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

Abstract

Gymnastics is a judgement-based sport whereby the decision-making processes of judges are expected to lead to valid and reliable outcome scores. The concurrent Think Aloud method has been used to study decision-making amongst coaches and athletes in previous sport-related studies but never in judging-based studies. Hence, this project has two aims: 1) to explore decision-making underpinning the judging process in Women's Artistic Gymnastic (WAG) by using a concurrent verbal report, Think Aloud (TA) and: 2) to examine the utilisation of TA as a means to facilitate judging education with Malaysian WAG judges. 10 qualified judges were required to verbalise (TA) their thought processes whilst judging a balance beam routine. Follow-up interviews investigated the prospective utilisation of TA within judging education. During the judging process participants verbally reported most frequently as to lack of balance, bending of arms and knees, pointing of feet, confidence, rhythm and tempo, and personal style as focal points for scoring. Overall TA was reported prospectively as an appropriate tool for use within judging education, however, some participants reported performance in the primary task of judging was affected by TA. Study outcomes reported the potential utility of TA to study the decision-making process amongst judges to enable deduction scores to be applied objectively. This study will inform future research to investigate the decision-making processes of both expert and novice judging extending to that of all four WAG apparatus.

Key words: Gymnastics, Think Aloud, Judging, Education.

Introduction

Women's Artistic Gymnastics (WAG) judges are expected to evaluate four apparatus of balance beam, floor, vault and uneven bars. They are required to evaluate these apparatus accurately, consistently, quickly, objectively and fairly and understand the intent, purpose, interpretation, and application of each rule for the current cycle. Within each current cycle, a set of rules and regulations governing WAG, the Code of Points (COP), is revised, updated, and approved every four years after an Olympic Games by the Fédération Internationale de Gymnastique (FIG) (Fédération Internationale de Gymnastique, 2016b).

WAG is a sport event combining a series of acrobatic movements with artistry. Gymnasts are expected to perform their routines with 1) maximum elements allowed for a particular apparatus within permitted time, 2) highest element difficulty with connections, 3) minimal execution deductions. Both execution and artistry scores are applied when judging. Execution deductions are applied when there is deviation from the required standards within the COP by E-panel judges. An E panel has four to six judges according to competition requirements. However, each E-panel judge is responsible for their own judgment and discussion is not permitted. The sum of the E-score awarded to a gymnast is the average score provided by E-panel judges excluding the highest and lowest deductions to reduce the 'halo' effect (McFee, 2013). All judges are required to watch the gymnast's performance whilst recording movements as symbols on notation sheets. Extra notes with symbols are also marked on the notation sheets such as a fall from an apparatus alongside execution deduction score markings (see appendix A). Therefore, WAG judges are required to have multi-tasking abilities in order to record the movements in symbol form, whilst also watching the performance and analysing the movements and comparing to the standards provided by COP (Ste-Marie, 2000). Further artistry scores are applied to apparatus such as the Balance Beam. The Balance Beam is an artistic performance whereby the gymnast must demonstrate

creativity, confidence of performance, personal style and perfect technique i.e. not "what" the gymnast performs, but "how" she performs. Composition is based on the movement vocabulary, both gymnastic and artistic, of the gymnast, as well as the choreography of these elements in relationship to the Balance Beam, while establishing a strong sense of rhythm and modulation of pace. Routines must show balance of elements of difficulty with artistic components in order to create a continuous flow, a cohesive whole; rhythm and tempo (speed/pace) must be varied, sometimes lively, sometimes slow. However, the routines must be predominately dynamic and above all uninterrupted and movement transition should be smooth and fluent, without unnecessary stops or prolonged preparatory movements before elements. Creative choreography is the originality of the composition of elements and movements. This means that the exercise has been constructed and is performed using new ideas, forms, interpretations and originality, thereby avoiding monotony (Fédération Internationale de Gymnastique (2012). The score calculations of a gymnast for an apparatus will be completed after the routine has finished and typically within 60 seconds.

