Title: The future of classification in wheelchair sports; can data science and technological advancement offer an alternative point of view?

Submission type: Original article

Authors: Rienk M.A van der Slikke[a,b,*], Daan J.J. Bregman[b],, Monique A.M. Berger[a], Annemarie M.H. de Witte[a,c], Dirk-Jan (H.) E.J. Veeger[b,c]

Affiliations:

[a] Human Kinetic Technology, The Hague University of Applied Sciences, Johanna Westerdijkplein 75, 2521EN The Hague,The Netherlands

[b] Department of Biomechanical Engineering, Delft University of Technology, The Netherlands

[c] Department of Human Movement Sciences, Vrije Universiteit Amsterdam, The Netherlands

Corresponding author:

The Hague University of Applied Sciences
Faculty of Health, Nutrition and Sports
Human Kinetic Technology
Johanna Westerdijkplein 75
2521EN The Hague
The Netherlands

Phone:

Mail: r.m.a.vanderslikke@hhs.nl

Running head: The future of classification in wheelchair sports

Abstract: 254 words

Manuscript: 3308 words (incl. headers & titles)

Number of figures: 3

Number of Tables: 6 (including 1 in appendix)

Abstract

Purpose: Classification is a defining factor for competition in wheelchair sports, but it is a delicate and time-consuming process with often questionable validity.[1] New inertial sensor based measurement methods applied in match play and field tests, allow for more precise and objective estimates of the impairment effect on wheelchair mobility performance. It was evaluated if these measures could offer an alternative point of view for classification. **Methods:** Six standard wheelchair mobility performance outcomes of different classification groups were measured in *match* play (n=29), as well as *best* possible performance in a field test (n=47). **Results:** In *match*-results a clear relationship between classification and performance level is shown, with increased performance outcomes in each adjacent higher classification group. Three outcomes differed significantly between the low and mid-class groups, and one between the mid and high-class groups. In *best* performance (field test), a split between the low and mid-class groups shows (5 out of 6 outcomes differed significantly) but hardly any difference between the mid and high-class groups. This observed split was confirmed by cluster analysis, revealing the existence of only two performance based clusters. **Conclusion:** The use of inertial sensor technology to get objective measures of wheelchair mobility performance, combined with a standardized field-test, brought alternative views for evidence based classification. The results of this approach provided arguments for a reduced number of classes in wheelchair basketball. Future use of inertial sensors in match play and in field testing could enhance evaluation of classification guidelines as well as individual athlete performance.

Key words: Paralympic sports, wheelchair basketball, classification, , inertial sensors, big data

Introduction

In most Paralympic sports, a classification system is used to attain fair competition between athletes with various levels of impairment. The Paralympic classification systems aims to promote sports participation of people with disabilities by minimizing the impact of eligible types of impairment on competition outcome.[1] Ideally, the classification should *only* cover the effect of impairment on game performance. Evidently, the magnitude of that effect is hard to estimate accurately given the number of confounding factors.[2] To determine the level of impairment itself, most classification systems categorize based on function levels rather than on pathology.[3] Functional assessment is either based on isolated function tests, with assumptions about their effect on game performance, or the classification system is based on match observation. Given the diversity of functions, it is nearly impossible to determine the effect of each impairment level on game performance. The latter argument pledges for the use of match observation based classification, but for those systems match related confounders (field position, opponent, tactics) affect the functional assessment.

