1-1-2014

A method to optimize a typology-based classification system

Christopher Schnitzler

James Croft
Edith Cowan University

Chris Button

Mats Ulmers

Keith Davids

10.1016/j.proeng.2014.06.003
Schnitzler C., Croft J., Button C., Ulmers M., Davids K. (2014). A method to optimize a typology-based classification system. Procedia Engineering. 72(), 9-13. Sheffield. Elsevier Ltd. Available here

Available online at www.sciencedirect.com
ScienceDirect
Procedia Engineering 72 (2014) 9 – 13
Procedia Engineering
www.elsevier.com/locate/procedia

The 2014 conference of the International Sports Engineering Association

A method to optimize a typology-based classification system

Christophe Schnitzler[a], James Croft[b,c], Chris Button[c], Mats Ulmers[a], Keith Davids[d,e]

[a]LISEC - EA 2310 -ESPÉ - Univer1sité de Strasbourg- France
[b]School of Exercise and Health Sciences, Edith Cowan University, Australia
[c] School of Physical Education, Sport and Exercise Sciences, University of Otago, New Zealand
[d]Centre for Sports Engineering Research, Sheffield Hallam University, UK[e]FiDiPro Programme, University of Jyvaskyla, Finland

Abstract

This study sought to provide guidelines for implementing typology-based qualitative analysis of human movement patterns. Fifteen participant-analysts were instructed how to classify treading water behaviours into eight different categories using a *training* set of videos. They were later provided with two additional sets of videos called *validation*, and *test* sets. Results first identified reliable (n=9), and not reliable (n=6) analysts. A decision study outlined that one analyst was sufficient to reliably categorize the behaviours in the 'reliable' analyst group, whereas up to four were necessary in the 'unreliable' group. These data provided new insights into more objective qualitative analysis methods for understanding human movement behaviours.

Selection and peer-review under responsibility of the Centre for Sports Engineering Research, Sheffield Hallam University

Keywords: Generalizability theory, clinical education, expertise

1. Introduction

The capacity to recognise a behavioural pattern is a general requirement of human performance in many different domains such as sport science and performance analysis (Mather and West, 1993). For example, it has been proposed that the ability to extract higher order predicates (e.g. tactical information) from positional data and temporal relationships between individuals on field is an important part of skilled motion perception. Smeeton et al. (2004) showed that once the ability to recognise patterns is acquired, in the context of a specific sport or physical activity, it is transferable to activities that are 'structurally' similar. An interesting issue concerns how to train individuals to identify and respond to spatial-temporal features of human movement for the purposes of performance analysis. Typologies are conceptual frameworks organised as potentially useful tools to help analysts in classifying human behaviours. In swimming, for example, Chollet et al. (2000) identified three pattern type using qualitative criteria: catch-up, opposition and superposition models. They showed that at high swimming

1877-7058
Selection and peer-review under responsibility of the Centre for Sports Engineering Research, Sheffield Hallam University
doi:10.1016/j.proeng.2014.06.003

speeds, only experts were able to use the superposition model. However, typologies have received some criticism (Bailey, 1994) since they are often based on somewhat arbitrary criteria, are essentially static and can pose significant reliability challenges.
The aim of this study is to examine whether dividing a set of video clips into training, validation and test sets can improve the objectivity of qualitative analyses of human movement patterns.