Training and subsequent examination to reaccredit a WAG judge occurs every 4 years and is aligned to the Olympic cycle. Here, judges are updated with the latest rules and regulations, to ensure integrity of decisions, competency to apply the COP and FIG rules, and demonstrate moral and ethical behaviour (Fédération Internationale de Gymnastique, 2016a). Research has identified seven common biases in gymnastics judging including patriotism bias, halo bias, memory-influenced bias, reputation bias, order bias, bias emerging from social comparison processes, and conformity bias (Boen, van Hoye, Auweele, Feys, & Smits, 2008; Leskošek, Čuk, Pajek, Forbes, & Bučar-Pajek, 2012). Further Pajek, Kovač, Pajek, and Leskošek (2014) reported poor inter-rater reliability and substandard validity in their study based on 194 gymnasts in the World Championship in Tokyo 2011 and subsequently suggested further research to improve the reliability and consistency of judging. However, to

date, the mechanisms by which score reliability and consistency could be explored are yet to emerge. It is therefore important to explore decision-making process of judges to inform a training method appropriate to develop and assess score consistency.

Decision-making is defined as the ability to use information from the current situation and associated knowledge possessed so as to plan, select, and execute an appropriate goal-directed action or set of actions (Williams & Ford, 2013). MacMahon and Mildenhall (2012) highlighted the challenge that a sport official faces, given that they have to possess perceptual-cognitive skills for processing incomplete, intentionally deceptive, and fast-paced information under time pressure during a competition. WAG judges are required to judge a series of fast-paced gymnastic movements whilst also under time restriction. There is presumed sufficient information and processing time for judging a slower moving apparatus, such as balance beam, hence fewer 'gaps' to significantly impact on judgement. In contrast, fast moving apparatuses, such as vaulting and uneven bars, may be influenced by the within-event context of previous decisions, time, and score. Within current judging education E-panel judges are briefed on general deductions applicable for each apparatus followed by specific deductions for each apparatus, moving from theory into practical judging using competition videos. Therefore, trainee judges engage in video simulated training prior to the experience of in competition judging.

Verbal reporting has been previously used in other forms of education and training, specifically that of Think Aloud (a form of verbal reporting), within fields such as nursing education (McRobert, Mercer, Raw, Goulding, & Williams, 2017), self-regulated reading (Hua & Gao, 2017), and motor-learning in rehabilitation (Kleynen, Moser, Haarsma, Beurskens, & Braun, 2017). Think Aloud (TA) requires continuous verbalisation of thoughts during the performance of a task (Ericsson & Simon, 1983). Ericsson and Simon identified three levels, whereby 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 therefore they needed to be transformed before being verbalised. Level 3 verbalisation involves further explanations of thoughts, ideas, hypotheses, or motives and hence requires additional cognitive processing beyond that of verbalisation (Boren & Ramey, 2000; Whitehead, Taylor & Polman, 2016) and therefore may alter concurrent and retrospective processes. Level 3 has been used to elicit further detail regarding participants decisions in elite snooker players and therefore explanations were collected within this research (e.g. Welsh et al., 2018). As snooker is a self-paced activity, Welsh et al, (2018) were able to determine that snooker players freely verbalised and explained their thoughts ideas and actions in their own environment. Using Level 3 TA, Welsh et al, (2018) were able to demonstrate how stress and coping is a transactional process.

Within sport ,TA has been used to investigate real-time thought processes of runners during a long-distance run (Samson, Simpson, Kamphoff, & Langlier, 2015), cognitive focus in Cyclists (Whitehead et al., 2018), golfers decision-making processes (Whitehead, Taylor, & Polman, 2015), decision-making and thought processing among poker players with varying skill-levels (St. Germain & Tenenbaum, 2011), skilled perception processes and skilled problem solving in chess (Gobet & Charness, 2006) and expert performance in scrabble (Tuffiash, Roring, & Ericsson, 2007) to explore decision-making processes. Furthermore, TA has been used to aid self-awareness and reflection-in-action with coaching practice (Whitehead, Cropley, Huntley, Miles, Quayle & Knowles, 2016).

TA, has however, received some criticism based on its reliability for participants to verbalise accurate thought processes. More specifically, participants may report additional descriptions or explanations that are not part of their actual thought process at the current time of TA (Eccles, 2012) and verbal overshadowing (Chin & Schooler, 2008; Ericsson,

2003; Meissner & Brigham, 2001; Schooler, 2011), whereby during TA, individuals may be distracted to perform the primary task. Therefore, when considering TA by gymnastics judges it may limit the extent to which data representative of decision-making can be collected. Indeed TA may inhibit the ability for non-conscious processing to emerge into consciousness. However, Whyte IV, Cormier, and Pickett-Hauber (2010) revealed concurrent verbal reports provided a more complete record of cognition when compared to retrospective verbal reports in a simulated task environment. Eccles (2012) highlighted the strengths of using both concurrent and immediate retrospective TA, which allowed complementary comprehensive analysis of decision-making in cognitive roots, and overcame that of incomplete verbal reports. Participants were advised explicitly that verbalisation of thoughts should be secondary to performance of the main task (Ericsson & Simon, 1993). Therefore, whilst TA might be inappropriate during actual competition, whereby attentional focus may be affected, it may be more appropriate to use within a simulated situation.