Wheelchair basketball was the first disability sport to use a functional classification system. Although functional classification is now a common practice, the wheelchair basketball system still stands out since the function level assessment is based on *match* observation of "volume of action", instead of isolated function tests. The wheelchair basketball classification system (IWBF; www.iwbf.org) started out as a medical based system (3 classes), but with the conversion to a function based system, the number of classes was extended to 8, in order to take the increasing heterogeneity of participants into account. Classifications range from 1 (most impaired) to 4.5 points (no functional limitation), with a team of five athletes composed of maximal 14 points. Although used since 1982,[4] there is an ongoing quest to provide scientific knowledge for more evidenced based classification guidelines.[2,5,6] The advantage of a *match* observation based classification is that the assessments are made in an ecologically valid way, but observation methods also have their flaws and limitations. Actions like ball handling are well observed, but estimations of speed, acceleration and force, cannot be assessed accurately on observation alone. Another contaminating factor in the current observations is that match specific factors like field position (guard, forward, centre), opponent and coach instructions are known to interact on performance[2]. Indeed, more impaired players (low classification) are often positioned in physically less demanding field positions, possibly masking their potential *best* performance levels. Therefore, assessment of performance in a *match* alone provides a narrowed image, possibly disregarding *best* possible performance levels. On the other hand, testing *best* performance in an isolated field test or lab setting alone, does not provide information on how well an athlete is able to make use of his performance capacities during the course of a match. Therefore, research on the relationship between *match* and *best* condition is needed to determine if measurements in only one condition are sufficient for well-founded classification.

Several researchers investigated the effect of impairment on performance as expressed in the current classification, both in *match* conditions as well as in a field test to measure *best* possible performance. Vanlandewijck et al. [5] assessed the wheelchair basketball performance of differently classified players during a match based on the Comprehensive Basketball Grading System (CBGS), next to the physical fitness in a laboratory test. Based on their results they considered a reduced number of classes viable. In a similar study by Vanlandewijck et al. [2] based on the CBGS scores of match performance, the relationship between class and position in the field was appointed as one of the factors for the absence of significant performance differences between two adjacent classes. In a study by Molik et al. [7] a Wingate Anaerobic Test was used to assess indexes of upper extremity anaerobic performance, which also led to the conclusion that a reduced number of classes was recommendable. So, in research a relationship between classification and different performance measures is acknowledged in various conditions. Yet, to identify the true effect of impairment on performance and to explore the relationship between *match* and *best* performance, a single outcome measure should be used in both conditions.

A recently introduced method based on inertial sensors, allows for objective performance estimations in both *match* and *best* condition, in a reliable and unobstructive way.[8] This method quantifies the *wheelchair mobility performance*, that is the ability to manoeuvre the wheelchair. This measure for the combined wheelchair-athlete combination is one of the most important performance aspects [9] contributing to the overall game performance as described by Byrnes et al.[10] In elite wheelchair basketball, van der Slikke et al.[11] confirmed the clear relationship between classification and wheelchair mobility performance, but so far only in *match* conditions not yet in *best* conditions (field test). In this study, wheelchair basketball athletes were measured in a sport specific wheelchair mobility performance field test,[12] that was first tested for reliability. Once the reliability had been ascertained, fourty-seven elite athletes of all classifications were tested for *best* wheelchair mobility performance in this field test, to rule out possible *match* related confounding factors on wheelchair mobility performance.

The present study explores the relationship between wheelchair mobility performance in both *match* and *best* condition and its interaction with classification. The current classification is then compared to clusters derived from wheelchair mobility performance analysis in *best* conditions, to outline a suitable number of performance based classes. Finally, we will evaluate whether such clustering may provide an alternative point of view to classification systems.

Methods

Subjects

Wheelchair mobility performance was measured in a match [11] for the first group of elite wheelchair basketball athletes (n=29) and in a standardised field test for a second group of athletes (n=47, Table I). Part of the athletes (n=12) were measured in both conditions, forming a third dataset for analysis of the relationship between match and field test performance. For the purpose of reliability testing, twenty-three of the athletes performed the field test twice. Results of this test-retest analysis are described in Appendix II. This study was approved by the ethical committee of the department of Human Movement Sciences: ECB-2014-2. All participants signed an informed consent after being informed on the aims and procedures of the experiment.

++ Please insert Table 1 here

Methodology

Each athlete's own sports wheelchair was equipped with three inertial sensors (xIMU for match, X-IO technologies; Shimmer3 for field test, Shimmer Sensing, Figure 1), one on each rear wheel axis and one on the rear frame bar. The frame sensor was used for measuring forward acceleration as well as rotation of the frame in the horizontal plane (heading direction). The combined signals of wheel sensor acceleration and gyroscope were used to estimate wheel rotation, which in turn provided frame displacement given the wheel circumference.