2. Method

Sets of video sequences of individuals treading water were used to train a group of 18 novice analysts. A conceptual typology (Schnitzler et al., 2014) was first explained to the analysts. This typology identifies eight pattern types as a function of their efficiency: simple patterns are characterized by in-phase, up and down movements, whereas complex patterns exhibit lateral and anti-phase movements. Analysts were trained using a set of training videos which contained clips of an expert synchronized swimmer who mimicked the eight pattern types of the conceptual typology (Schnitzler et al. 2014). During this period, they were free to ask any question about the classification. In a second step, analysts were exposed to a “validation set” of videos which contained 72 filmed clips (nine different versions of the eight distinct pattern-types) also mimicked by two national-level water polo players. The third and last step consisted of classifying a “test set” of videos that contained the spontaneous behaviours of eight different recreational -swimmers treading water. Here, those behaviours were presented three times to the participant analysts, who therefore had to examine 24 video clips. The participants’ level of categorisation accuracy was examined at the end of the second step with the “validation set” of videos. For the purposes of this study, ‘reliable’ analysts were those deemed to score on average 80% or more levels of categorisation accuracy during the final three trials. The other analysts were labelled ‘not reliable’. The reliability of the typology itself was examined on the test set using generalizability theory (G-theory). A crossed design with Videos (P, 8 levels) × Analyst (A, 15 levels) × Occasion (O, 3 levels) facets was used to obtain an absolute (phi, or ϕ) reliability coefficient, which represents the consistency of the rater’s judgment. Decision studies (D studies) were subsequently used to estimate which setting would be suitable to reach a ϕ coefficient above 0.8. Finally, the weight of evidence (Ev) was calculated for each video using the equation: $Ev(p)=\log_2 p/(1-p)$ (bits)
The typology was considered as providing confusing guidelines when $Ev(p)<0$, which meant that a specific video clip could not be classified correctly more than 50% of the time.

3. Results

3.1. The validation set

The evolution of participant accuracy over trials, highlighted in Figure 1, showed a logarithmic-shaped curve, and helped us to determine between ’reliable’ (n=9) and ‘not reliable’ (n=6) analysts.

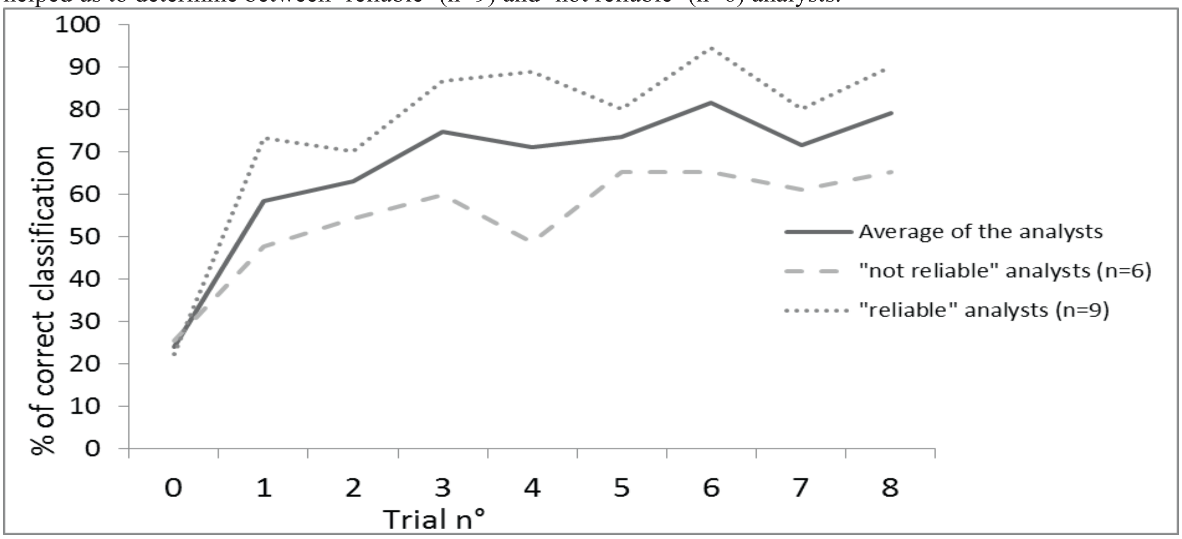


Figure 1: Evolution of the accuracy score in classifying treading water patterns over trials

3.2. The Test set

The role of this test set was twofold: to outline different source of variance (Figure 2) and to identify the number of analyst necessary to conduct a reliable analysis (Figure 3).

