A new qualitative typology to classify treading water movement patterns

Christophe Schnitzler
Chris Button
James L. Croft
Edith Cowan University
Ludovic Seifert

Follow this and additional works at: https://ro.ecu.edu.au/ecuworkspost2013

Part of the Biomechanics Commons

Recommended Citation

This Journal Article is posted at Research Online.
https://ro.ecu.edu.au/ecuworkspost2013/1238
A New Qualitative Typology to Classify Treading Water Movement Patterns

Christophe Schnitzler 1✉, Chris Button 2, James L. Croft 3 and Ludovic Seifert 1
1 C.E.T.A.P.S. Laboratory EA 3832: University of Rouen, Faculty of Sports Sciences, France
2 School of Physical Education, Sport and Exercise Sciences, University of Otago, Dunedin, New Zealand
3 Centre for Exercise and Sports Science Research, School of Exercise and Health Sciences, Edith Cowan University, Western Australia, Australia.

Abstract
This study proposes a new qualitative typology that can be used to classify learners treading water into different skill-based categories. To establish the typology, 38 participants were videotaped while treading water and their movement patterns were qualitatively analyzed by two experienced biomechanists. 13 sport science students were then asked to classify eight of the original participants after watching a brief tutorial video about how to use the typology. To examine intra-rater consistency, each participant was presented in a random order three times. Generalizability (G) and Decision (D) studies were performed to estimate the importance variance due to rater, occasion, video and the interactions between them, and to determine the reliability of the raters’ answers. A typology of five general classes of coordination was defined amongst the original 38 participants. The G-study showed an accurate and reliable assessment of different pattern type, with a percentage of correct classification of 80.1%, an overall Fleiss’ Kappa coefficient $K = 0.6$, and an overall generalizability $\varphi$ coefficient of 0.99. This study showed that the new typology proposed to characterize the behaviour of individuals treading water was both accurate and highly reliable. Movement pattern classification using the typology might help practitioners distinguish between different skill-based behaviours and potentially guide instruction of key aquatic survival skills.

Key words: Generalizability theory, clinical education, lifesaving.

Introduction

Treading water is an important survival skill in many near drowning incidents. This behavior enables the immersed individual to keep their head above water which avoids inspiration of water, allows them to survey the environment, and to make a decision about whether to wait for rescue or to swim to safety (Golden and Tipton, 2002). In aquatic environments, two different kind of forces can be generated; drag force, based on Newton’s law of action-reaction and; lift force, based on Bernouilli’s principle (Toussaint and Truijens, 2005). Drag force is a product of propulsive movement, the direction of which is the opposite of the body’s displacement. In contrast, lift force is produced with movements that are perpendicular to the axis of displacement (Toussaint and Truijens, 2005). Sanders (1999) showed that expert water-polo players and synchronized swimmers adopted a movement pattern characterized by large lateral movements of the legs in anti-phase, which enabled them to generate multiple lift force impulses throughout a leg cycle. As a consequence participants were able to adopt a stable position with their upper body high above water level. This movement pattern has been commonly referred to as the “egg-beater kick”. While the egg-beater kick is considered a mechanically efficient skilled technique, the behavioral adaptation of less proficient participants has not yet been characterized. Hence, the first aim of this study is to create a classification scheme with which to differentiate different behavioral types in treading water.

The capacity to recognize a behavioral pattern is a general human characteristic (Mather and West, 1993) that is attributed to brain areas specialized in that task (Campitelli et al., 2007; Greenlee et al., 2000). Compared to instrumental and social actions, locomotory actions are best recognized (Dittrich, 1993), especially the relative motion of each element to other configurual elements (Cutting and Proffitt, 1982). Furthermore this accuracy is maintained even in the presence of a wide range of frequencies and temporal asynchronies (Mowafy et al., 1990). On the other hand, the relative position and motion of each element are recalled more effectively when in a functional state, that is, if underlying rules govern the organization of the elements. For example, Williams et al. (2006) showed that in soccer, skilled players could accurately recognize pattern of play presented in point-light form, which was not the case anymore when several key players were occluded. In other words, patterns of play could only be recognized when representing a ‘meaningful’ game state.

Williams et al. (2012) suggested that the conceptual understanding provided by a typology helps analysts to accurately recognize the functional role of locomotory actions. Observational grids which are characteristic of typology-based methodologies (Bailey, 1994) are often created to help raters classifying phenomenon subjectively. The second aim of this study was to evaluate an observational grid set used by analysts to classify different types of skilled behavior.

