THE USE OF THINK ALOUD METHOD TO INVESTIGATE DECISION-MAKING PROCESSES OF WOMEN'S ARTISTIC GYMNASTICS JUDGES

JENNE LEE LING HUEY

A thesis submitted in partial fulfilment of the requirements of

Liverpool John Moores University for the degree of Doctor of Philosophy

FEBRUARY 2020

ABSTRACT

Decision-making of officials in aesthetic sports and more specifically in Women's Artistic Gymnastics (WAG) has received some attention within the sport psychology literature. However, research conducted has mainly utilised post-competition scores and has adapted retrospective methods of data collection to investigate judges' thoughts and decision-making during judging. Think Aloud (TA) method (Eccles & Arsal, 2017) proposed by Ericsson and Simon (1993) is a tool to collect concurrent data of cognitive processes, and therefore could be an alternative method to collect judge decision concurrently. As a result, this thesis aimed to investigate the robustness of a novel TA method in collecting thought processes of WAG judges and to explore the decision-making differences between expert and novice judges. There were three studies included in this thesis. Study One was conducted in Malaysia 2016 utilising the Code of Points (COP) 2012-2016, whereas Study Two and Study Three were conducted simultaneously in the United Kingdom from 2017-2018 utilising the updated COP 2017-2020.

Study One

Study One aimed to explore the decision-making underpinning judging processes by using both concurrent and immediate retrospective methods. This examined the utilisation of TA method as a training tool to develop Malaysia based WAG judge education. Ten qualified national judges were required to verbalise their thought processes in applying execution deductions and artistry deductions by using *Level 2* TA when judging a fix-sequenced video clip consists of ten routines on a singular apparatus, i.e. Balance Beam. Immediate follow-up interviews were conducted to investigate the judge's perceptions of using TA method whilst judging. Data collected in the TA sessions and follow-up interviews underwent thematic analysis (Braun & Clarke, 2014). During the judging process, participants verbally reported most frequently on *lack of balance, bending of arms and knees, pointing of feet, confidence,*

rhythm and tempo, and *personal style* as focal points on deductions. Overall TA method was reported as an appropriate tool for use within judge education to enable deduction scores to be applied objectively. However, some participants reported performance of the primary task on judging was adversely affected by verbal overshadowing. This study informed Study Two and Study Three to investigate the decision-making differences between expert and novice judges in addition to the viability of TA method extending to all four WAG apparatus.

Study Two

The aim of this study was to examine decision-making differences between ten expert (international and national judge) and eight novice (regional and club judges) WAG judges based in the United Kingdom in evaluating Balance Beam (BB), Floor Exercise (FX), Uneven Bars (UB), and Vault (VT) routines using fixed-sequenced competition video clips. Participants using *Level 2* TA method to verbalise all execution deductions concurrently where possible and artistry deductions by immediate retrospective whilst judging video-based routines that resembled actual competition with execution scores calculated at the end of each routine. Execution scores and verbalised deduction counts were tabulated into IBM® SPSS Statistics 24 and Microsoft Excel® for data analyses. Results showed that expert judges applied more deductions across all apparatus when compared to novice judges ($p < .05$). Further, verbalisation on deduction for all three types of deductions, that of general execution faults, specific apparatus deductions, and artistry deductions, were higher in counts by expert judges compared to novice judges ($p < .05$). These identified there were expert-novice differences in judging execution scores and applying deductions across all four apparatus in WAG as hypothesised. The highest count for deductions verbalised was recorded on the BB, followed by FX, UB, and least on VT revealing there were different deduction applications according to respective apparatus characteristics. Based on these findings, it was suggested to

further explore judges' perceptions of using TA method when judging to inform the future use of TA for both research and education purposes.

Study Three

This study conducted simultaneously with Study Two, aimed to investigate the perceptions of expert-novice WAG judges in using a novel TA method whilst judging video-based competition routines. Semi-structured interviews were conducted with the same participants from Study Two to collect their perceptions of using TA method to verbalise execution deductions whilst judging video-based routines across all four WAG apparatus concurrently and immediate retrospectively using *Level 2* TA to inform viability of TA method into future judge education development. Interview data underwent inductive and deductive thematic analyses. Five themes were generated, which were feelings of using TA method whilst judging WAG routines, perceptions of TA method viability within WAG judging, and TA method challenges within WAG judging, TA deductions across different WAG apparatus, and further consideration for learning resources adapting TA method. Both expert and novice judges reported initial apprehensions to verbalise deductions whilst judging, however, they reported become more comfortable and more used to the TA method across the sessions demonstrating skill acquisition. Furthermore, both expert and novice judges reported an increased awareness in applying more accurate execution deductions concerning of schemata and prior performance by previous gymnasts when using TA method without disrupting the natural thought process of judging in slower-pace apparatus. In addition, participants suggested TA as a viable method to collect in-event thought processes of WAG judges in slower-paced apparatus of that BB, FX, and UB concurrently. Fast-paced apparatus that of VT required the use of immediate retrospectively recall, despite concerning verbal overshadowing of TA method and existing multitask judging. Participants further suggested that a progressive judge education module adapting TA method beginning with

generic training, into single skill/element evaluation, then series of gymnastics skills and dance elements towards full routines to be included in future judge education. Therefore, TA method adapted in judge education module with utility to extend beyond current course delivery and ‘paper and pen’ assessment as well as providing a learning source by which to refresh and retrain judges after accreditation examinations.

The findings of these empirical studies suggest that TA method is viable to collect in-event data of cognitive processes among WAG judges when judging video-based routines. Subsequently, findings suggest that there are decision-making differences between expert and novice WAG judges related to experience and training acquired according to the level of judge accreditation. Expert judges applied significantly more execution deductions compared to novice judges across all four apparatus in WAG, i.e. BB, FX, VT, and UB. Therefore, TA method been suggested viable to scale-up novice judges in understanding their thought processes to make accurate decisions when evaluating gymnastics skills and dance elements from a practical perspective therefore suggested for TA method adaption into a pilot judge education module.

DECLARATION

I declare that the work contained in this thesis is my work.

Publication Output from the PhD

Study One:

Lee, J., Knowles, Z., & Whitehead, A. E. (2019). Exploring the use of think aloud within Women's artistic gymnastics judging education. *Psychology of Sport &* Exercise, 40, 135-142. doi:10.1016/j.psychsport.2018.10.007

ACKNOWLEDGEMENT

I would like to thank my Director of Study, Professor Zoe Knowles and supervisors, Dr Amy Whitehead and Dr Joe Causer for the continuous guidance, advice and support with patience throughout the course of my PhD. Your on-going encouragements throughout the course of PhD were greatly appreciated.

I would like to express heartfelt gratitude to the Ministry of Education, Malaysia in offering the Federal Training Scholarship to fund this PhD that provided me an opportunity to pursue this doctorate study.

I really appreciate the opportunities involving in the British Gymnastics, United Kingdom as an accredited National judge during the course of pursuing my PhD. The judging experience accumulated to judge competitions across club, regional, national, and invitational competitions for the past four years provided me an insider view to complete this thesis. Thank you The City of Liverpool Gymnastics Club and British Gymnastics.

I would like to express gratitude to the School of Sport and Exercise Sciences and The Doctorate Academy of Liverpool John Moores University, United Kingdom for continuous supports throughout the duration to facilitate my postgraduate student journey with trainings and assistance.

Last but not least, I would like to extend my gratitude to my family and friends who provided me with endless support, understanding, and love. None of this would have been possible without you. Thank you!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY OF TERMS

International judge	Women's Artistic Gymnastics Judge accredited by the Federation of International Gymnastics (FIG) to judge in any competition internationally. International judges are also known as "Brevet Judge".
National judge	Women's Artistic Gymnastics Judge accredited by the Malaysia Gymnastics Federation (Study One) and British Gymnastics (Study Two and Study Three) to judge at nationwide competitions in Malaysia and United Kingdom respectively.
Regional judge	Women's Artistic Gymnastics Judge accredited by British Gymnastics to judge at club and regional level competitions with opportunity to lead the judging panel.
Club judge	Women's Artistic Gymnastics Judge accredited by British Gymnastics to judge at club and regional level competitions as a member of the judging panel.
TA method	Defined as a research tool "for studying thinking that qualitative researchers from any disciplinary background can consider as an option for understanding thought" (Eccles & Arsal, 2017, p. 2). TA method was originally stated as "thinking-aloud protocols" (Ericsson & Simon, 1993) and has been subsequently cited as "think-aloud protocol" and "think aloud protocol" within the academic literature.
Time restriction	60 seconds provided for WAG judges to complete evaluation processes for a routine during a competition.
Verbal overshadowing	Performance of primary task distracted by verbalisations, with possible to slow down or even withhold solution attempts, even leading to longer completion time on primary task.

LIST OF ABBREVIATIONS

Abbreviation	Full-term
TA	Think Aloud
WAG	Women's Artistic Gymnastics
OG	Olympic Games
COP	Code of Point
FIG	Federation of International Gymnastics
BB	Balance Beam
FX	Floor Exercise
VT	Vault
UB	Uneven Bars
D-judge	difficulty panel judge
D-score	difficulty score
E-judge	execution panel judge
E-score	execution score
BG	British Gymnastics
MGF	Malaysia Gymnastics Federation
LTM	long-term memory
STM	short-term memory
WM	Working memory
LT-WM	Long-term working memory

CHAPTER 1

Introduction

1.1 Decision-making in Women's Artistic Gymnastics

The aesthetic sport of Women's Artistic Gymnastics (WAG) requires judges to evaluate performance of artistry in a subjectively way in conjunction with the difficulty of skills and elements presented according to the rules and regulations (Edgar, 2013; McFee, 2013). Decisions made by judges from local club and regional competition to international competitions have been reported as being ambiguous, which is reflected by variation of scores across a judging panel. Judges make their decisions based on what they see during the performance and from the sitting arrangement located at different distances and angles from the apparatus (FIG, 2016b). Therefore, an average deduction score across an execution deduction panel that excludes the highest and the lowest deduction scores is taken into a final score to avoid bias (McFee, 2013). However, statistical analyses studies conducted using post-competition scores from international competitions (Mercier & Klahn, 2017; Pajek, Kovač, Pajek, & Leskošek, 2014; Sacchi, 2018a, 2018b) revealed there were decision-making difference among judges. In addition, several studies examining judging performance (Pizzera, Möller, & Plessner, 2018; Ste-Marie, 2000) indicated there were differences between expert and novice judges when judging video-based skills and elements from the same angle of view. Experienced judges were able to utilise motor or visual experience together with knowledge from their gymnastics background to judge more accurately (Campo & Gracia, 2017; Heinen, Vinken, & Velentzas, 2012).

Expert-novice differences have been widely studied throughout the world of sports and sport psychology in addition to other domains such as medicine and education. Identifying differences and bridging the gap between expert and novice is of contemporary research interest from both a theoretical and practical perspective in all areas of sports. It is

important to facilitate less-skilled performers to become experts in a more efficient manner by understanding expert learning (Williams et al., 2017). According to Ericsson (2017), literature has consistently reported that experts with superior decision-making abilities had been exposed to long hours of sport specific activity as well as other related activity. According to the decision-making model (Memmert, 2015), central cognitive performance factors involve working memory, anticipation, perception, attention, and intelligence to support the expert superior performance. Ste-Marie (1999) proposed that stored memory of gymnastics elements, that of working short-term memory, both divided and selected attention capabilities, as well as detecting and identifying complex movement patterns were required by WAG judges to undertake an accurate evaluation. Studies examining differences between experts and novices in laboratory-settings using video clips from the same angle of view (Campo & Gracia, 2017; Pizzera et al., 2018) have shown experts judged more accurately than novices. This corroborates with another observational study conducted by Ste-Marie (2000) which explored difference between expert and novice judges at *actual* competitions whereby novice judges were less able than expert judges to engage in the multitask demands of gymnastics judging. Expert judges were better ability to anticipate upcoming gymnastics elements in a routine and possessed superior declarative knowledge about rules-based information due to their effective information retrieval from memory (Ste-Marie, 1999, 2000) to circumvent the processing limitations of actual competition. Therefore, it is timely and important to conduct research to understand the decision-making processes of expert judges in aesthetic sports set against a reasonable amount of literature focused on about decision-making more broadly.

One of the main criticisms of current literature is that majority of research was conducted using statistical scoring data at post-competitions, which does not reflect the real-time in-event decisions of judges in addition to the different angled view of judges when in

the competition setting. Another criticism within the decision-making of judges is that studies were in laboratory setting that allow judges to report their decisions in self-pace. Research is warranted that recreates the actual competition environment (McRobert, Ward, Eccles, & Williams, 2011; Williams & Ford, 2013) using videos captured from the unique perspective of judge's view angle and that mimics actual competition to test the specific cognitive and processing strategies developed by the expert judges (Ste-Marie, 1999).

The methodology of collecting in-event decision-making data with the retrospective recall of events has a number of limitation (Ericsson & Kintsch, 1995; Ericsson & Simon, 1993; Nicholls & Polman, 2008; Whitehead, Taylor, & Polman, 2015). Information provided is influenced by memory decay after completing the domain- specific tasks (Bernard, Killworth, Kronenfeld, & Sailer, 1984; Ericsson & Kintsch, 1995; Ericsson & Simon, 1993; Nicholls & Polman, 2008) and also experience with memory (Miron-Shatz, Stone, & Kahneman, 2009). Moreover, Whitehead et al. (2015) studied the congruence of verbal data collected on decision-making in golf putting using concurrent verbal reports and cued retrospective recall at different time intervals, that of 10 minutes, 24 hours, and 48 hours after which performance showed low level of congruence that of 38-41% between thoughts verbalised during in-event and retrospective recall in interviews. Therefore, more paradigms that are realistic need to be developed which reduce time delays between actions and assessment of decision-making with specific concerns of the time constraint when judging WAG in actual competitions. In addition, the use of qualitative research methodologies may provide more rich information on the decision-making processes than that of numerical data of scores indicating the final decisions made by judges alone. In an attempt to respond to this issue, Think Aloud (TA) method (Eccles & Arsal, 2017) has been increasingly used to study cognitive processes and thinking within sports (Fox, Ericsson, & Best, 2011). Data collected using TA method provides a sequence of observations *over time* rather that a single

observation at the end of a task when compared to using retrospective recall. It has been argued by some researchers (Boren & Ramey, 2000; Welsh, Dewhurst, & Perry, 2018; Whitehead et al., 2018) that the level of explanations or descriptions provided by TA method changed the sequence of thoughts, whereby additional information retrieval from memory were required. However, vocalisation of inner speech without providing explanation or description of that TA method allow researchers to collect data during performance of a task thus minimise the event-recall period and increases the likelihood of collecting accurate data did not alter performance compared to individuals completing the same task silently (Ericsson & Simon, 1993; Fox et al., 2011).

To date, there is no research investigating expert-novice judge differences when judging complete routines of all four apparatus in WAG by collecting their in-event decisions. Therefore, the proposed research aims to investigate the decision-making processes underpinning expert and novice judges to validate the use of TA method for collecting decision-making data during actual performance.

1.2 Introduction to the Thesis

This thesis is empirical and novel by way of investigating in-event decision-making of Malaysia and United Kingdom accredited WAG judges in applying execution deductions using TA method when judging video-based routines. Study One participants were a singular group of ten qualified national judges accredited by the Malaysia Gymnastics Federation to explore the decision-making underpinning judging processes of a singular apparatus of Balance Beam, in addition to examine the utilisation of TA method to adapt into future judge education. Focal points of deductions across judges were determined whilst TA method was reported viable for use within judge education despite verbal overshadowing affecting the judging performance. This informed Study Two and Three conducted simultaneously in examining the decision-making differences between expert and novice judges in evaluating

routines across all apparatus in WAG in addition to collect their perceptions in using TA method whilst judging. Ten expert judges with international and national accreditations and eight novice judges with regional and club accreditations from United Kingdom participated in these studies to evaluate all four apparatus in WAG using TA method and fixed-sequence competition videos to examine the score differences by verbalising the deductions applied in the Study Two. Follow-up interviews were conducted with each participant to collect their perceptions in using TA method across different apparatus in the Study Three.

1.3 Organisation of the Thesis

Chapter Two (Literature Review) provides a comprehensive review and critique of the current literature relating to decision-making within the aesthetic sport of WAG. The key findings of previous studies reporting TA method is viable to collect the data of cognitive processes, therefore adapted into this thesis with participants of WAG judges whilst evaluating video-based routines. This review highlight 'gaps' within the current literature, which have subsequently provided the rationale for the thesis. This chapter also states the aims and objectives of this thesis and the methodological approaches that have been employed. Chapter Three presents the Study One exploring decision-making underpinning judging processes involving ten Malaysia WAG judges with use of a singular apparatus, Balance Beam (BB) in addition to examine the utilisation of TA method as training tool for consideration into future judge education. Findings from this study have been published in the Journal of Psychology of Sport and Exercise (Lee, Knowles, & Whitehead, 2019). Therefore, this chapter report the original published article, whereby several terminologies used within this chapter are difference from the other chapters, such as FIG was denoted as 'Fédération Internationale de Gymnastique' in this specific chapter. Study Two, reported in Chapter Four, examines decision-making differences between expert and novice judges in evaluating routines across all four apparatus in WAG. Chapter Five reports Study Three to

investigate the perceptions of expert-novice WAG judges in using a novel TA method whilst judging video-based competition routines. Chapter Six provides a synthesis of the findings from the three studies and their implications in relation to the major themes of the thesis, as well as providing recommendations for future research and practice including directions for judge education.

1.4 Ethical Approval

All studies contained within this thesis received full ethical approval from the Research Ethics Committee within Liverpool John Moores University:

Study 1: 16/SPS/011

Study 2: 17/SPS/008

Study 3: 17/SPS/008

1.5 Thesis Study Map

For the purposes of the readers, a thesis study map outlining the aims, objectives, and key findings of each study in this thesis is presented below. The aim of presenting the "map" is to provide an overall view into each study that fits in with the overall thesis.

Thesis Study Map

Study 1		
Aim	Objective	Key Findings
To explore decision-making underpinning Malaysian WAG judges.	To identify focus of execution deductions in BB through TA verbalisation.	• Most frequently verbalised execution deductions were *lack of balance*, *bending arms and knees*, *feet relax*. • Most frequently verbalised artistry deductions were *confidence*, *rhythm and tempo*, and *personal style*. • Different deductions were applied by judges on a same routine.
To examine utilisation of using TA method into future WAG judge education.	To collect perceptions of using TA method during judging BB competition videos.	• TA increased cognitive process and awareness during judging. • TA method reported as an appropriate training tool for Malaysian based WAG judges, however, judging performance informed being affected by verbal overshadowing.

Study 2		
Aim	Objective	Key Findings
To explore decision-making differences between expert and novice WAG judges in UK.	To measure deduction score differences in apparatus BB, FX, VT, and UB.	• Expert judges applied higher deductions across all apparatus than novice judges ($p < .05$).
	To measure differences in execution deductions verbalised of apparatus BB, FX, VT, and UB.	• Expert judges TA more deductions across all apparatus than novice judges ($p < .05$). • Deductions were verbalised most on BB, followed by FX, UB, and VT.

Study 3		
Aim	Objective	Key Findings
To explore perceptions of expert and novice WAG judges in using TA method whilst judging video-based competition routines of apparatus BB, FX, VT, and UB.	To identify difference in TA execution deductions across BB, FX, UB, and VT routines.	• Utilisation of TA was easy on slow-pace apparatus (BB) but harder on fast-pace apparatus (VT). • Immediate retrospective TA applied to verbalise execution deductions of VT and artistry deductions of BB and FX.
	To explore utilisation of TA method in UK WAG judging.	• TA training reported viable to increase score objectivity extending beyond current writing materials. • Perceived TA benefits: increase awareness, reassurance. • Perceived TA disadvantages: initial apprehension, verbal overshadowing, multitask judging.

CHAPTER 2

Literature Review

2.1 Officiating in Aesthetic Sport

Reid (1970, p. 246) stated that "all sports are physical activities in which there is some definite practical aim or end to be achieved". According to writings within the 1970s there were two types of sports, which are 'purposive sports' and 'aesthetic sports' (Best, 1974). Aesthetics are relatively unimportant in 'purposive sports', which have objective measures for winning, such as scoring goals for football and crossing line in athletics. On the other hand, 'aesthetic sports' such as gymnastics, figure-skating, and high-board diving are sports that incorporate aesthetic evaluation whilst performing a series of movements in accordance with specific governed rules, whereby winning is determined by the subjective judgement. Best (1974, p. 198) further stated that "science is not the only way of considering what is available to sensory perception", which supports claims by Reid (1970, p. 257) that a judge is needed for aesthetic sports evaluation, as "no machine could assess original artistic value", indeed even computer-cum-slow-motion-filming machine is unable to mark movements objectively. This notion is contradicted by way of the emergence of real-time judging support systems to capture athletes' movement with technology then analyse movements as numerical data to achieve fair and accurate scores (FIG, 2019). Implementation of competition rules in aesthetic sports requires judges to be acquainted with the rules in addition to their perceptions towards aesthetic values within (Edgar, 2013) to determine scores and thus winners.

Previous studies investigating performance of judges across aesthetic sport competitions have confirmed the presence of several biases that reduce evaluation objectivity. The *patriotism effect* (Boen, Hoye, Auweele, & Feys, 2006; Leskošek, Čuk, Pajek, Forbes, & Bučar-Pajek, 2012) revealed that one judge would favour athletes from their own country,

whilst the *halo effect* was discovered to influence judge's evaluations based on an athlete's global impressions, physical appearance, and cooperate sponsorships (Nufer & Alesi, 2018). Furthermore, *order effect* was found to affect judge's expectations towards an athlete's performance level according to rank order during a competition (Kramer, 2017; Plessner, 1999). Likewise, *reputation effect* was found to influence evaluations based on the reputation of that athlete (Findlay & Ste-Marie, 2004) and *memory-influenced effect* leading to perceptual bias of judges during the evaluation process due to prior information-processing (Ste-Marie & Valiquette, 1996; Ste-Marie, Valiquette, & Taylor, 2001). *Conformity effect* i.e. that of causing judges to adapt scores to a panel, whereby normative influence affecting judges with the fear of standing out from that of the judging panel (Boen, Auweele, Claes, Feys, & Cuyper, 2006), whilst informational influence affecting judges with deficient knowledge in making accurate decisions (Auweele, Boen, Geest, & Feys, 2004; Boen, van Hoye, Auweele, Feys, & Smits, 2008). Artistic gymnastics was widely cited across these studies when discussing score objectivity. Whilst these biases have been identified, it is less understood in decision-making processes of judges (Raab, MacMahon, Avugos, & Bar-Eli, 2019) or indeed understanding the formulation of such biases which may help to promote score objectivity and fairness across competitions.

2.2 Judging Women's Artistic Gymnastics

Women's Artistic Gymnastics (WAG) began competitively in the Olympic Games (OG) from 1928 (International Olympic Committee, 2015). WAG judges have to reaccredit once in every four years following the OG cycle to ensure they possess updated knowledge for the latest Code of Point (COP), which has stated rules and regulations for the entire community within WAG, including that of the gymnasts, coaches, judges, and officials. There are several judging levels within WAG according to countries and nations. International judges are accredited by the Federation of International Gymnastics (FIG),

which qualifies judges to officiate competitions organised by the FIG internationally and events with international participants, including OG, Youth Olympic Games, World Championships, regional and intercontinental competitions. Each country and nation are responsible to accredit their national, regional, and club judges adapting the COP of current cycle with amended requirements to fit their roles. In Malaysia and United Kingdom, WAG judges are required to attend a judge course and pass an examination inclusive of both theoretical and practical sessions. Theoretical examination assesses knowledge regarding written rules within the COP, including general regulations, general deductions, recognition of gymnastics elements, as well as detailed deductions for complete routines. The practical examination assesses the accuracy of judges in evaluating routines by watching video clips representative of a competition situation.

In WAG, there are four apparatus, that of Balance Beam (BB), Floor Exercise (FX), Vault (VT), and Uneven Bars (UB). The scoring system in WAG consists of two parts, that of difficulty and execution. There are usually two difficulty judges (D-judge) in a panel, who are responsible to credit difficulty score (D-score) to a gymnast upon agreement of both judges. D-judges are to familiarise all rules and regulations embedded in the latest COP and respective competition. They also function as 'Head of Panel' to communicate with the Women's Technical Committee, who oversee the entire competition, when needed. D-judges are also responsible to monitor the warm-up time for each team and individual gymnast, as well as general deductions including attire and behaviour in a competition. During a routine performance, D-judges are to give signal to a gymnast to start, watch the routine and at the same time write down the gymnastics skills, dance elements, and specific requirements performed for the respective apparatus, such as connections of skills and elements on the notation sheet as stated in the COP or as provided by the competition organiser. D-score of a routine is awarded by agreement of both D-judges, whereby discussions among them are

allowed. The D-score includes maximum gymnastics skills and dance elements permitted for a specific apparatus, connection bonus and composition requirement for each apparatus. Specific bonus and additional rules may be applicable to specific competitions. Gymnastics skills and dance elements will be counted for D-score if they were performed up to standards as coded in the COP, otherwise they will be devalued or credited as another gymnastics skill or dance element as guided in the COP. For singular gymnastics skill perform for VT, there is a specific score table (FIG, 2016b, p. 170) provided for D-judges to ensure that the *exact* skill has been performed otherwise another element will be credited as stated in the COP. There is no dance element perform for UB, therefore only a maximum of eight gymnastics skills with highest difficulty that have fulfilling requirements will be counted for the D-score. For BB and FX, requirements are a maximum of three gymnastics skills, three dance elements, and another two optional gymnastics skills and/or dance elements with highest difficulty to be counted towards the D-score. Extra marks will be awarded to the gymnast who fulfils a maximum of four composition requirements specifically for BB, FX, and UB. Bonus marks for gymnastics skill and dance element connections are also to award gymnasts who had performed their routines according to the rules as stated in the COP as well as in respective competitions.

Table 2.1 Maximum gymnastics skills and dance elements counted for WAG D-scores

Apparatus	**Max gymnastics skill**	**Max dance element**	**Optional skill/element**	**Composition Requirement**
BB	3	3	2	4
FX	3	3	2	4
UB	8	-	-	4
VT	1	-	-	-

There are usually four to six execution judges (E-judge) in an execution panel for an apparatus, unless otherwise adjusted according to the competition requirements. E-judge is individually responsible for their own deduction scores whereby discussion is not allowed to

prevent influence and bias. The E-judge is required to record deductions applied for each gymnastics skill and dance element during a routine by evaluating deviation of a particular move performed by a gymnast compared to the standard and guidance as stated in the COP and respective competition rules. The execution score (E-score) includes three types of deductions, which are general execution faults that applicable across all four apparatus (FIG, 2016b, p. 29), specific apparatus deductions that applied to respective apparatus, and artistry deductions that applicable only to BB and FX. Average E-score across the execution panel excludes the highest and lowest deductions is then added to the D-score to award a final score, which to be published during a competition pause when a gymnast has finished her routine and before the next gymnast commences. All scores given by all judges within a competition will be monitored and oversee by the Women's Technical Committee for transparency and to avoid bias. Sanctions are given to judges who perform poorly in a competition as stated in the COP (FIG, 2016b).

In the United Kingdom, British Gymnastics (BG) is the governing body to organise national competitions as well as accredit national, regional, and club WAG judges. Current BG judge education involves traditional classroom teaching and culminates in a 'pen and paper' assessment at the last day of a four-day course. Course assessments include 1) practical judging, that includes element recognitions and observations of gymnasts performing their skills/routines showed in video clips to provide appropriate D-scores and E-scores, and 2) theoretical understanding, that includes multiple choice and open-ended questions assessing knowledge in the COP, as well as gymnastics skill and dance element recognition and symbol notation. The course syllabus is amended accordingly to meet the level of a judge course, whereby a higher-level judge is expected to be able to judge more advanced routines involving higher difficulty gymnastics skills and dance elements. There are pre-requisites for possessing higher level judge qualifications, for example, one must possess

at least two years of regional judge qualification for eligibility to enrol into a national judge course (British Gymnastics, 2017b). Meanwhile, in Malaysia, Malaysia Gymnastics Federation (MGF) is the governing body to accredit national and lower level judges. Similar 'pen and paper' classroom training and examinations are held in every OG cycle to accredit judges to ensure they possess updated knowledge of the published COP. The syllabus also amended accordingly to meet the level of a judge course besides complying local competition requirements.

The FIG reviewed the performance of judges both during and after major competitions held during 2013-2016 Olympic cycle with consistency to identify effective judges and sanction bias judges. An improved statistical engine, the Judge Evaluation Program was designed and developed in collaboration with Longines® and the University of Neuchâtel, Switzerland to provide constructive feedback to judges, executive committees and national federations, to assign the judges who evaluated performance most accurately to the most important competitions, and to detect bias and outright cheating (Mercier & Klahn, 2017). Recent reports published by the FIG (Sacchi, 2018a, 2018b) stated that D-judges had applied objective scores consistently across all gymnasts, whilst E-judges had worked 'generally well' with several score deviations within panels that indicated score difference among judges for both 2018 Youth Olympic Games and 2018 World Championships. However, more details are required to show the E-score deviations concerning score accuracy for determining the winners of a competition to subsequently inform judge education. Both empirical researches have highlighted recent FIG analytics of in-event scores suggested that understanding on decision-making process of gymnastics judges is warranted.

2.3 Decision-making Model

In WAG competitions, judges are required to multitask whilst judging, with both D-judges and E-judges having to simultaneously watch the routine performance and write down

gymnastics skills and dance elements performed by a gymnast in symbol notations on the judging sheet. By comparing the skills and elements performed to the standards as coded in the COP and specific rules in respective competitions, D-judges are making decisions to award the difficulty value of moves and movement connection for bonuses. On the other hand, E-judges are making decisions to apply execution deductions by evaluating the deviations of each element or skill performed from the expected standards, in addition to retrospectively evaluate the artistry performance at the end of a BB or FX routine to apply artistry deductions. These decision-making processes underpinning WAG judging been concerned looking at several decisions are made with time constraints, which supported by Kaya (2014) stated there was no systematic way of making decisions in high temporal sports.

A decision-making model proposed by Memmert (2015) illustrates a classification framework of central cognitive performance factors that involves several information-processing steps within cognitive psychology in sports (see *Figure 2.1*). The model underpins all actions in team and racket sports involving tactical creativity of working memory, anticipation, perception, attention, and intelligence, were mapping to the requirements of WAG judging when evaluating a gymnastics performance. This model supports decision-making within WAG, whereby Ste-Marie (1999, p. 2) stated that "gymnastics judging requires stored memory of gymnastics elements, working short-term memory, both divided and selective attention capabilities, as well as detecting and identification of complex movement patterns". Moreover, locally organised competitions were requiring judges to complete both D-scores and E-scores due to lack of judges, unlike that of the judging role for international competitions organised by the FIG where these are separate roles. Therefore, central cognitive performance of judges was a concern regarding to outcome fair and accurate scores.

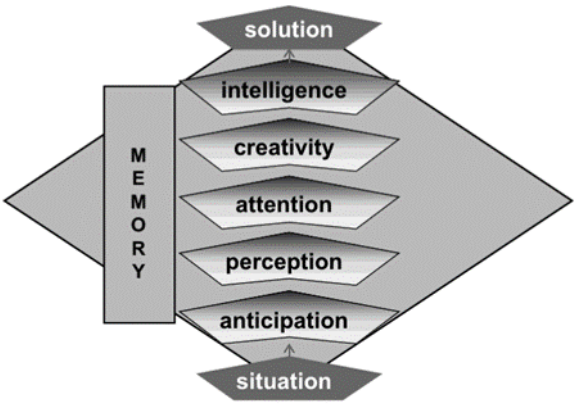


Figure 2.1 An overview of the central cognitive performance factors that underlie all actions in team and racket sports, adapted from Memmert (2015, p. 365)

2.4 General Theories of Memory

Previous studies (Ste-Marie, 1999, 2000) revealed that expert WAG judges possess superior declarative knowledge on rules-based information in the COP when compared to novice judges due to effective information retrieval from memory. This includes writing down the specific symbols notating gymnastics skills and dance elements effortlessly without referring the COP whilst simultaneously evaluating the performance aspects of each element. Given the time constraints in judging WAG, a higher ability of memory recall reduces information-processing demands when making decisions in the judging process that required multitasking.

According to the human memory model (Atkinson & Shiffrin, 1968), short-term memory (STM) is that of incoming sensory information that resides for a period of about 30 seconds, then decays completely overtime, whereas long-term memory (LTM) is a fairly permanent repository of information that transfers from STM through rehearsal, a control process to maintain a limited amount of information. Besides, working memory (WM) stated as an individual's short-term store that receives selected inputs from both STM and LTM. Baddeley (2012) further categories LTM into three types of memory: procedural memory, semantic memory, and episodic memory according to different retrieval processes.

Procedural memory is activated at subconscious level to allow automatic stimulus-response action, such as driving car. Semantic memory is encyclopaedic knowledge possessed by an individual that is affected by the frequency of retrieval, whereas episodic memory requires the deepest level of conscious thought for retrieval. In WAG judging, gymnastics skills and dance elements are learned as symbol notations that in the COP and stored as LTM by practice rehearsal during the judge education course. Frequent retrieval in judging practice and competition judging enables the storage conversion from semantic memory in LTM into procedural memory, therefore allowing judges to symbol skills and elements simultaneously whilst evaluating the performance. Nevertheless, fast-access automatic retrieval of semantic memory occurs with training of frequent retrieval (Chase & Ericsson, 1982), and in the case of gymnastics judges that enable the COP knowledge and information to be recalled via procedural memory when watching the gymnastics routines.

Besides that of higher ability of information retrieval from memory, studies have also investigated domain-specific knowledge possessed by expert in sports. Pioneering research work on the superior memory by Chase and Simon (1973) revealed chess experts stored a large number of specific chess pieces patterns in LTM that allow them to rapidly recognise several patterns in a presented chess position, which enable them to encode and recall by relying on chunks in STM. This shows the importance of superior memory and for WAG judges that of recognising patterns for frequently performed series of fast-moving acrobatic skills and dance elements to allow efficient judging from STM ‘chunks’. Research conducted on chess expertise found that chess masters could virtually reproduce the entire position of their thought processes during selection of the move in verbal reports (de Groot, 1978). Experts were therefore able to encode and recall information once they have completed the tasks even related stimuli been removed from view. This process is proposed as similar to the demands of WAG judging, where the E-judges are expected to recall the entire BB or FX

routine to apply artistry deductions at the end of performance immediately following execution deductions applied concurrently whilst watching the performance.

2.4.1 Working Memory

WAG judges are time constrained when evaluating routines, therefore they are required to constantly watch and evaluate gymnastics skills and dance elements at the same time as retrieving information from both LTM and STM. Judges are required to apply artistry deductions for BB and FX at the end of a 90-second routine and thus involves WM that enables the decision-making process. The term 'working memory' (WM) was adapted by Baddeley and Hitch providing a framework for an individual's ability to retain and manipulate information over short period of time (Baddeley, 1986). According to Baddeley (2012), WM was evolved from the concept of STM, implies a combination of storage and manipulation. The theoretical concept of WM assumes a limited capacity system that temporarily maintains and stores information and supports human thought processes by providing an interface between perception, LTM and action (Baddeley, 2003, p. 829). Here, "information is maintained in readily accessible storage for only a short period without rehearsal or reactivation" (Ericsson & Kintsch, 1995, p. 1). Baddeley (2012) outlined the flow of information from perception to WM involving central executive, episodic buffer, and two storage systems of that visuospatial sketchpad and phonological loop, as illustrated in the Figure 2.2.