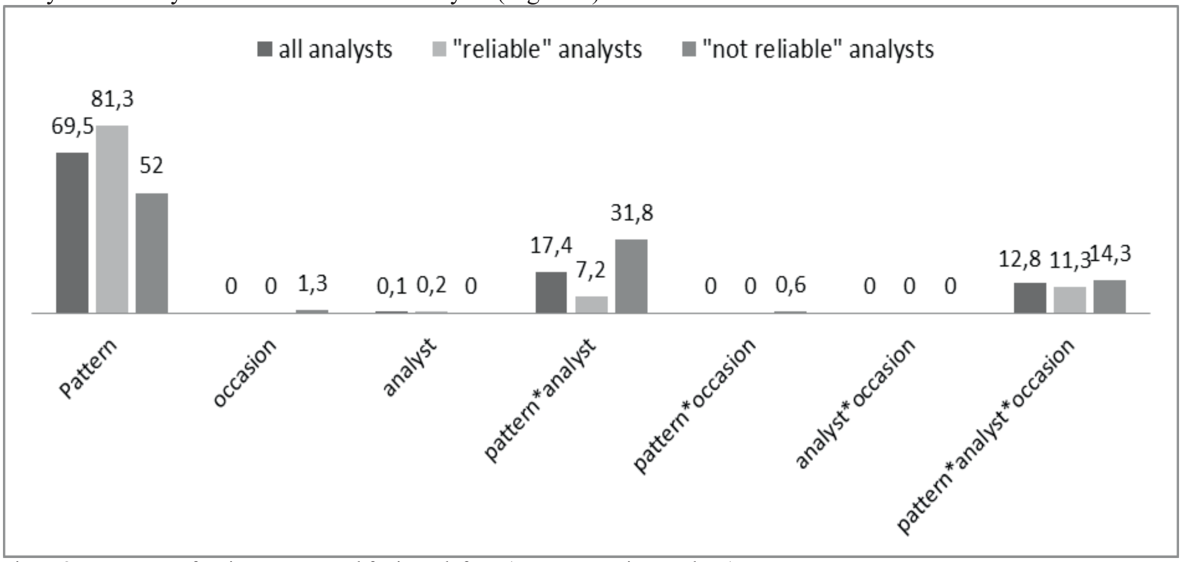


Figure 2: Percentage of variance accounted for in each facet (pattern, occasion, analysts)

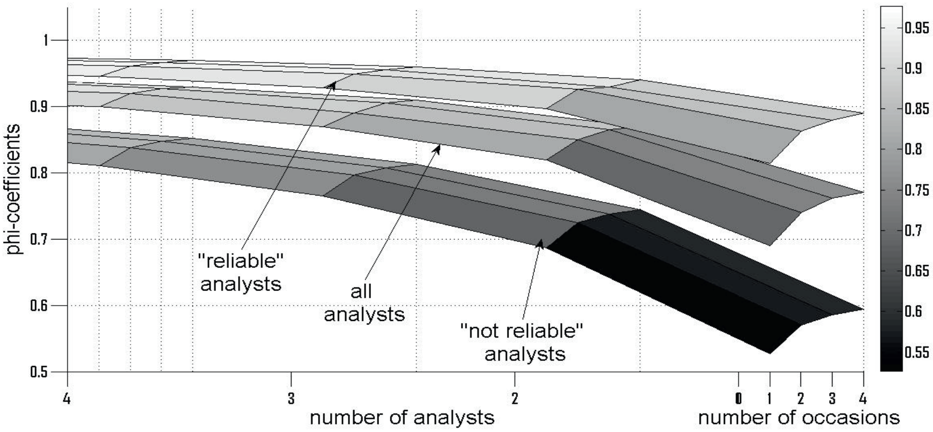
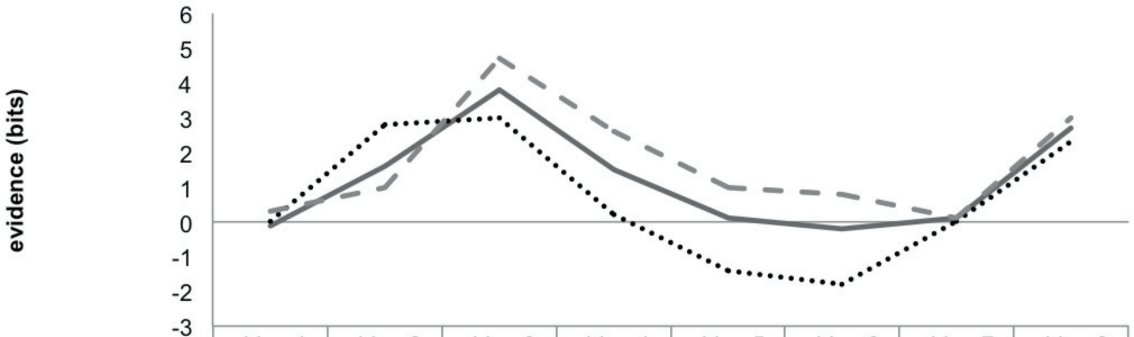