Estimates of frame rotations in the horizontal plane were used to correct the wheel gyroscope signal for wheel camber angle, as described by Pansiot et al.[13], Fuss et al.[14] and van der Slikke et al.[8] Furthermore, a skid correction algorithm was applied to reduce the effect of single or concurrent wheel skidding.[15]

++ Please insert Figure 1 here

Based on inertial sensor outcomes for each measurement a wheelchair mobility performance plot was generated, showing the six key outcomes of wheelchair performance.[11] The outcomes included are: average speed; average best speed (of best 5 in a match and of best 2 in the field test); average acceleration in the first 2m from standstill; average rotational speed during forward movement; average best rotational speed during a turn on the spot (of best 5 in a match and of best 2 in the field test) and average rotational acceleration.

Statistical analysis

To test for classification effects on wheelchair mobility performance, athletes were split into three classification groups: low (1 -1.5), mid (2 – 3) and high (4 – 4.5). These classification group boundaries were chosen in line with earlier research regarding wheelchair mobility performance. In the paper by van der Slikke et al.[11] they chose to separate the class I (1 – 1.5) in a single group, given their distinct performance levels [2,5] and to separate class IV (4 -4.5) from the class II & III athletes, since they also show (to a lesser extend) distinct performance levels.[2,5] Visual inspection of the distribution, followed by a Kolmogorov-Smirnov test was applied to test for normal distribution[16] of all six wheelchair mobility performance outcomes, to verify for the use of parametric statistics. A one-way ANOVA was used to test for group differences in the six standard mobility performance outcomes. For both field test (n=47) and match data (n=29), post-hoc Bonferroni tests were applied to identify between which groups significant differences occurred.[17] The magnitudes of the classification group differences in the field test were also expressed in the Smallest Detectable Difference (SDD 95%) as determined by the test-retest reliability (appendix II). For the 12 athletes measured in both field test and match, a Pearson correlation was calculated for all six outcomes of the wheelchair mobility performance, combined with a paired samples T-Test to verify if there were structural differences.

TwoStep clustering analysis was applied,[18-20] to the complete field test performance dataset, without the split in classification groups (appendix III). The TwoStep method is an exploratory tool designed to reveal natural groupings within a dataset that would otherwise not be apparent.[21] Given the small sample size, a log-likelihood distance measure was used combined with the Schwartz's Bayesian Criterion.[22] Since the maximal number of clusters is arbitrary, it was set in alignment to the current classification system (n=8).

Results

For the tweenty-nine athletes measured in *match* play, classification group averages are displayed in the standardized wheelchair mobility performance plot (Figure 2).[11] The plot range was slightly enlarged to allow display of the *best* wheelchair mobility performance outcomes per classification group of the fourty-seven athletes measured in the field test (Figure 3).

++ Please insert Figure 2 here

++ Please insert Figure 3 here

The differences of wheelchair mobility performance outcomes in the field test are also expressed in a factor of the SDD 95% (Table 2). The lowest factors of SDD 95% appear between the mid and high classification group (0 -1.0) and the highest factors show between the low and high classification group (1.3-6.5).

++ Please insert Table 2 here

Classification groups showed significant ($p<0.05$) differences in all six wheelchair mobility performance outcomes in the match and in 5 in the field test measurements (Table 3). Post-hoc Bonferroni tests revealed that in the *match* 3 out of 6 outcomes differed significantly ($p<0.05$) between the low and mid classified athletes and only best forward speed differed between the mid and high classified group (Table 3). For *best* performance as measured in the field test, five wheelchair mobility performance outcomes differed significantly between low and mid classified athletes and no outcomes differed between mid and high classified athletes.