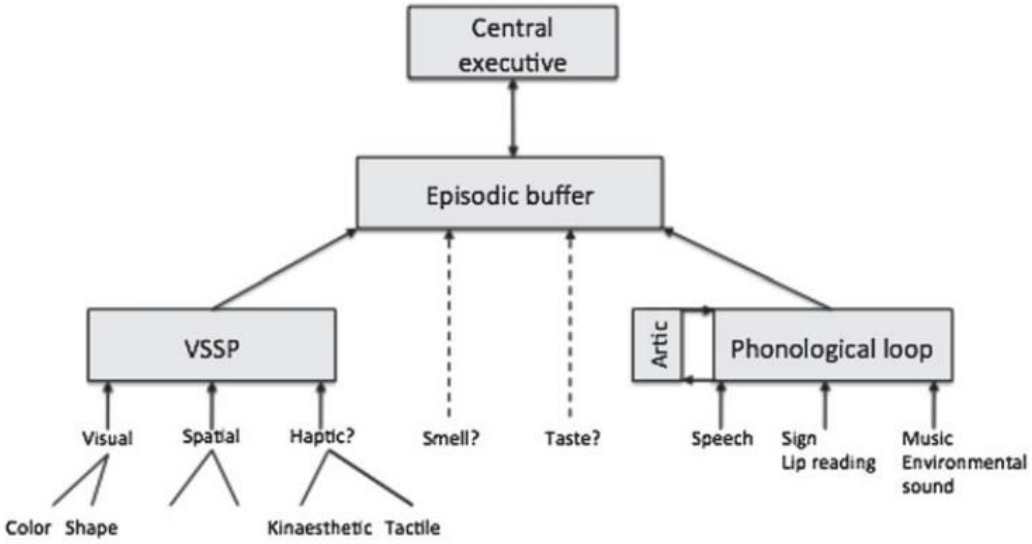


Figure 2.2 A speculative view of the flow of information from perception to working memory. VSSP, visuospatial sketchpad, adapted from Baddeley (2012, p. 24)

The visuospatial sketchpad is based on visual, spatial, and haptic, whilst the phonological loop is based on sound and language. The phonological loop comprises of a phonological store, which can hold memory traces for few seconds before they fade, whereby memory traces are refreshed by being retrieved and rearticulated. Likewise, visual WM is limited in capacity, typically to about three or four objects, and results in the phenomenon of change blindness, when objects in scenes can change colour, move or disappear without people noticing (Baddeley, 2003). Applying this framework into WAG, judging the FX involves the phonological loop, whereby musicality is one of judging criteria for artistry evaluation. Meanwhile, all four apparatus in WAG require the visuospatial sketchpad during judging process to evaluate the gymnastics skills and dance elements movements. Therefore, there is perhaps concern when individual judges are required to evaluate a series of fast-pace movements in WAG with limited WM capacity for both the visuospatial sketchpad and phonological loop. It is therefore questionable if that judges are able to retain watched information effectively over 90-seconds BB and FX routines as well as fast-pace VT and UB routines.

The episodic buffer is assumed to be a limited capacity buffer store that binds together information from integrated episodes as well as linking WM to perception and LTM. Consciousness is assumed to serve as a mechanism for binding stimulus features into perceived objects (Baddeley, 2012, p. 15). Therefore, consciousness and awareness among WAG judges when officiating a competition is vital to ensure gymnasts' performance are perceived accurately besides effective COP information retrieval from LTM to maintain effective and objective judging throughout the competition. However, the level of consciousness and awareness may depreciate after unexpected pause, such as that of a fall from the apparatus and thus a maximum 30-second delay for the gymnast to restart, as well as interruption by breaks during a typical whole-day judging schedule.

The central executive is an attentional system (Baddeley, 2012) that is assumed to be able to 1) focus attention that enables reduced attention on complex tasks; 2) divide attention between two important tasks, such as managing verbal and visuospatial tracking at the same time; 3) switch between tasks with specific control system; 4) interface with LTM. This indicates the impact of central executive functioning in WAG judging, whereby the judging process requires attentional focus to watch the routine performance, at the same time as being able to multitask including writing symbols notating elements and skills performed whilst making decisions to award difficulty or apply execution deductions through retrieving information from memory. Attentional focus and division by central executive also explains how WAG judges are able to switch tasks between decision-making of that routine evaluation into mathematical calculations to sum the total score awarded to a gymnast at the end of a routine.

Chase and Ericsson (1982) acknowledged that the STM capacity limits WM capacity of the human information-processing system, however, the retrieval process involves STM activates memory traces in LTM. Skilled individuals were able to associate information

recalled from their comprehensive knowledge base within the domain of expertise that was stored in LTM for later retrieval by virtue of proper index, whereby frequent storing and retrieving information sped up these processes. Chase and Ericsson (1982) proposed memory capacity could be expand over extended training in which an individual must be able to store information rapidly in LTM and this requires a large body of relevant knowledge and patterns for the particular type of information involved. Further, the activity must be very familiar to the individual to anticipate future demands accurately for relevant information retrieval.

Moreover, the individual must associate the encoded information with appropriate retrieval cues, whereby the association allows them to activate a particular retrieval cue later and thus partially reinstates the conditions of encoding to retrieve the desired information from LTM. 'Retrieval structure' refers to a set of retrieval cues organised in a stable structure. This approach has been applied mainly in mnemonists and mental calculations (Gobet, 1998) when individuals are required to repeatedly perform the same task in direct succession in most of skilled activities, showing experts generate and change intermediate results and products in WM within an activity (Ericsson & Kintsch, 1995). Looking into WAG, retrieval structure as skilled activities by experienced judges corroborate previous research findings (Mercier & Klahn, 2017; Ste-Marie, 1999, 2000) that revealed trained expert judges were able to judge more efficiently compared to novice judges, whereby expert judges frequently repeat the COP information retrieval concerning higher frequency in attending judging tasks.

2.4.2 Long-Term Working Memory (LT-WM)

Following memory theories of STM, LTM, and WM, Ericsson and Kintsch (1995) developed another memory theory, the long-term working memory (LT-WM), which makes primary distinction between immediate recall of activated information in STM and the additional step required to access other information in LTM. They proposed individuals recall

a small number of items in STM immediately during free recall, however, there is a long pause until additional items in LTM are accessed and reported. In LT-WM task, storage of STM overlaps completely with the rapid and reliable accessibility of information, therefore requiring shorter time for retrieval. On the other hand, it is estimated to take recognition times for accessing information from LTM compared to just-seen items that are retained in active form of STM. Therefore, individuals are unable to access a large amount of information on sequential free recall in skilled activities immediately. Nevertheless, if WAG judges frequently attend competitions for judging so as to actively recall COP knowledge stored in LTM, they should be able to access information with a speed and reliability with expanded effective WM capacity (Ericsson & Kintsch, 1995) when compared to that of accessing information from STM as explained in skilled memory theory (Chase & Ericsson, 1982).

Ericsson and Delaney (1999) suggested that experts acquire sophisticated and complex skills to expand their WM for acquiring knowledge and skills to rapidly encode information from LTM efficiently with retrieval cues of LT-WM that enable them to successfully complete task. Skilled memory theory (Chase & Ericsson, 1982) proposes that acquired memory skills allow stable states of cognitive processes to be stored in LTM and are kept directly accessible by means of retrieval cues in STM when conducting skilled activities. Throughout the judging process, WAG judges are required to recognise gymnastics skills and dance elements according to the COP through encoding information from LTM. The just-seen skills and elements performed stored in STM are compared to the specific standards as stated in the COP by retrieving information from LTM to evaluate deviations before awarding difficulty or applying execution deductions. LT-WM provides durable storage for sufficient retrieval cues in attention for accessing information in LTM, than that of the traditional models of WM involving temporary storage and must be extended to include WM

based on storage in LTM in order to account for the large demands on WM during text comprehension and expert performance (Ericsson & Kintsch, 1995). Therefore, LT-WM enables judges to overcome the demands of continuous incoming retrieval cues when watching a routine for accurate decision-making. Information storage in LT-WM implies that most types of accessible information in WM will remain in LTM during an interruption of skilled activities, whereby access to them is reinstated easily by reactivating of necessary retrieval cues. In WAG, there are unplanned interruptions during a routine performance, i.e. an injury or fall form the apparatus as well as planned interruptions at the end of a routine and before commencement of next gymnast that require judges to switch judging task from skills and elements evaluation into mathematical calculations to award the final score for a gymnast.

2.5 Expert Performance Approach

Expert memory capacity is developed over extended training or accumulation of judging experience that allows WAG judges to retrieve COP information in addition to anticipate frequent patterns of gymnastics acrobatic and dance series more accurately than novice judges (Heinen et al., 2012; Ste-Marie, 1999). However, outcomes from previous WAG studies revealed there were memory-influenced biases due to prior information processing by watching repeated moves from a same gymnast during warm-up time (Ste-Marie & Lee, 1991) and repeated or similar moves by different gymnasts across competition (Ste-Marie & Lee, 1991; Ste-Marie et al., 2001). Therefore, on-going studies investigating expert performance with a specific focus on perceptual-cognitive expertise (Williams et al., 2017) were deemed appropriate to explore the decision-making process underpinning WAG judging. Furthermore, judges with gymnastics background claimed they had the ability to anticipate gymnastics skills and elements that enable them to evaluate more accurately akin to expert judges. Indeed ex-gymnasts (Gentsch, Weber, Synofzik, Vosgerau, & Schütz-

Bosbach, 2016; Heinen et al., 2012; Pizzera & Raab, 2012) and coaches (Campo & Gracia, 2017) reportedly utilised such motor experience in judging. Therefore, there is a need to develop representative task to understand and examine changes in expert's learning within specific domains to study expert performance in capturing, identifying, and enhancing perceptual-cognitive skills.

A three-stage systematic framework of expert performance approach (Ericsson & Smith, 1991) (see *Figure 2.3*) deemed appropriate to be adapted for studying expert WAG judge's learning and performance, which begins with understanding that of methods used to capture expert performance in tasks, followed by processes to identify mediating mechanisms of superior performance, and finally the expertise development. Figure 2. outlines examples in using the expert performance approach to evaluate expertise across domains.

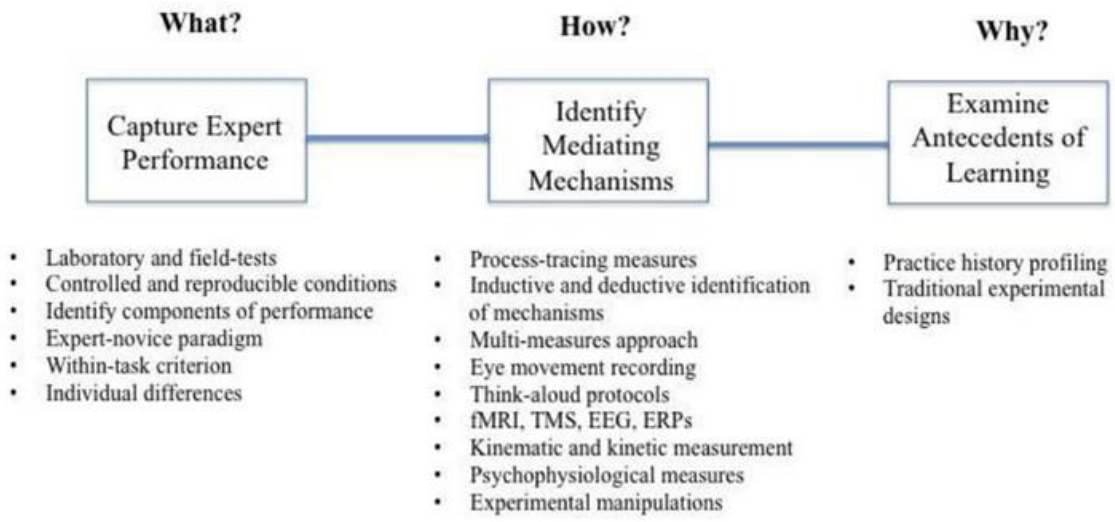


Figure 2.3 The expert performance approach by Ericsson and Smith (1991), adapted from Williams et al. (2017, p. 141)

Stage 1

Stage 1 of the expert performance approach recommends that representative tasks are used to provide precise and reproducible measurements to objectively evaluate and vary expertise under controlled and duplicable conditions (Williams et al., 2017). However,

according to Williams and Ericsson (2005), there were difficulties in assessing the behavioural constructs such as anticipation and decision-making, therefore, novel tasks required unique solutions that offered experts and novices the same amount of preparation time were suggested to remove experimental effects that advantaged experts in using regular method to solve a particular problem. Further, representative perceptual-cognitive tasks with realistic simulations were suggested to enhance measurement sensitivity within the sport domain to increase the possibility of identifying meaningful and important differences between expert and novice performers. Hence, advantages of using film and video were emphasised to "enable sequences of action to be reproduced in a consistent manner from trial to trial, providing an objective method of evaluating performance" (Williams & Ericsson, 2005, p. 287). These data corroborate findings from MacMahon, Starkes, and Deakin (2007), who examined 44 highly experienced basketball referees at a moderately elite level of performance that using 22 video clips to examine the effects of knowledge warm-up and instructions on infraction detection and naming performance utilising three tasks, of that rules, signals, and infraction detection. Outcomes of the study found the performances of officials were influenced by the specific format sequencing of video clips, hence suggested that decision-making tools of video clips and tasks to progress in perceptual difficulty to provide exposure of time-pressured decisions simulating actual competitions for faster and more accurate responses. This approach has been adapted in current WAG judge education to simulate actual competition judging for capturing perceptual-cognitive expertise during officiating, whereby video and film were widely used to evaluate singular gymnastics skill and dance element as well as a full routine (Campo & Gracia, 2017; Heinen et al., 2012; Pizzera, 2012; Pizzera et al., 2018; Ste-Marie, 1999, 2000). Same sequence video clips captured from the unique perspective of judge's view angle that mimic real-life competition are more likely to test the specific cognitive and processing strategies that have been

developed by the experts to investigate the expert-novice difference (Ste-Marie, 1999) although insight into how this occurs is not yet explored.

Moreover, sport expertise as a function of role was explored by examining decision-making skills and their relation to deliberate practice in elite football referees using video-based clips (MacMahon, Helsen, Starkes, & Weston, 2007). The test film consisted of 20 clips of tackles, with the participants required to make the most appropriate decision based on the laws of the game. Findings show that seven referees achieved a decision-making accuracy of that was 25.5% higher than 34 football players. This study evidenced there is domain-specific knowledge acquired with role-specific skills, even though both groups of participants accumulated equivalent playing experience. Further findings from this study showed that officials developed and engaged in greater volumes and types of training demonstrating superior performance in core role-related skills and role specificity within one sport. Therefore, referees become more diverse with a greater variety of training activities as they become more expert. However, football referees required to remain physically active whilst officiating, which is different from the role specificity of WAG judges with stationary sittings.

Besides team sport officials, the expert performance approach has also been utilised to examine general knowledge structures regarding baseball batting preparation and competition (McPherson & MacMahon, 2008) engaging 17 expert baseball players and 18 non-players. Participants watched three different edited video sequences of a baseball competition under different task conditions following verbal reports to provide insight into the acquisition of batting skill. Outcomes of the study indicated that players used sport-specific strategies to encode and retrieve pertinent game events from LTM to develop tactics for their upcoming times at bat and to recall as much information as possible. Further recommendations by the authors were to use competition videos for exploring player's tactic development through

retrospective think aloud, whereby players calls upon a variety of specialised processes such as encoding, retrieval, updating and monitoring of current events that are embedded in this sport-specific knowledge in developing batting tactics. These data and recommendations support the expert performance approach to examine the expert-novice differences within sports.

Stage 2

Expert performance approach stage 2 identifies the mediating mechanisms of how experts demonstrate superior performance on the domain-specific task when compared to less expert individuals (Williams et al., 2017). Previous WAG studies (Bard, Fleury, Carriere, & Halle, 1980; Heinen et al., 2012; Pizzera et al., 2018; Ste-Marie, 2000) utilise process-tracing measure of eye movement recording to study gaze behaviour of expert judges. Results of these studies revealed that expert judges acquired systematic visual search strategies supported with anticipation and memory recall in processing information facilitate them to detect more execution errors than novice judges (Ste-Marie, 2003). The study by Bard et al. (1980) involved four certified national judges and three uncertified local judges to evaluate four routines on the Balance Beam. Certified judges had 27% less fixations than the uncertified judges, whilst uncertified judges detected only half the execution errors as compared to the certified judges. There was a tendency by the participants for experience to influence their search patterns. The anticipation of the movement sequences reduced the number of fixations of the expert judges that enabled fine discrimination capacity in the judging task of gymnastics, that had enhanced memory contents serves as a criterion for comparison of incoming perceptual signal stimulus from the skill and element performance by a gymnast and the COP information stored in the long-term memory. However, this study utilised “reference score” evaluated by expert judge, who allowed to analyse the four routines using slow motion for twice and normal speed film projection for three times were unable to

present the actual competition scenario whereby judges were request to judge based on what were seen during the competition. Later, Heinen et al. (2012) explored the *biomechanical sources of information* with 23 gymnastics judges, who exhibited specific visual experience and 23 laypeople, who by virtue had no specific visual experience but were able to perform the handspring on Vault by themselves without any guidance technique. They were requested to rate a total of 30 handsprings on Vault videos based on a 9-point scale rating, that of a maximum of eight points for perfect mastery of the handspring and a minimum of zero points for major movement errors of an execution. However, this rating scale was unrepresentative of actual competition requires judges to evaluate a Vault performance throughout all four phases despite of allowing participants in this study to evaluate a Vault performances on "good", "bad", and "average" for the entire Vault performance. Nevertheless, results from this study indicated that judges could be facilitated by either own motor experience or specific visual experience, whereby judges showing difference on organisation and activation of judging knowledge by rating scores in average one point higher than the scores rated by laypeople. Findings revealed that motor experience led to a better perception of the movement execution thus enhanced the movement evaluation of laypeople without judging knowledge. Moreover, another recent study conducted by Pizzera et al. (2018) utilised judges with differing motor and judging expertise and their gaze behaviour examined on superior judging performance to explore the underlying mechanisms of decision-making process among gymnastics judges. Judging performance and eye movements of 35 judges that of experts with a higher-level license and novice with a lower-level license, utilising an eye-tracking device whilst judging 21 video-based handsprings forward with a half turn on/half turn off the Vault were compared. This study concluded that superior judging performance was reflected in a specific gaze behaviour, whereby judges with a higher-level license had *more fixations* on the gymnast during the whole skill and the landing phase, specifically on

the head and arms of the gymnast as compared with judges with a lower-level license. This outcome contradicted with previous literature (Bard et al., 1980) who stated experts had *less fixations* on gaze behaviours compared to novice judges. Despite, the importance of examining the performance of judges to include the underlying mechanisms in explaining how judges developed a specific memory representation been reinforced to further investigate that of movement templates through judge education as well as intensive training during competitions to acquire processing strategies and visual experience. The superior judging performance concluded to be reflected in a specific gaze behaviour for expert judges to specifically focused more on very specific body parts of the gymnast, which were relevant in terms of the judging criteria for the skill indicating a strong link between perceptual, cognitive, and action processes in sports. This is congruent with findings of the study conducted by Ste-Marie (2000) to investigate differences in behaviours of expert and novice gymnastic judges while they were engaged in judging actual competitions involved 10 expert and 10 novice gymnastic judges, who were videotaped while judging at actual competitions. Findings from that study showed that expert judges spent less time looking at the scoring paper and less time looking at the gymnast performance as compared to novice judges further stated that expert judges were able to engage in the dual-task demands required in gymnastic judging. This reinforced the expertise development can be viewed as learning to circumvent the processing limitations of a given task. Another more recent study conducted by Campo and Gracia (2017) found a stronger relationship between visual and verbalisation, when a WAG judge asked to verbalise two most important visual locations when evaluating gymnastics skills whilst exploring gaze behaviour on three singular gymnastics skills on Vault, Uneven Bars, and Floor Exercise respectively involving a judge, a coach, and a gymnast. The gymnast showed different search behaviours when compared to the judge and the coach in this study. Furthermore, the judge was able to show a stronger relationship

between the gaze and verbal reporting higher percentage of concordance between the visual and verbalised behaviour suggested that the judge perceived the visual information with a greater amount of perceptual cognitive skill to process specific information when conducting the judging task. However, there was no specific visual search pattern utilised by expert judges identified. Therefore, Think Aloud method stated in the expert performance approach are deemed appropriate to accurately report mediating thought process concurrently alongside task completion.

Stage 3

The final stage of expert performance approach aimed to improve acquire skills for experts to demonstrate reliably superior performance through adaptive learning (Williams & Ericsson, 2005). Expertise could be developed over time and training (Chase & Simon, 1973; Ericsson, 2017; Ericsson & Simon, 1993; Ste-Marie, 1999) to facilitate and enhance perceptual-cognitive skills. Questionnaires and interviews, including questions such as experience and hours engaged in domain-specific activities were found to be most efficient to examine historical profiles of an individual leading to expertise development (Ericsson, 2003). This corroborates with studies investigating expertise within WAG judging that often complemented with questionnaires (Campo & Gracia, 2017; Pizzera, 2012; Pizzera et al., 2018) to explore the background of participants for demographic purposes. Interventions were suggested to recreate the performance environment using video or film with instructions included as to the important cues underpinning performance that coupled with practice and feedback (Williams, Ford, Eccles, & Ward, 2011). In current WAG judge education, competition video clips were selected to mimic actual competition for judging practice in addition to facilitation by expert judges to provide guidance and feedback. Furthermore, Williams and Ericsson (2005) asserted that limits imposed by basic information-processing capacities, such as visual reaction time and STM were overcome by skilled performers

(Ericsson & Kintsch, 2000) through specific adaptations from practice and experience. This perhaps explains how expert WAG judges with accumulated judging experience were able to anticipate fast-moving gymnastics skills with their ability to pick up advance postural cues to overcome time constraints (Williams, Ward, & Chapman, 2003). Therefore, the expert performance approach provides important knowledge structures for WAG judge education to provide deliberate practice (Coughlan, Williams, McRobert, & Ford, 2014) in developing interventions to facilitate the acquisition of perceptual-cognitive expertise.

2.6 Perceptual-Cognitive Skill in Judging Gymnastics

In addition to studies of memory recall, visual search, and anticipation concerning decision-making within WAG judges, there are investigations exploring perceptual-cognitive skills to process environmental information during competitions. Perceptual-cognitive skills have been recognised to facilitate anticipation in exploring nature of expert performance across domains in recent years (Williams et al., 2011), with identified processes underpinning anticipation: 1) recognising and utilising task-relevant, postural information provided; 2) recognising and familiarity of structure in the patterns; and 3) applying probabilities and expectations in a situation. The ability to anticipate is crucial in sport performance and has become an increasingly important research area within sport psychology (Loffing & al-Bruland, 2017), especially to allow information processing in a time-constrained environments. Therefore, acquired knowledge structures and cognitive processes enable anticipation to overcome the unique constraints of the task (Williams et al., 2011) that may contribute to a decision-making (Raab et al., 2019) are vital for multitasking WAG judging assignments with time constraints.

Findings from a study (Campo & Gracia, 2017) with a gymnast, a coach, and a judge as participants to explore the influence of sensory motor experiences on visual search patterns and judgement performance suggested that the judge had perceived visual information with a

greater amount of perceptual-cognitive skill to process specific information when judging (Williams & Ericsson, 2005). Pattern recognition and use of situation as part of perceptual-cognitive skills was predicted to enable judges to anticipate what will happen next, which facilitate them in a better position to make decisions during evaluation (Memmert & Roca, 2019). This corroborates with earlier works (Ste-Marie, 1999; Ste-Marie & Lee, 1991) that found that expert judges with accumulated judging experiences were significantly better at perceptually anticipating upcoming gymnastics skills and dance elements based on advance visual cues to reduce information-processing demands. Therefore, expert judges were able to overcome time constraints during judging to outcome more accurate evaluations on skills and elements.

MacMahon and Mildenhall (2012) further suggest that sport officials to acquire perceptual-cognitive skills to process incomplete, intentionally deceptive and fast-paced information under time pressure during a competition. In WAG, judges are required to judge a series of fast-moving gymnastics skills and dance elements, such as vaulting, multi-skill acrobatic and dance series in BB and FX that are typically completed within five seconds (Pajek, Cuk, Pajek, Kovac, & Leskosek, 2013). Nevertheless, anxiousness and fatigue (Dosseville & Laborde, 2015) were found to influence perceptual-cognitive skills when making anticipation (Williams & Elliott, 1999; Williams et al., 2011), whereby visual search rate was reduced and focus of attention narrowed steering decrement capacity to use peripheral vision to capture different sources of information. These factors were concerned to influence perceptual-cognitive skills of WAG judges, whereby long-hour competitions require judges to evaluate routines consistently and effectively.

2.7 Research Methods and Limitations in Previous studies of WAG judge

A growing body of literature has explored decision-making underpinning cognitive processes in sports involving athletes, coaches, and officials. Several studies in the past

decade have utilised statistical analyses to investigate scoring objectivity on actual WAG competition data. Recent FIG Women's Technical Committee post-competition report of 2018 World Championships (Sacchi, 2018b) reviewed scores officiated by 89 international WAG judges from 62 federations judging on a total of 230 gymnasts across all four apparatus of that BB, FX, VT, and UB. The report found scoring consistency in both D-scores and E-scores, which reflected the correct rank order for gymnasts. However, there were a few E-scores reported that required intervention of reference score from the reference judges. This corroborates with another post-competition report of 2018 Youth Olympic Games by FIG Women's Technical Committee (Sacchi, 2018a) reviewing scores officiated by 19 international judges on 35 junior gymnasts, which reported there were 'important' E-score deviations. Both official reports from the FIG indicates E-score variations when judging an official international competition yet details that elucidate score deviations were not disclosed.

Score data of major international competitions held during the 2013-2016 Olympic cycle were analysed utilising a judge evaluation programme to investigate performance of international gymnastics judges (Mercier & Klahn, 2017). Results had revealed there were significant performance differences among well-trained international judges despite nearly all judges were former gymnasts. Judges were raised suspicion of bias in favouring gymnasts from their own country due to inconsistency and unreasonably high or unreasonably low scores given to gymnasts that corroborate with findings about patriotism effect (Boen, Auweele, et al., 2006). Nonetheless, an erratic judge could be unbiased whilst a precise judge could be biased when concerning the drawback of the marking scores that compare scores among judges but not based on objective performance of routines. Furthermore, Pajek et al. (2014) analysing artistry deductions officiated by five international judges across 194 gymnasts at World Championships in Tokyo 2011 revealed neither reliability nor validity of

artistry judging with relatively large variability of average artistry deduction scores received from separate judges in the same components of artistry. Findings from analysing actual data of competitions indicates there is much uncertainty about the relationship between the score deviations and decision-making underpinning the judging process.

Despite statistical analyses of post-competition scores, there were several studies investigating superior judging performance of experts by exploring their gaze behaviour. Pizzera et al. (2018) assessed judging performance and eye movements of 35 WAG judges with mixture of higher-level and lower-level judge licence with a total of 21 video-based handspring forward with a half turn on-half turn off the vault (see *Figure 2.4*). Judging performances and gaze behaviours were investigated across five phases in vaulting, which were take-off, first flight phase, repulsion phase, second flight phase, and landing phase according to the latest COP (2017-2020) (FIG, 2016b). Outcomes of the study confirmed that higher-level judges had better judging performance with scores deviated less from reference score compared to lower-level judges with larger deviation from reference score. Further, higher-level judges revealed more eye fixations when watching a gymnast performing a whole skill and landing phase when compared to lower-level judges, which corroborate to findings of Ste-Marie (2000) indicated superior judging performance was reflected in a specific gaze behaviour (Mann, Causer, Nakamoto, & Runswick, 2019). Visual search behaviours and expert perceptual judgements acquired through judging experience accumulated over intensive training and judging competitions to develop processing strategies and visual experience (Ste-Marie, 2003). This further reveals expert-novice differences among WAG judges when making decisions to apply execution deductions, in addition to that of judges who facilitated by specific motor or visual experience with gymnastics background were able to judge gymnastics skills more accurately (Campo & Gracia, 2017; Heinen et al., 2012). However, only a single study (Campo & Gracia, 2017)

enquired participants to verbalise the two most important visual locations where they thought they had fixed their gaze for longer, which was then compared with the information recorded by the eye tracker system. This study aimed to explore the influence of previous domain-specific visual and motor experiences onto perceptual judgments was participated by a judge, a coach, and a gymnast, whereby all participants possessed respective expertise in gymnastics. Verbal report method was adapted in this study in addition to the utilisation of eye tracker system to analyse the influences of previous sport-specific experiences on visual and motor experiences in shaping judging performance gained during years of training and practice to make perceptual judgments. There were total of 18 fixed-sequence video recordings of six trials each for Vault, Uneven Bars and Floor Exercise demonstrating three different gymnastic skills. After viewing each video, participants were request to verbalise what they considered the two most important visual locations in the video, whereby the information was compared with the information recorded by the eye tracker system. Findings of the study showed differences in visual search patterns between participants, whereby the gymnast participant focused more often and longer on the hips and next to the legs, whilst judge and coach participant spent less time fixating on gymnast body locations. Furthermore, the judge participant reported higher judgement accuracy when compared to the reference score, followed by the coach and the gymnast, suggested that the participants fixated on different information sources when making judgements on performance scores based on their different previous visual and motor experiences in evaluating gymnastic skills. These outcomes corroborate with previously reported findings by Bard et al. (1980) stated that expert judges in gymnastics were able to detect more errors than novices. This study showed the judge demonstrated a stronger relationship between the gaze and verbal reporting with higher percentage of concordance suggested that the judge perceived the visual information with a greater amount of perceptual cognitive skill to process specific information about the

task (Williams and Ericsson 2005). Nevertheless, it remains unclear whether differences between the anticipatory and decision-making performance of experts can be explained by advantages in their visual search behaviours (Mann et al., 2019). Therefore, it is important to consider appropriate methodologies to explore the actual decisions made during judging, whereby verbal report of Think Aloud protocol been suggested in the expert performance approach (Williams et al., 2017) to identify the mediating mechanisms for improving understanding of expert learning.

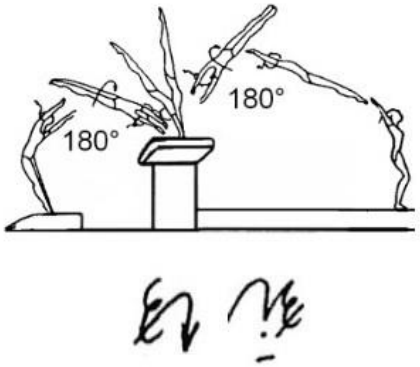


Figure 2.4 Description and symbol notation of handspring forward with a half turn on-half turn off from Vault, adapted from FIG (2016b, p. 61)

2.8 Verbal Report

Verbal report protocol (Ericsson & Simon, 1981, 1993) has been utilised to study cognitive processes in many areas of psychology, education, and cognitive science, including that of sport and exercise psychology (Eccles, 2012) comprehensive of concurrent and immediate verbal report, delayed retrospective report, and interviews. Comparison between different types of verbal reports were listed in the Table 2.2.

Table 2.2 Comparison between concurrent TA, immediate retrospective TA, retrospective TA, and retrospective interview

Method	Definition	Advantages	Disadvantages
Concurrent verbal report	Respondent verbalises thoughts at the same time watching a routine/performance	Provides in-event thought verbalisations concurrent with domain-specific tasks without delay	Missing information due to time restriction and verbal overshadowing
Immediate retrospective verbal report	Respondent verbalises thoughts within 7±2 seconds after watching a routine	Provides longer time than concurrent verbal report for thoughts verbalisation through STM	Memory decay to recall details of in-event cognitive thoughts chronologically
Delayed retrospective verbal report	Respondent verbalises thoughts within 60 seconds after watching a routine	Provides longer timeframe than concurrent and immediate retrospective verbal report for thoughts verbalisation through LTM	Memory decay to recall details of in-event cognitive thoughts chronologically and inaccurate information recall
Retrospective interview	Face-to-face interview conduct with individual respondent immediately after the verbalisation session beyond 60 seconds	Rich information from social cues Overcome time restrictions to provide verbal report Enable in depth responses concerning subject related discussions No time delay between questions and answer allowing respondents to react directly on questions	Memory decay to recall details of in-event cognitive thoughts chronologically and inaccurate information recall

Verbal report provides opportunity for research participants to verbalise their thoughts whilst performing cognitive tasks. However, earlier studies (Ericsson & Kintsch, 1995; Ericsson & Simon, 1993; Nicholls & Polman, 2008; Whitehead et al., 2015) reported retrospective verbal report of post-trial interviews that aimed to highlight key thought processes during an event were lack of accuracy affected by memory decay due to depletion of memory after completing domain-specific tasks (Bernard et al., 1984). Whitehead et al. (2015) examined the congruence of verbal data collected on decision-makings in golf using concurrent verbal reports and cued retrospective recall at different time intervals, 10 minutes, 24 hours, and 48 hours after performance. Results revealed there was only low level of congruence that of 38-

41% between thoughts verbalised during in-event and retrospective recall in interviews. These indicate that retrospective verbal report may not accurately represent the situation of in-event cognition processes. However, this problem might overcame by instructing participants to 'think aloud' their thoughts concurrently through verbalisation whilst working on a task consciousness to reveal normal sequence of cognitive processes, whereby Newell and Simon (1972) postulates that cognitive process is a sequence of internal states successively transformed by a series of information processes. The use of both concurrent and retrospective verbal reports were recommended by Whyte IV, Cormier, and Pickett-Hauber (2010), whereby they provided more comprehensive analysis of the cognitive roots in decision-making without adding extensive additional steps in the research processes. The study conducted in a simulated task environment participated by fifteen nurses from a nursing college to compare the content of concurrent and retrospective verbal reports during and after administering care indicated concurrent verbal reports provided the most complete representations of task performance and task related cognitions. Furthermore, a more complete record of high-level cognitions as compared to retrospective verbal reports during the simulated task environment were recorded, whereby these data were more closely followed the accurate sequencing of events as they occurred within the simulated task environment, that could not expect during a retrospective report. However, the duration up to 30 minutes for the concurrent verbal report was reported too lengthy to facilitate high accurate retrospective report that corroborate to outcomes of Whitehead et al. (2015) indicating the memory decay in recall information from LTM. Nevertheless, retrospective reports provided unique data, which were the reflective statements that offer important inferences into cognition during the performance of nursing case that are not present in concurrent verbal reports. These demonstrate the importance of collecting cognitive thoughts

using both concurrent and retrospective verbal reports concerning advantages and disadvantages of both methods.

2.8.1 Think Aloud Method

In order to circumvent limitations associated with retrospective recall and other methodologies applied in earlier WAG studies, the TA method (Eccles & Arsal, 2017), originally cited as 'thinking-aloud protocol' (Ericsson & Simon, 1993), which is a concurrent verbalisation that delivers information about the cognitive processes and thoughts mediating solutions under silent conditions while performing a task been adapted into research studies. In recent years, there has been an increasing amount of literature on TA method to study decision-making within sports, whereby Fox et al. (2011, p. 317) stated that "concurrent verbalisation has become a popular tool for studying cognitive processes and thinking".

