Primary video and data analysis: the initial collection and examination of video footage and data by an APA.
4. Pre-analysis: Utilising video and data to provide insights in preparation for the performance or match.
5. Post-analysis: Utilising video and data to provide insights following the performance or match.
6. Secondary data and video analysis: refers to the analysis of video footage and data that is not directly collected by the primary analyst but is sourced from other providers.
7. Live analysis: Analysts observe and record events in real-time during the performance or match.

### References

1. Mehta S, Furley P, Raabe D, et al. Examining how data becomes information for an upcoming opponent in football. *Int J Sports Sci Coach* 2024; 19: 978–987. Epub ahead of print 2023.
2. Francis JW, Owen A and Peters Derek M. Making every “point” count: identifying the key determinants of team success in elite men’s wheelchair basketball. *Front Psychol* 2019; 10: 1–14.
3. Martin D, O Donoghue PG, Bradley J, et al. Developing a framework for professional practice in applied performance analysis. *Int J Perform Anal Sport* 2021; 21: 845–888.
4. Thomson E, Lamb K and Nicholas C. The development of a reliable amateur boxing performance analysis template. *J Sports Sci* 2013; 31: 516–528.
5. Araújo D. Variables characterising performance and performance indicators in team sports. In: P Passos, D Araújo and A Volossovitch (eds) *Performance analysis in team sports*. Abingdon: Routledge, 2017, pp.38–52.
6. Watson N, Durbach I, Hendricks S, et al. On the validity of team performance indicators in rugby union. *Int J Perform Anal Sport* 2017; 17: 609–621.
7. Lames M. *Performance Analysis in Game Sports: Concepts and Methods*. Switzerland: Springer Nature, 2023.
8. McKenna M, Cowan DT, Stevenson D, et al. Neophyte experiences of football match analysis: a multiple case study approach. *Res Sports Med* 2018; 26: 306–322.
9. Nelson L, James N, Nicholls S, et al. The importance of appearing trustworthy in the workplace: performance analysts’ perspectives. *Sport Educ Soc* 2023: 1–17. Epub ahead of print 2023. DOI: 10.1080/13573322.2023.2288844.
10. Bateman M and Jones G. Strategies for maintaining the coach–analyst relationship within professional football utilizing the COMPASS model: the performance analyst’s perspective. *Front Psychol* 2019; 10: 1–12.
11. Clemente FM, Praça G, Oliveira R, et al. A systematic review of the criterion validity and reliability of technical and tactical field-based tests in soccer. *Int J Sports Sci Coach* 2022; 17: 1462–1487.
12. O’Donoghue P and Hughes M. Reliability issues in sports performance analysis. In: M Hughes, IM Franks and H Dancs (eds) *Essentials of performance analysis in sport*. New York, NY: Routledge, 2020, pp.143–160.

13. Almeida CH, Volossovitch A and Duarte R. Penalty kick outcomes in UEFA club competitions (2010-2015): the roles of situational, individual and performance factors. *Int J Perform Anal Sport* 2016; 16: 508–522.
14. Campos F, Pellegrinotti Í, Pasquarelli B, et al. Effects of game-location and quality of opposition in futsal league. *Int J Perform Anal Sport* 2015; 15: 598–607.
15. Hughes M, Caudrelier T, James N, et al. Moneyball and soccer: an analysis of key performance indicators of elite male soccer players by position. *Journal of Human Sport and Exercise* 2012; 7: 402–412.
16. Jayal A, McRobert A, Oatley G, et al. Sports analytics applications in soccer. In: *Sports analytics: analysis, visualisation and decision making in sports performance*. Abingdon: Routledge, 2018, pp.220–244.
17. Cullinane A, Davies G and O'Donoghue P. *An introduction to performance analysis of sport*. London: Routledge, 2024. Epub ahead of print January 18, 2024. DOI: 10.4324/9781003375463.
18. Pajczyk H, Hendricks S and Kraak W. Intra and inter-reliability testing of a South African developed computerised notational system among Western Province Club Rugby Coaches. *Int J Sports Sci Coach* 2018; 13: 1163–1170.
19. Liu H, Hopkins W, Gómez M-Á, et al. Inter-operator reliability of live football match statistics from OPTA sportsdata. *Int J Perform Anal Sport* 2013; 13: 803–821.