The purpose of this study was to explore the use of TA as a training method with Malaysian based WAG judges, through examining the content of the judge decision-making using TA and gathering WAG judge perceptions of using TA when judging a video simulated single apparatus (balance beam).

Method

Researcher position

In order to explore the subjective experiences of using TA, a qualitative approach was employed and underpinned by a relativist ontology and a constructivist epistemology. Consequently, 'knowledge' considered as socially constructed involves a double hermeneutic whereby the subjectivities of the researcher interpret participants sense making in relation to using TA. Hence, this qualitative position has the potential to identify not only depth and

breadth of knowledge than traditional quantitative approaches reduce to simplistic representations but also new knowledge that maybe tacit in nature.

To explore the depth of participants' experiences, requires the researcher to make explicit the 'biases' (e.g. values, beliefs and experiences) that are inherent within the decision-making processes throughout the study. As such, the first author was an accredited Women's Artistic Gymnastics (WAG) in Malaysia (7 years of experience) and hence had an intimate knowledge of the Malaysian gymnastic sporting context. Thus, the adopted insider epistemology, together with the researcher's values and beliefs in relation to improving the quality of training for WAG judges positively influenced the design of the research and help build rapport with the participants. As the first author has prior knowledge and experience of WAG judging, both a deductive and inductive approach was used. Where a deductive approach was used (testing of theories and hypothesis), this was based on the first authors prior knowledge and rules governing the judging decision making process. In addition, where an inductive approach was used, this allowed new ideas to be identified.

Participants

The participants were 10 female Malaysian Women Artistic Gymnastics judges with international (n=4) and national (n=6) accreditation for Cycle 13th (2012-2016). Participants had a range of years of experience, from 1-9 years (M = 6.60, SD = 2.31). Participants were recruited through email forwarded by the gatekeeper (Malaysia Gymnastics Federation) with participant information sheet and a demonstration clip attached. This was followed by convenience sampling and word-of-mouth methods used to recruit the judges who were attending the 18th Malaysian Games in Sarawak, Malaysia. University Ethics Committee approval was gained and within the consent of the Malaysian Gymnastics Federation (MGF) as gatekeeper.

Data Collection

To ensure quality of the video clips, both demonstration video clip and TA on balance beam routine video clip were sent to three non-potential participants as a pilot study for social validation. Participants confirmed that the video clips were sufficient and provided them with an overall understanding of how to use TA when observing the balance beam routines. They had also responded both entry and exit countdown timer were helpful to prepare for TA and to sum the total deductions.

General 'Think Aloud' (TA) training. Prior to data collection all participants were sent, via email, the video clips of general TA training, TA demonstration (https://www.youtube.com/watch?v=t1-UobhvxoM&t=1s), and TA on Balance Beam. A video clip previously used by McRobert, Williams, Ward, and Eccles (2009) was adapted and used to train participants to use TA specifically Level 2 TA. The training video required the participant to say out loud what is the next alphabet after "A" and calculating how many dots appeared on screen followed by further gymnastics specific examples, which involved balance beam clips. The clips asked participants to solve generic tasks and they were provided with the instructions to verbalise using Level 2. Therefore participants were instructed to 'please Think Aloud anything that comes to your mind, but do not try to explain this', (Ericsson & Simon, 1993). Level 2 TA was employed rather than Level 3, as previous research has established that instructing participant to verbalise their thoughts using Level 2 does not alter performance, whereas directing participants to provide explanations for their thoughts (Level 3) may alter performance (Fox, Ericsson and Best, 2011). Asking participants to explain their thoughts whilst judging during a fast-paced activity may disrupt the judge's ability to provide reliable or 'real life' verbalisations.