Pattern recognition is based upon subjective assessment, and consequently determining the reliability of assessment (in terms of stability, dependability, trustworthiness, and consistency) is also important (Worthen et al., 1993). Generalizability theory is a useful framework to examine reliability. In the case of inter-rater reliability, generalizability theory provides an estimate of multiple sources of error variance (Brennan, 1983). Depending on the nature of the pattern to be analyzed, they established how many raters or viewing occasions were necessary to
reach an acceptable level of reliability (Gronlund, 1988). For example, in the context of a problem-based learning approach Sluijsmans et al. (2001) used generalizability theory to examine the reliability of peer assessment in the classroom by asking groups of students to grade their teammates on various criteria using a 5-point likert scale. They established that reliability was acceptable when at least three students were grading one teammate. In assessing motor behaviors, Marty et al. (2010) required 13 raters to assess 10 videos of subjects performing three psychomotor skills. Thus generalizability theory has been applied, predominantly in the educational literature, to evaluate the accuracy and reliability of different raters in assessing motor behaviors, Marty et al. (2010) required 13 raters to assess 10 videos of subjects performing three psychomotor skills. Thus generalizability theory has been applied, predominantly in the educational literature, to evaluate the accuracy and reliability of different raters in identifying patterns. In this exploratory study, we aimed to establish a qualitative typology of treading water based on biomechanical principles observed in aquatic environments. We then applied generalizability theory to evaluate the accuracy and reliability of the typology.

Methods

This experiment was split into two distinct parts; 1) establishing the typology, and; 2) applying generalizability theory to evaluate the typology.

Part 1: Establishing the typology

For the first part of the study, thirty-eight individuals volunteered to be filmed while treading water. Their main characteristics are summarized in Table 1. Participants wore similar bathing suits and their head could not be seen on the video, thereby reducing the chance that each individual could be clearly identified by personal characteristics.

Table 1. Main characteristics of the swimmers. Data are means (±SD).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Women (n=19)</th>
<th>Men (n=19)</th>
<th>Overall (n=38)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>23.1 (7.2)</td>
<td>22.4 (6.3)</td>
<td>22.7 (6.7)</td>
</tr>
<tr>
<td>Height (m)</td>
<td>1.68 (.05)</td>
<td>1.76 (0.06)</td>
<td>1.72 (.07)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>66.6 (9.2)</td>
<td>75.5 (8.1)</td>
<td>71.2 (9.6)</td>
</tr>
<tr>
<td>BMI (kg∙m⁻²)</td>
<td>23.3 (2.8)</td>
<td>24.5 (2.4)</td>
<td>23.9 (2.7)</td>
</tr>
<tr>
<td>SM Mass (kg)</td>
<td>27.2 (3.6)</td>
<td>37.5 (3.6)</td>
<td>32.5 (6.3)</td>
</tr>
<tr>
<td>Fat Mass (%)</td>
<td>16.9 (6.4)</td>
<td>10.3 (4.9)</td>
<td>13.5 (6.6)</td>
</tr>
</tbody>
</table>

The filming took place in a temperature controlled aquatic flume (StreamLInNZ, E-Type, Invercargill, New Zealand) with underwater viewing windows through which participants could be filmed. Four camera perspectives (Figure 1) were used to synchronously record the video footage with digital HD cameras. All participants wore a light harness (Delta™ Repel™ Technology Riggers Harness, Capital Safety, Red Wing, MN) attached to a rope which allowed experimenters to remove them from the water quickly if they wished to end the trial. The harness did not interfere in any way with the participants’ movements. The depth of the flume was 2 m hence participants had to tread water unsupported and could not touch the floor of the flume. The participants were asked to keep their head above the water level and remain in the same location for at least 150 s.

In order to establish the typology, video recordings of each participant treading water were examined independently by two experienced Biomechanists (authors 1 and 2, i.e., Schnitzler and Button). Based on the video footage each analyst had to decide which was the dominant source of propulsion (drag vs. lift force), and what type of inter-limb coordination was used (synchrony vs. asynchrony, Ouiller et al., 2006). Table 2 summarizes the criteria which the analysts used as a basis to examine each participant’s skill level. The analysts also provided brief descriptions of the treading water technique used by each participant. The analysts were permitted to view each participant’s video as many times as required in order to categorize each individual swimmer within the Typology. In cases where numerous coordination patterns were adopted by participants during the 150 s trial, the dominant pattern (used most often) was selected for analysis. Finally, the videos footage from eight of the original 38 participants were selected for further analysis in Part 2. These eight participants were chosen as examples of each pattern type identified by the analysts in the Typology.