++ Please insert Table 3 here

For the twelve athletes measured in both match and field test conditions, the Pearson correlations for all six wheelchair mobility performance outcomes are displayed in Table 4. Three outcomes were significantly ($p<0.05$) higher in the field test compared to the match performance, and two outcomes were higher on average, but not significant. The average best speed was significantly lower in the test compared to the match performance.

++ Please insert Table 4 here

The TwoStep analysis revealed two clusters, from a model that was considered "good" based on the cluster quality (silhouette of cohesion and separation ≥0.5). Most important model predictors were all forward movement based outcomes (factor 0.93 – 1), whereas the importance of rotational outcomes ranged from a factor 0.35 - 0.51. If analysed for class allocation (Table 5), the first cluster (A) shows clear agreement with the low classified group, although 6 athletes of the higher-class groups are included as well. The second cluster (B) corresponds very well to the mid/high classified groups, with only one athlete of the low-class group included. The differences in performance outcomes between clusters, as expressed in the factor of SDD 95%, are quite similar to the ones shown between classification groups (low-mid & low-high, Table 2).

++ Please insert Table 5 here

Discussion

This study was aimed at exploring the relationship between *match* and *best* wheelchair mobility performance and to what extend that relationship is affected by impairment level as expressed in the current classification. In general, it is clear that wheelchair mobility performance is clearly affected by the athlete's impairment level. This effect is shown in the *match* results, with increased performance outcomes for each successive classification group. Of the six wheelchair mobility performance outcomes, three differ significantly between the low and mid-class group and one between the mid and high-class group. Once the *match* related factors are expelled, a different pattern emerges as shown by the *best* results (field test measurements). Rather than a gradual incline of performance with classification (Figure 2), a clear performance separation shows with the most prominent difference between low and mid-class group outcomes . The wheelchair mobility plot (Figure 3) neatly shows that in the field test, only the low-class group deviates from the performance of the other athletes. Five of the wheelchair mobility performance outcomes differed significantly between these class groups, whereas no significant differences showed between mid and high classified athletes.

A relationship between classification and wheelchair mobility performance was anticipated in *match* and *best* condition. Indeed, low-class athletes show the lowest performance outcomes and high-class athletes the highest wheelchair mobility performance values in both conditions, but the patterns of mid-class athletes differ between conditions. So only moderate correlations between *match* and *best* performance were expected due to those differences in the mid-class group. Moderate to high correlations (0.62-0.76) showed for the performance of the twelve athletes measured in both conditions. Given the unrestrained nature of the field test (no opponent or other obstructions), it was anticipated that wheelchair mobility outcomes would equal or exceed those of match conditions. Indeed, three out of six outcomes were significantly higher in that condition. Only average best speed appeared to score significantly lower in the field test. In the field test, the longest continuous run is 12 meter, where in a match -although not frequent- longer continuous runs occur, with corresponding higher speeds.

The impairment effect on performance should shape the classification system, so the International Paralympic Committee (IPC) is committed to the development of selective classification systems, not performance classification systems.[1] It is vital that athletes who improved their performance by training are not competitively disadvantaged by being placed into a less impaired class. Nevertheless, since performance level seems more dominated by impairment level rather than athlete training status or competition level,[11] performance clusters could be used to outline the number of classes needed in a particular system.

Once extracted from the *match* specific confounders, field test wheelchair mobility performance data could be enforced to argue for a reduced number of classifications. Based on TwoStep clustering, only two performance clusters appeared. In clustering, outcomes related to forward speed and acceleration showed to be dominant factors. The two clusters show much similarity with the current classification of athletes, with only one athlete of the low-class group assigned to cluster B. The remaining athletes of the low classified group were assigned to cluster A, but this cluster also comprised four athletes of the mid-class and two of the high-class group. In the population measured, athletes from both international and national competition level were included. The mid and high classified athletes assigned to cluster A were national males (n=4) and international females (n=2). In future research, a more homogenic group of athletes regarding competition level might slightly alter TwoStep cluster analysis outcomes.