TA method (Eccles & Arsal, 2017) is a research methodology introduced and subsequently refined by Ericsson and Simon (1993) that requires one to continuously verbalising his/her thoughts during the performance of a task. There are three levels of TA verbalisations. *Level 1 verbalisation* is simply the vocalisation of inner speech and need not to be transformed before being verbalised whilst *Level 2 verbalisation* involves the verbal encoding and vocalisation of an internal representation that is not originally in verbal code that needs to be transformed before being verbalised. A review of 40 studies found no evidence that giving concurrent verbal expressions (*Level 1* or *Level 2* TA) of one's thoughts altered performance when compared to individuals who completed the same tasks silently (Ericsson & Simon, 1993). This corroborate with a meta-analysis consists of a total of 94 studies that comparing the TA model performance while giving concurrent verbalisations to a matching condition without verbalisation. This meta-analyses involved about 3500 participants (Fox et al., 2011) indicate sequence of thoughts or accuracy of task performance did not change when instructing participants to verbalise their thoughts using *Level 1* or *Level*

2 TA without trying to explain or describe the cognitive processes. *Level 3 verbalisation* requires an individual to explain his or her thoughts, ideas, hypotheses, or motives alongside verbalisation of inner speech, hence require additional cognitive processing beyond that verbalisation as well as information retrieval from LTM. There were critics on *Level 3* TA that requires individuals to continuously explain or describe their thoughts that appeared unnatural and out of STM thus potentially impacting task performance by altering sequence of thoughts (Boren & Ramey, 2000; Welsh et al., 2018; Whitehead, Cropley, et al., 2016). This is contradicted to the findings in self-paced sport of golf putting (Whitehead et al., 2015) that *Level 3 verbalisation* did not impair task performance but provides richer verbal data regarding decision-making than cued retrospective recall and *Level 2* verbalisations.

TA has been used to avoid potential distortions of the retrospective nature required by self-report and questionnaire methods whereby it is presumed that thinking translates easily into words by focussing on thoughts whilst completing domain-specific task, neglecting inner experiences such as emotions or sensations (Dickens, Raalte, & Hurlburt 2018). However, it is important to acknowledge some identified limitations of TA despite recognise TA is a viable method to collect verbalised contents of thoughts whilst participants focus on completing challenging tasks. Fox et al. (2011) stated that TA procedures has limits and does not assure a complete record of participant's thoughts. TA has received some criticism based on its reliability for participants to verbalise accurate thought processes, for example, Eccles (2012) suggested that individuals may report additional descriptions or explanations that are not part of their actual thought process at the current time of TA. This also links to criticism raised by Nisbett and Wilson (1977) that participants may verbalise 'more than what they know'. In addition, verbal overshadowing (Chin & Schooler, 2008; Ericsson, 2003; Meissner & Brigham, 2001; Schooler, 2011) during TA reportedly distracted individuals to perform the primary task, whereby it was possibly to slow down or even withhold solution attempts and

thus perform below one's best performance alongside verbalisation (Ericsson, 2003), even leading to longer completion time (Ericsson & Simon, 1993). Chin and Schooler (2008) further mentioned that verbal overshadowing effect is not necessarily verbal description itself but possible to interfere memory by inducing people to provide very detailed descriptions. Therefore, it is becoming increasingly important to thoroughly describe procedures for instructing, familiarizing, and reminding participants to TA (Fox et al., 2011), such as warm-up exercises on easy-to-verbalise tasks and reminders to continue talking throughout the data collection process to facilitate participants in focusing on the domain-specific task whilst giving verbal reports on their thoughts as a secondary task. TA might be inappropriate during actual competitions due to the restriction of verbalisation among execution judges according to the COP (FIG, 2016b). Hence, TA method was suggested to be adapted into WAG judge's education to train novice judges within simulated situations in enhancing judging accuracy and objectivity, whereby the consequence of TA may that prolong completion time caused by verbal overshadowing does not influence the judge performance during judge's training.

2.8.2 'Think Aloud' based studies

The TA method was adapted in sports to explore cognitive processes and decision-makings among athletes to improve sport performance as well as allied fields including education, medicine, and nursing. Sport based studies have included cyclists (Whitehead et al., 2017; Whitehead et al., 2018), snooker players (Welsh et al., 2018), tennis players (McPherson & Kernodle, 2007; Swettenham, Eubank, Won, & Whitehead, 2018), long-distance runners (Samson, Simpson, Kamphoff, & Langlier, 2015), golf players (Arsal, Eccles, & Ericsson, 2016; Calmeiro & Tenenbaum, 2011; Kaiseler, Polman, & Nicholls, 2013; Nicholls & Polman, 2008; Whitehead, 2015; Whitehead, Taylor, & Polman, 2016), football players (Coughlan et al., 2014), cricket batters (McRobert et al., 2011), baseball players (McPherson & MacMahon, 2008), and volleyball players, (McPherson & Vickers,

2004; Ram & McCullagh, 2003) as participants. Collectively, these indicate the increasing popularity of using the TA method in collecting verbal reports of athletes whilst performing domain-specific sport tasks. Furthermore, TA has also been adapted into a framework to facilitate reflective practice among rugby league coaches (Whitehead, Cropley, et al., 2016) so moving beyond that of the athletes themselves. Coaches reported became more familiar and educated in the process of TA thus developed an increased awareness and enhanced communications to improve coach learning.

A recent study (Whitehead et al., 2017) involved fifteen male and three female cyclists to explore changes in their cognitions over a 16.1 km cycling time trial using *Level 2* TA method. This study reflected in-event decision-making processes in an endurance outdoor sport whilst comparing real-time concurrent thought processes of skilled and less-skilled athletes, alongside performance data of heart rate, speed, time, and power to compare performance between skilled and less-skilled cyclists. Findings of the study support TA a viable method to collect real-time concurrent cognitive thoughts corroborate with findings from another cycling study (Whitehead et al., 2018) involving a total of 30 cyclists grouped into trained and untrained cyclist to investigate the relationship between concurrent cognitive processes and pacing behaviour during endurance cycling performance using a novel TA method. Results of the study showed the number of verbalisation did not significantly differ between trained and untrained cyclists to verbalise their cognitive process throughout the time-trial. The semi-structured telephone post time-trial interviews with participants exploring their perceptions of using TA within 48 hours following completion of the time-trials supported the use of TA to collect concurrent data in endurance sport.

To date there have been several studies in self-pacing golf (Arsal et al., 2016; Calmeiro & Tenenbaum, 2011; Kaiseler et al., 2013; Nicholls & Polman, 2008; Whitehead, Taylor, et al., 2016) using *Level 2* TA for collecting cognitive processes to measure stress and

coping strategies among golfers. Findings of the studies (Arsal et al., 2016; Whitehead, Taylor, et al., 2016) showed more thoughts were verbalised overall by more-skilled golfers than less-skilled golfers using TA method. Nevertheless, Whitehead et al. (2015) utilising both *Level 2* and *Level 3* TA revealed higher volume of verbal data collected in *Level 3* TA compared to *Level 2* TA. However, the thought sequences were altered in *Level 3* TA when participants were requested to explain their thoughts by retrieving information out of that STM (Boren & Ramey, 2000; Welsh et al., 2018; Whitehead, Cropley, et al., 2016). Participants were performed slightly better in TA conditions than that of being silent (Fox et al., 2011; Whitehead et al., 2015), however, participants took longer time to complete domain-specific primary tasks with verbalisations (Arsal et al., 2016; Ericsson & Simon, 1993; Whitehead et al., 2015). These studies provided an understanding of using *Level 2* and *Level 3* TA in capturing in-event cognitive processing differences between expert-novice and across genders. This is despite Nicholls and Polman (2008) stating that TA method was less appropriate in capturing more complex form of cognitive information that require time for retrospection to take place due to memory decay. *Level 2* TA method has also been adapted in a study of tennis (Swettenham et al., 2018) and long-distance running (Samson et al., 2015), whilst *Level 3* TA method adapted into a study in snooker (Welsh et al., 2018) to investigate stress and coping strategies. Therefore, there appears to be potential for utilising TA to investigate in-event decision-making of that WAG judging to reduce external bias that can alter judge's recall of experience. Indeed, there are no naturalistic studies examining the real-time thoughts of WAG judges to date.

WAG judges who are qualified subsequently develop to that of experts through a combination of formalised education and accumulation of judging experience over years. Nevertheless, previous studies reported there were score deviations across a judging panel even among higher-level judges in international competitions (Mercier & Klahn, 2017; Pajek

et al., 2014; Sacchi, 2018a, 2018b). MacMahon and Mildenhall (2012) therefore highlighted a need to discover *how* sport officials perform the judging task and mechanisms driving decision-making behaviours. Previous research (Catteeuw, Gilis, Jaspers, Wagemans, & Helsen, 2010; MacMahon & Ste-Marie, 2002; Ste-Marie, 1999) advocate the use of tasks that mimic real-life to study expert-novice differences in officials. Meanwhile, TA method been utilised to investigate expert-novice differences of athletes (Arsal et al., 2016; Calmeiro & Tenenbaum, 2011; Welsh et al., 2018; Whitehead et al., 2018; Whitehead et al., 2015; Whitehead, Taylor, et al., 2016) by collecting in-event cognitive processes whilst completing domain-specific tasks. Hence, TA deemed appropriate as a viable method to report mediating thought processes underpinning decision-making of WAG judges when evaluating gymnastics skills and dance elements. TA therefore allow judges to verbalise their thoughts concurrently when applying execution deductions whilst evaluating gymnastics skills and dance elements by watching competition videos resemble competition setting. Therefore, this PhD work was to explore if TA method is viable to access decision-making of WAG judges across all four apparatus, that of BB, FX, VT, and UB to inform future judge education.

2.9 Aims and Objectives

This PhD work consists of three studies across two data collection points to explore the use of a TA method in current WAG judge education. As this PhD been funded by the Ministry of Education Malaysia, it was expected to include participants from both Malaysia and United Kingdom. Study 1 was conducted in 2016 with application of the Code of Point (COP) 2012-2016 (FIG, 2012) and involved Malaysian national WAG judges as participants. Study 2 and Study 3 was conducted from 2017 to 2018, therefore the Code of Point (COP) 2017-2020 (FIG, 2016b) was in used involving British WAG judges inclusive of international, national, regional, and club judges.

Study One in this PhD aimed to explore decision-making underpinning the judging process in WAG by using both concurrent and immediate retrospective verbal report to examine the utilisation of TA method. The purpose of this study was to explore the use of TA method as a training method with Malaysian based WAG judges. Specifically this was an examination of judge's decision-making by requesting participants to concurrently and immediate retrospectively verbalise all execution deductions applied whilst judging video simulated routines of a single apparatus (BB) using the Code of Point (COP) 2012-2016. In addition, this study also collect judge's perceptions in using TA whilst judging for the purpose to inform future judge education. Further, this also served as a pilot study for Study Two.

Study Two aimed to explore decision-making differences between expert and novice WAG judges using TA method with findings informed by Study One. To further explore the viability of adapting TA method into WAG judging, this study involving 'expert' and 'novice' WAG judges ranged from international, national, regional, and club levels accredited by British Gymnastics in the United Kingdom using the Code of Point (COP) 2017-2020. International and national judges were grouped as expert, whilst regional and club judges were grouped as novice. Participants were required to individually judge fixed-sequence video routines of all four WAG apparatus inclusive of BB, FX, UB, and VT, as far as practically possible, to resemble competition situation. E-scores and TA session data on execution deductions applied whilst judging video routines were statistically compared to explore the decision-making differences between expert and novice judges.

Based on outcomes from previous studies, Study Three aimed to investigate the perceptions of expert-novice WAG judges in using TA method whilst judging video-based competition routines. Semi-structured interviews were conducted with same participants from Study Two to collect their responses in using TA method to verbalise execution deductions

whilst judging video-based routines across all four WAG apparatus to inform viability of TA method into future development of WAG judge education.

CHAPTER 3

Exploring the use of Think Aloud within Women's Artistic Gymnastics Judging Education

The information presented in this chapter has been reported in the paper:

Lee, J., Knowles, Z., & Whitehead, A. E. (2019). Exploring the use of think aloud within Women's artistic gymnastics judging education. Psychology of Sport & Exercise, 40, 135-142. doi:10.1016/j.psychsport.2018.10.007

CHAPTER 4

Exploring Decision-making Differences between Expert and Novice Women's Artistic Gymnastics Judges using Think Aloud Method

4.1 Introduction

Following outcomes from Study One (see *Chapter Three*) that participated by ten Malaysian national and international WAG judges which had taken place in Malaysia that focused on a singular apparatus of the BB, data collected and measured by the researcher revealing there were different deduction scores by judges from evaluating same video routines. The combination of using TA method and interview in the Study One aligns with Garcı´a-Gonza´lez, Moreno, Moreno, Gil, and del Villar (2013) methods whereby to provide a structure on knowledge development using video feedback, reflection and questioning within a single study. In the present study demonstration of explicit and implicit learning processes by participants to adapt TA method into WAG judging was demonstrated through concurrent verbalization and detailed elaboration concerning perceptions of using TA method whilst judging. This informed the viability of TA method in accessing in-event cognitive processes when making decisions. The purpose of Study One was to explore the in-event decision-making with Malaysia based WAG judges on a single apparatus of the BB and to examine perceptions of utilising TA method whilst judging video simulated routines to adapt into future judge education. The data collection sessions conducted with a single participant with researcher. Ten international level BB routines with a range of execution scores and execution errors of gymnasts from several nations globally and across several competitions were randomly selected from publically available sources were judged according to the FIG COP 2012-2016 without knowing the competition scores in advance. However, it is accepted judges may have attended the competition where the BB routines were recorded or

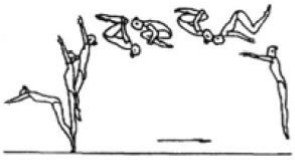
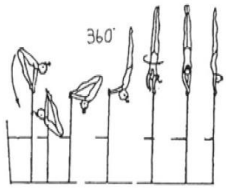
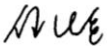
subsequently viewed the video clips. They were not in a contracted judging role for the events. Hence, decision-making differences between judges in all four WAG apparatus of that BB, FX, VT, and UB were concerned. Gymnasts perform their routines consisting elements allowed for a particular apparatus within permitted time composed of maximum elements with highest difficulty and connections in addition to performing the routine with minimal execution deductions. General execution faults are applied when gymnastics skills and dance elements performed deviate from the expected specific technical standards according to the COP and requirements of a competition, whilst specific apparatus deductions are applied according to the exact requirements of respective apparatus. Artistry deductions applicable only for BB and FX are evaluations on performance of a gymnast in demonstrating their ability to transform a routine from a well-structure composition, which are perceptions towards the routine composition and choreography in addition to rhythm and tempo during performance. E-judges are responsible for their own judgment and discussion is not allowed during the judging process. The sum of the E-score awarded to a gymnast is the average deduction score provided by the panel of E-judges deducted from the maximum deduction score of ten points (10.0), excluding the highest and lowest scores to reduce the "halo-effect" (McFee, 2013) bias.

Decision-making can be defined as the ability to use information from current situation and associated knowledge possessed to plan, select and execute an appropriate goal-directed action (Williams & Ford, 2013). In this study, the action indicates the decision-making of E-judges to apply execution deductions when judging WAG routines. Accurate decisions are applied after accounting for all available information by watching and evaluating a routine, which requires the integration of perceptual and cognitive information. Researchers (MacMahon & Starkes, 2008; Summerfield & de Lange, 2014) have highlighted

challenges faced by sport officials, that they have to possess perceptual-cognitive skill for processing incomplete, intentionally deceptive and fast-paced information under time pressure during a competition (MacMahon & Mildenhall, 2012). The importance of trained officials with experience for domain-specific decision-making has been shown, whereby expert officials are able to provide clear criteria for decision-making as compared to players and experienced observers (Woods, Kranjec, Lehet, & Chatterjee, 2015). However, findings from Nevill, Hemingway, Greaves, Dallaway, and Devonport (2016) revealed even qualified officials make biased decisions due to presence of stimuli, such as home and away supporter. These studies revealed accredited sport officials might undertake poor or inaccurate decision-making during officiating due to subjective elements besides interpreting rules and regulations governing the competition. Although somewhat limited in scope, several common biases in judges were found across aesthetic sports, such as patriotism (Boen, Hoye, et al., 2006; Leskošek et al., 2012), halo effect (Nufer & Alesi, 2018), memory-influenced (Ste-Marie & Valiquette, 1996; Ste-Marie et al., 2001), reputation effect (Findlay & Ste-Marie, 2004), order effect (Kramer, 2017), and conformity bias (Auweele et al., 2004; Boen et al., 2008). Poor inter-rater reliability and substandard validity have also been reported in a study based on 194 gymnasts in the World Championship in Tokyo 2011 (Pajek et al., 2014). There was relatively large variability of average total artistry deduction applied across expert judges within the same components of artistry denoting poor inter-rater reliability, ranging from 0.18 to 0.39 points. Further, substandard validity demonstrated by large deviations in reliability from the monitoring the artistry of competitors and significant values of systematic under- or over-rating. The dispersion of mean deductions was relatively large in addition to some calculation mistakes in the summation of artistry deductions were also noted. However, this study did not explore the intra-rater reliability concerning the artistry scores obtained during an actual competition. The poor inter-rater reliability and substandard validity could be

explained by the application of new rules on artistry evaluation and national biases from judges. Subsequently, the further research to improve the accuracy and consistency of judging scores been suggested.

Table 4.1 Descriptions and symbol notations of some gymnastics elements, adapted from FIG (2016b)

Element Description	Double Twist *Yurchenko* on VT	Double jump backward tuck on FX	Piked sole circle with backward turn on UB
Figure as in the COP			
Symbol Notation			

It has been shown that expert judges with accumulated judging experience over ten years were better at perceptually anticipating upcoming gymnastic elements and therefore those elements were judged more accurately (Ste-Marie, 1999). In WAG, only the higher-level accredited judges, categories 1 and 2 international judges, are eligible to judge at World Championships (FIG, 2016a), thus by virtue they have accumulated more judging experience compared to lower-level accredited judges. A study in rhythmic gymnastics (Flessas et al., 2015) reported that international judges outperformed national and novice judges in detecting execution faults during routine performances, therefore suggested that experienced judges probably make use of cognitive strategies to increase their overall execution detection efficiency. Therefore, they were able to process information within restricted time limit whilst multitasking the judging task. A recent study (Pizzera et al., 2018) examined the superior judging performance by exploring specific gaze behaviour between experienced judges and gymnasts. This study investigated where do experienced judges and gymnasts look while judging also supported that the higher accredited international judges showed

higher score accuracy in judging performance, which deviated less from the reference score as compared to lower-level accredited international judges when asked to judge a skill on the VT. Furthermore, Campo and Gracia (2017), who explored visual patterns and judgement on single-skills (see *Table 4.1*), reported that a judge possessing more than 10 years of judging experience was able to judge more accurately as compared to coach or gymnast, especially on gymnastics elements involving fast and complex movements in the dynamic performance environment.

The COP has altered significantly over the years, including elements difficulty expansion, judges scoring system, competition format, and education and certification courses (Grossfeld, 2014). Indeed, researchers in recent years have explored score consistency and accuracy among different level of judge's expertise in all four apparatus of BB, FX, VT, and UB across competitions (Pajek et al., 2013; Pajek et al., 2014). Based on the increasingly complex and dynamic nature of the WAG judging process, it is deemed appropriate to explore the accuracy of decision-making among judges in applying deductions, as well as calibration of judge's score following an updated version of COP at the beginning of every Olympic cycle (Mercier & Klahn, 2017). In previous studies, methods used to collect decision-making in judging included gaze behaviour analyses between higher and lower-rank level judges on vault skills in a video-based test (Pizzera et al., 2018), score accuracy analyses after major competitions (Mercier & Klahn, 2017), and visual search patterns analyses among judges, coaches, and gymnasts (Campo & Gracia, 2017). More recently, Think Aloud (TA) method were used to understand a judge's cognitive focus during the process of judging (Lee et al., 2019). Results of the study identified that the *Level 2* TA method request judges to verbalise all deductions applied whilst judging a series of BB competition videos has potential to capture judge's cognitions throughout the judging

process, in addition to reveal that there were decision-making differences among judges to apply deduction scores objectively.

Therefore, the aim of this study was to explore decision-making differences between expert and novice WAG judges accredited for Olympic Cycle 2017-2020 on all execution deductions, inclusive of general execution faults, specific apparatus deductions, and artistry deductions by using TA method to further developing the understanding of expert-novice WAG judge decision-making. Based on previous research, it was anticipated that expert judges would apply more deductions across all apparatus as compared to novice judges, driven by their own previous experience in performing the judging task. Specifically, it was hypothesized that the expert judges would able to notice more execution errors across all apparatus, as well as higher expectation towards artistry performance in a BB and FX routines, therefore more deductions applied by expert judges whilst judging routines due to the combination of previous judging experience and higher ability of information-processing.

4.2 Method

4.2.1 Participants

Participants were three international and seven national WAG judges in the expert group (n=10), and four regional and four club WAG judges in the novice group (n=8). Data collection taken place in the UK with researcher as the data collector, whereby the British Gymnastics acted as gatekeeper contacting all eligible participants accredited for Olympic Cycle 2017-2020. Demographic data of participants are summarised in Table 4.2. Judging experience of participants taken into account by the BG is inclusive of club, regional, and national level, or international level accreditation by the FIG. Coaching experience of participants is taken into account if they are an accredited coach by the BG, regardless of level. Previous studies involving only a judge and a coach have reported that there were no

differences between judge and coach in visual search strategies, judgement accuracy, and correspondence between the visual and verbal behaviours in reporting the two most important visual location (Campo & Gracia, 2017). However, the study had with a limited participant sample and revealed the judge had a stronger relationship between gaze and verbal reporting due to higher percentage of concordance between the visual and verbalisation behaviour despite of demonstrating superior performance to spot execution errors. Hence, the current study includes participant demographic information to understand their judging and coaching experiences without further exploration of previous visual and motor experiences in shaping judging performance and visual search strategies.

Table 4.2 Minimum, maximum, mean, and standard deviation of participants' judging and coaching experiences

	Expert Judges				Novice Judges			
	Min	Max	*M*	*SD*	Min	Max	*M*	*SD*
Judging experience (years)	4	6	5.30	0.82	1	5	2.88	1.13
Coaching experience (years)	0	50	17.30	15.66	0	8	3.00	3.51

4.2.2 Materials and Procedures

Similar procedures and video clips for general TA training and TA demonstration were adopted from a previous study (Lee et al., 2019). Prior to the TA tasks, participants were trained for *Level 2* TA in verbalising thoughts by say out loud activities adapting instructions from original protocols of Ericsson and Simon (1993, pp. 375-379) and McRobert et al. (2017), such as verbalising what is the next alphabet after "A" and calculating how many dots appeared on screen. These verbalisation warm up tasks were followed by domain-specific task with gymnastics specific examples, which had involved verbalising deductions on BB elements. A fixed sequence video montage of five BB, five FX, two VT, and two UB routines were compiled for participants to verbalise deductions whilst

judging. Video clips of BB and FX routines were pre-selected by British Gymnastics Women's Technical Committee to resemble local competition levels. VT and UB routines were edited from publicly available sources, whereby the gymnasts from several nations globally and across several competitions were selected randomly to meet international competition standards. The aim of this study is to explore decision-making differences between expert and novice judges participants within this study by using both deduction scores and verbalised execution deductions, therefore, competition scores of these routines were not in use. Routines were selected at random without standardisation of score, nationality, nor execution error to enable participants to verbalise execution deductions using TA method whilst judging to resemble actual competition. However, some judges who previously judged the competition where the routines taken place or involved in a training centre where the gymnasts trained might recognised some gymnasts or had viewed the routines before data collection, which is beyond the control of the researcher. Montages were created using the Window Movie Maker® following requirements of respective apparatus as stated in the COP (FIG, 2016b), where BB and FX were set as maximum 90-second routines, whilst there were no time limits for VT and UB routines. All routine video clips for BB, VT and UB were muted to exclude background noise, except for FX, where musical interpretation is part of judging process. The footage angles were set from the side angle akin to the judge's typical angle of view during an actual competition, as compliant with judge education training protocols. The construct validity of using domain-specific video clips to explore decision-making of officials were supported with findings from previous studies (Catteeuw et al., 2010; Larkin, Mesagno, Berry, Spittle, & Harvey, 2017). A Sony® Dictaphone (model ICD-PX240) was used to record all audio responses verbalised by participants during the TA sessions. An Olympus® AS-2400 transcription kit was use to process verbatim transcription.

4.2.3 Data Collection

Participants engaged in TA method by continuously verbalising all types of deductions concurrently and immediate retrospectively, including general execution faults, specific apparatus deductions and artistry deductions, applied onto each element and skill performed on BB, FX, VT and UB routines in fixed sequences. Participants were instructed to "please think aloud and verbalise everything that come into your mind whilst judging the routines showed in videos without further explaining the reason behind", whilst judge's notation sheets were provided for writing down symbols representing elements performed as well as recording deduction scores resembling actual judging. Participants were prompted by the researcher to think aloud at the beginning of routine after the second element performed by the gymnast in a routine, if they remained silent. At the end of each routine, respondents were prompted to verbalise artistry deductions if they remained silent for 10 seconds after they had completed calculating the deduction scores and were waiting for next routine without verbalise any artistry deduction. Verbalisation during the TA session was dictated whilst all written judging sheets were collected at the end of sessions for later analyses.

4.2.4 Data Analyses

E-scores of each routine were tabulated into the IBM® SPSS Statistics 24 and Microsoft Excel® for data analyses to report descriptive sample statistics. Tests of normality using Shapiro-Wilk statistics indicated that parametric analyses were appropriate with all value >.05, except for general deductions for VT in the TA verbalisation counts. Participants (n=14) were divided into two groups according to their judging expertise (expert, novice). The first analyses compared the expertise differences (n=14) in applying deductions across the fixed-sequence routines, whereby the independent variables were the fixed-sequence routines and dependent variables were the deduction scores. The outcome demonstrates only

the differences between expert and novice judges without evaluating their judging accuracy performance. Deduction score differences were determined by a 2 group (expert, novice) x 4 apparatus (BB, FX, VT, UB) mixed-design ANOVA. In total, there were 18 participants involved in this study, however, only 14 sets of deduction score data (8 experts, 6 novices) were used for this analysis as 4 participants did not provide deduction scores for every routine. Therefore, the incomplete data sets were excluded as missing data. Next, we examined the expertise differences (n=18) in verbalising specific deductions applied whilst judging the routines, with TA verbalisation counts for general execution faults, specific apparatus, and artistry deductions as dependent variables respectively and fixed-sequence routines as independent variables in general. Separate 2 group (expert, novice) x 4 apparatus (BB, FX, VT, UB) mixed design ANOVAs were conducted to analyse TA verbalisation counts on general execution faults and specific apparatus deductions. In addition, a 2 group (expert, novice) x 2 apparatus (BB, FX) mixed design ANOVA was used to analyse TA verbalisation counts on artistry deductions. Bonferroni adjustment post-hoc analyses were performed, where significances were found to explore the impact of judge's expertise on deduction scores and TA verbalisation counts in all WAG apparatus. Two-tailed significance was accepted as $p < 0.05$ and effect sizes were reported using partial eta squared (eta^2).

TA verbalisations of all types of deductions for all four apparatus of BB, FX, VT, and UB by all participants (n=18) were transcribed verbatim and provided a total of 102 pages of font Arial size 12 with double line spacing text verbatim transcription. Consistent with previous research conducted using TA method to understand cognition (Arsal et al., 2016; Swettenham et al., 2018; Whitehead et al., 2017) a post-positivist epistemology informed this study. Therefore, a content analysis approach was used to analyse the TA data, using a coding framework, which was adapted from Lee et al. (2019) and the COP (FIG, 2016b). Therefore,

a deductive analysis was initially conducted to identify differences between expert and novice judge's deductions. Following this process, co-authors acted as critical friends reviewing 10% of transcripts independently using the framework provided to increase inter-rater reliability and avoid data interpretation bias from first author. Although inter-rater reliability has been previous criticised (Smith & McGannon, 2017), in that different authors as coders may assign a different meaning to the same code. MacPhail, Khoza, Abler, and Ranganathan (2016) suggested that when a set of guidelines or a coding framework have been developed, where the framework offers a set of firm coding rules, this will reduce the ambiguity of the coding and allow for a more reliable, however not perfect, method of ensuring for reliability. A discussion was conducted after 82% agreement was found and agreement was made for the remaining 18% difference.

Table 4.3 Coding framework used to analyse verbalisation of general faults and penalties in WAG

Secondary Theme	Primary Theme	Description	Example Raw Data Extracts
Execution Fault	Bend arms or bent knees	Any verbalisation relating to bend arms or bent knees in execution	"free walkover also .3 [deduction] on the bent legs"
	Leg or knee separations	Any verbalisation relating to leg or knee separations in execution	"knees apart .1 [deduction]"
	Legs crossed during elements with twist	Any verbalisation relating to legs crossed during elements with twist in execution	".1 [deduction] for legs crossed in the twist"
	Height of elements	Any verbalisation relating to insufficient height of elements from external amplitude	"free walkover .1 [deduction] for the [lack of] height"
	Exactness of tuck or pike position	Any verbalisation relating to insufficient exactness of tuck or pike position in single somersault	"tuck front .1 [deduction] for lack of the tuck shape"
	Stretch body posture	Any verbalisation relating to failure to maintain stretch body posture, e.g. piking too early	".1 [deduction] for body alignment"
	Hesitation of elements & movements	Any verbalisation relating to hesitation during performance of elements & movements	".1 [deduction for] hesitation"
	Deviation from straight direction	Any verbalisation relating to deviation from straight direction	"it was stepping out [from straight line] as well, so that was direction [deduction]"
	Body and/or leg position in elements (non-dance)	Any verbalisation relating to body and/or leg position in non-dance elements, includes body alignment, feet not pointed/relaxed, and insufficient split in acrobatic non-flight elements	"handstand flexed foot"
	Failure to fulfil technical requirements in dance elements	Any verbalisation relating to failure to fulfil technical requirements in dance elements, i.e. body shape	"body shape .3 [deduction] in the change leg leap"
	Precision	Any verbalisation relating to precision of angle in element execution	"Lack of spin [under-rotation on full turn]"
	Dismount too close to the apparatus	Any verbalisation relating to performance of dismount too close to the apparatus, only applicable for UB & BB	"the distance from apparatus .1 [deduction]"
Landing Faults	Legs apart on landing	Any verbalisation relating to legs apart in landing	"landing with feet [legs] apart, .1 [deduction]"
	Extra arm swings	Any verbalisation relating to extra arm swings in landing	"too many arm movements which [are] not allow in this code"
	Lack of balance	Any verbalisation relating to lack of balance in landing	"full spin big wobble .3 [deduction]"
	Extra steps, slight hop	Any verbalisation relating to extra steps and/or slight hop in landing	"back handspring layout on both foot small step, .1 [deduction]"
	Body posture fault	Any verbalisation relating to body posture fault in landing	"dismount [element] obviously her shoulder was forward [than expected]"
	Deep squat	Any verbalisation relating to deep squat in landing	"double back big step [on landing]"
	Fall	Any verbalisation relating to fall in landing, includes support on mat with 1 or 2 hands, fall on mat to knees or hips, fall on or against apparatus, failure to land feet first on landing from element	"flick into layout with fall"

Table 4.4 Coding framework used to analyse verbalisation of specific apparatus deduction in WAG

Apparatus	Faults	Description	Example Raw Data Extracts
BB	Poor rhythm in connection	Any verbalisation relating to poor rhythm in element connection	".1 [deduction] for rhythm"
	Excessive preparation	Any verbalisation relating to excessive preparation, including adjustment of unnecessary steps & movements, excessive arm swing before dance elements, and pause applied at 2- second	".1 [deduction] for tapping at the end of the beam"
	Poor body posture/amplitude throughout	Any verbalisation relating to poor body posture/amplitude throughout, include head, trunk, shoulder and arm positions, feet not pointed/relaxed/turn in, lack of work in relevé, insufficient amplitude of leg swings/kicks	"toes [feet not pointed] .1 [deduction]"
FX	Excessive preparation	Any verbalisation relating to excessive preparation, include pause applied at 2-second, adjustment with unnecessary steps, and excessive arm swing before dance elements	"a long standing [pause] in the corner"
	Poor body posture/amplitude throughout	Any verbalisation relating to poor body posture/amplitude throughout, include head, trunk, shoulder and arm positions, feet not pointed/relaxed/turn in/flat, insufficient amplitude of leg swings/kicks	"horrible feet [feet not pointed] throughout"
	Lack of variety in choreography into corners	Any verbalisation relating to distribution of elements whereby lack of variety in choreography moving into corners	"that wasn't work very much [variety of movement] into the corner, probably take .1 [deduction] for that"
VT	First Flight Phase	Any verbalisation relating to first flight phase, include missing degree of longitudinal axis during flight phase and poor technique, such as hip angle, arch, bent knees, and leg or knee separation	"angle of take-off from the springboard wasn't quite right, so .1 [deduction] for the body shape"
	Repulsion Phase	Any verbalisation relating to repulsion phase, include poor technique, such as staggered hand placement, bent arms, shoulder angle, failure to pass through vertical, prescribed longitudinal axis turn begun too early on table	".1 [deduction] for shoulder angle in the repulsion phase"
	Second Flight Phase	Any verbalisation relating to second flight phase, include excessive snap, insufficient height, body positions, bent knees, leg or knee separations, under-rotation of salto, insufficient length from vault table, deviation from straight direction, lack of dynamics	"lack of height was a .5 [deduction]"
	Landing deductions	Any verbalisation relating to landing deductions	"the landing was a deep squat, .5 [deduction], and a big step forward, .1 [deduction]"
UB	Body alignment in handstand and cast to handstand	Any verbalisation relating to body alignment in handstand and cast to handstand	".1 [deduction] for short in handstand"
	Adjusted grip position	Any verbalisation relating to grip position adjustment	"adjusted grip .1 [deduction]"
	Poor rhythm in elements	Any verbalisation relating to poor rhythm in performing elements	"the *mo shoot* lack of some rhythm"
	Insufficient height of flight elements	Any verbalisation relating to insufficient height of flight elements	".3 [deduction] on height of the swing at the back on the legs in the Pak"
	Insufficient extension in kips	Any verbalisation relating to insufficient extension in kips/upstart	".1 [deduction] on her extension of her long upstart"
	Angle of completion of elements	Any verbalisation relating to incomplete angle of completion in element execution	"angle of completion on the Pak [element], .3 [deduction]"

Table 4.5 Coding framework used to analyse verbalisation of artistry deduction in WAG

Apparatus	Secondary Theme	Primary Theme	Description	Example Raw Data Extracts
BB	Artistry performance	Confidence	Any verbalisation relating to lack of confidence	".1 [deduction] for [lack of] confidence"
		Personal style	Any verbalisation relating to lack of personal style	"she does have her own personal style"
		Rhythm & tempo	Any verbalisation relating to insufficient variation in rhythm & tempo	"rhythm and tempo .1 [deduction] because there was a lot of pauses"
		Disconnected elements & movements	Any verbalisation relating to lack of fluency in executing performance, i.e. a series of disconnected elements & movements	"a lot of stops and starts so it's disconnected elements"
	Composition	Length of beam	Any verbalisation relating to insufficient use of the entire length of beam	"she used the [entire] length of the beam"
		Dide movements	Any verbalisation relating to lack of side movements	"nothing side-ways [movements]"
		Close to the beam movements/elements	Any verbalisation relating to missing combination of movements/elements close to the beam, with part of torso touching the beam	"I don't see any close to the beam moves [elements/dance], so .1 [deduction] for lack of that"
		Complexity or creativity in the movement	Any verbalisation relating to insufficient complexity or creativity in the movement	"it was nothing complex or creative [movement] in the routine"
		One-sided use of elements	Any verbalisation relating to more than one half-turn on 2-feet with straight legs throughout the exercise	"another turn on 2-feet"
FX	Artistry Performance	Expressiveness	Any verbalisation relating to lack of expressiveness	"didn't really expressive with the music"
		Failure to engage the audience	Any verbalisation relating to gymnast unable to engage the audience	"she didn't engage the audience"
		Play a role or a character throughout	Any verbalisation relating to gymnast unable to play a role or a character throughout the performance to reflect the musical theme	"inability to play a role"
		Disconnected elements & movements	Any verbalisation relating to performance of the entire exercise as a series of disconnected elements & movements	"here again a bit of disconnected movement"
	Composition	Insufficient complexity or creativity of movements	Any verbalisation relating to insufficient complexity or creativity of movements	"nothing to complex [movement] so .1 [deduction]"
		Missing movement touching floor	Any verbalisation relating to missing movement touching floor, including minimum trunk, or tight, or knee or head	"down to the floor [with dance movements]"
	Musical and Musicality	Background music	Any verbalisation relating to exercise is connected to the music only at the beginning and end of the exercise	"incorrect of selection of music for movements yes she didn't dance"
		Lack of synchronisation between movement and musical beat	Any verbalisation relating to lack of synchronisation between movement and musical beat, including during a part of exercise, and at the end of exercise	"she looks a bit out from the music on the way of it"

4.3 Results

4.3.1 Deduction scores

The deduction scores showed in Figure 4.1 were to be taken-off from a maximum of ten points to award the final E-score. There was a significant main effect for group, $F_{1,12}$ = 7.303, p = .019, ηp^2 = .38. On average, expert judges applied more deductions across all four apparatus (*M* = 3.17, *SD* = 1.22) than novice judges (*M* = 2.53, *SD* = 1.12). There was also a significant main effect for apparatus, $F_{3,36}$ = 95.346, p < .001, ηp^2 = .89. Results showed that the mean deduction score for BB and FX were significantly higher than VT (p < .001) and UB (p < .05). Furthermore, scores on UB were significantly higher compared to VT (p < .001). There was also a significant apparatus x group interaction, $F_{3,36}$ = 2.815, p = .05, ηp^2 = .19. Expert judges applied higher deductions on FX (*M* = 3.61, *SD* = 0.95) than BB (*M* = 3.49, *SD* = 1.28), whilst the novice judges applied higher deductions on BB (*M* = 2.91, *SD* = 1.18) than FX (*M* = 2.76, *SD* = 1.06).