Pre data collection training. During face-to-face data collection, study participants were re-oriented with the TA process whereby the general training exercises were replayed (McRobert et al., 2009; Ericsson & Simon, 1993). Participants were oriented with the TA video format including the entry and exit countdown timer provided with the latter post routine to complete the E-score calculation within the 60-second time allocated. Participants were briefed to be ready for prompts by the researcher such as "please think aloud" to verbalise all execution deductions and artistry deductions whilst judging balance beam routines and in particular if they remained silence for more than 10 seconds. Van Someren, Barnard and Sanberg (1994) recommend that the training task is similar to the target task, or as they state "in general it is wise to look for a task which is not too different from the target task" (p. 43). Therefore, participants were given the opportunity to practice TA on practice videos provided previously. Sony Dictaphones (model ICD-PX240) were used to record all audio responses verbalised by participants during the TA sessions and interviews. Olympus AS-2400 transcription kit was use to process verbatim transcription.

Data collection of TA on balance beam. Participants were instructed to TA and verbalise their thoughts that were relevant to all execution deductions applied onto each element performed on balance beam routines played in a 26 minute TA balance beam video montage (https://www.youtube.com/watch?v=nzWgjxmC4RQ). This footage comprised of 10 balance beam routines from publicly available sources with gymnasts from several nations globally and across several competitions and was created using Window Movie Maker and uploaded to a privately accessed YouTube account created by the researcher. A balance beam routine is set as less than 90 seconds as coded in COP (Fédération Internationale de Gymnastique, 2015) while a 5-second entry countdown timer was added on screen before the routine began to serve as preparation time with a green light flashing during actual competition. A 60-second exit countdown timer was added at the end of each routine for

participants to calculate execution scores and to TA on artistry deductions. All routine video clips were muted to exclude background noise whilst the footage angle was set from an angle akin to the judges perspective during actual competition. Participants were instructed to use Level 2 verbalisation whilst writing down usual notations and/or symbols on the judging sheets provided as in the COP (see appendix 1 for an example). Participants were prompted by researcher to TA 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 balance beam routine, respondents were prompted to TA on artistry deductions if they remained silence for 10 seconds after they had completed calculating the execution deduction scores and were waiting for next routine. Verbalisation during the TA session was recorded whilst all written judging notation sheets were collected at the end of session (see appendix 1 for an example).

Interview questions. Immediately following the completion of TA on balance beam session, participants (n=10) took part in an interview exploring the use of TA into current WAG judging education. A post TA data collection semi-structured interview was developed to gain an understanding of participants' individual experiences of engaging in the TA process. The interview (available on request) consisted of questions aimed to explore participant reflection on phases before, during, and after the TA data collection session and assessed the appropriateness to adapt TA into current WAG judging education. In addition, participants were asked to comment on the potential for TA outcomes and how it may inform coaches and gymnasts understandings of judging process/decision-making.

Data Analysis

A total of 227 minutes of TA audio clips were collected and were transcribed verbatim to make up a total of 38 pages of font Arial size 12 with double line spacing text. All transcripts underwent translation checks from the multiple languages used, including

Malay, Chinese, and Cantonese to English. Data was analysed using both deductive and inductive approaches. Firstly, the first author's knowledge and experience of gymnastics judging together with the judging Code of Points where employed deductively to create a list of commonly used 'judging terms'. A content analysis approach aligned to the 'judging terms' was subsequently used to identify the number of matching terms expressed by the judges using TA. Exploring the experiences of judges using TA beyond the deductive framework also allowed inductive themes to be identified and in doing so presented gymnastics judging as a socially constructed reality. In order to make sense of this reality the first author – who is immersed in the field of gymnastics judging – identified themes that were consistent across the participant data. Following this, authors 2 and 3 acted as critical friends in order to provide an 'opportunity for dialogue and the reflexive acknowledgement of multiple truths, perspectives and results in the research process (Smith & McGannon, 2017, p17). The combination of these approaches facilitated a process of 'meaning making' between the judges shared cognitive expressions relating to decision making and the researchers interpretations of these meanings (Lofland & Lofland, 1996). In doing so we acknowledge that whilst each participant judges according to personal interpretations informed by experience, this approach to data analysis was done not to promote individual differences, but to highlight the shared meaning across the group.