Only regarding wheelchair mobility performance, a single separation between the current class 1-1.5 athletes and the rest would be adequate. Subsequently, the 2+ class athletes could be divided into two groups given the effect of their impairment regarding ball handling. Such a reduced number of classes is in line with the conclusion of Vanlandewijck et al.[5] and Molik et al.[7], pinpointing the viability of a reduction in the number of classes. A reduction in classes is also in line with the idea that the range of activity limitation within a class should also be as large as possible without disadvantaging those most severely impaired.[1] The wheelchair basketball specific field test used, is more closely related to match mobility performance than general performance measures (such as a physical fitness test or Wingate Anaerobic test) frequently used in earlier research, so it provides more match specific functional outcomes.

The aim of this study was to provide insight in the relationship between impairment and mobility performance in both *best* and *match* condition, and to demonstrate the additional value of objective measures as provided by new technologies. Although the current classification system functions, with athletes and coaches generally satisfied,[23] there still remains some controversy about the best approach to determine function level. The International Wheelchair Basketball Federation does not want to discard a reasonable well-functioning classification system based on years of gradual improvement, whereas the IPC seeks unity in systems over all sports, with selective classification based on “physical and technical assessment” off court. Given that aspiration, the wheelchair mobility performance method used in this research seems unsuitable as a direct classification tool. Still, the need for sport specific test batteries to aid the classifiers in objective decision making is emphasised by Tweedy et al.[1] They state that current classification systems are still based on the judgement of a small number of experienced classifiers, rather than on empirical evidence, making the validity of the systems often questionable. In wheelchair basketball, the classification method is also time consuming and complicated. The use of objective measurement methods and sport specific field tests can aid classifiers in their decision making. Results of the present study show the significance of on court mobility performance measurements, whereas the ease of use of the inertial sensor based method enables big scale measurements in the future. By using the same method in both conditions, results of continued measurements in *match* play will also approximate *best* performance (field test), reducing the effect of random factors typical to the observation of only a few matches as in the classification current system. Indeed, it also brings to light whether athletes intentionally show a misrepresentation of their abilities in the classification tests, a major issue in

Paralympic sports.

Practical Applications

The wheelchair basketball specific field test used in this study,[12] proved to be reliable combined with the inertial sensor based method for measuring wheelchair mobility performance. In that sense, it complies to the IPC appeal to develop sport specific test batteries for classification support. Next to use for classification support, the field test is also a useful tool for individual athletes and coaches. Given the magnitudes of the smallest detectable differences for all 6 outcomes, the field test is expected to be sensitive enough to detect performance changes as a result of training or interventions regarding wheelchair settings. Additional body fixed inertial sensors could be used for more profound insight in the relationship between body movement (“volume of action”) and wheelchair mobility performance.

Conclusion

Technological advancement, especially application of inertial sensors, allows for easy to use, large scale, objective and increasingly precise measurement of performance. Those benefits enable data science in adapted sports research that is traditionally characterized by small participant numbers. Such a big data approach with continued measurements in all conditions might offer an alternative point of view for classification outlining in Paralympic sports. Future research with additional body fixed inertial sensors might reveal more insight in the relationship between impairment and performance, bridging the gap to the selective classification envisioned by the IPC.

Acknowledgement

The authors would like to thank all athletes participating in this research project, all coaches involved in measurement planning, Yves Vanlandewijck, Kees van Breukelen and Frank van der Meulen for their contribution to this research and manuscript.

Conflict of interest statement

None.