20. Goodman J, Ferguson K and Paradis K. Corner kick characteristics: a case study of the 2020/21 Northern Ireland Football League (NIFL) premiership season. *Int J Perform Anal Sport* 2024; 1–24. Epub ahead of print 2024. DOI: 10.1080/24748668.2024.2316334.
21. Francis JW, Owen A and Peters DM. A new reliable performance analysis template for quantifying action variables in elite men's wheelchair basketball. *Front Psychol* 2019; 10: 1–16.
22. Gao H, Liu W and Zhang H. Analysis of players' stroke performance relevance in tennis matches. *Int J Perform Anal Sport* 2024; 24: 535–556. Epub ahead of print 2024.
23. Barbero JR, Lapresa D, Arana J, et al. An observational analysis of kicker–goalkeeper interaction in penalties between national football teams in international competitions. *Int J Perform Anal Sport* 2023; 23: 196–212.
24. Martin D, O'Donoghue PG, Bradley J, et al. Identifying the characteristics, constraints, and enablers to creating value in applied performance analysis. *Int J Sports Sci Coach* 2024; 19: 832–846.
25. Hughes M. Analysis of notation data: reliability. In: M Hughes and I Franks (eds) *Essentials of performance analysis in sport*. Abingdon: Routledge, 2015, pp.169–179.
26. O'Donoghue P. Reliability issues in performance analysis. *Int J Perform Anal Sport* 2007; 7: 35–48.
27. Bradley P, Lago-Peñas C, Rey E, et al. The effect of high and low percentage ball possession on physical and technical profiles in English FA Premier League soccer matches. *J Sports Sci* 2013; 31: 1261–1270. Epub ahead of print 2013.
28. O'Donoghue P and Longville J. Reliability testing and the use of statistics in performance analysis support: a case study from an international netball tournament. In: P O'Donoghue and MD Hughes (eds) *Performance analysis of sport 6*. Cardiff: CPA UWIC Press, 2004, pp.1–7.
29. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; 20: 37–46.
30. Cohen J. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull* 1968; 70: 213–220.
31. Robinson G and O'Donoghue P. A weighted kappa statistic for reliability testing in performance analysis of sport. *Int J Perform Anal Sport* 2007; 7: 12–19.
32. Bland M and Altman D. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999; 8: 135–160.
33. Cooper SM, Hughes M, O'Donoghue P, et al. A simple statistical method for assessing the reliability of data entered into sport performance analysis systems. *Int J Perform Anal Sport* 2007; 7: 87–109.
34. Thomson E and Lamb K. The technical demands of amateur boxing: effect of contest outcome, weight and ability. *Int J Perform Anal Sport* 2016; 16: 203–215.
35. Worsfold P and Macbeth K. The reliability of television broadcasting statistics in soccer. *Int J Perform Anal Sport* 2009; 9: 344–353.
36. Stanway B. Normalising data: definitions and differences. In: USW FLSE research conference, Treforest, 2018.
37. O'Donoghue P. Editorial: special issue on reliability. *Int J Perform Anal Sport* 2007; 7: i–ii.
38. FIFA. The future is now: FIFA bringing performance analytics to a whole new level, <https://www.fifa.com/tournaments/mens/arabcup/arabcup2021/media-releases/the-future-is-now-fifa-bringing-performance-analytics-to-a-whole-new-level> (2021, accessed 29 July 2022).
39. Wright C, Atkins S, Jones B, et al. The role of performance analysts within the coaching process: performance analysts survey 'the role of performance analysts in elite football club settings'. *Int J Perform Anal Sport* 2013; 13: 240–261.
40. Francis JW, Kyte J and Bateman M. The role of the analyst: comparative analysis of applied performance analyst job advertisements in the UK and Ireland (2021-2022). *Int J Perform Anal Sport* 2024; 24: 331–360. Epub ahead of print 2023.
41. University of Worcester. Applied sports performance analysis MSc. *University of Worcester*, <https://www.worcester.ac.uk/courses/sport-performance-analysis-msc> (2024, accessed 3 May 2024).
42. University of Derby. Performance analysis and coaching science BSc (Hons), <https://www.derby.ac.uk/undergraduate/sport-exercise-science-courses/performance-analysis-coaching-science-bsc-hons/> (2024, accessed 19 June 2024).
43. Parker C, Scott S, Geddes A. Snowball sampling. In: P Atkinson, S Delamont, A Cernat, et al. (eds) *Sage research methods foundations*. Sage. Epub ahead of print 2019. DOI: 10.4135/9781526421036831710.