Interviews lasted between 13-17 minutes and provided a total of 73 pages of font Arial size 12 with double line spacing text were transcribed for the follow-up interviews. Both a deductive and inductive approach was taken when analysing the interview data (Scanlan, Stein & Ravizza, 1989). The first author's previous knowledge and experience of being a WAG judge and implementing TA within this study was used to analyse the data from a deductive perspective. Given that, this study is the first to consider participants perceptions of TA during gymnastics judging, inductive reasoning was also employed with a

view of allowing themes to be generated from the raw data, through a process of thematic analysis (Braun and Clarke, 2006). Thematic analysis offered a "theoretically flexible approach" (Braun & Clarke, 2014, p.1), and involved the following stages; 1) familiarising ourselves with the data, reading and re-reading transcripts and noting initial themes, 2) generating initial codes and collecting data relevant to each code, 3) searching for themes by collating codes into potential themes, 4) reviewing the themes and 5) defining and naming themes, where clear definitions and names for themes were generated. To ensure for rigour, a double hermeneutic was undertaken, where the researchers tried to make sense of the participants own sense making, regarding their experiences of using TA. As with the TA data, a critical friend was used in the same manner (Smith & McGannon, 2017).

Results

Content of TA Verbalisations

Table 1 shows the thematic structure of major deductions on both general execution and artistry focused by judges during the TA session. Data revealed that all judges were able to take note of the major deductions such as a "fall" which penalised the gymnast a whole point (1.0) deduction. All participants also focused on "insufficient height of elements" executed by gymnasts with a total of 144 quotes were found across the study. These were followed by 126 quotes of "lack of balance", which also known colloquially as "wobble", and 76 quotes of "bend arms or bend knees" quotes verbalised by 90% of the participants. A total of 64 quotes of "relaxed feet" or "feet not pointed" were verbalised by participants (n=8) while 60 quotes of "confidence" been mentioned by 90% of the participants showing judges were concerned with the artistry executed by gymnast's despite of general execution deductions. 90% of participant's verbalisations linked to a gymnast's rhythm and tempo in

movement while 80% of verbalisations link to a gymnast's personal style whilst performing the routine showed that artistry deductions were of highly priority.

Table 1. Themes verbalised during execution deductions during the TA session (P refers to participant number)

Theme	Sub-Theme	Raw Data Extracts
General Deduction	Fall (*n*=10)	"fall down, deduct 1.0 point" (P1)
	Insufficient height of elements (*n*=10)	" split leg back leg not high enough, .1" (P4)
	Bend arms/knees (*n*=9)	"front pike with the bend knees, but she managed to catch up" (P10)
	Turn (*n*=9)	"a bit slack, short of the turn, she didn't complete the full turn" (P5)
	Landing (*n*=9)	"deep squat landing...let me give her a maximum .5 towards landing" (P5)
	Leg/knee separation (*n*=9)	"double twist, legs apart" (P6)
	Pause (*n*=9)	"she pause again before she do a skill" (P10)
	Wobble/lack of balance (*n*=9)	" ouu wow, big wobble, .3 deducted..." (P2)
	Extra steps (*n*=8)	"round-off two and a half... with a large step, so .3..." (P7)
	Feet not pointed/relaxed (*n*=8)	"...combine with switch leg side aerial, didn't point toes, .1" (P7)
	Insufficient split in dance (*n*=7)	"there was not a good split leap, there was a deduction for both leg a little bit below horizontal" (P5)
Artistry Deduction	Confidence (*n*=9)	"overall I think she needs to boost up her confidence, especially for her dance elements, she pause like 3-5 seconds before she really did on the beam" (P10)
	Rhythm and tempo in movements (*n*=9)	"she had confidence, personal style, not really tempo and rhythm" (P5)
	Personal style (*n*=8)	"in terms of artistry, no confidence, no personal style…" (P8)
	Lack of side movements (*n*=6)	"lack of side movement, 0.1..." (P10)
	Missing combination of movements/ elements close to beam(*n*=3)	"missing combination (of movements/elements close to beam) 0.1" (P1)

Post TA Session Interview

Table 2 provides an overview of the main themes that emerged through participant interviews after using TA whilst judging balance beam videos. Within this, 60% of participants expressed positive perceptions of TA in applying TA within WAG judging education. Participants (40%) reported tangible benefits from TA, for example participant 2 noted: "it (TA) helps you to speak out what's inside your mind". Moreover, 40% of participants reported that the TA could assist WAG judging course instructors to access to thoughts/decision-making of novice judges in particular to correct errors such as invalid execution deductions. For example, participant 5 said course instructors could understand the *reason for deductions* through TA thus corrections could be made immediately while participant G noted: "I feel that this [TA] is very good for judging purposes, especially for training the new judges because a lot of the time they do not actually know *how* they arrive at the deductions".