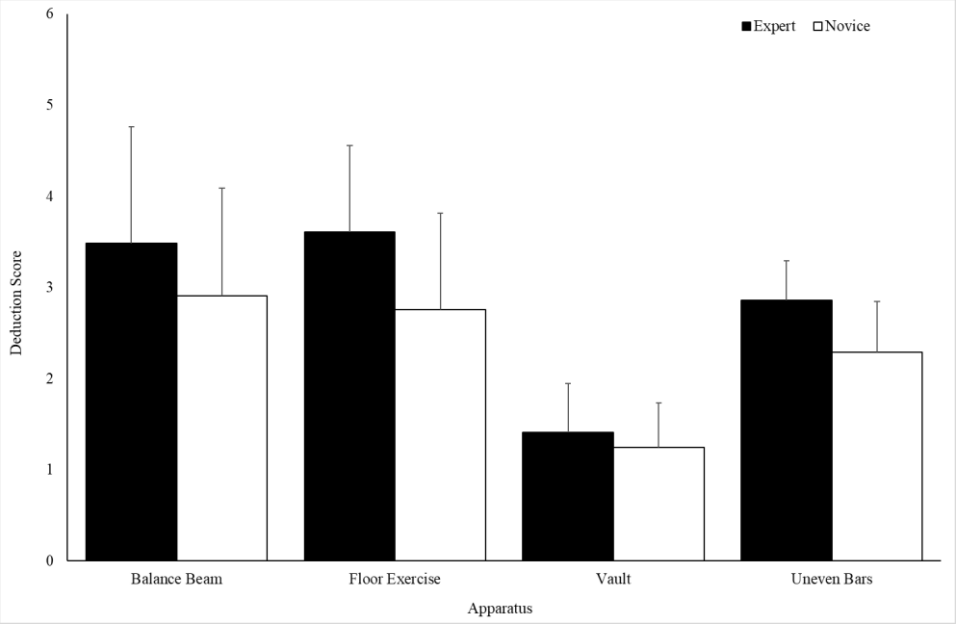


Figure 4.1 Mean (*SD*) deduction scores of expert and novice judges for the BB, FX, VT, and UB apparatus

4.3.2 TA Deduction Counts

There are three types of deductions as stated in the COP (FIG, 2016b). General execution faults were applicable to all four apparatus, whilst specific apparatus deductions were applicable to respective apparatus. Artistry deductions were applicable only to BB and FX, whereby lists of deductions with descriptions for respective apparatus were in the COP. Figure 4.2 illustrates all deductions TA by expert and novice judges across all four WAG apparatus.

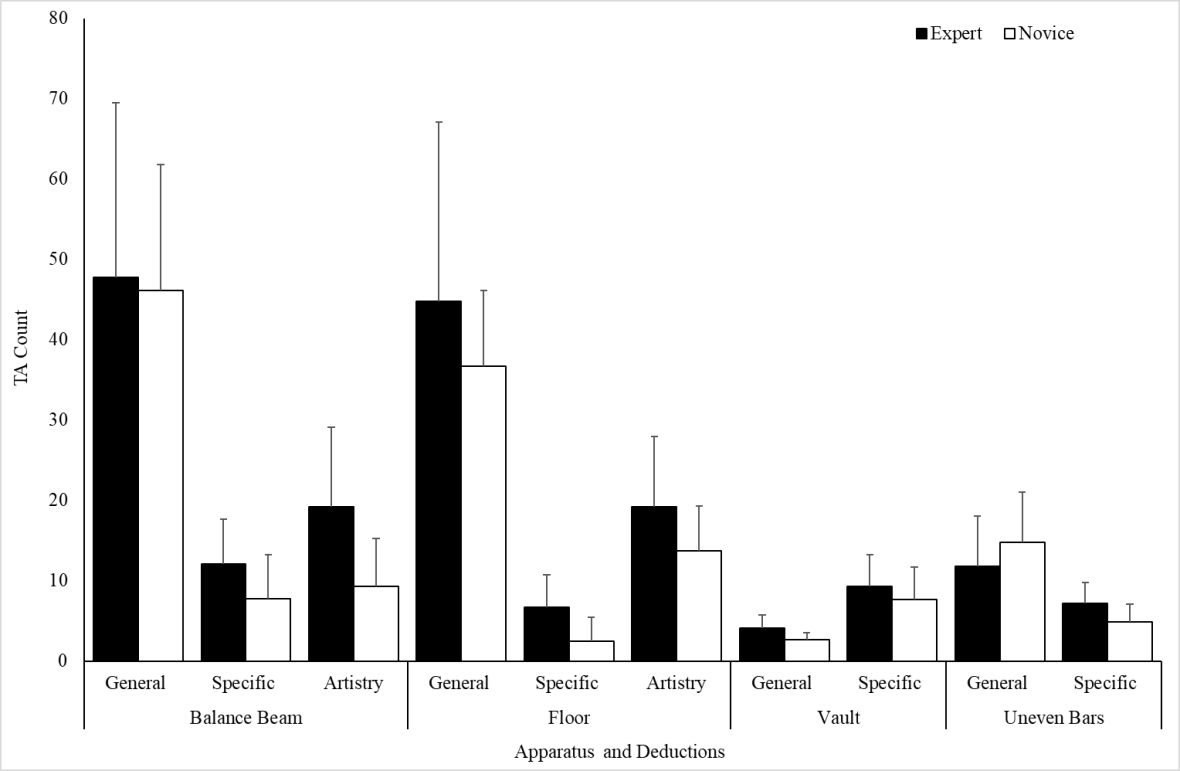


Figure 4.2 Mean (*SD*) TA counts of expert and novice judges for *general execution faults* and *specific apparatus deductions* for BB, FX, VT and UB, and *artistry deductions* for BB and FX

General execution faults

There was no significant group main effect for general execution faults, $F_{1,16} = .172$, $p = .684$, $\eta p^2 = .01$. However, there was an apparatus main effect, $F_{3,48} = 80.652$, $p < .001$, $\eta p^2 = .83$. The TA deduction counts were higher on BB ($M = 47.06$, $SD = 18.77$) and FX ($M =$

41.22, *SD* = 17.79), compared to both UB (*M* = 13.11, *SD* = 6.26; p < .001) and VT (*M* = 3.44, *SD* = 1.54; p < .001). TA deduction counts on general execution faults for UB were also significantly higher than VT (p < .001). There was no significant group x apparatus interaction, $F_{3,48} = .930$, p = .433, $\eta p^2 = .06$.

Specific apparatus deductions

There was a significant group main effect for specific apparatus deductions, $F_{1,16} = 9.845$, p = .01, $\eta p^2 = .38$. The TA deduction counts were higher for experts (*M* = 8.83, *SD* = 4.52) compared to novices (*M* = 5.69, *SD* = 4.23). There was also an apparatus main effect, $F_{3,48} = 6.470$, p = .001, $\eta p^2 = .29$. The TA deduction counts were significantly higher on BB (*M* = 10.17, *SD* = 5.79) compared to FX (*M* = 4.83, *SD* = 4.09; p < .05). There was no significant group x apparatus interaction, $F_{3,48} = .510$, p = .677, $\eta p^2 = .03$.

Artistry deductions

There was a significant group main effect for artistry deductions, $F_{1,16} = 4.648$, p = .05, $\eta p^2 = .23$. The TA deduction counts were higher for experts (*M* = 19.20, *SD* = 9.08) compared to novices (*M* = 11.50, *SD* = 6.02). There was no significant apparatus main effect, $F_{3,48} = 3.364$, p = .085, $\eta p^2 = .17$, or group x apparatus interaction, $F_{3,48} = 3.364$, p = .085, $\eta p^2 = .17$.

4.4 Discussions

This study aimed to investigate expertise differences in WAG judging execution deductions. TA method was introduced to analyse decision-making underpinning WAG judging officials by verbalising all deductions (general execution faults, specific apparatus deductions, and artistry deductions) applied on gymnastics skills and dance elements whilst judging fixed-sequence competition video routines in each apparatus. This study has

extended our knowledge of adapting the TA method to explore cognitive processes of WAG judges, who are officials in an aesthetic sport towards decision-making differences in applying execution deductions.

In this study, we hypothesized that the expert judges with accumulated judging experience over years would apply more deductions compared to novice judges due to more refined perceptual-cognitive skills that enable them to identify more execution errors across routines in all apparatus (Campo & Gracia, 2017; Pizzera et al., 2018). Outcomes of the study show that deduction scores of VT was significantly different from other three apparatus, likewise for UB. The deduction scores of VT were significant lower than other three apparatus of BB, FX, and UB besides there was no significant group differences between experts and novices in judging VT. These were supposed to the specific characteristics of VT, whereby it is a fast-pace apparatus completed within 5 seconds (Pajek et al., 2013) involving only a single gymnastics skills that evaluated across five phases, which were take-off, first flight phase, repulsion phase, second flight phase, and landing phase (FIG, 2016b). Therefore, both expert and novice judges were evaluating the performance deviation based on STM after the entire performance provided shorter time to process information concurrently. Furthermore, there was only a single gymnastics skills performed on VT compared to other three apparatus that performing minimum eight skills and elements, therefore less deductions were spotted leading to lower deduction scores compared to other apparatus were expected. However, there was no significant difference in deduction scores between BB and FX, nevertheless deduction scores of both BB and FX were significantly different from VT and UB respectively. BB and FX have three lists of deductions (general execution faults, specific apparatus deductions, and artistry deductions), whereas UB and VT only have two lists of deductions (general execution faults and specific apparatus deductions), which explains the

findings that BB and FX had higher deduction scores than UB and VT. A longer list of applicable execution deductions in BB and FX provided judges higher chances to spot skill and element executions that had deviated from specific standard executions during a routine performance that leading to higher deduction scores.

Further, the results of this study show that expert judges applied more deductions on FX than BB, whilst novice judges applied more deductions on BB than FX. The higher deduction scores applied on BB compared to FX by novice judges corroborate post-competition scoring analyses for World Championship 2018 (Sacchi, 2018b) and Youth Olympics 2018 (Sacchi, 2018a), whereby deductions applied in BB were higher than FX in overall. There are several possible explanations for these results concerning common gymnastics skills and dance elements were performed on both BB and FX. The higher repetitions in retrieving the general execution faults listed in the COP (FIG, 2016b) whilst judging BB and FX may have strengthened memory (Chase & Ericsson, 1982) to recall the deduction list resulting higher deductions on BB in addition to that of a longer time for information-processing in the slower pace BB. Moreover, routine performances of BB were on a single line required less perceptual-cognitive skills to evaluate the precision of degree completion for elements when compared to FX routines performed in 360 degrees reduced their information-processing demands to evaluate the deviation of skills and elements. Novice judges with lower deductions applied in FX routines were supposed lack of awareness and conscious thoughts in applying execution deductions for series of movements with faster-pace requiring decision-making with time constraint. On the other hand, findings show the expert judges applied more deductions on FX routines than BB routines contradicted to previous findings (Sacchi, 2018a, 2018b). There are more subjective evaluations on FX concerning artistry deductions and musicality with supposed lower deductions from expert

judges in applying the rule could be attributed to "when in doubt, give the benefit of that doubt to the gymnast" (FIG, 2016b). Another explanation for FX deductions were higher than BB in this study, whereby participants were judging FX routines after BB routines, thus they applied deductions on common gymnastics skills and dance elements automatically by schemata (Baddeley, 2012) that demand little of attentional control.

As predicted, data from the current study demonstrated that the expert group verbalised significantly more deductions across all four WAG apparatus video routines, when compared to the novice group. Furthermore, expert judges reported higher deduction counts on all three types of deductions compared to novice judges. These data indicated that expert judges were able to manage multitask judging (Ste-Marie, 1999) whilst having better 'judging eyes' (Campo & Gracia, 2017) to perceive visual information in detecting more errors than novice judge within same allocated time. These data corroborate previous literature reporting more refined information-processing (Ste-Marie, 2000, 2003) and anticipation for up-coming elements (Loffing & al-Bruland, 2017; Ste-Marie, 1999), which enable expert judges to process domain-specific information (Williams & Ericsson, 2005) and verbalise more deductions by retrieving information from LTM in a shorter time. We suggested expert judges were able to expand their effective working memory capacity according to the skilled memory theory (Chase & Ericsson, 1982), which allowed stable states of cognitive processes to be stored in LTM and kept directly accessible by means of retrieval cues in LT-WM within short time that leading to more TA verbalisation deduction counts. Expert judges showed their effective working memory capacity in writing the symbols representing skills and elements whilst processing information to evaluate performance deviations from the COP at the same time verbalising specific execution deductions applied onto each skill and element (see *Appendices D, E, and G*). On the other

hand, novice judges were unable to write the symbols whilst verbalising deductions applied when judging routines (see *Appendices H-K*) within the same allocation time. These demonstrated that expert judges with enhanced judging experiences outperformed novice judges during judging assuming frequent information retrievals had expand working memory capacity of experts that enable them to utilise working memory more than novices.

For general execution faults, there was no statistically significant difference found between expert and novice judges. It was suggested that the same list of general execution faults stated in the COP (FIG, 2016b, p. 29) were common faults applied to all apparatus had received equal attention regardless of expertise. Frequent storage and recall of general execution faults from STM (Atkinson & Shiffrin, 1968) and high revision repetition enhanced the information retrieval (Chase & Ericsson, 1982) when judging regardless of apparatus judging task assigned. Unlike for specific apparatus deductions and artistry deductions, expert judges verbalise significantly more deductions than novice judges. There were different lists of specific apparatus deductions for each apparatus, which perceived lower recall frequency from STM as respective information only needed when judging specific apparatus. Expert judges with accumulated judging experience had watched more routines compared to novice judges over years. We therefore suggested they possessed higher ability to anticipate routine pattern with advance cues of a gymnast's movement that enhanced the use of pattern recognition and situational probabilistic to reduce the demand on working memory by using structured and systematic visual search patterns (Mann et al., 2019). These were unlikely for novice judges that required longer time to retrieve information from LT-WM (Ericsson & Kintsch, 1995), which results in fewer deductions compared to expert judges. Meanwhile, this study was conducted at the beginning of cycle with recently updated COP (2017-2020), whereby changes applied on specific apparatus deductions. We

therefore suggested higher TA deduction counts verbalised by expert judges were due to conscious awareness in expertise to retrieve COP knowledge via supervisory attentional system (Norman & Shallice, 1986), which had informed them to apply specific apparatus deductions accurately in each singular apparatus according to latest COP despite of general execution faults that retrieved from schemata.

This study also revealed highest TA deduction counts verbalised by both expert-novice judges were recorded for BB, followed by FX, subsequently on UB, and least for VT. These data were congruent with previous findings reporting verbal overshadowing in TA method, whereby more time is required to complete verbalisations in addition to ordinary multitasking judging (Chin & Schooler, 2008; Ericsson, 2003; Fox et al., 2011; Lee et al., 2019; Meissner & Brigham, 2001; Schooler, 2011). Judging fast-paced apparatus in WAG required judges to attentively observe the routine, and at the same time evaluate the deviation of elements performed from required technical specification as stated in the COP whilst writing down symbols representing elements in addition to deductions, where WM was required to manage these multitask judging. MacMahon and Mildenhall (2012) stated challenges for a sport official to possess perceptual-cognitive skill in processing incomplete, intentionally deceptive, and fast-paced information under time pressure during a competition. Information-processing and memory are required for WAG judges to perform the judging task efficiently, whilst thinking aloud converting what is in the STM (Baddeley, 2012), which store 7±2 seconds bits of information to verbalisation proved challenging, especially to novice judges. Therefore, slowest-pace BB routines had recorded most TA deductions verbalised concurrently whilst judging the video routines.

In contrast to the slowest pace BB, VT received least verbalised deductions overall, which could be explained due to it is the fastest pace apparatus with singular element

performance in addition to different judging criteria than other apparatus. VT had recorded higher TA deduction counts on specific apparatus deductions as compared to other apparatus due to a longer list stated in the COP (FIG, 2016b). A singular element VT was judged in four phases, which are *first flight* phase, *repulsion* phase, *second flight* phase, and *landing* phase with specific deductions applied in respective phases (FIG, 2016b, p. 42). This is different to other apparatus, which must include at least eight elements in a routine of BB, FX, and UB. There are time constraints for judges to verbalise all deductions concurrently whilst watching the high-speed singular vaulting element, therefore evaluations were done retrospectively. However, accuracy of deductions verbalised were concerned due to limited visual STM capacity that depreciates over time (Phillips, 1974). It was explained that anticipation enabled experienced judges to detect and recognise execution errors through selective attention (Zhao, Al-Aidroos, & Turk-Browne, 2013), which facilitates perception by prioritising sensory inputs (Summerfield & de Lange, 2014) to recall related information in performing task (Desimone & Duncan, 1995) within a short time. These were consistent with previous studies (Tenenbaum, Levy-Kolker, Sade, Liebermann, & Lidor, 1996; Thomas & Thomas, 1994) reporting that expert performers possessed advanced anticipatory skills are higher ability than novices in using advance visual information to bypass potential limits of information-processing thus quickly access knowledge structures (Tenenbaum et al., 1996). Experts' visual search activities under control of LT-WM (Ericsson & Kintsch, 1995) enable them to direct attention to cue a larger area around the visual fixation point, thereby eliminating irrelevant information from being elaborated upon in LTM to allow for faster decision-making (Mann et al., 2019). Therefore, results in this study suggest that experts were able to spot more execution errors than novice was. However, the objectivity of the execution scores remain arguable due to limitations in human memory capacity. Therefore, the real-time judging support (FIG, 2019) currently under development by collaboration

between FIG and Fujitsu Limited using technology in capturing the gymnasts' movements with a 3D laser sensor then analysing them as numerical data could overcome limitations of gaze behaviour to award fair and accurate final score to fast-paced singular element VT occurred in short time.

Looking at both BB and FX shared similar judging criteria, there was no significant difference of deduction scores and TA verbalisation counts on general execution faults between BB and FX in this study. We suggest these were due to shared gymnastics skills and dance elements in both apparatus using same symbol notations and execution deductions, which had increased familiarisation of judges in applying deductions caused by schemata and declarative knowledge (Ste-Marie, 1999). Furthermore, both BB and FX shared another similarity, whereby there is an extra judging criterion of artistry deductions. Judges applied artistry deductions at the end of a routine of BB or FX, with perceptions towards the routine composition and choreography in addition to rhythm and tempo during performance by referring to the artistry deduction list for respective apparatus. Therefore, more deductions were applied onto BB and FX compared to VT and UB. These were in addition to general execution faults and specific apparatus deductions likewise listed for VT and UB. Moreover, higher TA deduction counts reported on artistry deductions compared to specific apparatus deductions in BB and FX, perhaps due to the similarities of artistry deductions in both apparatus, alongside the artistry deduction implementation since 1996 (Sengupta & Paul, 2018), which had received high concern over the pass few cycles on reliability and objectivity (Pajek et al., 2014). We further explained this had made FX recorded least TA deductions in specific apparatus deductions, as attentions been distracted to artistry deductions in additional to musical components.

This study revealed there were systematic difference between expert and novice judges in evaluating same video routines, whereby expert judge applied higher deduction scores across all four WAG apparatus. Findings of the study show that deduction scores and TA deduction counts for all three types of execution deductions across all four apparatus were different, which showed differences underpinning decision-making process based on experience possessed by judges. To our best knowledge, judging experience accumulated by expert judges increases multitask ability aforementioned whilst officiating in WAG. Further, engagement of correct type and amount of practice activities facilitates the acquisition of perceptual-cognitive skills by matching particular domain-specific information processing within that specific apparatus (Williams et al., 2011). Therefore, this study provided implications to WAG community regarding decision-making differences between expert and novice judges in officiating video routines resemble actual competitions. Therefore, suggestion was made for future study to include history questionnaire to explore gymnastics background of expert and novice judges. Besides, adapting TA method into studies exploring gaze behaviour when evaluating routines provides rich information to reveal in-event decision-making made by judges. Awareness raised to coaches and gymnasts as outcomes of the study had reported highest deductions were recorded for BB, followed by FX, subsequently on UB, and least deductions for VT, therefore expectation on final scores should adjust accordingly with respective to apparatus. Furthermore, it is deemed appropriate for future judge education to increase awareness in judging different WAG apparatus of BB, FX, VT and UB, looking at the unique criteria of respective apparatus, in addition to the latest updated deduction lists in respective cycle. Furthermore, a more consistent training environment mimic actual competition situation (McRobert et al., 2011; Williams & Ford, 2013) provided during judges training facilitated them to circumvent time constraint in completing multitask judge assignment. Further research should investigate perceptions of

expert and novice judges in using TA method within WAG judge education across all four apparatus to further developing an effective judge education towards achieving score accuracy and objectivity, as well as developing anticipation contribute to decision-making (Raab et al., 2019).

In order to complement previous studies (Lee et al., 2019; Pajek et al., 2014; Pizzera, 2012; Pizzera et al., 2018) investigating decision-making difference between expert and novice WAG judges by exploring singular apparatus, we included all four apparatus of BB, FX, UB, and VT. According to outcomes of the study, we therefore suggested deduction score of each apparatus likely to be independent and different from other apparatus due to specific requirements set for each apparatus hence are not comparable.

4.5 Conclusion

Findings of the current study suggest that there were systematic expertise differences in judging deduction scores and verbalised deduction counts across all four apparatus in WAG indicating differences between expert and novice judges, excluding VT deduction scores and general execution faults across all four apparatus. Using the novel method of TA, which has yet to be used to explore expertise differences in WAG judging, these data suggest that experts possess more domain-specific knowledge and more refined perceptual-cognitive skills in order to effectively and efficiently process complex and dynamic information under severe temporal constraints. Expert judges were able to spot more execution errors that deviated from the standard COP in respective apparatus thus applying more deductions whilst judging routines compared to novice judges indicating they possessed domain-specific knowledge. Moreover, expert judges were showing higher ability in verbalising specific execution deductions applied onto each skill and element within the same allocation time demonstrating their perceptual-cognitive skills and ability to retrieve COP knowledge within

shorter time. These differences highlight future considerations for WAG judge education in supporting novice judges to achieve objective and reliable judging scores through deduction verbalisations, with considering the unique judging criteria of respective apparatus.

CHAPTER 5

Investigating the perceptions of expert-novice Women's Artistic Gymnastics judges in using Think Aloud method whilst judging video-based competition routines

5.1 Introduction

WAG judges reaccredit their qualifications by way of examination once in every four years corresponding with Olympic cycles. Accreditation examinations update judges on rules and regulations in that current cycle, in addition to ensure ability in judging routines across all four WAG apparatus accurately and objectively. There are currently four levels of WAG judges in the United Kingdom, whereby the highest rank of judge, international judges, are accredited by the FIG, whilst national, regional, and club judges are accredited by the BG. However, different countries and nations have different accreditation systems. Post-competition scoring analyses reports revealed there were score variations even by expert judges officiating international competitions held during 2013-2016 Olympic cycle (Mercier & Klahn, 2017). It is necessary to enforce the COP application from the beginning of Olympic cycle to avoid over generous or strict judgements, in order to ensure the correct ranking orders of gymnasts in a competition. Official post-competition report of the 2018 Youth Olympic Games (Sacchi, 2018a) and 2018 World Championships (Sacchi, 2018b) reinforced there were score deviations across judges within the execution panel.

Outcomes from Study Two (see *Chapter Four*) revealed there were statistically significant differences of deduction scores applied by expert and novice WAG judges when evaluating fixed sequence competition video routines on BB, FX, and UB. These demonstrate that expert judges were able to spot more execution errors than novice judges within same amount of time during a routine performance. Moreover, judges were verbalising specific deductions applied onto each skill and element concurrently whilst evaluating a routine by

using TA method and thus revealed the cognitive processes on decision-makings in applying execution deductions. These corroborate with studies (Pizzera, 2012; Pizzera et al., 2018) concerning judgement difference across expert and novice WAG judges. In general, expert judges were verbalising *more* execution deductions in all routines across all four apparatus, except for general execution faults on UB. However, there were significant group main effect found on specific apparatus deductions and artistry deductions. Therefore, it was hypothesised that judging experience accumulated over time had enabled expert judges to note more execution errors during routine performance thus the higher volume of execution deductions. However, more detail needed to inform the decision-making during in-event judging to reveal the cognitive processes by acknowledging the drawback of retrospective verbal reports due to memory decay over times after completing domain-specific tasks. (Bernard et al., 1984; Ericsson & Kintsch, 1995; Ericsson & Simon, 1993; Nicholls & Polman, 2008; Whitehead et al., 2015). In Study One (see *Chapter Three*), immediate retrospective recall interviews following the completion of TA tasks explored participants perceptions of using TA method in respective roles. Social validation through interview enable participants to subjectively expand their thoughts in addition to the concurrent tasks to strengthen external validity towards future delivery of the effective TA method intervention (Page & Thelwell, 2013; Whitehead, Cropley, et al., 2016; Whitehead et al., 2018). Moreover, interviews provided time for participants to express their feelings and opinions when using the TA method whilst judging therefore enable rich information concerning their perceptions of using TA method. In addition, participants were also provided opportunity to feedback the viability of TA method across all apparatus in WAG.

The purpose of this study was to explore the perceptions of expert and novice WAG judges accredited by the BG with experience of using *Level 2* and immediate retrospective

TA method to verbalise execution deductions whilst judging competition videos across all four apparatus, i.e. BB, FX, VT, and UB. This study extends the Study One by involving two groups of participants, i.e. expert and novice judges to collect their perceptions in using the TA method across all four apparatus, whereby Study One includes only a singular apparatus of that BB. Besides, this study also extends the Study Two that statistically analyses the deduction score and verbalisation deduction counts by conducting follow-up interviews to collect their perceptions to report the viability of using the TA method whilst judging.

5.2 Method

5.2.1 Researcher Position

A qualitative interpretativist approach underpinned by relativist ontology and a constructive epistemology was adapted to explore the subjective experiences of using TA method in judging WAG competition video routines in all apparatus via interview. The first author was a BG accredited national judge with three years of judging experiences in the United Kingdom (UK). Further, she attended national, regional and club judge courses for Cycle 2017-2020 organised by the BG that had offered detailed knowledge of current judge education in the UK. This prior knowledge and shared experience enabled an 'insider view' towards WAG judging, which has helped building rapport and interactions with other BG judges across judging local, regional and national competitions within the UK. Being an 'insider' in the research process was also an advantage for the first author as she has a shared knowledge and understanding of the WAG culture and language for data interpretation (Kerstetter, 2012; Saidin & Yaacob, 2016). As such this subsequently influenced the research design to include an inductive approach of constructivist epistemology to co-create understanding (Lee , 2012) in perceptions of using TA method whilst judging WAG video routines when analysis interview transcripts.

5.2.2 Participants

As this study involved all participants from Study Two (see *Chapter 4*), participants were three international and seven national WAG judges in the expert group (n=10) and four regional and four club WAG judges in the novice group (n=8). BG acted as gatekeeper contacting all eligible participants accredited for Cycle 2017-2020 with approval from the University Ethics Committee. Demographic data of participants recruited with convenience sampling based on the criteria set for dependability are summarised in the Table 5.1 below.

Table 5.1 Minimum, maximum, mean, and standard deviation of the judges' experiences and their coaching experiences comparing expert and novice judges

	Expert				Novice			
	Min	Max	*M*	*SD*	Min	Max	*M*	*SD*
Judging experience (years)	4	6	5.30	0.82	1	5	2.88	1.13
Coaching experience (years)	0	50	17.30	15.66	0	8	3.00	3.51

5.2.3 Materials and Procedures

Video clips for general TA training and TA demonstration on BB were adopted from previous study (Lee et al., 2019) and a video montage containing fixed sequence routines of BB, FX, VT, and UB (see *Study 2*) were used to provide TA sessions for participants. Participants were asked to concurrently verbalising their thought processes when applying execution deductions onto each singular gymnastics skill and dance element whilst judging routines based on the latest rules as stated in the COP 2017-2020 (FIG, 2016b). The construct validity of using domain-specific video clips to explore decision-making of officials were supported with findings from previous studies (Catteeuw et al., 2010; Larkin et al., 2017).

Immediately following completion of their engagement in the TA sessions, one-to-one interviews with participants to collect their perceptions of using TA method whilst judging. Face-to-face interviews provided the researcher more information with regard to

social cues (Dialsingh, 2008; Opdenakker, 2006), including body language and tone of verbalisation. In addition, the interviews were conducted immediately after the TA sessions was completed to increase reliability of reporting thoughts related to the TA session and in overcoming the limitations of verbal reports (Ericsson & Simon, 1993), such as memory decay (Bernard et al., 1984; Whitehead et al., 2015). Furthermore, interviews provided the interviewer the opportunity to explore responses more widely concerning subject related discussions that spontaneously introduced by the respondent (Dialsingh, 2008). Therefore, face-to-face interview without significant time delay between questions and answer allow respondents to react directly on questions raised by interviewer (Opdenakker, 2006) deemed appropriate to explore participant's unique experiences engaging in the TA sessions. A semi-structured interview guide informed by previous studies (Lee et al., 2019; Whitehead et al., 2018) was developed and used in the interviews. The use of semi-structure interview guide allow researcher to stay focused concerning the aims and objectives of the current study when asking participants questions to extract data for that purpose despite not to steer participants to obtain inductive data (Elo et al., 2014). A Sony® Dictaphone (model ICD-PX240) was used to record all verbalisation during the interviews. This is the first study to explore the perceptions of WAG judges using TA method to verbalise all deductions applied whilst judging competition routines across *all four* apparatus. Interview questions focussed primarily on participants' experiences in using TA method whilst judging. For example, "What were your thoughts of using TA whilst judging deductions?", "Was there any differences to TA deductions across different apparatus?", "How do you think if TA could have effect on judging ability?", "How do you think of the feasibility to include TA into current WAG judge education?", "What were the challenges you had face when using TA whilst judging?" were included in interviews. An Olympus® AS-2400 transcription kit was used to process verbatim transcription.

5.2.4 Data Analyses

Interviews with each participants last between 25 and 80 minutes. A total time of 14-hour interviews with 18 participants created 356-page font Arial size 12 with double line spacing text transcribed verbatim by the lead author. Both inductive and deductive approach of thematic analysis was applied in this study for the researcher (coder) to identify, analyse, and interpreting patterns of meaning as 'themes' within qualitative data from interviews (Clarke & Braun, 2017). The verbatim transcriptions were inputted into NVivo® 11 for content analysis then underwent familiarisation process through repeated active reading (Braun & Clarke, 2006, p. 12). to identify and interpret data relevant to the research questions to generate initial codes. A semantic approach (Braun & Clarke, 2006, p. 84) was conducted to deductively identifying pattern of codes from description across transcripts according to participants' perceptions in judging competition routines across all four WAG apparatus using TA method, followed by summarising significant patterns and meanings to generate themes and sub-themes. Inductive reasoning also employed by first author with extensive WAG judging knowledge to generate themes from raw data, besides applying "theoretically flexible approach" thematic analysis (Braun & Clarke, 2014, p. 1). Themes were reviewed, defined and named before reported in the result sections below. These were followed by member check by participants to reduce the potential of researcher bias and to increase the credibility of analyses (Birt, Scott, Cavers, Campbell, & Walter, 2016; Elo et al., 2014; Lee et al., 2019; Varpio, Ajjawi, Monrouxe, O'Brien, & Rees, 2017), however, no subsequent adjustments was made to the transcripts. A critical friend approach was applied to offer critical feedback to encourage reflexivity knowledge on deduction keywords interpretation (Smith & McGannon, 2017) and checking for the representativeness as a whole to increase credibility of analyses (Elo et al., 2014), specifically on words commonly used within judging

field, such as “wobble” denoted “lack of balance” execution deduction as stated in the COP. Hence, shared meaning on the deduction keywords and common phrases were discussed across the supervision team that included of an ex-elite performer who possessed WAG coaching knowledge, a TA method experts enriched in experience of using TA method, and an active sport psychology researcher, whilst personal interpretation into judging were acknowledge to interpret the interview contents.

5.3 Results

Throughout the analyses of verbatim transcripts for interviews, five themes were generated, which were feelings of using TA method whilst judging WAG routines, perceptions of TA method viability within WAG judging, and TA method challenges within WAG judging, TA deductions across different WAG apparatus, and further consideration for learning resources adapting TA method. There were ten sub-themes generated within these themes. This study involved both expert-novice BG judges and all four WAG apparatus, that of BB, FX, VT, and UB compared to the previous study (Lee et al., 2019) that involving only a single expert group of judges with only BB, hence, these outcome some similarity and differences of themes and sub-themes. The only two common sub-themes were the perceptions of participants on the TA method viability within WAG judging to extend beyond written materials in judge education, besides challenges of using TA method within WAG judging to overcome verbal overshadowing in multitask judging. Meanwhile, this study provide more details in exploring the perceptions of expert-novice WAG judges in using TA method whilst judging video-based competition routines. This study had informed the feelings of using TA method including initial apprehension and confidence development, perceptions of TA method viability to increase awareness and reassurance within judge education, differences to TA deductions across different apparatus concerning pace, element

recognition, deduction verbalisation, and artistry deduction of respective apparatus, and consideration to adapt TA method in future judge education.

Table 5.2 Theme and sub-theme of perceptions by expert-novice WAG judges in using TA method

Theme	Sub-theme
Feelings of using TA method whilst judging WAG routines	Initial apprehension
	Developing confidence
Perceptions of TA method viability within WAG judging	Increasing awareness
	Reassurance
	Extending beyond written materials in judge education
Challenges of using TA method within WAG judging	Verbal overshadowing in multitask judging
TA deductions across different WAG apparatus	Pace
	Element recognition
	Deduction verbalisation
	Artistry Deduction
Further considerations for learning resources adapting TA method	

Theme 1: Feelings of using TA method whilst judging WAG routines

Sub-theme: Initial apprehension

Within this theme, it was evident that both novice and expert participants initially experienced feelings of apprehension, worry and nervousness when using TA method. For example, expert 6 stated “before [starting the data collection] I was a little bit apprehensive what you [myself] were going to say. Well, I'm not used to saying it [deductions] out loud.” whilst expert 4 reported “I was worried that I maybe won't recognise something [skill/element] I've been judging for ever”. Likewise, novice 1 reported that “before [the TA session], I was a little bit nervous because I wasn't sure how to do it [verbalising deductions], how to sort of say things out loud”. Novice judges also reported they were under pressure to

verbalise deductions and had concerns as to ‘being judged’ if they had applied incorrect deductions.