44. Hughes M. Performance analysis – a 2004 perspective. *Int J Perform Anal Sport* 2004; 4: 103–109.
45. Martin D, Swanton A, Bradley JJ, et al. The use, integration and perceived value of performance analysis to professional and amateur Irish coaches. *Int J Sports Sci Coach* 2018; 13: 520–532.
46. Gratton C and Jones I. *Research methods for sport studies*. Abingdon: Routledge, 2010.
47. Ball HL. Conducting online surveys. *J Hum Lact* 2019; 35: 413–417.
48. Nosek P, Brownlee TE, Drust B, et al. Feedback of GPS training data within professional English soccer: a comparison of decision making and perceptions between coaches, players and performance staff. *Sci Med Football* 2021; 5: 35–47.

49. Hopkins W. A spreadsheet for deriving a confidence interval, mechanistic. *Sportscience* 2007; 11: 16–20.
50. Braun V and Clarke V. *Thematic analysis: a practical guide*. London: Sage Publications Ltd, 2022.
51. Marton F. *Necessary conditions of learning*. Routledge, 2014, Epub ahead of print June 20, 2014. DOI: 10.4324/9781315816876.
52. Anders Ericsson K. Deliberate practice and acquisition of expert performance: a general overview. *Acad Emerg Med* 2008; 15: 988–994.
53. Larkin P, Mesagno C, Berry J, et al. Development of a valid and reliable video-based decision-making test for Australian football umpires. *J Sci Med Sport* 2014; 17: 552–555.
54. Lucas D, Stokes K, McGuigan P, et al. Consensus on a jockey's injury prevention framework for video analysis: a modified Delphi study. *BMJ Open Sport Exerc Med* 2022; 8: e001441. Epub ahead of print December 15, 2022.
55. Osório A. Performance evaluation: subjectivity, bias and judgment style in sport. *Group Decis Negot* 2020; 29: 655–678.
56. Silva H and Marcelino R. Inter-operator reliability of InStat Scout in female football games. *Sci Sports* 2023; 38: 42–46.
57. Larkin P, O'Connor D and Williams M. Establishing validity and reliability of a movement awareness and technical skill (MATS) analysis instrument in soccer. *Int J Perform Anal Sport* 2016; 16: 191–202.
58. Coutts AJ. In the age of technology, Occam's razor still applies. *Int J Sports Physiol Perform* 2014; 9: 741.
59. Bruce L, Dwyer D and Fox A. Reliability of live and video-based coding in netball using the NetballStats application. *PLoS One* 2022; 17: e0269330. Epub ahead of print June 1, 2022.
60. Butterworth A. *Professional practice in sport performance analysis*. London: Routledge, 2023.
61. Wonwoo J, Lewis CJ, Evans M, et al. The validity and reliability of an integrated approach for quantifying match physical-tactical performance. *Biol Sport* 2022; 39: 253–261.
62. Barlow L, Partington M, Doncaster G, et al. The strategic delivery of video-based coaching: a season-long investigation in one English youth soccer academy. *Int J Sports Sci Coach* 2024; 19: 1897–1904. Epub ahead of print May 21, 2024.
63. O'Donoghue P. *Research methods for sports performance analysis*. London: Routledge, 2010. Epub ahead of print 2010. DOI: 10.4324/9780203878309.
64. McLean S, Salmon PM, Gorman AD, et al. What's in a game? A systems approach to enhancing performance analysis in football. *PLoS One* 2017; 12: 1–15.
65. Tenga A, Kanstad D, Ronglan LT, et al. Developing a new method for team match performance analysis in professional soccer and testing its reliability. *Int J Perform Anal Sport* 2009; 9: 8–25.
66. Ericsson KA. Development of elite performance and deliberate practice: an update from the perspective of the expert performance approach. In: JL Starkes and A Ericsson (eds) *Expert performance in sports. Advances in research on sport expertise*. Champaign, IL: Human Kinetics, 2003, pp.49–83.
67. Soroka A, Duda H, Stuła A, et al. Ball possession as an indicator of differences in game effectiveness of football teams during the FIFA World Cup Qatar 2022. *Journal of Kinesiology and Exercise Sciences* 2023; 33: 9–20.
68. Araújo D, Couceiro M, Seifert L, et al. *Artificial intelligence in sport performance analysis*. London: Routledge, 2021.