Participants commented on the utility of TA within judging and suggested that by sharing thoughts/views between expert judges and novice judges as well as between novice judges themselves may help to improve application of correct execution deductions for a particular movement by way of appreciation of views from other judges. Participant 3 stated "it will be more useful if there is a pair or more than one person looking into the video and TA together so that all of the judges can share their thoughts on the gymnasts performance and from this, one can learn from each other on the deductions and also the execution." while participant 7 said "It's always easier if there is someone more qualified to sit with them [novice judges] because sometimes they do not understand where the deductions come from, that's why actually I think doing this [TA session] is very good to train new judges." However, participant 10 shared her experience in previous judging whereby novice judges who served in execution panel might be correct sometimes as compared to expert judges who

used to judge both difficulties and execution at the same time which might distract them: "We also need to reflect on ourselves [experienced judges] because we can't see because we are the experience judge that means we are always right, there could be a possibility that actually we have overlook certain things…".

Findings showed that 40% of participants expressed that they experienced restrictions to using TA whilst judging the balance beam routine video clips. Participant F noted "the mind is faster than the mouth…" while participant 7 said "we [judges] can't multi-task so much by talking and writing and recording whatever we need to do. If we are doing it all together, most of the time actually we might miss out one or two of the deductions". In addition, the participants noted that the TA using video clips could be further applied within 'live' training sessions to provide feedback to coaches and gymnasts in situ. Participant A, both a coach and a national judge, suggested gymnasts themselves via the judges TA data could understand faults such as 'wobble', 'lack of height' of elements and under rotation with turns better when they, as is typical, are simply shown the video replays recorded during training. Indeed, seven out of ten participants offered support to the use of TA whilst coaching and its potential influence therefore on coaching. Participant 7, an international judge and also a coach expressed that holding a 'dual role' as a coach and judge could influence decision-making process:

> '…As judges actually, we only focus on looking at how well they (gymnasts) can perform the skills, how they execute the skills and what are the deductions that we should actually…like… look at. But as a coach, they [coaches] are more into technique where they sometimes… you know they actually didn't look at how judges *judge* the routine. By working together, the gymnast actually if they work together with the judges, they can do better and they will score higher….' (Participant 7)

348

349 **Table 2. Perceptions of TA by WAG officials (P refers to participant number).**

Theme	Sub-Theme	Example Raw Data Extracts
Perception of TA use	TA in judge training programme	"If this [TA] is applied in the judging training courses I think it's quite good so that it can help judges to understand at which point… exactly at that point of time where [the deductions are] happening." (P6)
	Speak out	"I feel that I learnt that somehow we [judges] need to speak out loud more instead of you just keep in inside in your heart and then when you speak it out, you can share more your judging experience with others because maybe our judging experience and other people's judging experience is different." (P9)
	Accessing thoughts of novice judges	"For new judges I would think it's good for them to speak out about their own deductions and from there actually they will know whether what they are deducted is actually correct or it's actually they need maybe more training." (P7)
	Accessing thoughts of expert judges	"It [TA] will be more useful if there is a more experienced judge to TA together with you so you can know what did other judges (did) looked at the gymnasts, what their deductions on their elements." (P3)
	To correct judging errors	"A lot of time the new judges do not actually know what to deduct... or maybe it's a wrong deduction instead of a wobble 0.1 they might take a 0.3 or 0.5 (deduction) so that is a big difference if you actually take a wrong deduction." (P7)
	Training new judges	"I feel that this [TA] is very good for judging purposes, especially training the new judges because a lot of the time they do not actually know how to arrive with all these deductions." (P7)
	Helpful to gymnast	"When we record then show our gymnasts, then only they understand where is their mistake, … if you never show them the video they don't understand. Like for wobble, if you record down (and) show them then they only know… okay, okay… leg not high enough, never jumped, never turned properly… the details can (be) seen very clear from the video. Only (by) talk, they [gymnast]… sometimes they don't understand because they never see, they don't know." (P1)
Limitation of TA	Multi-tasking	"You want to speak out what you see then you want to write down, so you can't do a lot of things at one time." (P4)
Experiences in TA	Previous TA experience	"We [judges] didn't speak aloud like that… like we just go through …I mean like not whole routine, maybe like a certain skill only. We didn't like speak aloud like… play the whole routine" (P4)