References

1. Tweedy SM, Vanlandewijck YC. International paralympic committee position stand--background and scientific principles of classification in paralympic sport. *Br J Sports Med*. 2011;45(4):259-269.
2. Vanlandewijck YC, Evaggelinou C, Daly DJ, Verellen J, Van Houtte S, Aspeslagh V, Zwakhoven B. The relationship between functional potential and field performance in elite female wheelchair basketball players. *J Sports Sci*. 2004;22(7):668-675.
3. Pickering Francis L. Competitive sports, disability, and problems of justice in sports. *Journal of the Philosophy of Sport*. 2005;32(2):127-132.
4. DePauw KP & Gavron SJ. Disability and Sport. *Champaing, Il.: Human Kinetics*. 1995.
5. Vanlandewijck YC, Spaepen AJ, Lysens RJ. Relationship between the level of physical impairment and sports performance in elite wheelchair basketball athletes. *Adapted Physical Activity Quarterly*. 1995;12(2):139-150.
6. Altmann VC, Hart AL, Vanlandewijck YC, van Limbeek J, van Hooff ML. The impact of trunk impairment on performance of wheelchair activities with a focus on wheelchair court sports: A systematic review. *Sports medicine-open*. 2015;1(1):22.
7. Molik B, Laskin JJ, Kosmol A, Marszałek J, Morgulec-Adamowicz N, Frick T. Relationships between anaerobic performance, field tests, and functional level of elite female wheelchair basketball athletes. *Human Movement*. 2013;14(4):366-371.
8. Van Der Slikke R, Berger M, Bregman D, Lagerberg A, Veeger H. Opportunities for measuring wheelchair kinematics in match settings; reliability of a three inertial sensor configuration. *J Biomech*. 2015;48(12):3398-3405.
9. Mason BS, van der Woude, Lucas HV, Goosey-Tolfrey VL. The ergonomics of wheelchair configuration for optimal performance in the wheelchair court sports. *Sports Medicine*. 2013;43(1):23-38.
10. Byrnes D, Hedrick B, Hedrick B, Byrnes D, Shaver L. Comprehensive basketball grading system. *Wheelchair Basketball (edited by B.Hedrick, D.Byrnes and L.Shaver)*. 1994:79.
11. van der Slikke R, Berger M, Bregman D, Veeger H. From big data to rich data: The key features of athlete wheelchair mobility performance. *J Biomech*. 2016;49(14):3340-3346.
12. de Witte AM, Hoozemans MJ, Berger MA, van der Slikke, Rienk MA, van der Woude, Lucas HV, Veeger D. Development, construct validity and test–retest reliability of a field-based wheelchair mobility performance test for wheelchair basketball. *J Sports Sci*. 2017:1-10.
13. Pansiot J, Zhang Z, Lo B, Yang G. WISDOM: Wheelchair inertial sensors for displacement and orientation monitoring. *Measurement Science and Technology*. 2011;22(10):105801.
14. Fuss FK. Speed measurements in wheelchair sports–theory and application. *Sports Technology*. 2012;5(1-2):29-42.

15. van der Slikke R, Berger M, Bregman D, Veeger H. Wheel skid correction is a prerequisite to reliably measure wheelchair sports kinematics based on inertial sensors. *Procedia Engineering*. 2015;112:207-212.
16. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: A guide for non-statisticians. *International journal of endocrinology and metabolism*. 2012;10(2):486-489.
17. Field A. *Discovering statistics using IBM SPSS statistics.* Sage; 2013.
18. Fraley C, Raftery AE. How many clusters? which clustering method? answers via model-based cluster analysis. *The computer journal*. 1998;41(8):578-588.
19. Bacher J, Wenzig K, Vogler M. SPSS TwoStep cluster-a first evaluation. 2004.
20. Mooi E, Sarstedt M. *Cluster analysis.* Springer; 2010.
21. Chiu T, Fang D, Chen J, Wang Y, Jeris C. A robust and scalable clustering algorithm for mixed type attributes in large database environment. *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining*.2001:263-268.
22. Schuetz T. *A concise guide to market research: the process, data and methods using IBM SPSS statistics*. 2011.
23. Molik B, J Laskin J, L Golbeck A, et al. The international wheelchair basketball federation's classification system: The participants'perspective. *Kinesiology: International journal of fundamental and applied kinesiology*. 2017;49(1):21-22.
24. Bloxham L, Bell G, Bhambhani Y, Steadward R. Time motion analysis and physiological profile of canadian world cup wheelchair basketball players. *Sports Med , Train Rehabil*. 2001;10(3):183-198.