Sub-theme: Developing confidence

Both expert and novice judges reported on their confidence developed throughout the TA session. Expert 8 stated “I kind of got into the swing of it [verbalise deductions whilst judging], and so by the end I felt quite confident in just sort of talking through the routines and deductions”, whilst novice 2 said “as I got more used to it throughout each piece [WAG apparatus], it [verbalising deductions whilst judging] got easier”.

Theme 2: Perceptions of TA method viability within WAG judging

Throughout the interviews, both experts and novice expressed a range of positive views towards the use of TA method in WAG judging. Sub-themes generated within this theme include increasing awareness, reassurance, extending beyond written materials in judge education.

Sub-theme: Increasing awareness

Both expert and novice judges acknowledged their awareness had increased when using TA method whilst judging leading to more accurate decision-makings application in execution deductions. Expert 1 revealed that experts were unsure how often they did applied a specific execution decision across routines before using TA method, whilst expert 2 said “I think it [TA method] actually does help people to think more about why you're taking deductions and for what, so you can't just write something down without explaining why you've taken it”. Novice 6 stated that using TA method had increased her awareness before applying deduction onto any skill/element, whereby “you're sort of clarifying and explaining things [deductions] more in your own mind as you're going through [verbalisation] what

you're saying". This was further supported by novice 3, who said "You know what you're taking it [deductions] for, but you just take it, or you don't really think about it, whereas like now [using TA method], you've got to really think about exactly what you're taking it for".

Sub-theme: Reassurance

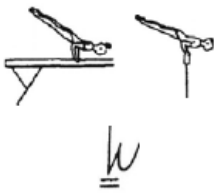
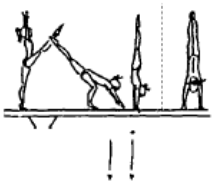
TA method requires judges to verbalise execution deductions applied onto each gymnastics skill and dance element concurrently whilst judging video-based competition routines. Artistry deductions applied at the end of BB and FX routines were immediate retrospectively verbalised by the end of routine performance. Participants reported TA method would allow judge course facilitators to access thought of judge candidates during an accreditation course, meanwhile novice judges would be able to reassure their decision-makings by listening to experts' verbalisations (novice 4) on deductions applied. Expert 3 further support reassurance from TA method, whereby verbalisations by experts in applying execution deductions "would help newer judges who aren't so used to analysing skills, so they kind of know what they're looking for, so it's kind of instinctive".

Sub-theme: Extending beyond written materials in judge education

Both expert and novice judge recommended TA method as a means by which to complement current judge education, whereby pictures and text descriptions as stated in the COP are deemed different from actual live presentations by gymnasts (expert 7). Therefore, it would be useful for an expert to verbalise and/or explain the deductions whilst judging looking at speed performance of a skill/element (novice 7). Hold-elements, for example Planche, clear pike support, handstand (see *Table 5.3Table 5.*), are credited if the element was held for 2-seconds. The '2-seconds' counts prompts varied across expert and novice judges, such as "one, two, three", "one and two and", "one gymnastics two gymnastics". In addition, some judges who were also a coach, reported they were not counting the '2-second'

in any of aforementioned ways as an actual timeframe period instead they credited the hold-element by looking at gymnasts showing body control during execution. However, these differences in interpreting the '2-second rule' were unwritten on the current COP.

Table 5.3 Description for hold-elements, adopted from FIG (2016b, p. 113, 131)

Description Symbol	*Planche with support on one or both bent arms (2 sec.), also legs in cross split position* 	*Clear pike support (2 sec.)* 	*Kick to side or cross hstd (2sec), lower to end position touching beam* 
Name	Planche	Clear Pike Support	Handstand

Further, participants both with and without gymnastics background perceived ex-gymnasts or coaches were able to judge 'better'. This was supported by expert 3 who stated that "coaching knowledge helps, because you know what the skills are, and you know what the skills are meant to look like, and you can recognise the skills easier, so I guess that helps if you've got prior knowledge of skills going into judging". Furthermore, novice 7 who also a coach revealed "when I'm coaching, you [I] can explain to the gymnasts and say, "OK, well, that's a .1 [deduction] right then, that's a .3 [deduction] if you take that [big] step", revealing coaches were using TA method when coaching which facilitated their deduction verbalisation whilst judging in the TA sessions. This is corroborate to previous findings (Campo & Gracia, 2017) revealed that judge and coach possessed superior performance than a gymnast to spot execution and demonstrated a stronger relationship between gaze and verbal reporting in judging same WAG skills.

Theme 3: Challenges of using TA method within WAG judging

Despite judges reporting some positives for using TA method whilst judging WAG routines, challenges were reported by both expert and novice judges, which made up sub-themes of verbal overshadowing in multitask judging.

Sub-theme: Verbal overshadowing in multitask judging

Both expert and novice judges reported that "writing and talking [verbalising deductions] and looking [watching routine] at the same time, then thinking [deduction application], was challenging" (expert 5). Novice 4 also mentioned that verbalising deductions was harder on faster-pace apparatus, whereby analysing the accuracy of a performed skill/element whilst making decisions to apply execution deductions based on deviations from standards as stated in the COP, was difficult with time constraint. Novice judges also reported that they were unable to perform judging task without referring to the COP when making decisions to apply deductions, besides admitted they were unable to look into a skill 'as a whole' and instead focused only on certain body part, such as knee, feet, or arm. Expert 10 reported instant decisions made for very quick skill/element was hard: "for example a twisting somersault, you know it doesn't quite look right, but to actually verbalise that [specific deduction] is quite hard. If they've done a straddle lever to handstand, so if they do a slower element, it's easier to verbalise it [execution deduction]". Challenges also faced by judges who multitask two roles in a panel to judge both difficulty and execution at the same time, whereby novice 8 stated "I can't think of those three things fast enough." She explained the difficulties to recognise and write down movement as symbols on judge notation sheet in addition to making decisions to award the element/skill difficulty, despite decision-makings to apply accurate execution deductions.

Verbal overshadowing challenged verbalisation speed and terminology recall during the use of TA method were mentioned by both expert and novice judges. Expert 5 said "I felt rushed [to verbalise deductions whilst judging], because obviously I think faster than I talk… I think in my own language, and just finding words was challenging. So once I found a word what I want to just say, the routine obviously went on, and I didn't say everything that happened in between", whilst novice 7 reported a similar problem where she missed out judging some elements when using TA method whilst evaluating routines. Expert 8 stated that she felt the use of TA method had 'slowed down' her judging and writing speed as her focus been distracted to verbalisation at the same time. Novice 1, who classified herself as a 'fairly newish judge' revealed lacking of familiarisations on terminologies stated in the COP increased difficulty in using TA method to verbalise deductions whilst judging. Novice 2 disclosed that she undertook some reflective thinking on her judging by thinking back the routine therefore verbalisation that requires her to think instant decision aloud was difficult for her.

Theme 4: TA deductions across different WAG apparatus

Looking at different skill/element pace across all four apparatus within WAG, Table 5.4 presented raw data extracts showing similarities and differences between expert-novice judges to verbalise deductions whilst judging video-based competition routines across all four WAG apparatus of that BB, FX, VT, and UB. Sub-themes generated includes element/skill pace, element/skill recognition, and deduction verbalisation to compare differences in using TA method across all four apparatus, whilst BB and FX had an extra comparison in artistry deduction verbalisation.

Table 5.4 Perceptions of expert-novice WAG judges in using TA method to verbalise deductions across apparatus of BB, FX, VT, and UB

Primary Theme	Description	Apparatus	Raw Data Extracts (Expert)	Raw Data Extracts (Novice)
Pace	Reference to the availability to TA deductions on each element according to the element pace on specific apparatus	BB	"It [elements] happens slowly, so you've got time [to TA deductions], and your brain's got time to think [the deductions]." (E6)	"…I think it's [the element pace] just a bit slower, so you get more time [to verbalise the deductions]." (N8)
		FX	"…especially in the tumbles [acro-lines], where it's happening so quickly [limited time to TA deductions]" (E1)	"…because they [FX routines] have lots more dance in between, I tend to think that it's a little bit slower [to allow TA]." (N4)
		VT	"Oh, vault's too quick, it's too quick [to TA deductions]" (E4)	"It's so quick. It's just so quick. I think vault is just so quick [that could not TA deductions], and particularly for a higher [international competition] level like that." (N1)
			"…I normally write the landing deductions as they landed, and then I think I probably then think back and I use the vault crib sheet a lot, so I can replay the different sections [but not TA deductions whilst watching]" (E1)	"You have to rely on your memory of how that vault went, so it's almost like when you watch the vault, you have to put it in short-term memory [but could not TA deductions whilst watching]" (N7)
		UB	"I think with bars it's [elements are] fast, so if you're trying to get your symbols and deductions down, you've just got to be [TA deductions] quick, you've just got to work so quickly, and I just literally try and like process the information as fast as I can." (E8)	"it's harder to judge [TA deductions] it because it's [element is] so fast, and it's hard to exactly know what [deductions] to take off, like you've got to just know it instantly, otherwise you'll miss the next part of the routine..." (N3)
			"…sometimes you even have to go back [reply in mind], and if you miss writing deductions for a skill, at the end of the routine you have to go back and think through what they were doing again." (E2)	"…it might be that at the end I'm still kind of going back over it and thinking [reply in mind]" (N1)
Element Recognition	Reference to the judge's capability to recognise the element in respective apparatus	BB	"…I just see the moves more often [so it's easier to recognise the BB elements]."(E3)	"…beam [elements were easier to recognise], because it's similar to floor [elements], but it's [BB elements were] slower …"(N6)
		FX	"Floor is ok when the jumps are clean, and the leaps, but when they're not good [deviated from required technical aspects], that becomes quite difficult just to judge the leaps." (E9)	"…because you've got all the dance [elements] in between. So the floor, even though you were verbalising as well, I think the floor was easiest for me to judge then."(N4)
		VT	"…I know it's broken down into phases, [but] it's one skill."(E3)	"I think vault happens very, very quickly, and there's a lot of stages to [apply deductions for the entire] vault that you have to deduct on" (N6)
		UB	"I think it's easier [for element recognition] because you are only really looking quite small space, from the concentration it's easier and I think the [deduction] decisions based on angles are easier than I found on the twisting decisions on floor and beams with the leaps [which] I think quite difficult."(E1)	"…as you're watching it, you can watch every move and almost give a comment on every move as it's happening."(N6)

Primary Theme	Description	Apparatus	Raw Data Extracts (Expert)	Raw Data Extracts (Novice)
Deductions Verbalisation	Reference to the deductions verbalisation in the TA session	BB	"Beam, in general, I think, is quite straightforward [to TA deductions]." (E1)	"…because they're [skills] nice and slow, so it's easy [to TA deductions]."(N2)
		FX	"I find it easier to analyse the skills. I look at the skills and know where the deductions are."(E3)	"…it is harder [to TA deductions] because it's [elements] quicker and there's more [deductions] in it, as regards to the tumbles and things, so there's much more [deductions] to look for [within short time]" (N6)
		VT	"Vault is possibly the worst judged piece [to TA the deductions]." (E9)	"I think the hardest to judge [TA deductions], I think is vault, because it goes so quick, and it's like trying to take a photograph in your mind of what it looks like..."(N4)
		UB	"…quite a few deductions [to be TA] in a short space of time, trying to sort of calculate it in my head [in normal judging protocol]." (E10)	"I don't like judging [TA deductions on] bars too much. I find it a bit difficult."(N5)
Artistry	Reference to the judging criteria on artistry performance specifically applicable on BB and FX	BB	"…the beams [elements] were far more connected..."(E9)	"Because it's almost like you're focused on the actual skills whilst you're marking, but you can't help but see the artistry, because it's throughout the whole routine. (N1)
		FX	"I think sometimes judges will do their personal opinion, not only if they've interpreted those facts, and it is hard because that is your opinion, because it's not black and white."(E10)	"…that's the thing I'm going to do at the end, and if they hadn't looked like they performed it as well, they'd probably get a lower artistry mark because I think that the other gymnast was a bit better [artistry]."(N1)

Sub-theme: Pace

In terms of pace in relation to specific apparatus for deduction verbalisation, participants stated BB was the easiest apparatus to verbalise deductions because elements performed were seen as more 'straightforward' and at a slower pace, whereby routines with well-structured compositions and intermittent dance elements had also provided judges more time for decision-making. This had included being able to have time to 'replay' previous elements in their mind, whilst providing more time for writing down the symbols and execution deductions at the same time of verbalising execution deductions. Participants perceived verbalising deductions easiness were attributed to common gymnastics skills and dance elements of BB and FX, where both apparatus sharing some same coded elements and technical aspects for execution deduction application, even elements/skills in FX were faster in pace compared to BB. Both expert and novice judges reported UB was a fast-paced apparatus, therefore, it was challenging to verbalise all deductions for each element as they had to process information 'instantly' whilst judging otherwise they may have missed out the entire routine. Participants had stated one-skill element VT performed within an extremely short timeframe was rated the most challenging apparatus for execution deductions decision-making and required immediate retrospective evaluation. Likewise all participants reported that STM was required to recall the entire vault immediately after a gymnast landed for the purposes of applying accurate execution deductions across different vaulting phases. Novice 7 stated that "it's over so quickly. You have to rely on your memory of how that vault went, so it's almost like when you watch the vault, you have to put it in STM, but quite often, what you're doing is, you're thinking back into the shape that you saw, the actions that you saw, as that gymnast was going over the [vault] table." Some participants revealed they applied instinctive deductions whilst judging VT, whereby expert 1 stated "I think when you've been

in the sport for long enough, if you like come out with a [deduction score] number, it tends to be right, just because you get [used to] it". Expert 1 further added that "it made me realise there's a lot [deductions] that I do [applied] without thinking, that is probably an experience thing of having seen so many of them [skill/element execution], [that] I know what I want it to look like, and how far away [deviation] from the normal [standard] it is." Expert 8 elaborated the use of instinctive deductions during judging "…when you've been in the sport for long enough, if you like come out with a number [deduction score], it tends to be right". There were novice judges who also supported the importance of experience in judging WAG. Novice 4 said "the more you judge, the more you can see what's wrong [execution error] with it [element/skill]", whilst novice 6 stated "the longer you've done it, the easier it gets [into the deduction scores]" further explained that someone with experience and seeing gymnastics more often tends to judge a skill/element according to the COP easier. Collectively, this indicated that judges probably make use of cognitive strategies in detecting execution errors during the judging process based on experiences.

Sub-theme: Element recognition

Element recognition was reportedly easier on BB and FX for both expert and novice judges. However, participants stated they felt challenging in using TA method to verbalise deductions whilst judging FX routines. The fast moving series of continuous tumbling gymnastics elements performed in acro-lines, as well as dance series connecting turns, hops, jumps, and leaps to achieve higher connection bonus were concerned by judges, whereby expert 4 reported "Did it go round once or did it go round one and a half? I have to really think quite hard about how many times I'd seen the back and front [face], and where did the legs split". Nevertheless, expert 1 reported the changes onto the latest COP required judges to

only award and recognise jumps and leaps which were performed under 30 degrees of completion may offer some explanation.

Sub-theme: Deduction verbalisation

Likewise with element recognition, concurrent deduction verbalisation were reported easier for BB and FX as they were performed in slower pace that providing more time for decision-making. For UB, participants mentioned it was easy to apply execution deductions due to the 'clear-cut' decision-making. Expert 1 stated that "I think it's easier because you are only really looking at quite a small [visual] space, from the concentration it's easier and I think the decisions based on angles are easier than I found on the twisting decisions on floor and beams with the leaps [that] I think quite difficult".

Sub-theme: Artistry deductions

Artistry deductions that specifically allocated for BB and FX were perceived difficult for decision-making due to the subjective interpretation perhaps according to personal preference. Expert 8 when asked how she applied artistry deductions stated "when you've done it regularly, get in the habit of kind of just coming up with an overall figure [deduction score] in your head, and sometimes it means I can miss things [deductions]". Moreover, both expert and novice judges revealed they usually make decisions to apply artistry deductions typically within 30 seconds after the routine ended to add up into the total deduction scores. Therefore, they replayed the routines by immediate retrospectively for an overall 'feeling' towards the artistry whilst some judges applied different reasoning, such as comparing the artistry performance of a gymnast with previous gymnasts (expert 4).

Theme 5: Further considerations for learning resources adapting TA method

Some novice judges mentioned resources that had facilitated them in developing judging expertise to pass the judge accreditation examinations, furthermore to progress from club judge to regional judge. In BG, a club judge is expected to recognise the FIG coded common elements of A- to C-value difficulty to execute the role of an E-judge fairly as a member of a judging panel, which focus is only that of execution deduction applications. They acknowledged the need for judging practice before attending a WAG judge course to pass the accreditation examination and whilst online resources were previously accessible in the BG GymNet®, an online software, was reported helpful, however currently no longer available. Despite passing the judge accreditation examination, participants stated the utility of online resources to refresh their judging knowledge and skills after completing a regional/club judge course, which conducted within two weeks over two weekends that deemed intensive. Furthermore, video clips used in the club/regional judge's course were reported as being of poor quality and not 'up-to-date' concerning routine constructions were not compliance with the current COP. They further mentioned the use of TA method in the TA sessions were similar to the teaching method in the judge course, whereby a course facilitator played a video clip asking candidates to make decisions on execution deductions before revealing 'standard answer' examined by the experts. Novice judges then compared their decisions to those of that from the experts to reflect for accurate deduction applications. Hence, online resources adapting TA method was suggested by several participants, as "it [TA method] helps the novice judges during their learning, in clarifying all the deductions, like how much and the weightage for each deduction" (novice 1) demonstrated the viability of TA method to collect in-event cognitive thoughts of WAG judges whilst judging.

5.4 Discussions

This is a novel study to explore perceptions of expert and novice judges in using TA method to verbalise execution deductions across all apparatus in WAG, that of BB, FX, VT, and UB when judging video-based competition routines, immediately after the TA sessions. Both expert and novice judges reported initial apprehension to engage TA method in verbalising deductions whilst judging, whereby the feelings of nervous and anxious arise in learning and applying a new skill of TA method. Previous literature (Eccles & Arsal, 2017; Stephenson, Cronin, & Whitehead, 2019; Whitehead et al., 2018) has recommended participants go through TA training to engage the method fully with domain-specific tasks. Novice judges, especially parent judges disclosed they were nervous and worried about verbalising deductions, with concern being judge on the accuracy of deductions applied. Hence, the feelings of apprehension may have decreased their attention and perceptual-cognitive skills (Dosseville & Laborde, 2015) and resulted in fewer deductions applied when compared to 'usual judging', which is to be further investigated in future research. However, participants reported that through the research participant experience they had develop confidence in using TA method across the TA session demonstrated skill acquisition in judges through adaptive learning (Williams & Ericsson, 2005). Indeed participants expressed they "got into the swing of it" towards the end of the TA session. Nevertheless, participants reported their awareness increased when they requested to verbalise specific execution deductions applied onto each singular skill or element. As such this avoids the deductions applied through schemata (Norman & Shallice, 1986) to ensure score accuracy. Expert 3 mentioned "sort of have a look at the skills as a whole and kind of collectively" in making decisions when judging, further added that "on vault, I just get there [deduction scores] somehow, but I'm not entirely sure how [what specific deductions applied] sometimes"

indicating some decisions were made instantly without concerning the exact execution faults according to the COP. Further, expert 1 stated that “I hadn't realised before I did it [TA sessions], how often a decision is made and I'm not really quite sure what I made it for… when it's [skill/element/movement] happening very quickly, something [decisions] that's very subconscious about”. These statements demonstrating TA verbalisation whilst judging had increased their awareness in making decisions to apply accurate deductions.

In addition, verbal overshadowing (Chin & Schooler, 2008; Ericsson, 2003; Meissner & Brigham, 2001; Schooler, 2011) reported in this study that affecting the performance of participants to judge in naturalistic environment of that actual competitions, whereby verbalisation had slowed down the judging processes by adding another speaking element into multitask judging (Ste-Marie, 1999) that corroborate findings from Study One. TA methods were expected to provide richer data on in-event cognitive processes according to sequences of skills and elements to circumvent limitations of retrospective verbal reports, whereby retrospective verbal reports had low congruence in recalling memory concerning memory decay (Ericsson & Kintsch, 1995; Ericsson & Simon, 1993; Nicholls & Polman, 2008; Whitehead et al., 2015). Participants within this study reported that verbalising deductions had slowed down their speeds in making decisions whilst completing other writing requirements. Participants stated that their decision-makings were faster in silent conditions, especially when judging fast-moving elements, such as continuous tumbling gymnastics elements in acro-line on FX, non-stop gymnastics elements continuing from low bar and high bars in UB, and high-speed single elements on VT. This corroborates with previous findings (Chin & Schooler, 2008; Ericsson, 2003; Lee et al., 2019; Meissner & Brigham, 2001; Schooler, 2011) who noted that by adding another speaking element into high cognitive demanding task may lower performance or take a longer time to complete the

actual task (Arsal et al., 2016; Whitehead et al., 2015). Nevertheless, these findings were contradicted to earlier studies (Ericsson & Simon, 1993; Fox et al., 2011; Whitehead et al., 2015), who found *Level 1* and *Level 2* TA had not alter performance of an individual completed the same task silently. Moreover, participants stated they were searching for 'textbook language' instead of verbalising common names for some elements and deductions, which had delayed the judging process. For example, "feet", "foot", "flat feet" been widely used to denote "feet not pointed/relaxed", "wobble" was mentioned widely to denote "lack of balance", whilst "chest down" denote "body posture fault" in the COP (FIG, 2016b, p. 29). The apparent issue in using TA method was a perceived need to suppress the use of common phrases representing execution deduction across all apparatus reportedly by both novice and expert judges. As a result, verbal overshadowing challenged the use of TA method despite the advantages reported. However, judging WAG routines are under time constraint and requires multitasking (Ste-Marie, 2000) and memory (Ste-Marie, 1999) to simultaneously watching routine whilst recording symbols notating skills and elements and at the same time evaluate each skills and elements comparing to standards as stated in the COP. Therefore, task-specific TA trainings were suggested for WAG judges to familiarise the TA method before application into cognitively demanding judging task for future study. Warm-up exercises to use TA method in general, followed by TA exercises starting with easy-to-verbalise tasks and reminders from researcher to participants during data collection in focusing on the domain-specific task whilst giving concurrent verbal report were important to reduce the impact of verbal overshadowing in influencing natural thought process and performance.

Table 5.5 were suggestions on types of verbal reports to TA deductions concerning there were different paces and deduction lists for respective apparatus.

Table 5.5 Types of verbal reports in judging WAG apparatus

Apparatus	General Execution Faults & Specific Apparatus Deductions	Artistry Deductions
BB	Concurrent TA	Immediate retrospective TA
FX	Concurrent TA	Immediate retrospective TA
VT	Immediate retrospective TA	-
UB	Concurrent TA	-

BB was reported the easiest WAG apparatus to adapt TA method for verbalising deductions whilst judging based on it being the slowest paced apparatus in WAG. Therefore, there were more time for judges to process information for decision-making and verbalising deductions during the judging process. Deductions verbalised by judges were inclusive of general execution faults, specific apparatus deductions, and artistry deductions. By using concurrent TA method to verbalise deductions applied whilst evaluating each skill and element, sequences of deductions verbalised following skill/element therefore self-explained the execution scores applied by judges. This answers the gap of reliability and objectivity of artistry deductions raised by Pajek et al. (2014) and bias investigation raised by Ste-Marie et al. (2001) by understanding the decisions made underpinning the judging process. By verbalising, each specific deductions applied on each element/skill concurrently using TA method whilst judging a routine were explained the reasons and weight of every execution deductions applied on each movements in a routine chronologically following the sequence of each skill/element. Therefore, the objectivity in applying execution and artistry deductions underpinning the judging processes increased thus reduced the bias in judging. A BB routine takes up to 90 seconds with maximum of eight skills and elements counted (FIG, 2016b) and as such provides more time for judges to make decision compared to that of FX, VT, and UB. FX and UB were reportedly easy to use TA method for deduction verbalisation after BB, with similar skills and elements share in both apparatus (Ste-Marie, 1999). This corroborates with

previous studies stated TA method is a viable tool to collect in-event cognitive processes of athletes in self-pace sports, that include of golf (Arsal et al., 2016; Calmeiro & Tenenbaum, 2011; Kaiseler et al., 2013; Nicholls & Polman, 2008; Whitehead et al., 2015; Whitehead, Taylor, et al., 2016), snooker (Welsh et al., 2018), endurance cycling (Whitehead et al., 2017; Whitehead et al., 2018), and long-distance running (Samson et al., 2015). Furthermore, previous studies exploring in-event cognitive processes underpinning decision-makings of athletes found the performance of task was not altered by using *Level 1* and *Level 2* TA (Ericsson & Delaney, 1999; Fox et al., 2011) in the self-paced sports, however, judging WAG routines required effective and objective judgement to be complete within restricted time. Therefore, challenges arise when adapting TA method in faster-paced WAG apparatus of that FX and UB, where novice judges expressed that they frequently missed out skills and elements during the multitask judging. Expert judges however stated they had speeded up the information-processing when requested to adapt TA method in judging to verbalise deductions whilst evaluating video-based routines demonstrating expert judges possessed superior declarative knowledge (Ste-Marie, 1999, 2000) to retrieve COP information with semantic memory (Baddeley, 2012). The acro-line on FX consisting of connecting tumbling elements were reportedly difficult to TA deductions for novice judges, however, dance elements and artistry performance embedded in the routine were perceived as 'slowing down' the pace and therefore provide more time for them to make decisions to apply deductions. Novice judges revealed judging deductions in UB routines was difficult, due to the limited time allocated for decision-making on connecting high-speed elements, particularly those between bars elements causing them to miss opportunities to observe and verbalise execution faults on the upcoming elements when using TA method. However, expert judges mentioned they overcame these challenges by voluntarily increasing practice before attending a competition that enabled them to recall information instantly, such as visiting practice venue

for watching trainings and watch competition video routines. Expert judges seemingly adapted STM and LT-WM through repeated exposure to the environment and the task to enhance their perceptual-cognitive expertise (Williams et al., 2017) in circumvent the time constraint. Specifically, they developed retrieval structure that encode and retrieve domain-specific knowledge more efficiently in LT-WM (Ericsson & Kintsch, 1995). Hence, they were able to write down symbols and apply execution deductions onto each element immediately when performed by gymnasts with superior procedural knowledge (Ste-Marie, 1999). Further, expert judges also reported that execution deductions for uneven bars elements were easier as most execution deductions involved angles of completion, against decision-making on the fastest-moving apparatus, VT. These practices demonstrated some differences between expert and novice judges and that are explained by expert performance approach (Williams et al., 2017), whereby experts enable to adaptive learning utilising expertise develop with experience accumulated over time (Ericsson, 2017; Ericsson & Simon, 1993; Ste-Marie, 1999). Frequent COP information storing and retrieving, i.e. writing down symbols notating skills and elements enhanced STM and LT-WM (Ericsson & Kintsch, 1995) of expert judges to retrieve COP information and thus reduce attentional resources when make decisions on execution more efficiently compared to novice judges as stated in the superior judging performance (Ste-Marie, 2000) and working memory (Baddeley, 2012).

VT routines usually complete within five seconds (Pajek et al., 2013) and required judges to evaluate the performance across phases of running, first flight, repulsion, second flight, and landing. Participants reported there were difficulty to verbalise deductions concurrently for VT with time constraint, therefore, immediate retrospective TA method (Eccles, 2012) is suggested to collect decision-makings on VT routines immediately after the entire vault exercise been performed. Surprisingly, some participants reported that they

applied deductions based on their ‘instinct’ without exactly specify deductions from the lists as stated in the COP (See *5.3 Theme 4: sub-theme Pace for quotations supported this statement*). The heuristic decisions were made to simplify the judging process due to limitations of time, information, and computational ability which might cause biases (Raab et al., 2019). This suggests future judge education may benefit from a focus on perceptual-cognitive skills development among judges to process incomplete, intentionally deceptive and fast-paced information with time constraint (MacMahon & Mildenhall, 2012) in judging routines on fast-pace apparatus more objectively and accurately. Furthermore, trained judges with developed highly specialised, sophisticated knowledge structures enable them to identify fast-pace skills and elements through pattern recognition and recall through extensive deliberate practice (Chase & Simon, 1973). These structures development enable experts to selectively attend to the most relevant features of the movement in contrast to those novice judges who are less able to distinguish between relevant and irrelevant cues, which leads to too much information or the wrong information coming in that leading to impaired decision-making. Therefore, extensive training to enhance anticipation of judges to be concern in future judge education to enhance WAG judging (Ste-Marie, 2003).

Expert judges reported they were able to anticipate upcoming skills and elements in series of connections that performed frequently across gymnasts in competitions that enable them to make decisions in applying execution deductions to speed up the evaluation process corroborate to findings from Ste-Marie (2003). Expert judges accumulated experience over time to enhance their perceptual-cognitive skill in recognising frequent chunk of patterns (Chase & Simon, 1973) and in addition to anticipate upcoming skills and elements to overcome time constraint in the judging process (Loffing & al-Bruland, 2017; Williams et al., 2003). Novice judges also perceived those who were ex-gymnasts (Gentsch et al., 2016;

Heinen et al., 2012; Pizzera & Raab, 2012) and coaches (Campo & Gracia, 2017) were advantaged with motor and visual experience to judge more efficiently and effectively, whereby they were better in perceiving techniques to execute skills and elements. This corroborates with findings (Ste-Marie, 2000, 2003) who noted expert judges were better in information processing to outcome more accurate decisions, whereby they possessed superior declarative knowledge retrieved from semantic memory. This perhaps suggests that extra effort needed by novice judges, who without gymnastics background to increase their learning and practice opportunities to become an expert judge by enhancing their LT-WM in recognising skill/element and retrieving COP information to circumvent time constraint during actual judging.

Participants supported the suggestion of adapting TA method into future judge education. Specifically this was to extend beyond of that 'paper and pen' format to that of an online learning module, which provided self-paced learning opportunity, enabled them to prepare for judge accreditation examination in addition to refresh their knowledge prior to judge for a competition. Expert 4 expressed her concern that "they [novice judges] can't see the deductions, so they can't see the bent leg because it was a bit quick, or maybe they don't recognise the skill that they're looking at, especially if we've got parents that have maybe come in, not with coaching experience, so they're seeing a routine from a parent's perspective". Therefore, TA method adaptation in judge education provides learning opportunity through expert's verbalisation. Novice judges were able to learn the base line in applying execution and artistry deductions through experts' concurrent verbalisation whilst watching the video routines. Further, expert judges were able to pause, slow-down, or replay the video routine to justify the execution deduction weight applied onto each single skill and element, in addition to the artistry deductions applied for the entire routine of BB and FX

following the TA deductions. During the judging education incorporating the use of video routines, tutors would be able to prompt novice judges to some blind spots that require more a cautious, analytical and comparative approach to elements through verbalising deductions concurrently to further explain the deduction application for knowledge transfer. Expert 6 stated “when we did our national judging course, they [facilitators] were saying to us, particularly on things like bars, "That swing didn't go up [to specific degree], and it didn't go over the bar, so that's .5 [deduction]", and thinking out loud and watching it like that, you go, "Oh, right", and then when you're in the competition, you see a swing like that, and you go, "Oh, the swing didn't go over the bar. That's .5 [deduction]", because you can remember by people talking about the deductions that they're taking”. Novice judges stated they could see how they would gain confidence by affirming their decisions made during the judge course when compared their decisions made to that of the facilitator, who typically expert judges appointed by the BG, whereby novice 3 stated “I think that'd [deduction verbalisation] be really useful, because then you can actually see what they're [experts] taking it [deduction] for, rather than just being like, where did that come from?” Furthermore, novice 6 revealed “I was always interested to see what they [experts] were getting [deduction score], what they were deducting on, because it gives you a bit of an idea whether you're in the right place or not”. This supported the viability of TA method in collecting decision-making underpinning judging process thus provide an opportunity for novices to feedback and correct inaccurate decisions if they were able to access thoughts and decisions by listening to expert’s verbalisation on deductions applications during the judging trials. When asked for specific inclusions, novice judges recommended to include guidance on counting ‘2-seconds’ for hold-elements in addition to guidance for execution deductions applications on those regional moves. Regional moves were specific regional competition permitted allowed elements, which were not coded in the FIG COP, such as *teddy bear roll, stretch jump, tuck jump,*

forward roll, *backward roll*. Specifically, participants reported lacked clarity for execution deduction application when evaluating regional moves, such as *caterpillar* (see *Figure 5.1*) designed specifically for local competitions, which was only available on BB for the BG National Development Plan Club Grade Six (British Gymnastics, 2017c, p. 36).

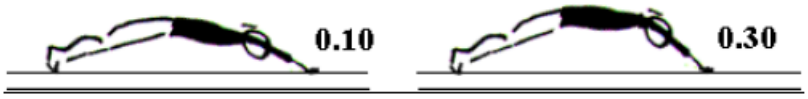


Figure 5.1 *Caterpillar* penalties, adapted from British Gymnastics (2017c, p. 36)

Video practice used in pre-competition judging has been found to refresh and synchronise decisions among international judges (Sacchi, 2018a) and indeed could be adapted into local competitions to ensure score objectivity and reduce differences between expert and novice when judging the same routine. Besides playing video-based competition routines in the judges meeting before attending actual judging task for decision synchronisation, the ecological validity to reproduce a routine under controlled and realistic conditions that mimic actual competitions (McRobert et al., 2011; Williams & Ford, 2013) enable judges to make decisions from the same angle of view (Williams et al., 2011). Outcomes from this study suggest that future WAG judge education would benefit from consideration to adapt TA method in enhancing the objectivity and accuracy of execution deductions applied. Further to the resource would simulate judging experience of novice judges by providing practices before attending an actual competition judging task through watching videos mimic competition situations that displaying accurate deductions applied by expert judges. There were recommendations in steps for using TA method as a reflection/learning tool into WAG judge education as stated in the Table 5.6 below in addition to the current "paper and pen" format. The current judge education providing the COP information, that of basic knowledge for skill and element recognitions and execution

deduction applications. The rule-based knowledge possessed by judges before proceed to utilisation of the TA method for expertise training believed to reduce the information-processing demands in memory recall on the COP information (Ste-Marie, 1999, 2000) when judging, especially novice judges who stated that they required more time during judging to refer rules and the COP before decisions made for execution deductions. The process of TA method familiarisation at the beginning of TA training is important for judges to verbalise inner speech (*Level 2* TA) continuously whilst judging to think aloud the decisions made onto each singular movement. The TA training is suggested to commence with general tasks to verbalise thoughts by say out loud activities adapting instructions from original protocols, such as verbalising the next alphabet after 'A' and the number of dots appeared on a page (Ericsson & Simon, 1993; Lee et al., 2019; McRobert et al., 2017; McRobert et al., 2009). Next, TA training could proceed with simple to complex sport and domain-specific warm-up tasks using concurrent TA. For example, WAG judge warming up to TA deductions on BB, which is the slowest pace apparatus of BB to provide more time for cognitive processes verbalisations, from a singular BB skill or element, before TA deductions for a full routine. The TA training for novice judges also suggested to include only verbalisation on execution deductions, neglecting other judging tasks of confounding variables, such as writing down symbols representing skill and element performed or verbalising the skill and element for recognition purposes to reduce the distractions but focus on speaking only task. The TA familiarisation training beginning with slow-pace movement enable novices to TA concurrently, which require only STM, before proceed progress with more challenging task, such as TA deductions on faster pace apparatus of that FX, UB, and VT, which require LT-WM to retrieve the COP information in judging by using immediate delayed TA. The training tasks that similar to the target task that are not too different from the actual competitions (Van Someren et al., 1994) enable the expertise training. The progressing TA

exercises that frequently retrieve the role-specific information of judges thus strengthening domain-specific knowledge of recalling LTM on the COP information when completing the judging tasks (Baddeley, 2012) hence enable them to store the COP information as LT-WM to increase the retrieval speed during judging (Ericsson & Kintsch, 1995). Besides, The TA training suggested to be guide by a TA expert who possessed domain-specific knowledge for the purpose to correct any inaccurate decisions made during judging as well as in reminding novice judges to provide TA verbal reports, which to verbalise execution deductions whilst watching video clips. TA method proposed to complement the current WAG judge education with the purpose to utilise the TA method as reflection tool for self-learning to enhance score objectivity and accuracy, whereby superior judging performance shall increase throughout the TA training on role and domain-specific tasks.