Current TA experience	"It's a good experience to [TA] actually to talk and write and record down everything… I think it's interesting. We should try [TA] again… I think if we just talk about the deductions I think it's even better." (P5)

Discussion

The primary aim of the research was to design, implement, and examine the utility of TA for use within women's artistic gymnastic judging education. Following design and implementation, participants expressed their acceptance towards the use of TA in WAG judging education with advantages such as sharing of thoughts to apply correct execution deductions through TA verbalisation. Participants also expressed the potential to apply TA within judging education courses and training for thought access and to ensure the objectivity of judging scores.

The TA video clip compilation of non-stop 10 balance beam routines according to the interview data may have overburdened the non-international respondent. Previous TA research has acknowledged that there is no guidance as to the length of time that a participant should TA for (Eccles & Arsal, 2017). Nicholls and Polman (2008), did consider the length of TA within their study noting that using TA for a prolonged period of time may become challenging for the participant. There was no scheduled break during the TA session and whilst suitable for training within judging education courses/post course development, it was deemed unrealistic as that of a typical competition setting may consist of only four to six gymnasts in a rotation. Therefore, the length of TA session for a single apparatus should be shortened in further research to mimic that of a real competition. Therefore, five routines for balance beam and floor exercise respectively was suggested for the TA session in future studies.

Study participants reported having to multi-task during judging (Ste-Marie, 2000) and therefore adding a 'speaking' element (TA) into the judging task may have adversely affected the judging process. This may be attributed to verbal overshadowing (Chin & Schooler, 2008). Schooler and Engdtler-Schooler (1990) provided evidence for verbal overshadowing and found that a disruptive effect occurs through verbally reporting as verbal-overshadowing occurs as the formation of a verbally biased memory representation that overshadows original visual memory. It is argued that by asking participants to TA may result in reactive effects on task process and can influence the performance of a task. Further, future research aiming to employ TA within WAG judging may consider removing, as part of a developmental process, the notation from the task in order to reduce potential verbal overshadowing and improve attentional focus.

Interestingly the content of TA verbalisations varied across judges, with deduction themes being reported by some judges and not others. For example, *Fall* and *insufficient height* were verbalised by all participants, however some themes were only verbalised by 7 (insufficient split - in dance element), 6 (lack of side movement - in routine construction) and 3 (missing combination of movements) participants. Although, not something that was investigated within this study, however, future research could seek to investigate the decision-making differences between different levels such as international-non-international during judging (Catteeuw, Gilis, Jaspers, Wagemans, & Helsen, 2010; MacMahon & Ste-Marie, 2002; Ste-Marie, 1999). Although some time ago, Ste-Marie (1999) found that expert judges with more than 10 years of judging experiences were better at predicting the upcoming gymnastic elements and judged more correctly on those elements. Indeed, since 1999 the COP has moved through several revised versions due to the increasing complexity of routines and skills now seen in competition. In WAG, only categories 1 and 2 international judges, are eligible to judge at World Championships (Fédération Internationale de Gymnastique, 2016a)

thus, by virtue, have more judging experience when compared to those holding lower levels of judging awards. As such, it could be said that they might be more able to verbalise their decision-making using TA during judging tasks by providing more objective and reliable judging scores than novice judges as a consequence of accumulated judging experiences. Further, such judges may be able to retrieve information in their memory more efficiently and cope with the multiple attentional demands. This insight may well inform further study to explore verbal overshadowing (Chin & Schooler, 2008) and the multi-tasking of judges (Ste-Marie, 2000) whereby adding another "speaking" element into existing judging task in the TA session may affect the subsequent reliability and objectivity of judging scores.

Data provided through TA from the WAG judges could be used to inform coaching practice as it allows the coach to understand the decision-making with regard to deductions and provide illustration beyond that gained from video replay for the gymnast themselves. Although not within judging, similar suggestions have also been provided in previous research, which relates to the coach and athlete, where TA could be used to inform coach and practitioner interventions and practice (Nichols and Polman, 2008; Samson et al, 2015; Whitehead et al., 2015; Whitehead et al., 2018). More specifically, through understanding athlete cognition coaches and practitioners may be able to provide more informed interventions when working with their athlete. In a similar fashion, the coach could learn from the WAG judge's decision-making process through the use of TA.