Table 2. Qualitative criteria to examine participant’s skill level.

<table>
<thead>
<tr>
<th>Propulsive mode</th>
<th>Coordination mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less efficient</td>
<td>Synchonization (between arms, legs, or between arm and legs)</td>
</tr>
<tr>
<td>More efficient</td>
<td>Asynchrony (between arms, legs, or between arm and legs)</td>
</tr>
</tbody>
</table>

Part 2: Evaluating the typology

A group of 13 sport science students enrolled in the third or fourth year at the Sport Science Faculty of Strasbourg (age: 21 to 26 years) were selected as trainee analysts to evaluate the typology established in Part 1. A 1-hour period of training was provided to the analysts to teach them how to recognize the different pattern types. They were first instructed about the different propulsion types and were given macroscopic descriptors to recognize lift-
and drag-dominant propulsion: lateral movements relative to the surface for the former, up-down movements for the latter. Also, they were taught to differentiate coordination types: synchronized / asynchronized / unsynchronized movements of left/right side, upper/lower part of the body. The video file in the appendix served as an illustration of the two concepts (propulsion and coordination) on which the typology was based, and the analysts were free to ask questions during the training period.

After training, the analysts were independently shown representative sample videos (each lasting 20 s in duration) of the eight participants selected from Part 1. The eight videos were presented in a random order three times (i.e., 24 videos in total). The analysts were not told that the same pattern type would be presented several times during the analysis. The analysts were asked to classify each video segment according to the typology and were allowed to watch the videos as many times as they wanted. The percentage of patterns that were correctly classified was assessed for significance using Fleiss’ Kappa (K) using A Matlab routine (Cardillo, 2007). K is a statistical measure to compute the extent of agreement between raters on nominally scaled data.

The reliability of the classification was examined using generalizability study (G study) techniques which are used to characterize the reliability of subjective assessment (Strube, 2006). A G study uses generalizability theory (G theory), an extension of classical test theory with a series of analyses of variance to determine the dependability of behavioral measurement. In the present case, the 13 analysts were presented with the video footage of eight coordination types on three occasions. Rater and viewing occasion factors represent here the facets of measurement. A completely crossed pattern type (P, 8 levels) × rater (R, 13 levels) × occasion (O, 3 levels) design was used to obtain a relative (G) or absolute (φ) reliability coefficient. Both represent the reliabilities of the scores across the 13 raters and the three occasions. We were interested in the absolute magnitude of error, that is, the consistency of the rater’s judgment. For this reason, only φ coefficients were taken into account.

Generalizability theory also allowed us to determine which conditions would improve the reliability of the analysis by performing decision studies (D studies). A D study typically answers questions such as “how many raters should be used” and “on how many occasions should the data be rated” to achieve acceptable reliability (Strube, 2006). D studies can also be used to estimate how φ, the dependability coefficient, might change if the number of analysts or viewing occasions were different. A φ coefficient was calculated as part of the D study, which indicated the extent to which patterns were correctly classified. A Matlab routine (Mushquash and O’Connor, 2006) was used to perform all G and D studies.

Results

Part 1

Independent analysis and subsequent discussion led the two Biomechanists to identify nine different coordination types, although one pattern was later removed from the Typology. It was decided to define four main pattern types based on the nature of force generated, each divided into two sub-patterns according to the nature of the interlimb coordination. Table 3 presents an explanation of eight pattern types established by the typology.

Examples of each pattern are presented in a video file (Available at http://www.jssm.org/video/jssm-14-530m.htm). The first class of patterns was characterized by the simplest mode of propulsion (vertical movements from the arms, anterior-posterior movements from the legs). The second and third classes were characterized by at least one propeller (arm or leg) exhibiting lateral movements (which was considered as producing lift force). The fourth class contained the participants who used an in-phase sculling pattern for leg propulsion, characteristic of the so called “eggbeater kick”, with the use of the arms for additional propulsion. A fifth and last class of pattern was also identified in one participant who exhibited the egg-beater kick pattern with the legs, and no active movement of the arms at all. This somewhat spurious pattern was not present in any of the other participants and was therefore removed to simplify the typology. Thus, the typology classified movement patterns from the putatively simpler and less efficient propulsion, to the more efficient (propulsion) and complex coordination pattern. The 38 participants (Type 1.1, n=5; Type 1.2, n=4; Type 2.1, n=9; Type 2.2, n=2; Type 3.1, n=7; Type 3.2, n=2; Type 4.1, n=7; Type 4.2, n=1) were classified according to the typology.