Figure 3: ϕ coefficients as a function of number of analysts and number of occasions.

3.3. The validation set

To examine the relevance of the typology itself, the weight of evidence was calculated for each video (Figure 4).



	video 1	video 2	video 3	video 4	video 5	video 6	video 7	video 8
all	-0.1	1.6	3.8	1.5	0.1	-0.2	0.1	2.7
not reliable	0	2.8	3	0.2	-1.4	-1.8	0	2.3
reliable	0.3	1	4.7	2.6	1	0.8	0.1	3

Figure 4: Weight of evidence of the different videos presented in the test set

4. Discussion

The aim of this study was to devise a procedure to improve the objectivity of qualitative analyses of human movement patterns. Principal results showed that by dividing the video material into training, validation and test sets, we were able to: (1) distinguish between 'reliable' and 'not reliable' analysts; (2) establish which setting would suit an analysis of motor behaviours; and (3) identify whether or not the typology itself should be refined.

Prior to training, analysts were asked to classify treading water behaviours from a putatively simpler and less efficient pattern, to a more efficient pattern. Unsurprisingly, the accuracy rate was low (i.e., 25%) between the raters and the typology, and this agreement score was similar in the 'reliable' and 'not reliable' analyst groups. During the eight successive trials that constituted the validation period, analyst accuracy improved in a logarithmic manner, recalling the learning curve of Adler and Clark (1991). A closer observation of the learning curve in Figure 1 indicates that the accuracy rate levelled off quickly until trial 4, then improved more gradually, which vindicated our reasoning to evaluate the overall accuracy of the analyst based on the average of the final three accuracy scores. This evaluation period allowed us to distinguish between 'reliable' and 'not reliable' analysts, using the 80% of accuracy rate as a boundary, in accordance with suggestions of Gronlund (1988). As Adler and Clark (1991) pointed out, learning is internally complex and may lead to large inter-individual differences in learning curves. Data on the validation set also suggested that analysts build their own models of pattern categorization at a very early stage, which remain stable thereafter, despite feedback provided by their questions and knowledge of their accuracy rate.