Table 5.6 Recommendation steps for using TA method as a reflection/learning tool into WAG judge education

Step	Details
1	Familiarisation of TA method: TA training with general task
2	Familiarisation of TA method: TA training with sport-specific/domain-specific task
3	Guidance from TA expert
4	Using TA method as a reflection/learning tool

5.5 Conclusion

The aesthetic sport of WAG requires judges to responsibly apply accurate and consistent scores across gymnasts in a competition according to current COP (FIG, 2016b) or respective rules of a competition to ensure objectivity of a competition. Findings of this study have extended that of previous studies to explore perceptions of WAG judges in using TA method whilst judging fixed-sequence video routines, which included both expert and novice judges as participants in all four apparatus of WAG, which were BB, FX, UB, and UB.

Additionally, understanding perceptions of expert and novice judges to verbalise deductions according to the pace of respective WAG apparatus also provided insights that the use of TA method was the easiest on BB, followed by FX and UB. This study suggested concurrent TA method a viable tool to collect in-event cognitive processes and decision-makings within slower-paced BB, FX, and UB, whilst immediate retrospective TA method used to collect decisions made by judges in the fast-pace VT. The verbalisation process increased awareness of judging during execution deductions application to avoid schemata. Nevertheless, appropriate training to use TA method whilst judging is required to reduce effect on information processing during judging. Therefore, educational resources might consider adapting TA method into future judge education for training purposes.

CHAPTER 6

Synthesis of Findings

6.1 Introduction

This thesis examined the suitability of using TA method to explore the decision-making process underpinning WAG judging and to investigate expert-novice differences within this domain. Findings from Study 2 (Chapter 4) found that expert judges were able to identify more execution errors across all four WAG apparatus and thus more deductions were applied compared to novice judges. Further, expert judges were able to verbalise more specific execution deductions applied onto each skill and element than novice judges whilst judging fixed-sequence competition video routines, demonstrating expertise perceptual-cognitive skills and ability of retrieving domain specific knowledge from the COP in managing multitask judging. Expert judges verbalised more general execution deductions, specific apparatus deductions, and artistry deductions than novice judges across all WAG apparatus, except for the general execution faults of UB, whereby the TA deduction counts by novices were reported more than experts. Outcomes from Study 3 (Chapter 5) revealed that expert judges were able to anticipate upcoming skills and elements, especially when judging fast-moving tumbling series in BB and FX, as well as fast pace UB routines that demonstrate the perceptual-cognitive expertise in WAG judging. On the other hand, novice judges reported they faced challenges when judging fast pace apparatus, especially in judging UB routines that require them to process information with time constrain in addition to apply execution deductions objectively. Moreover, expert judges reported undertaking judging practice to enhance retrieving domain-specific knowledge of COP, specifically when they were required to speed up the information processing – perhaps linked to the level of competition expected. Therefore, expert judges demonstrated their ability to manage

multitask judging that required them to write down symbols representing every skill and movements in a routine and at the same time making decisions to apply execution deductions whilst evaluating a routine, indicating the importance of superior procedural knowledge in WAG judging. Furthermore, some expert judges revealed they applied what they defined as "instinct" based on experience when evaluating a skill, element, and routine as a whole instead of referring to the deduction lists stated in the COP judging to apply every execution and artistry deductions. Thesis outcomes are intended to be of use to inform future judge education.

6.2 Review of Findings

Study One explored decision-making underpinning a single group participant involving ten Malaysian WAG judges when evaluating fixed-sequence BB routines using concurrent TA whilst also examining utilisation of TA as a technique to facilitate judge education with Malaysian WAG judges according to the COP 2012-2016 (FIG, 2012). General execution faults of *fall* and *insufficient height of elements* were mentioned by all participants across the study indicated the focal of participants in evaluating skills and elements. Further, the most frequently verbalised execution deductions were *lack of balance, bending arms and knees,* and *feet relax,* whilst most frequently verbalised artistry deductions were *confidence, rhythm and tempo,* and *personal style*. Follow-up interviews with participants exploring perceptions of using TA method revealed that TA method could inform coaching practice to understand decision-making of judges. In addition, results suggested that TA method was viable to collect decision-making research data within judging when applying execution and artistry deductions whilst evaluating a routine. Study conclusions stated that TA method could be used to explore the difference of decision-making between expert and novice judges across all four apparatus in WAG.

To develop findings from Study One, Study Two applied TA method to examine decision-making differences between ten expert and eight novice WAG judges accredited by British Gymnastics in evaluating fixed-sequence competition videos to resemble actual competition across all four apparatus that of BB, FX, VT, and UB, according to COP 2017-2020 (FIG, 2016b). Results showed how there were score variations informing decision-making differences between expert and novice judges. Results showed there were variations in deduction scores evaluated across expert-novice judges on the same fixed-sequence video routines informing decision-making differences between expert and novice judges. There were significantly more deduction scores applied by experts (international and national judges) compared to that of novices (regional and club judges) in evaluating the routines across all four WAG apparatus indicating more executions were spotted by experts during the judging processes within same allocated time. In addition, expert judges verbalised more deductions applied onto each skill and elements performed than novice judges when using *Level 2* and immediate retrospective TA whilst evaluating video-based routines. Expert judges reported higher frequency in verbalising all three types of deductions than novice judges, inclusive of general execution faults, specific apparatus deductions, and artistry deductions, except for general execution deductions on UB. The apparatus with the most frequent deductions verbalised was BB, followed by FX, UB, and least on VT, concerning the pace and time constraint of respective apparatus. Perceptual-cognitive skills (MacMahon & Mildenhall, 2012) possessed by judges enable them to circumvent time constraints for decision-makings, however, these skills were stretched to complete the multitask of judging within a short time. This study revealed decision-making differences between expert and novice judges with the use of TA method, which was informed by Study One. In addition, results of this study demonstrated there were differences in deduction scores across different apparatus imply that there might be decision-making differences between expert and novice

judges in applying deductions across four WAG apparatus. To further understand how TA method can be used and applied to judge decision-making, Study Three aimed to explore perceptions between expert and novice judges in using the TA method across all four WAG apparatus.

Study Three, was conducted simultaneously with Study Two, and aimed to explore expert and novice judge perceptions of using TA method whilst judging WAG routines across all apparatus. Semi-structure interviews were conducted immediately after judges completed the TA sessions in Study Two and indicated that participants found it easiest to use TA method in verbalising deductions on BB, followed by FX, UB, and reported difficulty on VT, which was linked to the pace of movement and reportedly used judging retrospectively due to the nature of the apparatus. Concurrent *Level 2* TA was perceived as viable to collect decisions made on deductions for general execution faults and specific apparatus deductions for BB, FX, and UB whilst immediate retrospective TA was viable to collect decisions made on artistry deductions for BB and FX and that of the entire routine evaluation on VT. Both expert and novice judges reported an incremental level of awareness over the course of using TA method in applying deductions when they were required to verbalise deductions thus reducing inaccurate deductions made automatically through schemata (Norman & Shallice, 1986). TA method requires judges to verbalise specific deductions applied onto each skill and element and reportedly increased judges consciousness during the decision-making process, thus it prevented judges from applying deductions subconsciously as reported by expert judges from Study Three (see *Chapter 5*), especially to that of a fast-moving series of movements. Initial apprehensions were reported by participants in learning and adapting the TA method into judging at the beginning of data collection, however both expert and novice judges reported that they gained confidence to

verbalise deductions over the course of the TA session. Findings from the studies revealed verbal overshadowing when applying TA method affected performances to complete multitask judging, therefore more training of TA method for familiarisation before the data collection session were required. Both expert and novice judges perceived TA method as viable to collect in-event cognitive processes for understanding decision-making, and thus deemed it appropriate for adaptation into future judge education, specifically to increase score objectivity. These corroborate to previous findings (Whyte IV et al., 2010) comparing the use of concurrent and retrospective verbal reports of nurse performance in a simulated task environment, whereby this study revealed concurrent and immediate delayed TA verbal report enable WAG judges to disclose their decision-makings on deductions applied in a routine. The concurrent TA revealed every execution deductions applied onto each movement following the sequence of element performance, in addition to the immediate delayed TA revealed every artistry deductions applied for the entire routine performance. These provide rich information representing cognition of WAG judges when evaluating routines in a simulated task environment mimic actual competitions, whereby no previous research investigating the details of deductions applied in complete WAG routines. Using both concurrent and immediate delayed TA method thus allow more comprehensive analyses of the decision-making among WAG judges during judge education without concerning the memory decay of using retrospective verbal report approach (Ericsson & Kintsch, 1995; Ericsson & Simon, 1993; Nicholls & Polman, 2008; Whitehead et al., 2015). However, TA during WAG judging is under time constraints but not a self-pace task, therefore it is recommendations made to adapt TA into judge education concerning discussion among E-judges during competitions are prohibited (FIG, 2016b). Furthermore, findings from the study indicated WAG judging is mapped to the expert performance approach (Williams et al., 2017), whereby expertise in WAG judging could be developed with evidence-based training

over time (Chase & Simon, 1973; Ericsson, 2017; Ericsson & Simon, 1993). Resources for this training could include high quality videos meeting the competition levels to mimic actual competitions (Ste-Marie, 1999) with progression from singular skill/element to series of movements then finally on full routines to provide judges to practice at own pace. These are to enhance perceptual-cognitive skill in evaluating WAG routines through video replay and review, and slowing down the video to ensure accurate deductions are applied for each movement.

6.3 Critical Evaluation and Limitations

TA has been adapted in previous sports research involving self-paced sports such as golf (Arsal et al., 2016; Calmeiro & Tenenbaum, 2011; Kaiseler et al., 2013; Whitehead et al., 2015; Whitehead, Taylor, et al., 2016), cycling (Whitehead et al., 2017; Whitehead et al., 2018), long-distance running (Samson et al., 2015), snooker (Welsh et al., 2018), and tennis (Swettenham et al., 2018). However, WAG judging is not self-paced but occurs with a time constraint, whereby BB and FX routines typically complete within 90 seconds (FIG, 2016b), whilst VT is completed within 5 seconds (Pajek et al., 2013). Even though there is no time limit for UB, according to the COP, it is by virtue a faster-paced apparatus compared to BB and FX. Therefore, concurrent verbalisation of judges to TA whilst judging a routine was limited. This is in addition to the effects of verbal overshadowing (Ericsson, 2003; Schooler, 2011), where verbalising inner speech prolonged the time to complete primary task as well as slowing down the decision-making process (Arsal et al., 2016; Fox et al., 2011) of that judging task. This is a novel study utilising concurrent and immediate delayed TA method as verbal reports to provide detail information on decision-makings mediating the execution deduction process for entire WAG routines, inclusive of general execution faults, specific apparatus deductions, and artistry deductions within 60 seconds after the routine completed to

mimic real competitions. Verbal reports previously used in similar time-constrained sports tasks indicated the viability of verbal reports in capturing cognitive thoughts. McRobert et al. (2011) previously examined differences between ten skilled and ten less skilled cricket batters in making anticipation judgements using combination of both concurrent and retrospective verbal reports with 24 video-based stimuli. Limited time of 10-20 seconds were allocated for concurrent verbal reports at the initiation of trials, which indicated time constraints for participants to provide responses using the concurrent TA in that study were complement by retrospective verbal report at the end of the trials. Outcomes of the study supported the viability of using multiple types of verbal reports during time-constrained events to collect cognitive thoughts concerning domain-specific tasks.

Participants reported initial apprehension in verbalising deductions whilst judging video routines. However, they reportedly overcame this across TA sessions and gained confidence towards the end of the session as reported in Study One and Three. More TA training before the actual data collection may be warranted for participants to familiarise themselves with concurrent verbalisation whilst completing multitask WAG judging (Ste-Marie, 1999), in addition to reducing the effects of verbal overshadowing. Judges could acquire verbalisation skill through adaptive learning (Williams & Ericsson, 2005) from the TA training and therefore would be able to concurrently TA their cognitive process reporting decision-makings in STM and LT-WM (Ericsson & Kintsch, 1995) with less hesitation to TA. Moreover, these enable to reduce information-processing demands of judges, especially novice judges to multitask in managing symbol notation, element recognition, and the COP information retrieval involving memory recall, which are confounding variables that influencing the decision-makings on execution and artistry deductions during routine evaluations. Therefore, detailed instructions to focus on the primary task of verbalising

deductions applied on skills, elements, and routines, despite reducing the confounding variables by providing controlled contextual information (McRobert et al., 2013), such as skills and elements within a routine were suggested. A similar investigation conducted in medical field participated by nine skilled and nine less skilled emergency medicine physicians (McRobert et al., 2013) demonstrated that skilled medical physicians outperformed less skilled medical physicians by showing higher diagnostic accuracy irrespective of context-specific information under emergency scenarios. However, both skilled and less skilled medical physicians were able to report higher percentage of correct diagnosis in high-contextual condition compared to low-context condition, indicating the importance of contextual information to ensure primary task of decision-making not affected.

Execution scores and deductions verbalised by judges across studies in this thesis were compared with participants from within studies without reference of "gold standard" that represent the exact scores evaluated by technical committee. Furthermore, verbalisation of execution judges during competition is not permitted according to the rules as stated in the COP (FIG, 2016b) to avoid bias and influence. In addition, WAG require judges to make decisions based on their angle of view, and so their sitting arrangements are at different locations and distances from the apparatus (FIG, 2016b, p. 19). Therefore, it is impossible to collect in-event data of execution judges when judging in a competition for comparison so as to provide information regarding scores differences and/or verbalisation between TA and actual competition. The ecological validity of TA in WAG judging therefore would be best served with experiments conducting trials in environments that mimic actual competitions (McRobert et al., 2011; Williams & Ford, 2013). Video recording competition of routines that mimic actual performance setting for judges in addition to reproduce a routine under controlled and realistic conditions could enable examination of expert performance in judging

WAG by exploring mediating process of decision-makings from the same angle of view (Williams et al., 2011). However, this 'set up' may increase prior processing effects (Ste-Marie & Lee, 1991) through watching a routine repetitively during training thus induce memory-influenced biases (Ste-Marie & Valiquette, 1996; Ste-Marie et al., 2001) during actual judging despite increase anticipation ability that facilitates WAG judging (Ste-Marie, 2003).

Finally, both male and female were found able to verbalise responses when using TA method (Kaiseler et al., 2013; Swettenham et al., 2018) whilst other studies using TA method to explore expert-novice differences (Arsal et al., 2016; Welsh et al., 2018; Whitehead et al., 2018; Whitehead et al., 2015; Whitehead, Taylor, et al., 2016) did not test gender. Gender was not a factor within this thesis therefore was not studied within this thesis, however, future research may consider including gender. However, study participants were all female apart from two participants who were male despite targeting both genders. Therefore, it is possible that the under representation of male WAG judges within this thesis constitutes as a confounding variable and reflects the nature of WAG judging which is dominated by female judges both internationally and locally.

6.4 Theoretical Implications

This thesis makes an original contribution to understand the decision-making of WAG judges when evaluating fixed-sequence video routines across all apparatus, including Balance Beam, Floor Exercise, Vault, and Uneven Bars. TA method appears viable to collect in-event data on decisions made among execution judges, including general execution deductions, specific apparatus deductions, and artistry deductions applied onto each gymnastics skill and dance element through concurrent and immediate retrospective verbalisations. This extends beyond previous WAG studies reporting gaze behaviour (Campo & Gracia, 2017; Omorczyk,

Nosiadek, Ambroży, & Nosiadek, 2015; Pizzera et al., 2018) and motor experience (Heinen et al., 2012; Pizzera, 2012) influencing judging decisions made among expert and novice judges. Outcomes of this thesis indicate there were deduction score differences across expert and novice judges through analyses on execution deductions verbalised onto each skill and element whilst judging a video routine that mimic actual competitions. Focus of expert and novice judges in applying deductions on complete routines across all apparatus were explored to extend beyond the most two important locations viewed by judges when judging singular skill on VT, UB, and FX (Campo & Gracia, 2017). Expert judges demonstrated superior judging performance to verbalise more specific apparatus deductions and artistry deductions that were applicable only on respective apparatus through semantic LTM retrieval, in addition to the general execution faults applied across all four apparatus corroborate to Ste-Marie (2000).

Further, expert judges demonstrated they possessed superior judging performance (Ericsson, 2017) with ability to verbalise more execution deductions on fast-moving acrobatic tumbling and dance series, as well as routines on fast apparatus of that UB. Skill and element performed in a routine were compared to the standards as stated in the COP for deviation evaluations by execution judges. Judges were required to multitask in writing down the symbols which represented every skill and element performed at the same time evaluating it. Findings from this thesis show that expert judges demonstrated developed perceptual-cognitive skills that facilitated them in recognising, familiarising, and anticipating gymnastics skills and dance elements (Ste-Marie, 2003) to reduce information-processing demands when conducting the evaluations, in addition to effective information retrieval from memory (Ste-Marie, 1999) to apply execution deductions. The perceptual-cognitive skills reduced information-processing time especially on fast-moving acrobatic tumbling and dance series

for decision-making with time constraints (Raab et al., 2019; Williams et al., 2003), typically within 60 seconds after a routine complete. Although a specific number of hours were not used as a measure within this thesis, it was assumed that those judges with international level accredited by the FIG and national level accredited by respective nations were demonstrating substantial and accumulated expertise in judging national and international level competitions than those lower-level judges. According to British Gymnastics (2017b), there are pre-requisites to enter a national WAG judge course: 1) Complete at least two-years of WAG regional judge qualification before moving into a national judge; 2) A national WAG judge of previous cycle; or 3) An international judge of previous cycle. A national WAG judge accredited by BG is expected to be able to accurately evaluate routines in UK wide competitions by applying the FIG rules and regulations and to check routine compositions and FIG tariff sheets, which resemble the role of a FIG accredited international judge. Those gymnasts who previously competed at senior international level competitions may be able to start the judging pathway at national judge qualification, in some circumstances, upon approval of the Women's Technical Committee, as they are deemed to possess knowledge of the FIG rules and routine composition (British Gymnastics, 2017a). On the other hand, the pre-requisite of at least one-year club judge accredited for current cycle as well as regional and national judge of past two cycles were eligible to enrol for a regional WAG judge course. Regional judges are required to judge only club and regional competitions that include basic FIG COP of common A up to D elements and BG only rules (British Gymnastics, 2017d). BG currently states no pre-requisite to enrol into a club judge course, whereby accredited club judges are required to recognise common A up to C elements as coded FIG COP in addition to apply deductions as part of the execution panel at club and regional competitions (British Gymnastics, 2017a). Therefore, outcomes from this thesis corroborate to findings from Ste-Marie (1999) that expert judges showed attention capability in detecting and

identifying complex movement patterns during the judging process with perceptual-cognitive skills. Results from this thesis showing expert judges were able to notice more execution errors during a routine performance across all apparatus than novice judges corresponding with expertise skill developed over extensive training. Furthermore, outcomes from this thesis show that the general execution faults, which is the same list applicable for all four WAG apparatus of that BB, FX, VT, and UB received equal attention during judging with no significant group difference found between expert and novice judges. This indicates that the general execution faults with higher retrieval frequency in judging regardless of apparatus could be easily applied demonstrating the information storage within LT-WM of both expert and novice judges, even the verbalised deduction counts by novice judges were less than expert judges. On the other hand, each apparatus has different list of specific apparatus deductions and artistry deductions, whereby judges were recalling those information when judging respective apparatus hence retrieval frequency were reduced. Therefore, novice judges were unable to verbalise specific apparatus deductions and artistry deductions concurrently whilst judging the video routines compared to expert judges resulting the significant difference of group main effect on the deduction count analyses. This corroborates literature whereby higher frequency of information storing and retrieval enhanced LT-WM to facilitate the WAG judging (Ste-Marie, 1999, 2000) in managing information-processing demands. International and national judges were eligible to judge more competitions in addition to club and regional levels thus providing them more judging opportunities for judging practice that enhance the COP information retrieval through semantic memory (Baddeley, 2012) as superior declarative knowledge. These were substantiated by findings from previous study (McPherson & MacMahon, 2008) that LTM structures were mediating the decision-makings to support judging performances. In Study Three, novice judges revealed they were unable to accurately evaluate skills and elements on faster-pace apparatus

with time constraints. On the other hand, expert judges mentioned they voluntarily increased practice volumes on faster-pace apparatus before attending a competition with the purpose to enable them in recalling COP information instantly to overcome the challenge of multitasking in noting down routine movements whilst evaluating the routine performance. These indicated the expert-novice differences in possessing sophisticated role-specific knowledge adapting STM and LT-WM through repeated exposure to the environment and tasks enhancing perceptual-cognitive expertise (Williams et al., 2017) in circumvent the time constraint. Specifically, expert judges possessed conditional knowledge with developed declarative knowledge to retrieve and encode domain-specific knowledge from the COP more efficiently in LT-WM (Ericsson & Kintsch, 1995), in addition to superior procedural knowledge (Ste-Marie, 1999) that enable them to write down symbols and apply execution deductions onto each element immediately when performed by gymnasts.

Results from Study Three had extends understanding surrounding perceptions of expert and novice judges in using TA method to judge each apparatus of WAG. The study also offers considerations for judge education by way of adapting learning of TA method according to specific characteristics of respective apparatus. The systematic framework of expert performance approach proposed by Ericsson and Smith (1991), reviewed by Williams et al. (2017), suggested to employ TA method in understanding the skill acquisition of experts' learning to enhance perceptual-cognitive skill in further to promote the using of instructional interventions and practice opportunities for expertise development. Outcomes of the studies revealed that expert judges were better able to judge faster-paced apparatus and series of movements, compared to novice judges, by effectively retrieving COP information effectively (Ste-Marie, 1999, 2000). Despite expertise typically accumulating over many years, systematic training and deliberate practice activities that refine the perceptual-

cognitive skills required for complex domain-specific skills can be developed for novice judges (Ericsson, 2017). Therefore, adaptation of TA method into judge education could enhance capability of novice judges to process incomplete, intentional deceptive and fast-paced information of fast-moving series of gymnastics skills and elements. As such this would be achieved using enhanced perceptual-cognitive skills, in addition to superior declarative knowledge of COP deduction applications for effective decision-makings under time constraint.

This been supported with results from this thesis showing expert judges were able to notice more execution errors during a routine performance across all apparatus than novice judges corresponding with expertise skill developed over extensive training. Furthermore, findings from this thesis demonstrated that novice judges, who are club and regional judges, were focusing more on general execution deductions that were applicable to all apparatus but less on specific apparatus deductions that applicable only to respective apparatus. On the other hand, expert judges, who are international and national judges, were able to apply more specific apparatus deductions in respective apparatus compared to novice judges. These corroborates literature whereby higher frequency of information storing and retrieval enhanced LT-WM to facilitate the WAG judging (Ste-Marie, 1999, 2000) in managing information-processing demands. International and national judges were eligible to judge more competitions in addition to club and regional levels thus providing them more judging opportunities for judging practice that enhance the COP information retrieval through semantic memory (Baddeley, 2012) as superior declarative knowledge.

Results from Study Three had extends understanding surrounding perceptions of expert and novice judges in using TA method to judge each apparatus of WAG. The study also offers considerations for judge education by way of adapting learning of TA method

according to specific characteristics of respective apparatus. The systematic framework of expert performance approach proposed by Ericsson and Smith (1991), reviewed by Williams et al. (2017), suggested to employ TA method in understanding the skill acquisition of experts' learning to enhance perceptual-cognitive skill in further to promote the using of instructional interventions and practice opportunities for expertise development. Outcomes of the studies revealed that expert judges were better able to judge faster-paced apparatus and series of movements, compared to novice judges, by effectively retrieving COP information effectively (Ste-Marie, 1999, 2000). Despite expertise typically accumulating over many years, systematic training and deliberate practice activities that refine the perceptual-cognitive skills required for complex domain-specific skills can be developed for novice judges (Ericsson, 2017).

6.5 Practical Implications

This thesis makes a novel contribution to WAG technical committee in planning future judge education by demonstrating differences between expert and novice judges in evaluating routines across all four apparatus, of that BB, FX, VT, and UB. Novices reported that more time was required to complete the judging tasks during a competition compared to expert judges, whereby novice judges were referring the COP, as well as additional rules and regulations for specific competitions before applying deductions or element recognitions. This indicates that novice judges require more practice to enhance for effective information retrieval on declarative knowledge to reduce information-processing time (Ste-Marie, 1999, 2000) judging for an actual competition. Preparations are needed to conduct multitask WAG judging efficiently to circumvent time constraints during actual competition, such as frequent revision on the COP knowledge, in addition to watching video routines for judging practices to enhance STM and LT-WM (Ericsson & Kintsch, 1995) for information retrieval. Judges

with enhanced expert memory capacity developed through training were able to anticipate frequent patterns with enhanced perceptual-cognitive skills (Williams et al., 2011). Hence, accurate execution deductions were applied onto skills and elements based on superior declarative knowledge (Heinen et al., 2012; Ste-Marie, 1999) from semantic memory (Baddeley, 2012) with caution of that biases induced by prior-information processing (Ste-Marie & Lee, 1991; Ste-Marie et al., 2001). TA deductions require judges to verbalise every deduction applied onto each skill and element in a routine, inclusive of general execution faults, specific apparatus deductions, and artistry deductions. This enable judges to review their decisions reflectively and as such reportedly increases awareness to avoid deductions applied by schemata (Norman & Shallice, 1986), whereby novice judges in this thesis reported they did apply repeated deductions onto different skills and elements.

Novices stated that they gained confidence by comparing their decisions with experts during the practical trainings in judge courses and thus reported it was useful to enhance their judging accuracy. Therefore, an online module with updated and high-quality video routines that enable replay, at variable pace for convenient practice for example before attending an accreditation examination or a judging assignment would be useful. This extends beyond the current offering of 'in course' resources designed to prepare judges for examination and not as an ongoing resource stream. This thesis also revealed different challenges faced by expert and novice judges when judging a singular apparatus in a competition. Therefore, interventions targeting different levels of judges could facilitate judge's expertise development. This includes inserting regional only allowed moves, such as *teddy bear roll* on FX and *caterpillar* on BB with specific descriptions to allow accurate evaluations. Currently no videos exist for these moves despite them being seen with low frequency in lower-level club and regional competitions. The viability of video-based online training to enhance sport

official's decision-making had substantiated by previous finding (Schweizer, Plessner, Kahlert, & Brand, 2011) found that decisions of participants improved over the course of the training program, whereby immediate feedback on the correctness of decisions without further explanations was sufficient for increasing decision accuracy. Therefore, providing reference answers by expert judges at the end of video clips enable WAG judges to practice at own pace and space, further to reflect and correct their decision-makings to outcome accurate and objective scores. Online intervention adapting TA method with utilisation of video clips thus provide opportunities for judges to practice judging to improve superior declarative knowledge with developed memory recall to enhance decision-makings (Garcı´a-Gonza´lez et al., 2013). Online intervention using video clips mimic actual competitions but without any time limitation to pressure judges compared to actual competitions thus enable them to increase comprehension of factors that influencing their performances. Moreover, WAG require judges to make decisions based on their angle of view, and so their sitting arrangements are at different locations and distances from the apparatus (FIG, 2016b, p. 19). Video clips recorded in competitions with controlled and realistic conditions enable examination of expert performance in judging WAG by exploring mediating process of decision-makings from the same angle of view (Williams et al., 2011). These allowed judges to practice their decision-making skills to facilitate the development of expertise thus best served with experiments conducting trials in environments that mimic actual competitions (McRobert et al., 2011; Williams & Ford, 2013).

Findings of this study revealed that deduction scores applied by novice judges were less than that of expert judges across all four apparatus. Moreover, verbalised TA deduction counts for all apparatus recorded for novice judges were less compared to expert judges, with the exception of the general execution faults on UB. Novice judges reduced their deductions

therefore when the pace of skills and elements of an apparatus increased. Novice judges are expected to enhance their perceptual-cognitive skills to anticipate upcoming frequent skills and elements to reduce information-processing time when evaluating fast pace apparatus (Williams et al., 2011). Data suggests how practicalities and content of judge education concerning the capability of novice judges should begin with BB, followed by FX, then UB and VT. Furthermore, participants report TA as an acceptable method to collect in-event data that enables examination of score objectivity and bias in addition to utilising post-competition scores (Bučar, Čuk, Pajek, Karacsony, & Leskošek, 2012; Mercier & Klahn, 2017; Pajek et al., 2013; Pajek et al., 2014; Sacchi, 2018a, 2018b). Nevertheless, applying respective deductions when judging different WAG apparatus were not reveal in previous studies investigating expert performance of WAG judges despite mentioning deductions in general. TA method requires judges to verbalise every deduction to reveal the specific deductions and weight of deduction scores applied in a routine chronologically.

In making recommendations for WAG education programme this should include online interventions to develop TA and as such extends beyond the current judge education of 'paper and pen' tasks aligned to judging level, which seems dated and reportedly ineffective from the participants in this thesis. TA method could increase expertise among WAG judges according to the expert performance approach (Williams et al., 2017), hence, future WAG judge education is proposed to include TA method as a progressive interactive programme to increase novice judge's awareness in own decision-making processes whilst capturing decision-making processes of experts. WAG judge education should include contemporary and clear videos, which contain elements and routines that are commensurate with the respective judge course level, be it of that club, regional, national. Results from data also suggested novice judges needed to first be familiar on single skills to skill and dance

elements, before progressing into series and full routines across all four apparatus with guidance or 'modal answer' from experts providing novice judges training opportunities. The TA method could enable novice judges to record their own responses when judging a routine then reviewed and for this to be compared with the TA responses from experts with reference scores towards enhancing score objectivity. Moreover, the use of video clips that mimic actual competitions within an online intervention are providing judges more opportunities to develop their judging efficacy through sport-specific decision-making scenario practices, whereby the potential benefits associated with video-based decision-making training for officials were explored (Catteeuw et al., 2010; Larkin et al., 2017). A greater amount of contextualised visual experiences without explicit instruction during a 12-week video-based decision-making training programme led to an improved decision-making performance of less experience Australian football umpires (Larkin et al., 2017). Therefore, novice officials with less experience demonstrated greater improvement capacity than the experienced officials, whereby a greater amount of learning occurs in the initial stages of practice according to the power law of practice and ceiling effect, as less experienced officials have fewer robust knowledge structures regarding decision-making situations at the commencement of training.

6.6 Recommendations for Future Research

Future studies are recommend to include questionnaire or interview questions to track the practice hours of participants involving in the judging specific hours in addition to history involving in gymnastics related activities. The practice history profiling (Williams et al., 2017) would enable researchers to further investigate the expertise development, which has been widely used on athletes but not much on officiating.

Further research could consider using TA in addition to existing gaze behaviour studies (Heinen et al., 2012; Omorczyk et al., 2015; Pizzera & Raab, 2012) to reveal decision-making mediating the evaluation process of WAG judges. This also enable further investigation of differences between expert and novice WAG judges (Campo & Gracia, 2017; Pizzera et al., 2018) to increase score objectivity and accuracy using videos which mimic actual competitions. This would involve investigations of longitudinal studies that includes statistical analysis of pre and post scores of TA training and which also included follow-up interviews to collect perceptions and feedback from judges who attended the TA intervention in addition to exploration of their attentional focus of visual. Judge education needs to be improved to meet the training needs of judges across level, especially novice judges who are, by virtue of how they enter judging, may not be able to recognise gymnastic skills and dance elements and judge regional only moves accurately. Follow up studies to explore decision-making of judges over time could reveal the effectiveness of judge education to develop expert judges in meeting expert performance approach (Williams et al., 2017).

It is impossible to collect in-event data of judge's decisions in actual judging during competitions, whereby discussion among WAG judges during a competition is prohibited as stated in the COP (FIG, 2016b) to avoid biases. Moreover, sitting arrangement of judges may vary across competitions, whereby individual sittings are allocated around an apparatus in most of international level competitions (FIG, 2016b, p. 19) whereas judges were sitting as a panel next to each other/side by side in most local competitions. These arrangements may result in different decisions made due to a different angle of view. Due to verbalisation is prohibited in actual competition, therefore, future research might investigate difference of decisions made by collecting scores applied during an actual competition then compared to

the verbalisation outcomes using TA method in the follow-up trials with videos captured from the competition.

Moreover, TA method enables data collection involving aesthetic evaluation of WAG judges, especially in judging BB and FX, whereby artistry evaluation is part of the execution scores. Aesthetic value in sports require subjective judgment from judges (Reid, 1970) to include their perceptions towards aesthetic values within (Edgar, 2013). Therefore, artistry deductions deemed inappropriate to be 'evaluated by machine' considering a real-time judging support system to capture gymnast's movement and then converts to numerical data to undergo analysis, which has been proposed for implementation from Olympic Tokyo 2020 (FIG, 2019). Further research comparing scores outcome by a system such as this and 'human' evaluations from judges is warranted to investigate the viability of the real-time judging system to replace judges in whole/part given the need to offer scores for artistry etc.

6.7 Conclusions

This thesis has extended the application of TA method of expert performance approach (Williams et al., 2017) in understanding decision-making differences between expert and novice WAG judges when evaluating fixed-sequence video-based routines across all four apparatus of that BB, FX, VT, and UB further to perceptions of using TA method whilst judging. This thesis found that expert judges were applying more deduction scores compared to novice judges across all four apparatus. Moreover, expert judges verbalised more TA deductions compared novice judges in all apparatus, inclusive of general execution faults, specific apparatus deductions, and artistry deductions, except for the general execution faults on UB. These revealing decision-making differences between expert and novice judges demonstrating the capability variances of perceptual-cognitive skills (Williams et al., 2011) and superior declarative knowledge (Ste-Marie, 1999, 2000) in managing multitask WAG

judging. This suggested future WAG judge education to concern the expert-novice differences in designing training modules to ensure the score objectivity across competitions besides memory and perceptual-cognitive skills refinements though deliberate practice could facilitate their judging ability. In addition, expert and novice judges expressed their perceptions towards TA method as a viable tool to collect in-event thought processes when judging routines, thus proposing future WAG judge education to include TA method as a training method to increase score objectivity.

REFERENCES

Arsal, G., Eccles, D. W., & Ericsson, K. A. (2016). Cognitive mediation of putting: Use of a think-aloud measure and implications for studies of golf-putting in the laboratory. *Psychology of Sport and Exercise, 27*, 18-27. doi:10.1016/j.psychsport.2016.07.008

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and Its Control Processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89-195). New York: Academic Press.

Auweele, Y. V., Boen, F., Geest, A. D., & Feys, J. (2004). Judging Bias in Synchronized Swimming: Open Feedback Leads to Nonperformance-Based Conformity. *Journal of Sport & Exercise Psychology, 26*, 561-571. doi:10.1123/jsep.26.4.561

Baddeley, A. (1986). *Working memory*. Oxford: Clarendon.