Part 2

All 13 raters assessed 20 s video segments of eight participants performing treading water on three separate occasions. The overall rate of correct classification, 80.1%,
with an overall Fleiss’ kappa coefficient, $K = 0.69$.

To perform the G study, two facets were taken into consideration: rater (R) was facet 1, and occasion (O) was facet 2. A two-facet-fully crossed design was used on the 8 pattern type (P). The overall dependability coefficient ($\phi$) was 0.98, which indicates a level of overall reliability much higher than the 0.8 limit proposed by Cronlund (1988). The largest component of variance was Pattern Type, which accounted for 86.7% of the total variance, which indicates that much of the variability in the participant’s pattern type classification was due to differences in participant behavior.

The largest source of undesirable variation was the Type × Occasion × Rater component (10.2%), which can’t be explained by the measurement facets and is considered to be random error (Marty et al., 2010). The second largest source of undesirable variance is Type × Rater (4.1%), which indicates that for specific pattern type, the raters had different interpretations of the behaviors. Contributions to overall variance of “Rater”, “Occasion”, “Type×Occasion” and “Occasion×Rater” were negligible.

The reliability coefficient produced in a D study can be used to determine how dependability might change if numbers of raters or occasions were changed (Table 4). The D study was conducted on the number of raters and the number of occasions. The reliability for one rater on one occasion was 85.6%, improving to 91.9% with two viewing occasions. Furthermore, reliability markedly improved to 97.1% with four raters, viewed on two conditions.

### Table 4. The $\phi$ coefficients of the D-study as a function of number of raters and occasions.

<table>
<thead>
<tr>
<th>Number of occasions</th>
<th>Number of raters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>85.6%</td>
<td>91.9%</td>
</tr>
<tr>
<td>92.2%</td>
<td>95.8%</td>
</tr>
<tr>
<td>94.6%</td>
<td>97.1%</td>
</tr>
</tbody>
</table>

### Discussion

There were two main findings in this study: First, that a typology based on the nature of the generated force and the coordination type can be used to characterize treading water behaviors; second, that when provided with appropriate training (i.e., 1 hour tutorial to recognize macroscopic pattern characteristic), inexperienced analysts are able to characterize treading water behaviors accurately and reliably according to the typology.

Examination of video footage revealed that despite the fact that the swimming participants all successfully managed to tread water for 150 s, the vast majority of them did not use the so-called ‘eggbeater-kick’ pattern described by Homma (2005) and Sanders (1999). The wide inter-individual variability is not surprising when one considers treading water behavior from a constraints-led approach. This approach considers that the interaction of constraints (task, environmental and organismic) potentially leads to the emergence of a wide range of functional patterns (Davids et al., 2004; 2008; Newell, 1986).

In this study, the task was clearly imposed (that is, treading water for 150 s) and the range of behavior limited by the properties of aquatic environment (i.e., the environmental constraint). So the behavioral variability might be best explained by organismic constraints. For example, the physical property of the swimmer (such as fat to lean body mass distribution) directly influences their relative buoyancy. Furthermore, his/her expertise level influences the amount and the efficiency of propulsive force generation.

Biomechanists have shown that propulsion in aquatic environment can be achieved by generating lift, drag or a combination of both of these forces (Counselman, 1971; Schleihauf et al., 1983; Toussaint and Trujens, 2005) and has been applied to the treading water patterns of expert swimmers (Homma, 2005; Sanders, 1999). When a solid surface (here, a hand or a foot) is moved through a fluid, a drag force is generated in the opposite direction of the displacement. This movement creates a difference of pressure between the posterior and the anterior side of the element, thus producing a force in the opposite direction from the axis of displacement. Lift force relies on Bernoulli’s principle that is often illustrated using the analogy of an aircraft wing (Counselman, 1971; Schleihauf et al., 1983). Aircraft wings are designed in such a way that the shape of the upper surface is more convex than the lower surface. So when in motion, the air molecules on the upper side of the wing have to travel further and therefore faster than those on the lower side. This creates a depression that generates a force perpendicular to the axis of displacement that “lifts” the aircraft. These concepts have served as a basis to explain the complex movement patterns of expert swimmers, especially the lateral movement of the hand which creates propulsion (Schleihauf et al., 1983). Of these two methods of creating propulsive force, lift force is considered to be more efficient than drag force (Homma, 2005; Zielinski, 2001). In addition, generating force more consistently throughout the leg cycle leads to less vertical displacement of the body. These principles serve to justify why, in this study, participants exhibiting lateral, asynchronous movement were considered more efficient than those exhibiting up-down, symmetric movements.