An important limitation to acknowledge could be due to the researcher collecting this data having a significant level of expertise in the area of gymnastics judging. This level of expertise could have created some sort of 'Hawthorne Effect' (Haessler, 2014), where the subjects awareness of being observed during their TA trial, may have affected their responses. In addition, due to the researcher having a high level of expertise within WAG

judging, this could also have had an impact on what is being reported. Future research may want to take this into consideration.

In conclusion, results suggest that TA could be an appropriate tool to include within current Women's Artistic Gymnastics judging education to explore the decision-making of judges when making general execution and artistry deductions. TA may support, in particular, the development process of novice judges by improving the cognitive processes and awareness of the execution deductions during routine performance. It is recommended that future research develops the use of TA as a training method to facilitate the development of WAG judges and to investigate the TA differences between experienced and less experienced judges to inform future practice. Further the utility of TA across apparatus could be explored beyond that of the balance beam exercise and thus becomes fully representative of the judging requirements within a competition

References

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. and Clarke, V. (2014) What can thematic analysis offer health and wellbeing researchers? *International Journal of Qualitative Studies on Health and Well-being*, 9

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.

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

Dewey, J. (2008). The influence of Darwinism on philosophy. In J. Boydston & L. Hahn (Eds.), The middle works of John Dewey, 1899-1924 (Vol. 4, pp. 1-215). Carbondale: Southern Illinois University Press. (Original work published 1910b)

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., & 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

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

Fédération Internationale de Gymnastique. (2015). Women's Artistic Gymnastics Help Desk 2013-2016. In. Fédération Internationale de Gymnastique. (2016a). FIG General Judging Rules: Cycle 2017-2020. In.

Fédération Internationale de Gymnastique. (2016b). Women's Artistic Gymnastics: 2017-2020 Code of Points In. 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

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, 316–344.

Gibson, K. (2016). Mixed methods research in sport and exercise. In B. Smith, & A. C. Sparkes (Eds.). Routledge handbook of qualitative research methods in sport and exercise (pp. 382–396). London: Routledge.

Haessler S. (2014). The Hawthorne effect in measurements of hand hygiene compliance: a definite problem, but also an opportunity. *BMJ Quality & Safety,* doi:10.1136/bmjqs-2014-003507.

Hua, J., & Gao, X. (2017). Using think-aloud protocol in self-regulated reading research. *Educational Research Review, 22,* 181-193. doi:10.1016/j.edurev.2017.09.004

Kleynen, M., Moser, A., Haarsma, F. A., Beurskens, A. J., & Braun, S. M. (2017). Physiotherapists use a great variety of motor learning options in neurological rehabilitation, from which they choose through an iterative process: a retrospective think-aloud study. *Disability and Rehabilitation, 39*(17), 1729-1737. doi:10.1080/09638288.2016.1207111

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

Lofland, J., & Lofland, L. (1996). Analyzing social settings (3rd ed.). Belmont, CA: Wadsworth.

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.

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, 198–212.

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

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., 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

Nicholls, A.R., & Polman, R.C.J. (2008). Think aloud: Acute stress and coping strategies during golf performances. *Anxiety, Stress, and Coping*, 21, 283–294.

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

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.

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

Schooler, J. W., & Engstler-schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, 22, 36-71.

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*, DOI: 10.1080/1612197X.2015.1069877

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*. http://dx.doi.org/10.1080/1750984X.2017.1317357.

Ste-Marie, D. M. (1999). Expert-Novice Diferences in Gymnastic Judging: An Information-processing Perspective. *Applied Cognitive Psychology, 13*, 269-281.

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

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., 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., Rowley, C., Quayle, L., Marchant, D., 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.

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.

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. *Front Psychol, 6,* 1083. doi:10.3389/fpsyg.2015.01083

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

Whyte IV, J., Hauber, R., & Ward, P. (2014). The effect of a structured internship on the clinical performance of novice nurses? *Journal of Nursing Education and Practice, 4*(11), 74-81. doi:10.5430/jnep.v4n11p74

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.

Appendix 1: Example of a judging symbol notation sheet