The test set consisted of unlabelled data in order to examine the relevance of the typology itself in describing actual, not idealised, motor behaviours. The focus was, therefore, on reliability rather than accuracy. Overall, the largest component of variance in the G-study was Pattern Type. This desirable source of variance indicates that much of the variability in a participant analyst's pattern type classification was due to differences in behaviour displayed on the videos. However, Pattern × Analyst and Pattern × Occasion × Analyst variations were also observed as undesired variance sources. Pattern × Analyst variations accounted for 17.8% of the variance, indicating that some videos in the set were more challenging to reliably classify than others, especially video 1 and 7 (see Figure 4) which were classified correctly about 50% of the case even by reliable analysts. Pattern × Occasion × Analyst interactions accounted for 12.8% of the total variance, a level which can be considered to represent random error (Marty et al., 2010). Contribution to the overall variance by 'Occasion' was negligible, thus confirming that the training programme helped each analyst to develop his/her own way of classifying the behaviour presented, and stick with that *modus operandi* when presented with it several times, rather than merely guessing.

Variance sources remained similar when 'reliable' and 'not reliable' analysts were distinguished. Only the

percentages changed, with more variance explained by the 'Pattern' facet for 'reliable' compared to the 'not reliable' analysts (81 vs 52%). Less variance was explained by the Pattern × Analyst factor (7.2 vs 31%). The consequence could be measured in the D-study, which showed that improving classification reliability is best achieved by changing the number of analysts. Results showed that two analysts would be sufficient to achieve an acceptable level of reliability in general, but this estimate could vary as a function of the skill of the analyst: one 'reliable' analyst is sufficient, whereas four 'not reliable' analysts are necessary to achieve the 80% reliability limit. This finding shows how important it is to first determine an analyst's skill level.

The last stage of the procedure was to examine the relevance of the typology itself. This could be achieved by examining the Pattern × Analyst variance in the G-study. The weight of evidence for each video was calculated, and we considered all videos with a weight of evidence close to or under zero to provoke a reliability issue. Using 'reliable' analysts only, our data showed that one of the video clips could not be easily identified with the typology (Video 7), which suggests how the typology itself could be refined.

This study raises some interesting practical implications. By dividing the observational material into a training, validation and test set, it shows how generalizability and evidence theory could to be used to outline 'reliable' analysts and improve qualitative typologies to evaluate motor patterns. The use of a typology in qualitative analysis provides practitioners with an objective framework upon which to base their pattern identification and fault correction methods. Therefore, this new method of analysing and classifying human movement patterns could be helpful in any contexts that require human identification of patterns. These techniques could be applied to clinician training (e.g., gait analysis), ethology, and sports coach education. Also, we believe that the procedure proposed in this article could address methodological issues that often prevent typology-based studies being published. Lastly, used in parallel with machine learning procedures, this method might help engineers to refine their algorithms by introducing features first identified qualitatively.

References

Adler, P. S., & Clark, K. B., 1991. Behind the learning curve: a sketch of the learning process. Management Science, 37(3), 267-281.

Bailey, K., 1994. Typologies and taxonomies. In *Typologies and taxonomies: An introduction to classification techniques.* In : M. Lewis-Beck (Ed.). Thousand Oaks, California 91320: Sage publications. pp. 1-16.

Chollet, D., Chalies, S., & Chatard, J. C., 2000. A new index of coordination for the crawl: description and usefulness. International Journal of Sport Medicine, 21(1), 54-59.

Gronlund, N. 1988. How to construct achievement tests (4th ed.). Englewood Cliffs: Prentice-Hall.

Marty, M. C., Henning, J. M., & Willse, J. T., 2010. Accuracy and reliability of peer assessment of athletic training psychomotor laboratory skills. Journal of Athletic Training, 45(6), 609-614. doi: 10.4085/1062-6050-45.6.609

Mather, G., & West, S., 1993. Recognition of animal locomotion from dynamic point-light displays. Perception, 22(7), 759-766.

Schnitzler, C., Button, C., Seifert, L., & Croft, J., 2014. Analyzing expertise through data mining: an example based on treading water. Journal of Sport Sciences.

Smeeton, N. J., Ward, P., & Williams, A. M., 2004. Do pattern recognition skills transfer across sports? A preliminary analysis. Journal of Sport Sciences, 22(2), 205-213. doi: 10.1080/02640410310001641494