These two principles (nature of the force and frequency of the force impulse) provided an interesting means to describe the macroscopic characteristics of the movements (up-down vs lateral, synchronized vs asynchronous movements). As we have explained propulsive forces generated in water rely on somewhat complex biomechanical principles not obvious from mere observation. However, in line with Williams et al. (2006), we expected that the relative motions of the limbs could be perceived from the video and easily classified using the macroscopic descriptors of the typology, a strategy which has proven efficient in past studies (Koedjiker et al., 2011). For this endeavor, the accuracy and reliability of the typology was evaluated. The rate of correct classification was 80.1%, with an overall Fleiss’ kappa coefficient of $K = 0.69$ which represents a substantial agreement between raters (Landis and Koch, 1977). These results
showed that scoring consistency (i.e. the ability to correctly identify different pattern types) was quite high.

The ability to identify correct patterns on different occasions (occasion-variance) it was necessary to use generalizability theory. According to Mushquash et al. (2006), classical test theory provides an assortment of useful individual methods for assessing reliability (e.g., test–retest, alternate forms, internal consistency, and inter-observer agreement), but it considers only one source of measurement error at a time. Generalizability theory allows one to separate different source of variance, and to outline the interactions between all variance sources. Thus, the power and complexity of generalizability theory increases when additional facets of measurement are included (Strube, 2006), which was the case in this study where two facets (rater and occasion) were assessed.

Pertinent studies in education research have reported index of dependability (φ) ranging from 0.6 to 0.96 (Marty et al., 2010; Sluijsmans et al., 2001). The φ coefficient found in our study (0.98) was very high, and suggests that raters were able to classify the pattern type with few errors despite the fact that they had limited experience of analyzing or watching this specific task. Also, the study indicated that just one rater on one occasion could correctly recognize the pattern type 85.6% of the time, which is above the 80% threshold generally accepted in such studies (Gronlund, 1988; Strube, 2006). This finding shows that raters could reliably assess the behavior when the evaluation occurred just once (Linn et al., 1975). However, this result differs from the study of Marty et al. (2010). Possibly, this is because raters who reviewed the video in our study had to categorize behaviors on the basis of simple macroscopic descriptors, which is apparently easier than grading psychomotor skills on a scale.

Conclusion

This study proposes a new qualitative method to assess the skill of treading water based on a conceptual typology. Our results showed that a relatively brief training period was sufficient for an inexperienced analyst on one single occasion to correctly characterize the behavior of a person treading water. Overall, this study provides an opportunity to show how generalizability theory could be used in the context of physical education or sport coaching.

It could be argued that the raters had the advantage of prior training in biomechanics and sports techniques (they were all enrolled in sports science in 3rd and 4th year in sport sciences) thereby potentially limiting the generality of findings. Indeed as shown by others, the type of training and clear guidelines on how to assess such psychomotor skill is influential for raters (Ladyshewsky and Gotjamanos, 1997). Perhaps our results would not be replicated if the population of raters originated from a different background, or were less mature. On the other hand, identifying different movement patterns is arguably most beneficial for those involved in sport and/or training, so the generalizability of the rating to a general population is probably not of primary importance.

This study raises some interesting practical implications. With relatively little formal training individuals can use a qualitative typology to accurately and reliably evaluate a learner’s movement patterns. This is a key skill for many sports coaches and officials and it is possible that simple typologies could be derived for other activities. The present study is also supportive of providing analysts with global, macroscopic prescriptions (in preference to detailed, micro-level) to learn how to identify patterns of movement (see also Maxwell et al., 2000).

Acknowledgements

This study was funded by Water Safety New Zealand, New Zealand. We also wish to thank Pr Mushquash who kindly provided the Matlab routine that was used to perform the analysis.

References


Key points

- Treading water behavioral adaptation can be classified along two dimensions: the type of force created (drag vs lift), and the frequency of the force impulses
- Based on these concepts, 9 behavioral types can be identified, providing the basis for a typology
- Provided with macroscopic descriptors (movements of the limb relative to the water, and synchronous vs asynchronous movements), analysts can characterize behavioral type accurately and reliably.