Baddeley, A. (2003). Working Memory: Looking Back and Looking Forward. *Nature reviews Neuroscience, 4*(10), 829-839. doi:10.1038/nrn1201

Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology, 63*, 1-29. doi:10.1146/annurev-psych-120710-100422

Bard, C., Fleury, M., Carriere, L., & Halle, M. (1980). Analysis of Gymnastics Judges' Visual Search. *Research Quarterly for Exercise and Sport, 51*(2), 267-273. doi:10.1080/02701367.1980.10605195

Bernard, H. S., Killworth, P., Kronenfeld, D., & Sailer, L. (1984). THE PROBLEM OF INFORMANT ACCURACY: The Validity of Retrospective Data. *Annual Review Anthropol, 13*, 495-517.

Best, D. (1974). The Aesthetic in Sport. *The British Journal of Aesthetics*, *14*(3), 197–213. doi:10.1093/bjaesthetics/14.3.197

Birt, L., Scott, S., Cavers, D., Campbell, D., & Walter, F. (2016). Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation? *Qualitative Health Research, 26*(13), 1802-1811. doi: 10.1177/1049732316654870

Boen, F., Auweele, Y. V., Claes, E., Feys, J., & Cuyper, B. D. (2006). The impact of open feedback on conformity among judges in rope skipping. *Psychology of Sport and Exercise, 7*, 577–590. doi:10.1016/j.psychsport.2005.12.001

Boen, F., Hoye, K. V., Auweele, Y. V., & Feys, J. (2006). Judging Bias in Artistic Gymnastics: The Pressure to Conform. In F. Boen, B. D. Cuyper, & J. Opdenacker (Eds.), *Current research topics in exercise and sport psychology in Europe* (pp. 123-138). Leuven: LannooCampus.

Boen, F., van Hoye, K., Auweele, Y. V., Feys, J., & Smits, T. (2008). Open feedback in gymnastic judging causes conformity bias based on informational influencing. *Journal of Sports Sciences, 26*(6), 621-628. doi:10.1080/02640410701670393

Boren, M. T., & Ramey, J. (2000). Thinking Aloud: Reconciling Theory and Practice. *IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION, 43*(3), 261-278.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. doi:10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2014). What can "thematic analysis" offer health and wellbeing researchers? *International Journal of Qualitative Studies on Health and Wellbeing, 9*(1), 26152. doi: 10.3402/qhw.v9.26152

British Gymnastics. (2017a). Women's Artistic Gymnastics Club Judge Course. Retrieved from https://www.british-gymnastics.org/courses/6308/womens-artistic-gymnastics-club-judge

British Gymnastics. (2017b). Women's Artistic Gymnastics National Judge Course. Retrieved from https://www.british-gymnastics.org/courses/6310/womens-artistic-gymnastics-national-judge

British Gymnastics. (2017c). Women's Artistic Gymnastics NDP Grades 2018-2021. United Kingdom: British Gymnastics.

British Gymnastics. (2017d). Women's Artistic Gymnastics Regional Judge Course. Retrieved from https://www.british-gymnastics.org/courses/6309/womens-artistic-gymnastics-regional-judge

Bučar, M., Čuk, I., Pajek, J., Karacsony, I., & Leskošek, B. (2012). Reliability and validity of judging in women's artistic gymnastics at University Games 2009. *European Journal of Sport Science, 12*(3), 207-215. doi:10.1080/17461391.2010.551416

Calmeiro, L., & Tenenbaum, G. (2011). Concurrent Verbal Protocol Analysis in Sport: Illustration of Thought Processes During a Golf-Putting Task. *Journal of Clinical Sport Psychology, 5*(3), 223-236. doi:10.1123/jcsp.5.3.223

Campo, V. L., & Gracia, I. E. (2017). Exploring Visual Patterns and Judgments Predicated on Role Specificity: Case Studies of Expertise in Gymnastics. *Current Psychology*. doi:10.1007/s12144-017-9572-1

Catteeuw, P., Gilis, B., Jaspers, A., Wagemans, J., & Helsen, W. (2010). Training of Perceptual-Cognitive Skills in Offside Decision Making. *Journal of Sport & Exercise Psychology, 32,* 845-861.

Chase, W. G., & Ericsson, K. A. (1982). Skill and Working Memory. *Psychology of Learning and Motivation, 16*, 1-58. doi:10.1016/S0079-7421(08)60546-0

Chase, W. G., & Simon, H. A. (1973). The Mind's Eye in Chess. In W. G. Chase (Ed.), *Visual Information Processing* (pp. 215-281). Pennsylvania: Elsevier Inc.

Chin, J. M., & Schooler, J. W. (2008). Why do words hurt? Content, process, and criterion shift accounts of verbal overshadowing. *European Journal of Cognitive Psychology, 20*(3), 396-413. doi:10.1080/09541440701728623

Clarke, V., & Braun, V. (2017). Thematic analysis. *The Journal of Positive Psychology, 12*(3), 297-298. doi:10.1080/17439760.2016.1262613

Coughlan, E. K., Williams, A. M., McRobert, A. P., & Ford, P. R. (2014). How Experts Practice: A Novel Test of Deliberate Practice Theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(2), 449–458. doi:10.1037/a0034302

de Groot, A. D. (1978). *Thought and Choice in Chess* (2nd ed. Vol. 4). The Hague, The Netherlands: Mouton Publishers.

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience, 18*, 193-222. doi:10.1146/annurev.ne.18.030195.001205

Dialsingh, I. (2008). Face-to-Face Interviewing. In P. J. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods* (pp. 276-261). Thousand Oaks: Sage Publications, Inc.

Dickens, Y. L., Raalte, J. V., & Hurlburt , R. T. (2018). On Investigating Self-Talk: A Descriptive Experience Sampling Study of Inner Experience During Golf Performance. *The Sport Psychologist, 32*(1), 66-73. doi:10.1123/tsp.2016-0073

Dosseville, F., & Laborde, S. (2015). Introduction to the special issue: Officials in sports. *Movement & Sport Sciences, 87*, 3-10. doi:10.1051/sm/2015006

Eccles, D. W. (2012). Verbal Reports of Cognitive Processes. In G. Tenenbaum, R. C. Eklund, & A. Kamata (Eds.), *Measurement in Sport and Exercise Psychology* (pp. 103-117). Leeds: Human Kinetics.

Eccles, D. W., & Arsal, G. (2017). The think aloud method: what is it and how do I use it? *Qualitative Research in Sport, Exercise and Health, 9*(4), 514-531. doi:10.1080/2159676X.2017.1331501

Edgar, A. (2013). The Aesthetics of Sport. *Sport, Ethics and Philosophy, 7*(1), 80-99. doi:10.1080/17511321.2013.761885

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative Content Analysis: A Focus on Trustworthiness. *SAGE Open*, 1-10. doi:10.1177/2158244014522633

Ericsson, K. A. (2003). Valid and Non-Reactive Verbalization of Thoughts During Performance of Tasks: Towards a Solution to the Central Problems of Introspection as a Source of Scientific Data. *Journal of Consciousness Studies, 10*(9-10), 1-18.

Ericsson, K. A. (2017). Expertise and individual differences: the search for the structure and acquisition of experts' superior performance. *Wires Cognitive Science, 8*(1-2), e1382. doi:10.1002/wcs.1382

Ericsson, K. A., & Delaney, P. F. (1999). Long-Term Working Memory as an Alternative to Capacity Models of Working Memory in Everyday Skilled Performance. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 257-297): Cambridge University Press.

Ericsson, K. A., & Kintsch, W. (1995). Long-term Working Memory. *Psychological Review, 102*(2), 211-245. doi:10.1037/0033-295X.102.2.211

Ericsson, K. A., & Kintsch, W. (2000). Shortcomings of generic retrieval structures with slots of the type that Gobet (1993) proposed and modelled *British Journal of Psychology, 91*(4), 571-590. doi:10.1348/000712600161998

Ericsson, K. A., & Simon, H. A. (1981). *Protocol Analysis*. Departments of Psychology University of Colorado and Carnegie-Mellon University.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data* (Revised Edition ed.). Cambridge: The MIT Press.

Ericsson, K. A., & Smith, J. (1991). Prospects and limits of the empirical study of expertise: An introduction. In K. A. Ericsson & J. Smith (Eds.), *Towards a General Theory of Expertise: Prospects and Limits* (pp. 1-38). New York: Cambridge University Press.

FIG, F. o. I. G. (2012). Women's Artistic Gymnastics: 2013-2016 Code of Points.

FIG, F. o. I. G. (2016a). FIG General Judge's Rules: Cycle 2017-2020.

FIG, F. o. I. G. (2016b). Women's Artistic Gymnastics: 2017-2020 Code of Points.

FIG, F. o. I. G. (2019). The International Gymnastics Federation and Fujitsu collaborate on building a judging support system for Artistic Gymnastics competitions. Retrieved from http://www.gymnastics.sport/site/pages/judges-support.php

Findlay, L. C., & Ste-Marie, D. M. (2004). A Reputation Bias in Figure Skating Judging. *Journal of Sport & Exercise Psychology, 26*, 154-166.

Flessas, K., Mylonas, D., Panagiotaropoulou, G., Tsopani, D., Korda, A., Siettos, C., . . . Smyrnis, N. (2015). Judging the Judges' Performance in Rhythmic Gymnastics. *Medicine & Science in Sports & Exercise, 47*(3), 640-648. doi:10.1249/MSS.0000000000000425

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do Procedures for Verbal Reporting of Thinking Have to Be Reactive? A Meta-Analysis and Recommendations for Best Reporting Methods. *Psychological Bulletin, 137*(2), 316-344. doi:10.1037/a0021663

Garcı´a-Gonza´lez, L., Moreno, M. P., Moreno, A., Gil, A., & del Villar, F. (2013). Effectiveness of a Video-Feedback and Questioning Programme to Develop Cognitive Expertise in Sport. *PLos ONE, 8*(12), e82270. doi:doi:10.1371/journal.pone.0082270

Gentsch, A., Weber, A., Synofzik, M., Vosgerau, G., & Schütz-Bosbach, S. (2016). Towards a common framework of grounded action cognition: Relating motor control, perception and cognition. *Cognition, 146*, 81-89. doi:10.1016/j.cognition.2015.09.010

Gobet, F. (1998). Expert Memory: A Comparison of Four Theories. *Cognition, 66*, 115-152. doi:10.1016/S0010-0277(98)00020-1

Grossfeld, A. (2014). Changes during The 110 Years of The World Artistic Gymnastics Championships. *Science of Gymnastics Journal, 6*(2), 5-27. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=s3h&AN=99664125&site=ehost-live

Heinen, T., Vinken, P. M., & Velentzas, K. (2012). Judging Performance in Gymnastics: A Matter of Motor or Visual Experience? *Science of Gymnastics Journal, 4*(1), 63-72.

International Olympic Committee. (2015). *GYMNASTICS: History of Artistic Gymnastics at the Olympic Games*. Olympic Studies Centre Retrieved from studies.centre@olympic.org

Kaiseler, M., Polman, R. C. J., & Nicholls, A. R. (2013). Gender differences in stress, appraisal, and coping during golf putting. *International Journal of Sport & Exercise Psychology, 11*(3), 258-272. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=s3h&AN=90215527&site=ehost-live

Kaya, A. (2014). Decision making by coaches and athletes in sport. *Procedia - Social and Behavioral Sciences, 152*, 333-338. doi:10.1016/j.sbspro.2014.09.205

Kerstetter, K. (2012). Insider, Outsider, Or Somewhere in Between: The Impact of Researchers' Identities on the Community-based Research Process. *Journal of Rural Social Sciences, 27*(2), 99–117.

Kramer, R. S. S. (2017). Sequential effects in Olympic synchronized diving scores. *Royal Society Open Science, 4*, 1-9. doi:10.1098/rsos.160812

Larkin, P., Mesagno, C., Berry, J., Spittle, M., & Harvey, J. (2017). Video-based training to improve perceptualcognitive decision-making performance of Australian football umpires. *Journal of Sports Sciences, 36*(3), 239-246. doi:10.1080/02640414.2017.1298827

Lee, C. G. (2012). Reconsidering Constructivism in Qualitative Research. *Educational Philosophy and Theory*, *44*(4), 403-412. doi:10.1111/j.1469-5812.2010.00720.x

Lee, J., Knowles, Z., & Whitehead, A. E. (2019). Exploring the use of think aloud within Women's artistic gymnastics judging education. [ ]. *Psychology of Sport & Exercise, 40*, 135-142. doi:10.1016/j.psychsport.2018.10.007

Leskošek, B., Čuk, I., Pajek, J., Forbes, W., & Bučar-Pajek, M. (2012). Bias of Judging in Men's Artistic Gymnastics at the European Championship 2011. *Biology of Sport, 29*(2), 107-113. doi:10.5604/20831862.988884

Loffing, F., & al-Bruland, R. C. (2017). Anticipation in sport *Current Opinion in Psychology, 16*, 6-11. doi:dx.doi.org/10.1016/j.copsyc.2017.03.008

MacMahon, C., Helsen, W. F., Starkes, J. L., & Weston, M. (2007). Decision-making skills and deliberate practice in elite association football referees. *Journal of Sports Sciences, 25*(1), 65 – 78. doi:10.1080/02640410600718640

MacMahon, C., & Mildenhall, B. (2012). A Practical Perspective on Decision Making Influences in Sports Officiating. *International Journal of Sports Science & Coaching, 7*(1), 153-165. Retrieved from <Go to ISI>://WOS:000302513600015

MacMahon, C., Starkes, J., & Deakin, J. (2007). Referee Decision Making in a Video-Based Infraction Detection Task: Application and Training Considerations. *International Journal of Sports Science & Coaching, 2*(3), 257-265.

MacMahon, C., & Starkes, J. L. (2008). Contextual influences on baseball ball-strike decisions in umpires, players, and controls. *Journal of Sports Sciences, 26*(7), 751-760. doi:10.1080/02640410701813050

MacMahon, C., & Ste-Marie, D. M. (2002). Decision-making by Experienced Rugby Referees: Use of Perceptual Information and Episodic Memory. *Perceptual and Motor Skills, 95*, 570-572.

MacPhail, C., Khoza, N., Abler, L., & Ranganathan, M. (2016). Process guidelines for establishing Intercoder Reliability in qualitative studies. *Qualitative Research, 16*(2), 198-212. doi:10.1177/1468794115577012

Mann, D. L., Causer, J., Nakamoto, H., & Runswick, O. R. (2019). Visual Search Behaviours in Expert Perceptual Judgements. In A. M. Williams & R. C. Jackson (Eds.), *Anticipation and Decision Making in Sport* (pp. 59-78). London & New York: Routledge.

McFee, G. (2013). Officiating in Aesthetic Sports. *Journal of the Philosophy of Sport, 40*(1), 1-17. doi:10.1080/00948705.2012.725910

McPherson, S. L., & Kernodle, M. (2007). Mapping two new points on the tennis expertise continuum: tactical skills of adult advanced beginners and entry-level professionals during competition. *Journal of Sports Sciences, 25*(8), 945-959. doi:10.1080/02640410600908035

McPherson, S. L., & MacMahon, C. (2008). How Baseball Players Prepare to Bat: Tactical Knowledge as a Mediator of Expert Performance in Baseball. *Journal of Sport and Exercise Psychology, 30*, 755-778.

McPherson, S. L., & Vickers, J. N. (2004). Cognitive control in motor expertise. *International Journal of Sport and Exercise Psychology, 2*(3), 274-300. doi:10.1080/1612197X.2004.9671746

McRobert, A. P., Causer, J., Vassiliadis, J., Watterson, L., Kwan, J., & Williams, A. M. (2013). Contextual information influences diagnosis accuracy and decision making in simulated emergency medicine emergencies. *BMJ Qual Saf, 22*, 478-484. doi:10.1136/bmjqs-2012-000972

McRobert, A. P., Mercer, S. J., Raw, D., Goulding, J., & Williams, M. A. (2017). Effect of expertise on diagnosis accuracy, non-technical skills and thought processes during simulated high-fidelity anaesthetist scenarios. *BMJ Simulation and Technology Enhanced Learning, 0*, 1-5. doi:10.1136/bmjstel-2016-000129

McRobert, A. P., Ward, P., Eccles, D. W., & Williams, A. M. (2011). The effect of manipulating context-specific information on perceptual-cognitive processes during a simulated anticipation task. *British Journal of Psychology, 102*(3), 519-534. doi:10.1111/j.2044-8295.2010.02013.x

McRobert, A. P., Williams, A. M., Ward, P., & Eccles, D. W. (2009). Tracing the process of expertise in a simulated anticipation task. *Ergonomics*, *52*(4), 474-483. doi:10.1080/00140130802707824

Meissner, C. A., & Brigham, J. C. (2001). A Meta-analysis of the Verbal Overshadowing Effect in Face Identification. *Applied Cognitive Psychology, 15*, 603-616. doi:10.1002/acp.728

Memmert, D. (2015). Development of Tactical Creativity in Sports. In J. Baker & D. Farrow (Eds.), *Routledge Handbook of Sport Expertise* (pp. 363-372). London & New York: Routledge

Memmert, D., & Roca, A. (2019). Tactical Creativity and Decision Making in Sport. In A. M. Williams & R. C. Jackson (Eds.), *Anticipation and Decision Making in Sport* (pp. 203-214). London & New York: Routledge.

Mercier, H., & Klahn, C. (2017). *Judging the judges: Evaluating the performance of international gymnastics judges*. Paper presented at the Sports Analytics Conference, Hynes Convention Centre.

Miron-Shatz, T., Stone, A., & Kahneman, D. (2009). Memories of Yesterday's Emotions: Does the Valence of Experience Affect the Memory-Experience Gap? *American Psychological Association, 9*(6), 885–891. doi:10.1037/a0017823

Nevill, A. M., Hemingway, A., Greaves, R., Dallaway, A., & Devonport, T. J. (2016). Inconsistency of decision-making, the Achilles heel of referees. *Journal of Sports Sciences*. doi:10.1080/02640414.2016.1265143

Newell, A., & Simon, H. A. (1972). *Human Problem Solving*: Prentice Hall.

Nicholls, A. R., & Polman, R. C. J. (2008). Think Aloud: Acute Stress and Coping Strategies during Golf Performances. *Anxiety, Stress & Coping, 21*(3), 283-294. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=s3h&AN=32996521&site=ehost-live

Nisbett, R. E., & Wilson, T. D. (1977). Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review, 84*(3), 231-259.

Norman, D. A., & Shallice, T. (1986). Attention to Action: Willed and Automatic Control of Behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and Self-Regulation* (pp. 1-18). Boston, MA: Springer.

Nufer, G., & Alesi, N. (2018). The Halo Effect in Sports *International Journal of Business and Social Science, 9*(2), 31-44.

Omorczyk, J., Nosiadek, L., Ambroży, T., & Nosiadek, A. (2015). High-frequency video capture and a computer program with frame-by-frame angle determination functionality as tools that support judging in artistic gymnastics. *Acta of Bioengineering and Biomechanics, 17*(3). doi:10.5277/ABB-00123-2014-02

Opdenakker, R. (2006). Advantages and Disadvantages of Four Interview Techniques in Qualitative Research. *Forum: Qualitative Social Research, 7*(4). doi:10.17169/fqs-7.4.175

Page, J., & Thelwell, R. (2013). The Value of Social Validation in Single-Case Methods in Sport and Exercise Psychology. *Journal of Applied Sport Psychology, 25*(1), 61-71. doi:10.1080/10413200.2012.663859

Pajek, M. B., Cuk, I., Pajek, J., Kovac, M., & Leskosek, B. (2013). Is the Quality of Judging in Women Artistic Gymnastics Equivalent at Major Competitions of Different Levels? *Journal of Human Kinetics, 37,* 173-181. doi:10.2478/hukin-2013-0038

Pajek, M. B., Kovač, M., Pajek, J., & Leskošek, B. (2014). The Judging of Artistry Components in Female Gymnastics: A Cause for Concern? . *Science of Gymnastics Journal, 6*(3), 5-12. Retrieved from http://web.a.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=d043c8a3-2df4-4f97-953d-300aa29c8a72%40sessionmgr4009&vid=1&hid=4107

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16*(2), 283-290.

Pizzera, A. (2012). Gymnastic Judges Benefit From Their Own Motor Experience as Gymnasts. *Research Quarterly for Exercise and Sport, 83*(4), 603-607. doi:10.1080/02701367.2012.10599887

Pizzera, A., Möller, C., & Plessner, H. (2018). Gaze Behavior of Gymnastics Judges: Where Do Experienced Judges and Gymnasts Look While Judging? *Research Quarterly for Exercise and Sport, 89*(1), 112-119. doi:10.1080/02701367.2017.1412392

Pizzera, A., & Raab, M. (2012). Perceptual Judgments of Sports Officials are Influenced by their Motor and Visual Experience. *Journal of Applied Sport Psychology, 24*(1), 59-72. doi:10.1080/10413200.2011.608412

Plessner, H. (1999). Expectation Biases in Gymnastics Judging. *Journal of Sport & Exercise Psychology, 21,* 131-144.

Raab, M., MacMahon, C., Avugos, S., & Bar-Eli, M. (2019). Heuristics, Biases, and Decision Making. In A. M. Williams & R. C. Jackson (Eds.), *Anticipation and Decision Making in Sport* (pp. 215-231). London & New York: Routledge.

Ram, N. N., & McCullagh, P. P. (2003). Self-modeling: Influence on psychological responses and physical performance. *The Sport Psychologist, 17*(2), 220-241. doi:10.1123/tsp.17.2.220

Reid, L. A. (1970). Sport, the Aesthetic and Art. *British Journal of Educational Studies,, 18*(3), 245-258. Retrieved from https://www.jstor.org/stable/3120582

Sacchi, D. (2018a). *WAG General Report: 2018 Youth Olympic Games – Buenos Aires, ARG*. Retrieved from http://www.fig-docs.com/website/newsletters/WAG_NL_43_en.pdf

Sacchi, D. (2018b). *Women's Artistic Technical Committee Report: 2018 World Championships – Doha, QAT*. Retrieved from http://www.fig-docs.com/website/newsletters/WAG_NL_43_en.pdf

Saidin, K., & Yaacob, A. (2016). *Insider Researchers: Challenges & Opportunities.* Paper presented at the International Seminar on Generating Knowledge Through Research, Universiti Utara Malaysia, Malaysia.

Samson, A., Simpson, D., Kamphoff, C., & Langlier, A. (2015). Think aloud: An examination of distance runners' thought processes. *International Journal of Sport and Exercise Psychology, 15*(2), 176-189. doi:10.1080/1612197X.2015.1069877

Schooler, J. W. (2011). Introspecting in the Spirit of William James: Comment on Fox, Ericsson, and Best (2011). *Psychological Bulletin, 137*(2), 245-350. doi:10.1037/a0022390

Schweizer, G., Plessner, H., Kahlert, D., & Brand, R. (2011). A Video-Based Training Method for Improving Soccer Referees' Intuitive Decision-Making Skills. *Journal of Applied Sport Psychology, 23*(4), 429-442. doi:10.1080/10413200.2011.555346

Sengupta, D. D., & Paul, S. (2018). A review on vital changes since 1996 in the evaluation system of artistic gymnastics 'code of points' *International Journal of Physical Education, Sports and Health, 5*(1), 93-95.

Smith, B., & McGannon, K. R. (2017). Developing rigor in qualitative research: problems and opportunities within sport and exercise psychology. *International Review of Sport and Exercise Psychology*. doi:10.1080/1750984X.2017.1317357

Ste-Marie, D. M. (1999). Expert-novice differences in gymnastic judging: An information-processing perspective. *Applied Cognitive Psychology, 13*(3), 269-281. doi:Doi 10.1002/(Sici)1099-0720(199906)13:3<269::Aid-Acp567>3.0.Co;2-Y

Ste-Marie, D. M. (2000). Expertise in Women's Gymnastic Judging: An Observational Approach. *Perceptual and Motor Skills, 90*, 543-546.

Ste-Marie, D. M. (2003). Expertise in Sport Judges and Referees: Circumventing Information-Processing Limitations. In J. L. Starkes & K. A. Ericsson (Eds.), *Expert performance in sports : advances in research on sport expertise*. Champaign, IL ; Leeds: Human Kinetics.

Ste-Marie, D. M., & Lee, T. D. (1991). Prior Processing Effects on Gymnastic Judging. *Journal of Experimental Psychology: Learning. Memory, and Cognition, 17*(1), 126-136.

Ste-Marie, D. M., & Valiquette, S. M. (1996). Enduring Memory-Influenced Biases in Gymnastic Judging. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(6), 1498-1502. doi:10.1037/0278-7393.22.6.1498

Ste-Marie, D. M., Valiquette, S. M., & Taylor, G. (2001). Memory-Influenced Biases in Gymnastic Judging Occur across Different Prior Processing Conditions. *Research Quarterly for Exercise and Sport, 72*(4), 420-426. doi:10.1080/02701367.2001.10608979

Stephenson, J., Cronin, C. J., & Whitehead, A. E. (2019). 'Suspended above, and in action': Think Aloud as a reflective practice tool. *International Sport Coaching Journal*. Retrieved from http://researchonline.ljmu.ac.uk/id/eprint/10875

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature reviews Neuroscience, 15*, 745-756. doi:10.1038/nrn3838

Swettenham, L., Eubank, M., Won, D., & Whitehead, A. E. (2018). Investigating stress and coping during practice and competition in tennis using think aloud. *International Journal of Sport and Exercise Psychology*. doi:10.1080/1612197X.2018.1511622

Tenenbaum, G., Levy-Kolker, N., Sade, S., Liebermann, D. G., & Lidor, R. (1996). Anticipation and Confidence of Decisions Related to Skilled Performance. *International Journal of Sport Psychology, 27*(3), 293-307.

Thomas, K., & Thomas, J. R. (1994). Developing Expertise in Sport: The Relation of Knowledge and Performance. *International Journal of Sport Psychology, 25*, 295-312.

Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes. *Information Processing and Management, 31*(6), 906-907. doi:10.1016/0306-4573(95)90031-4

Varpio, L., Ajjawi, R., Monrouxe, L. V., O'Brien, B. C., & Rees, C. E. (2017). Shedding the cobra effect: problematising thematic emergence, triangulation, saturation and member checking. *Medical Education, 51*, 40-50. doi:10.1111/medu.13124

Welsh, J. C., Dewhurst, S. A., & Perry, J. L. (2018). Thinking Aloud: An exploration of cognitions in professional snooker. *Psychology of Sport & Exercise, 36*, 197-208. doi:10.1016/j.psychsport.2018.03.003

Whitehead, A. E. (2015). *The Use of Think Aloud Protocol to Investigate Golfers Decision Making Process.* (PhD). University of Central Lancashire, Retrieved from http://clok.uclan.ac.uk/12118/

Whitehead, A. E., Cropley, B., Huntley, T., Miles, A., Quayle, L., & Knowles, Z. (2016). 'Think Aloud': Toward a Framework To Facilitate Reflective Practice Amongst Rugby League Coaches. *International Sport Coaching Journal, 3*(3), 269-286. doi:10.1123/iscj.2016-0021

Whitehead, A. E., Jones, H. S., Williams, E. L., Dowling, C., Morley, D., A., T. J., & Polman, R. C. (2017). Changes in cognition over a 16.1 km cycling time trial using Think Aloud protocol: Preliminary evidence. *International Journal of Sport and Exercise Psychology, 17*(3), 266-274. doi:10.1080/1612197X.2017.1292302

Whitehead, A. E., Jones, H. S., Williams, E. L., Rowley, C., Quayle, L., Marchant, E., & Polman, R. C. (2018). Investigating the relationship between cognitions, pacing strategies and performance in 16.1 km cycling time trials using a think aloud protocol. *Psychology of Sport & Exercise, 34*, 95–109. doi:10.1016/j.psychsport.2017.10.001

Whitehead, A. E., Taylor, J. A., & Polman, R. C. (2015). Examination of the suitability of collecting in event cognitive processes using Think Aloud protocol in golf. *Frontiers in Psychology, 6*, 1083. doi:10.3389/fpsyg.2015.01083

Whitehead, A. E., Taylor, J. A., & Polman, R. C. J. (2016). Evidence for Skill Level Differences in the Thought Processes of Golfers During High and Low Pressure Situations. *Frontiers in Psychology, 6*, 1-12. doi:doi: 10.3389/fpsyg.2015.01974

Whyte IV, J., Cormier, E., & Pickett-Hauber, R. (2010). Cognitions associated with nurse performance: A comparison of concurrent and retrospective verbal reports of nurse performance in a simulated task environment. *International Journal of Nursing Studies, 47*, 446–451. doi:10.1016/j.ijnurstu.2009.09.001

Williams, A. M., & Elliott, D. (1999). Anxiety, Expertise, and Visual Search Strategy in Karate. *Journal of Sport & Exercise Psychology, 21*(4), 362-376. doi:10.1123/jsep.21.4.362

Williams, A. M., & Ericsson, K. A. (2005). Perceptual-cognitive expertise in sport: Some considerations when applying the expert performance approach. *Human Movement Science, 24*, 283-307. doi:10.1016/j.humov.2005.06.002

Williams, A. M., Fawver, B., & Hodges, N. J. (2017). Using the 'Expert Performance Approach' as a Framework for Improving Understanding of Expert Learning. *Frontline Learning Research, 5*(3), 139 - 154. doi:10.14786/flr.v5i3.267

Williams, A. M., & Ford, P. R. (2013). 'Game Intelligence': Anticipation and Decision Making. In A. M. Williams (Ed.), *Science and Soccer: Developing Elite Performers* (pp. 105-121). London: Routledge.

Williams, A. M., Ford, P. R., Eccles, D. W., & Ward, P. (2011). Perceptual-Cognitive Expertise in Sport and its Acquisition: Implications for Applied Cognitive Psychology. *Applied Cognitive Psychology, 25*, 432–442.

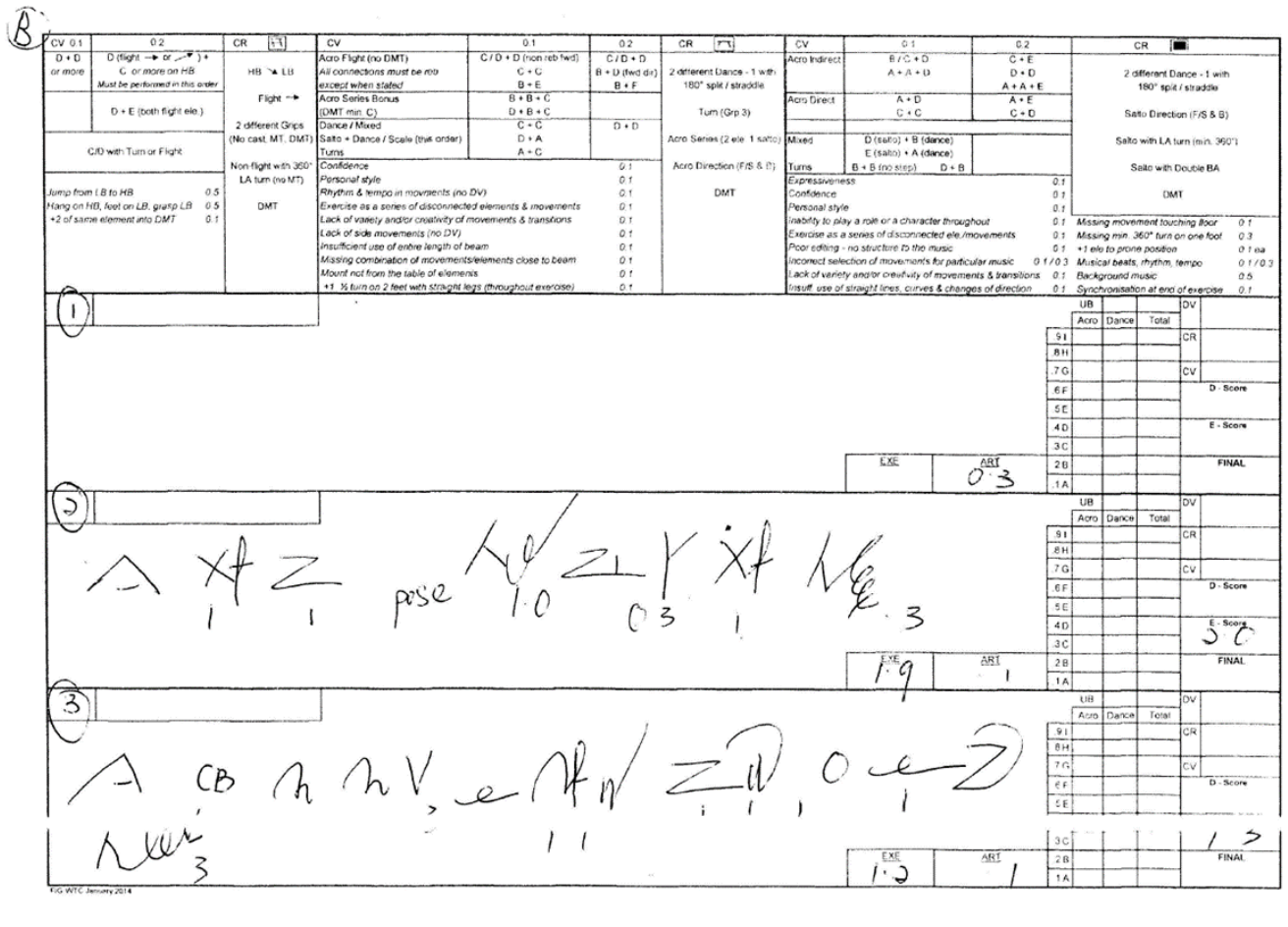
Williams, A. M., Ward, P., & Chapman, C. (2003). Training Perceptual Skill in Field Hockey: Is There Transfer from the Laboratory to the Field? *Research Quarterly for Exercise and Sport, 74*(1), 98-103. doi:10.1080/02701367.2003.10609068

Woods, A. J., Kranjec, A., Lehet, M., & Chatterjee, A. (2015). Expertise and decision-making in American football. *Frontiers in Psychology, 6*(994). doi:10.3389/fpsyg.2015.00994

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention Is Spontaneously Biased Toward Regularities. *Psychological Science*, 1-11. doi:10.1177/0956797612460407

APPENDICES

Appendix A: Judge's notation sheet (Study One)



Appendix B: Semi-structured Interview Guide (Study One)

Semi-Structured Interview Guide

Project Title*:*
Using a Think Aloud Protocol in Women's Artistic Gymnastics Judges: A Pilot Study

Introduction (3 minutes)

Thank you for taking part in this PhD project with Liverpool John Moores University for developing Women Artistic Gymnastics judges' education by using a Think Aloud Protocol.

For this stage of my project, I would like to understand your experiences judging a series of balance beam routines by applying the Think Aloud protocols.

There are no right or wrong answers for these interview questions. I am interested in your opinions as an individual, please use your own words to do this. I appreciate your comments as a qualified judge accredited by MGF (Malaysian Gymnastics Federation) or FIG (Federal International de Gymnastics). If you do not understand a question I have asked, please let me know so that I can repeat or rephrase it and you do not have to answer any question you do not wish to. You may answer in Malay language if you prefer.

There are 3 main research questions which will be explored during this interview and will take about 15 minutes. We may continue the discussion when the times up only if you would like to.

As in the Participant Information Sheet and Participant Consent form signed, this conversation will be recorded using a Dictaphone. This study is completely voluntary, therefore you do not have to take part if you do not wish to and you may withdraw at any time without having to give a reason.