69. Nicholls SB, James N, Wells J, et al. Performance analysis practice within Olympic and Paralympic sports: a comparison of coach and analyst experiences. *Int J Perform Anal Sport* 2022; 22: 343–351.
70. James N, Taylor JB and Stanley S. Reliability procedures for categorical data in performance analysis. *Int J Perform Anal Sport* 2007; 7: 1–11.
71. Stanway B and Boardman P. The analysis process: applying the theory. In: E Cope and M Partington (eds) *Sports coaching: a theoretical and practical guide*. London: Routledge, 2020, pp.93–108.
72. Williams AM, Ford PR, Eccles DW, et al. Perceptual-cognitive expertise in sport and its acquisition: implications for applied cognitive psychology. *Appl Cogn Psychol* 2011; 25: 432–442.
73. Hughes M, Cooper S-M and Nevill A. Analysis procedures for non-parametric data from performance analysis. *Int J Perform Anal Sport* 2002; 2: 6–20.
74. Williams J. Operational definitions in performance analysis and the need for consensus. *Int J Perform Anal Sport* 2012; 12: 52–63.
75. Hughes M, Cooper S-M and Nevill A. Analysis of notation data: reliability. In: M Hughes and I Franks (eds) *Notational analysis of sport: systems for better coaching and performance in sport*. Abingdon: Routledge, 2004, pp.189–204.
76. Cooper SM. *Statistical methods for resolving issues relevant to test and measurement reliability and validity in variables related to sport performance and physical fitness*. PhD Thesis, University of Wales Institute Cardiff, Cyncoed Campus, 2006.

Appendix

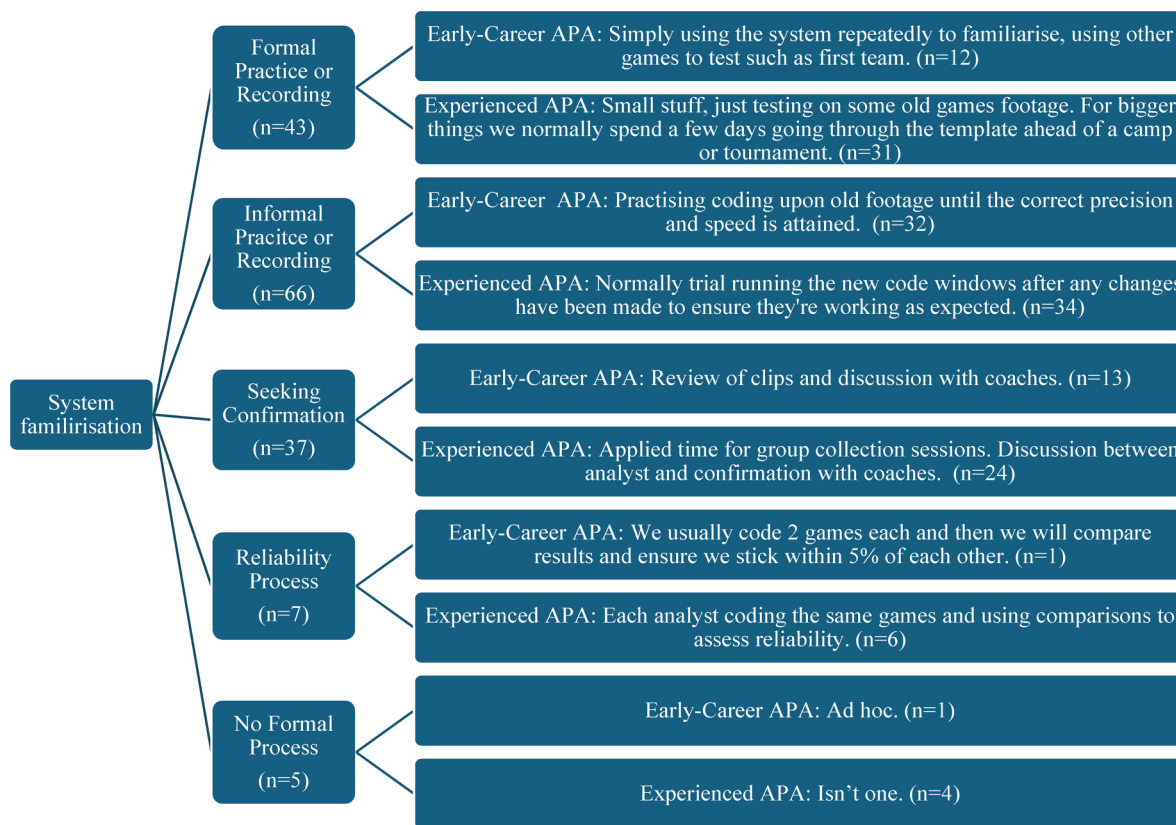
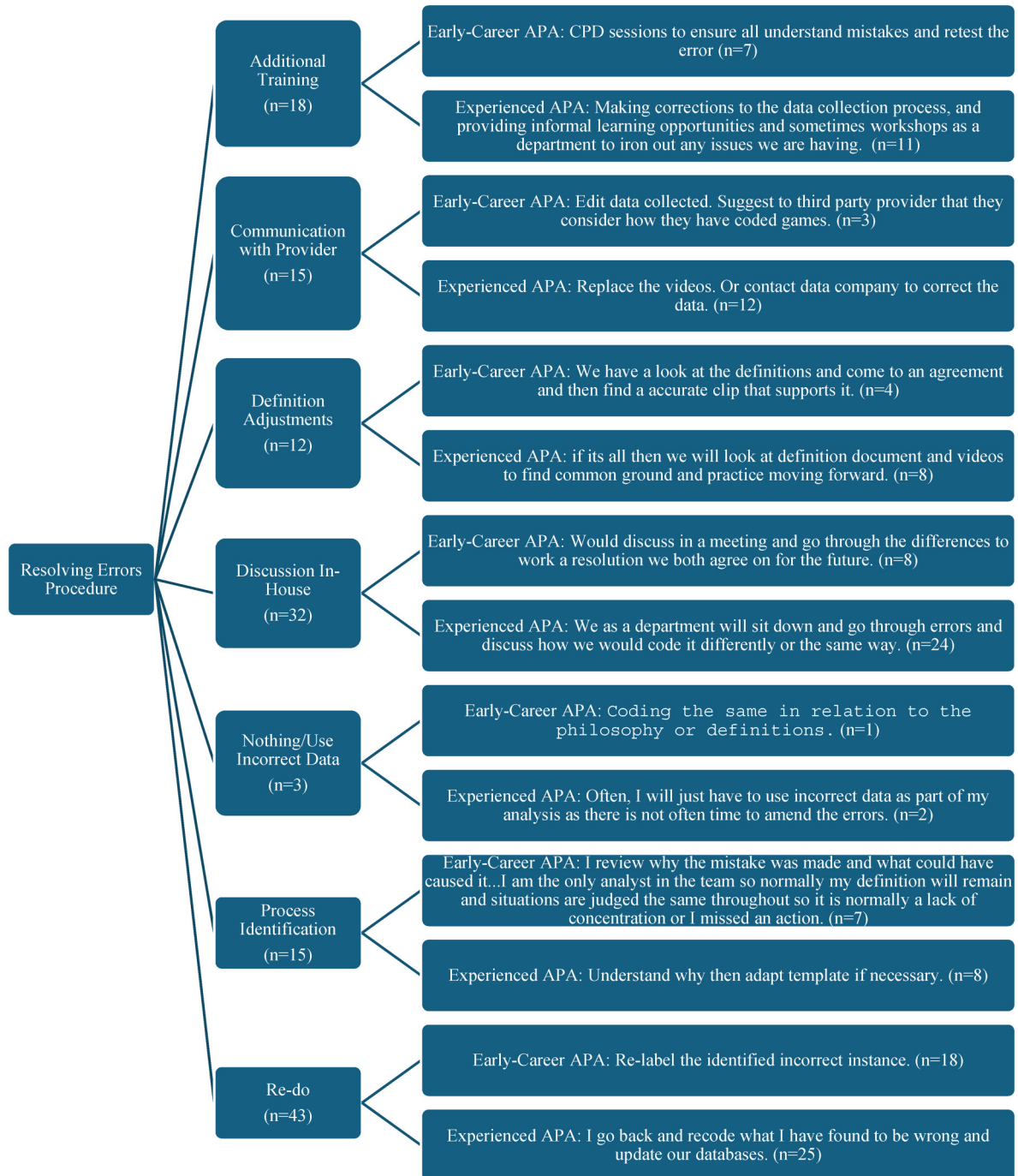


Figure A1. Thematic analysis of open-ended responses regarding system familiarisation processes.



**Figure A2.** Thematic analysis of open-ended responses regarding resolving error procedures.