Rationale	Research Question	Orientating Statement	Question	Prompt
Firstly I am going to ask your acceptability of a TA protocol to be applied in the WAG judging training.				
WAG judges require multi-tasking abilities to write down the movements into symbol form while watching the performance and analysing the movements comparing to the standards provided by COP (Ste-Marie, 2000). Thelwell & Greenlees (2001), Thelwell & Maynard (2003)	Explore judge's acceptability of a TA protocol within WAG judges.	I would like to first talk about your experiences in judging balance beam routine by using the TA protocol.	Can you please tell me your thoughts and feelings before judging? Can you please tell me your thoughts and feelings when judging? Can you please tell me your thoughts and feelings after your judging had finishes?	Did you enjoy TA? If so, why? How do you feel the differences before and after you TA on balance beam routines? How confident are you in TA? What challenges you had faced in the TA session? Any changes that happened to you after the TA session? Do you consider TA to be an acceptable means of assessing thoughts during judging?

(10 minutes)		I would like to know your opinion regarding TA protocols to be applied in the WAG judging education.	How you think the TA protocols will help a WAG judge to develop her judging capability?	How has the TA technique proved useful to you? Why do you think this is? Do you consider TA to be helpful?
Thank you for sharing your feelings and experiences. What I am interested next is your opinion on the relationship between WAG judges and gymnasts.				
Rationale	**Research Question**	**Orientating Statement**	**Question**	**Prompt**
(2-3 minutes)	Explore the relationship between judges and gymnasts to improve the performance.	I would like to know your view on how a judge could help to improve a gymnast's performance.	Would you be willing to share your TA scripts with a gymnast to improve her performance?	Can you tell me a little bit more what a judge can do to help a gymnast using TA? Have you ever as a judge helped a gymnast by using TA? How often do you take part in these activities?

Thanks again. Besides the role as a judge, are you a coach at the same time?
I am interested to know your view on the relationship between judges, coaches, and gymnasts.

Rationale	Research Question	Orientating Statement	Question	Prompt
(2-3 minutes)	Explore the relationship between judges and coaches to promote gymnasts performance achievement.	I would like to know your view on how judges can cooperate with coaches to improve a gymnast's performance.	Would you be willing to share your TA scripts with a coach to improve a gymnast's performance?	How important do you think a judge's TA scripts on a gymnast to be useful to a coach?

Closure:

This is the end of all my questions. Thank you very much for your time. I really enjoy the session with you and I appreciate all thoughts you had shared with me. All information or data (personal details/audio and video recording etc) collected throughout the study will remain strictly confidential. Data will remain anonymous and stored in locked systems (hard files) or in password-protected files (electronic files). Pseudonyms will be used in transcripts and written reports to protect the identity of individual and organisations.

Is there any questions you would like to ask or is there anything you feel I have not covered during this discussion which you feel is important?

If you would like any further information about the study, please feel free to ask me any questions or if you think of anything later you can contact me on my details given on the participant information sheet.

Thank you again for your participant.

Appendix C: Semi-structured Interview Guide (Study Three)

Semi-Structured Interview Guide

Project Title*:*
Investigation into decision-making of Women's Artistic Gymnastics judges using Think Aloud (TA) protocol

Introduction (3 minutes)

Thank you for taking part in this PhD project with Liverpool John Moores University to inform Women Artistic Gymnastics judges' education using a Think Aloud Protocol.

I would like to document your experiences in judging a series of balance beam and floor exercise routines by applying the Think Aloud protocol.

There is no right or wrong answer for these interview questions. I am interested in your opinion as an individual; please use your own words to do this. I appreciate your comments as a qualified judge accredited by BG (British Gymnastics) and/or FIG (Federal International de Gymnastics). If you do not understand a question I have asked, please let me know so that I can repeat or rephrase it and you do not have to answer any question you do not wish to.

There are 5 main research questions which will be explored during this interview and will take about 35 minutes. We may continue the discussion upon your discretion.

As stated in the Participant Information Sheet and Participant Consent form, this conversation shall be recorded using a Dictaphone. This study is completely voluntary. Therefore, you are entitled to decline your participation or withdraw at any stage of this study.

Rationale	Research Question	Orientating Statement	Question	Prompt
Your acceptability of a TA protocol applied in the WAG judging training.				
WAG judges require multi-tasking abilities to write down the movements into symbol form while watching the performance and analysing the movements comparing to the standards provided by COP (Ste-Marie, 2000). Thelwell & Greenlees (2001), Thelwell & Maynard (2003) **(10 minutes)**	Explore judge's acceptability towards a TA protocol into WAG judging education.	I would like to discuss about your experiences in judging balance beam and floor exercise routines using the TA protocol.	Could you please tell me your thoughts and feelings **before, during and after** judging by Think Aloud? Did it improve over time?	Did you enjoy TA? If so, why? How do you feel about the differences before and after you TA on those routines? What challenges you had faced in the TA session? How confident are you to TA before and after?
		There are some hold elements in balance beam required gymnast to hold an element for at least 2 seconds.	Could you please tell me how do you count for **2 seconds**?	Is there any specific methods you had used in counting 2 seconds for hold elements?
		I would like to know your opinions regarding TA protocols application into the WAG judging education.	Do you think TA protocols could help a WAG judge to improve his/her judging capability?	How has the TA technique proved useful to you? Why do you think this is applicable? How do you consider TA to be helpful as a WAG judge?
Thank you for sharing your thoughts and experiences. Besides Think Aloud on balance beam and floor exercise routines, I would like you to try Think Aloud on vault and uneven bars.				

Next, I am going to show you 2 different vault routines. Please Think Aloud (say aloud) the deductions you would apply if you were an E-panel judge.				
Play 2 Vault routines.				
Then, I am going to show you 1 uneven bar routines. Please Think Aloud (say aloud) the deductions you would apply if you were an E-panel judge.				
Play 1 Uneven Bars routine.				
Rationale	**Research Question**	**Orientating Statement**	**Question**	**Prompt**
(10 minutes)	Explore the differences to TA on slow and fast moving apparatus in WAG.	I would like to know if there is any difference to TA on balance beam, floor exercise, uneven bars, and vaulting routines.	Could you please tell me your thoughts and feelings when judging vault and uneven bars routines? How did you feel when judging different apparatus across balance beam, floor exercise, vault and uneven bars?	Any challenges you had faced when think aloud on vault and uneven bars? If so, what are they? Any difference when you TA on different apparatus? What challenge/difficulty to TA on certain apparatus? How did you overcome those challenges?
Thank you for trying Think Aloud on all WAG apparatus routines. Next, I would like to listen your opinion as a qualified WAG judge if applying Think Aloud protocol in the current WAG judging education.				

Rationale	Research Question	Orientating Statement	Question	Prompt
	Explore judges' acceptability towards feasibility of applying TA into current WAG judging education.	I would like to know your opinion if it is feasible to apply Think Aloud elements into current WAG judging education.	Could you please tell if you think it is appropriate for judges to think aloud during judge's course training?	What is your opinion if facilitator/instructor TA on routines? What is your opinion if judges who attending judges course for accreditation/re-accreditation to think aloud on routines?
			Do you think it would be any difference between expert and novice judges to think aloud on routines?	What do you think the difference would be between expert and novice judge?
(5 minutes)	Explore differences in judging during accreditation course and actual competition.	I would like to know how you think about the differences in judging during accreditation course and actual competition venue.	Could you please tell me if there is any difference judging video clips and actual competition?	How do you think about the crowd noise during competition will affect your judging?
Thank you for your thoughts. I am appreciative of your opinion as an experienced judge in UK. Next, I am interested in your opinion on the relationship between WAG judges and gymnasts.				

Rationale	Research Question	Orientating Statement	Question	Prompt
(5 minutes)	Explore the relationship between judges and gymnasts to improve the performance.	I would like to know your view on how a judge could help to improve a gymnast's performance.	How do you provide feedback to a gymnast in any form before this on her routine? How frequent you do these? What is your opinion if asking you to TA on a gymnast routine then share your audio feedback over the routine footage with the gymnast to improve her performance?	Could you please tell me a bit more what a judge can do to help a gymnast using TA? Have you ever as a judge helped a gymnast by using TA or any similar way?
This is the last question from me before ending this interview. Besides the role as a judge, you are [also a coach at the same time]/[not involve in coaching]? (Depending on questionnaire answer) I am interested to know your view on the relationship between judges, coaches, and gymnasts.				

Rationale	Research Question	Orientating Statement	Question	Prompt
(5 minutes)	Explore the relationship between judges and coaches to promote gymnasts performance achievement.	I would like to know your view on how judges can cooperate with coaches to improve a gymnast's performance.	Would you be willing to share your audio feedback over the routine footage with a coach to improve a gymnast's performance? Would you consider doing this 'live' with a coach alongside you?	How useful do you think a judge's TA feedback may be to to a coach?

Closure:

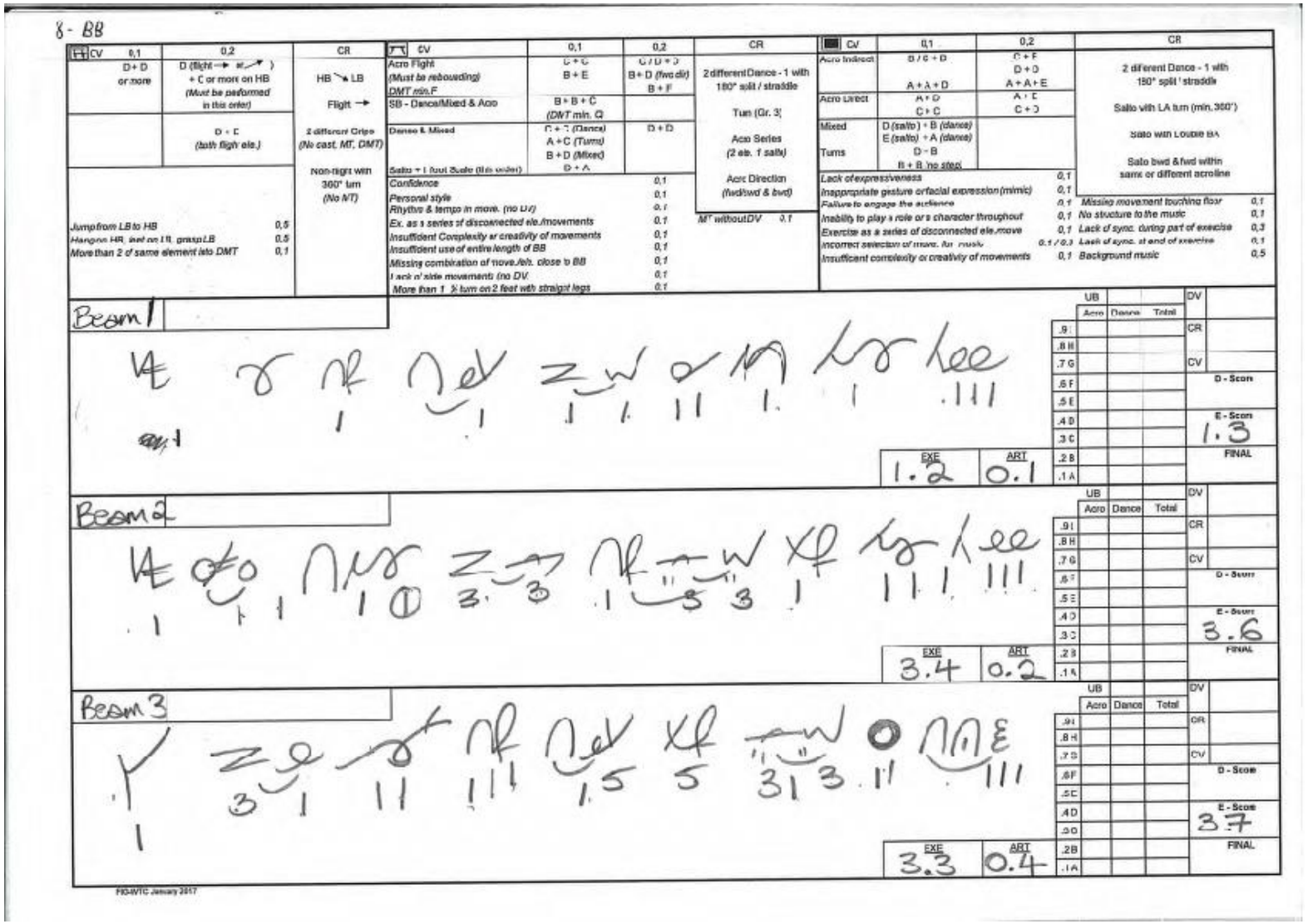
This is the end of all my questions. Thank you very much for your time. I really enjoy the session with you and I appreciate all of your thoughts. All information or data (personal details/audio recording) collected throughout the study will remain strictly confidential. Data will remain anonymous and stored in locked systems (hard files) or in password-protected files (electronic files). Pseudonyms will be used in transcripts and written reports to protect the personal and organisation identity.

Is there any questions you would like to ask or is there anything you feel I have not covered during this discussion, which you feel is important?

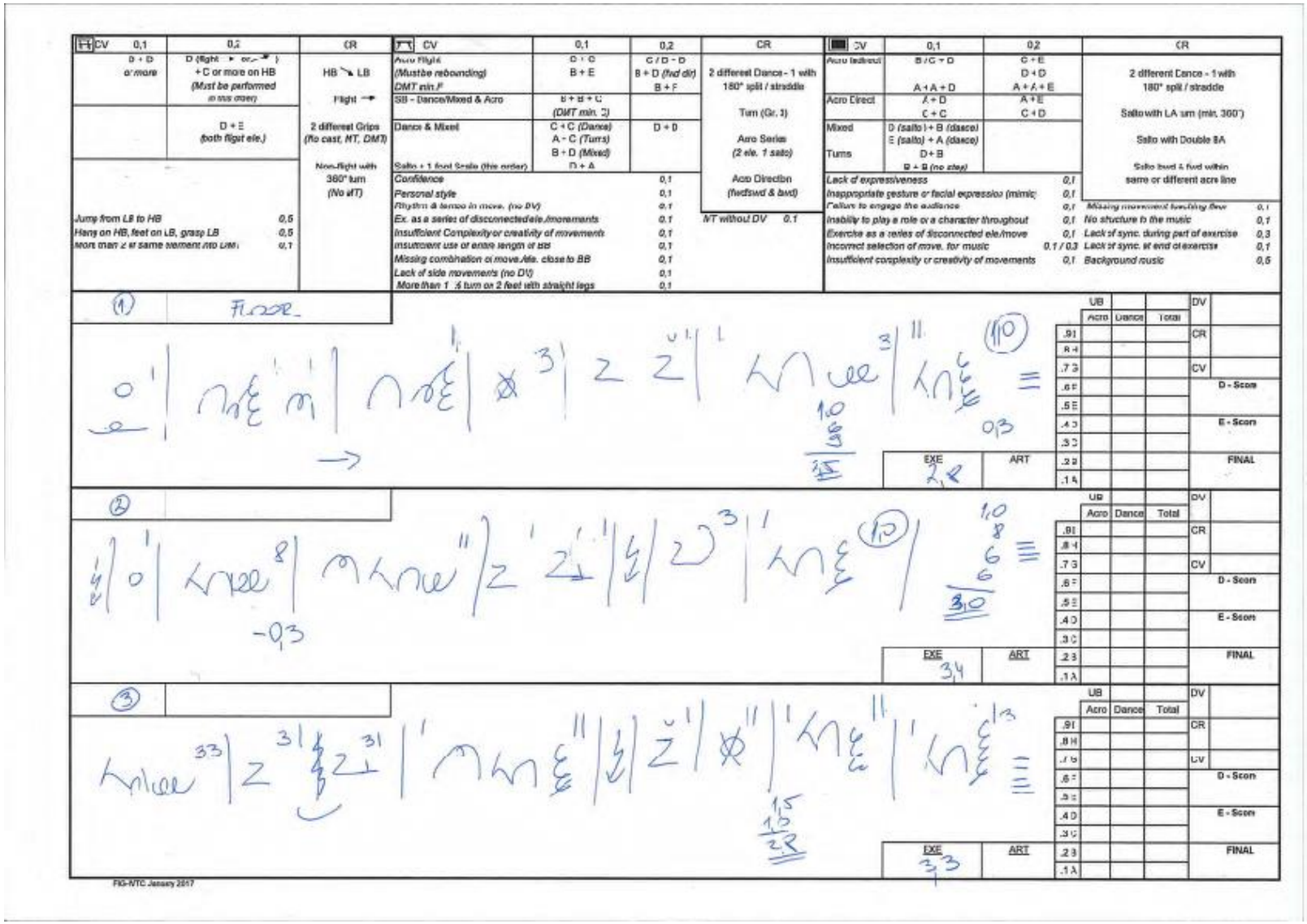
If you would like any further information about the study, please feel free to ask me any question or if you have think of anything, you can contact me in due course.

Thank you again for your participation.

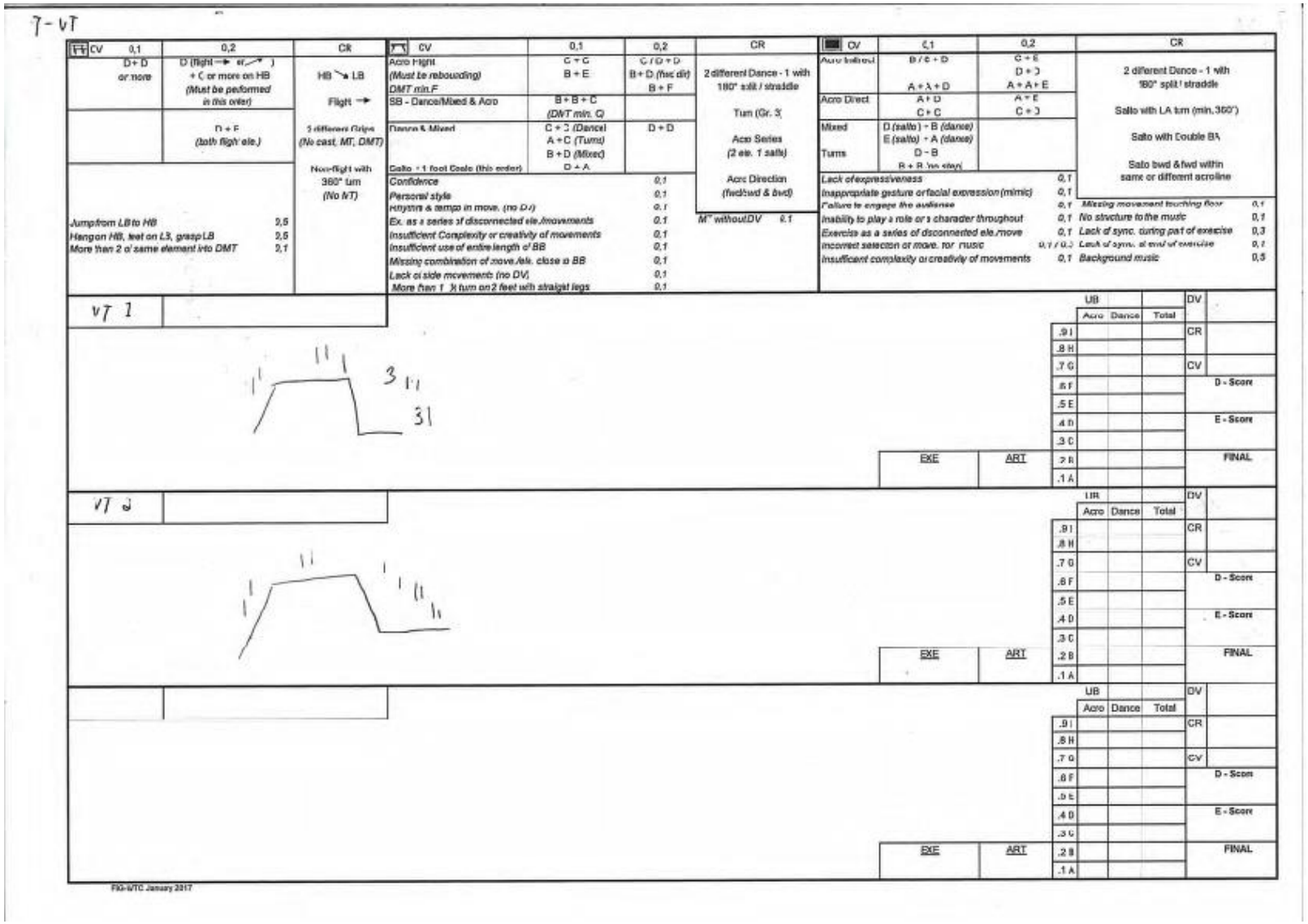
Appendix D: Expert judge's notation sheet for Balance Beam (Study Two and Three)



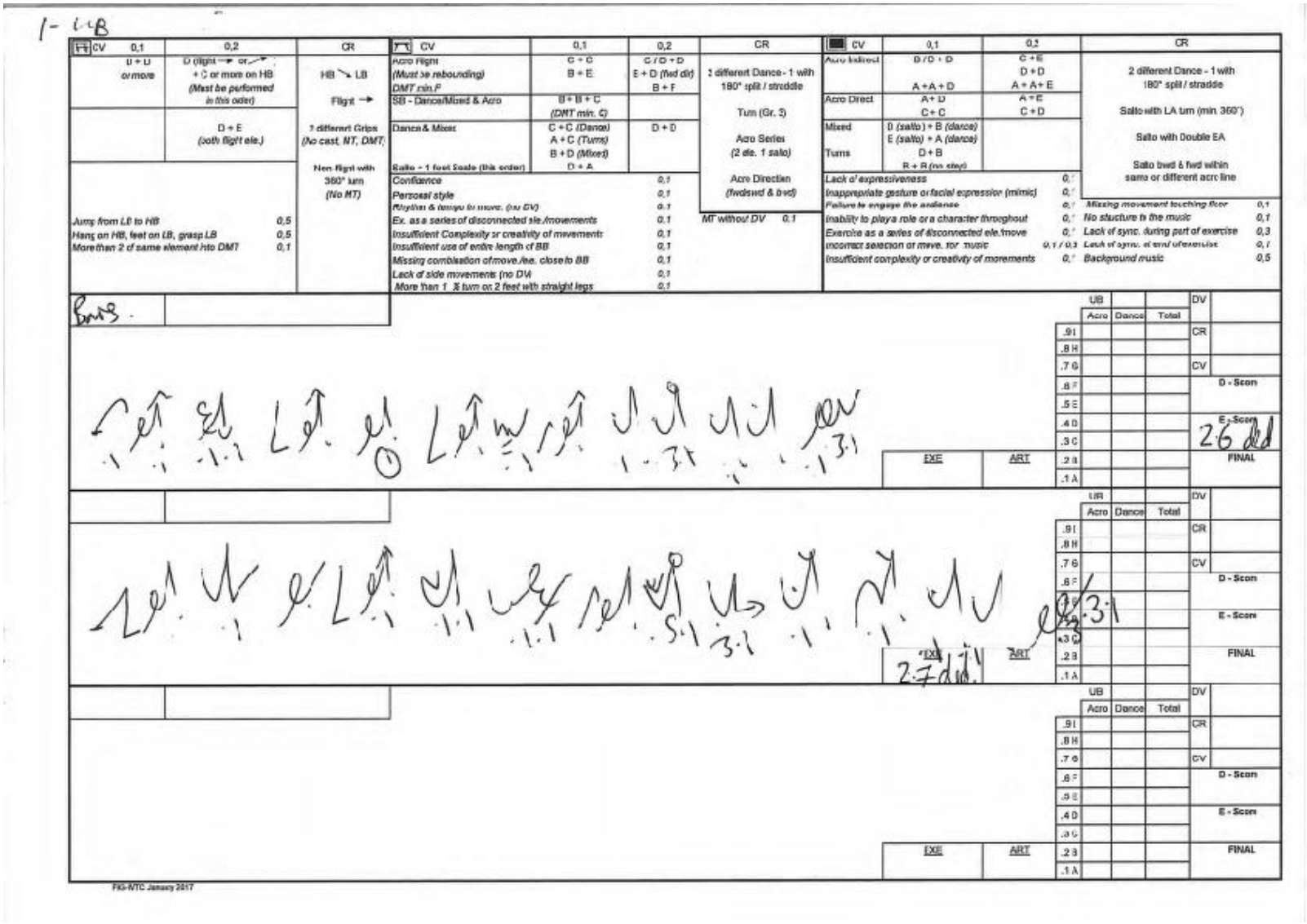
Appendix E: Expert judge's notation sheet for Floor Exercise (Study Two and Three)



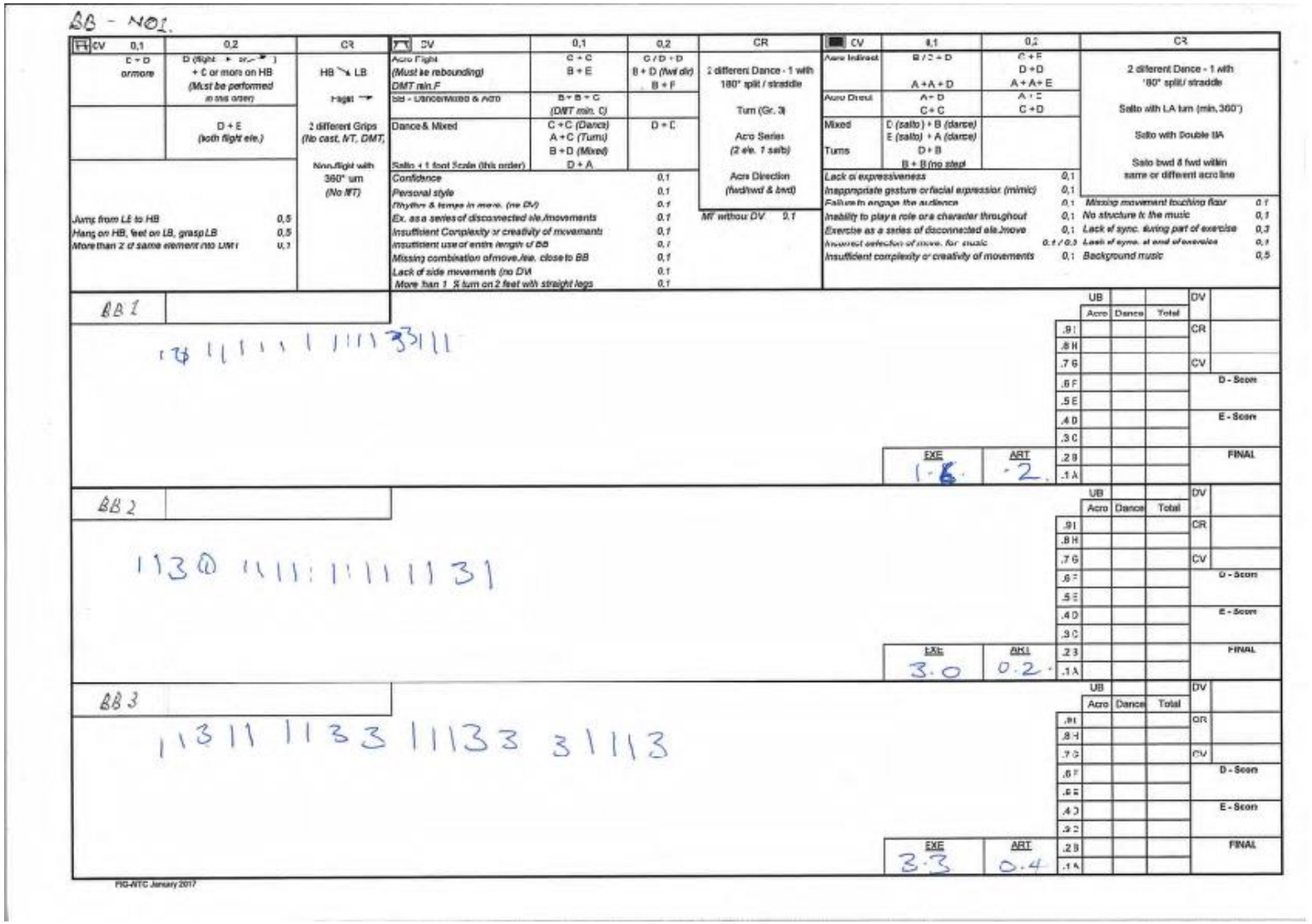
Appendix F: Expert judge's notation sheet for Vault (Study Two and Three)



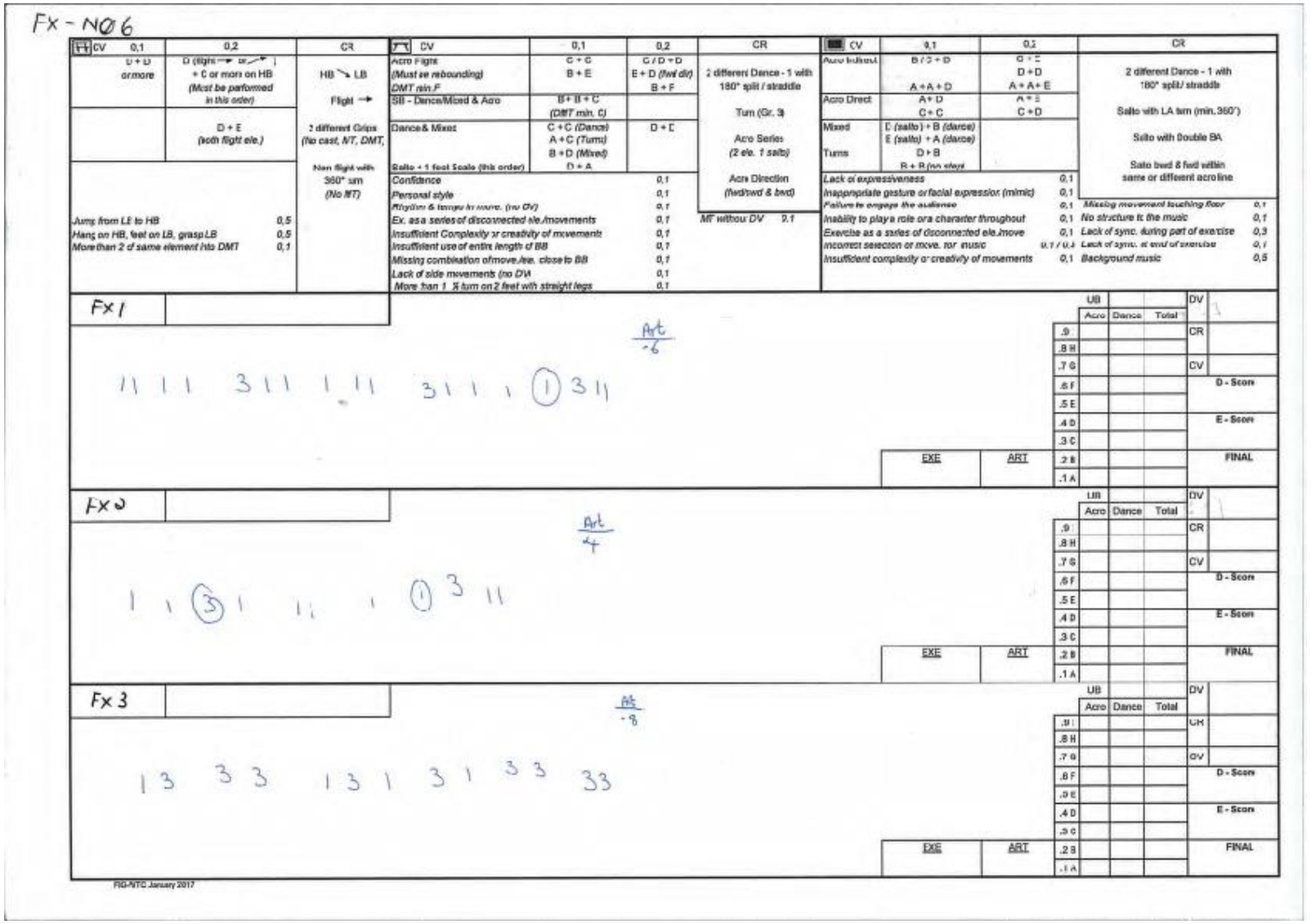
Appendix G: Expert judge's notation sheet for Uneven Bars (Study Two and Three)



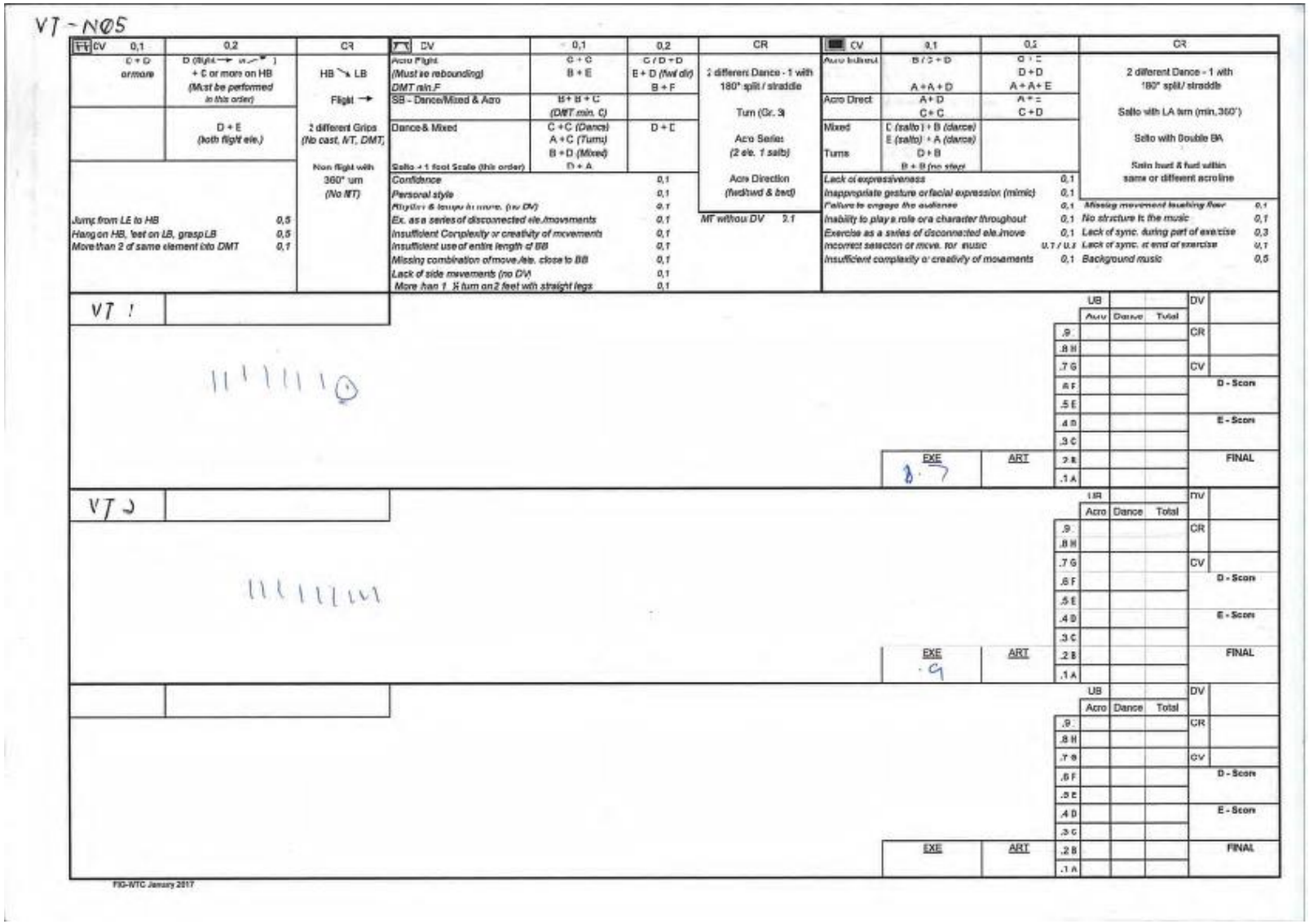
Appendix H: Novice judge's notation sheet for Balance Beam (Study Two and Three)



Appendix I: Novice judge's notation sheet for Floor Exercise (Study Two and Three)



Appendix J: Novice judge's notation sheet for Vault (Study Two and Three)



Appendix K: Novice judge's notation sheet for Uneven Bars (Study Two and Three)