Appendix I

The athlete's performance can be divided in physical performance, mobility performance and game performance. Physical performance only concerns the athlete, [24] whereas mobility performance is the measure for the combined wheelchair-athlete combination.[9] Therefore, although mobility performance is established by athlete exertion, it is often expressed in terms of wheelchair kinematics. Van der Slikke et al.[11] used a set of three inertial sensors to measure the wheelchair kinematics of 29 athletes in wheelchair basketball match play. To reduce the vast number of kinematic outcomes that could be measured with this configuration, principal component analysis was used to extract a set of six key features describing ***wheelchair mobility performance*** characteristics. Three of these outcomes describe forward motion and three describe the rotational aspect (manoeuvrability). All outcomes are plotted in a radar plot, with a scale relative to the group average and standard deviation.

Appendix II

Reproducibility of wheelchair mobility performance outcomes in the field test was tested by measuring 23 male athletes twice.[12] Re-tests were performed one week after, under the same conditions (same timeframe, day of the week and same location). For each of the six performance outcomes the Intra Class Correlation coefficient for consistency (ICC_c) between test and re-test was calculated (Table 6). Based on the ICC_c value and Standard Deviation (SD), the Standard Error of Mean for consistency (SEMc) and the Smallest Detectable Difference (SDD 95%) were calculated using:

$$SEM_c = SD * \sqrt{(1 - ICC_c)}$$

$$SDD\ 95\% = SEM_c * \sqrt{2} * 1.96$$

The SDD 95% for each of the six performance outcomes is used to describe the differences between average performance of classification groups. For each outcome, the difference is divided by the SDD 95%, resulting in a dimensionless factor.

++ Please insert Table 6 here

Appendix III

The TwoStep Cluster Analysis procedure is an exploratory tool designed to reveal natural groupings (or clusters) within a data set that would otherwise not be apparent. It has several unique features that makes it very versatile. The most important feature for application in this study is the fact that it is capable of automatic selection of the number of natural clusters.

The two steps can be summarized as follows: Step 1) The procedure begins with the construction of a Cluster Features (CF) Tree. The tree begins by placing the first case at the root of the tree in a leaf node that contains variable information about that case. Each successive case is then added to an existing node or forms a new node, based upon its similarity to existing nodes and using the distance measure as the similarity criterion. A node that contains multiple cases contains a summary of variable information about those cases. Thus, the CF tree provides a capsule summary of the data file. Step 2) The leaf nodes of the CF tree are then grouped using an agglomerative clustering algorithm. The agglomerative clustering can be used to produce a range of solutions. To determine which number of clusters is "best", each of these cluster solutions is compared using the Schwarz's Bayesian Criterion (BIC).

In this study, for each of the forty-seven athletes, six wheelchair mobility performance outcomes are included in the dataset for clustering. The TwoStep clustering procedure reveals the number of natural clusters and the assignment of each athlete to a cluster. To quantify the "goodness" of a cluster solution, the silhouette coefficient is used. This coefficient indicates how well the elements within a cluster are similar to one (cohesive) while the clusters themselves are different (separated). The TwoStep analysis also indicates which of the data (six wheelchair mobility performance outcomes) was of most importance for clustering. The factor for importance to the model prediction can range from 0 (unimportant) to 1 (most important). This information helps to gain insight in the bases for the clustering model, and the contribution of each performance outcome.

Table 1. The distribution of classification and age (years) per competition level group of athletes measured in the field test.

Level		Mean	SD	Classification 1.0	1.5	2.0	2.5	3.0	4.0	4.5
National Male (NM)	Class	3.3	1.2	2	1	1	1	2	7	4
	Age	23.7	10.1							
International Male (IM)	Class	3.0	1.2	2	1	1	4	3	2	4
	Age	26.4	7.8							
International Female (IF)	Class	2.8	1.2	1	2	1	2	3	1	2
	Age	32.9	8.0							
Total				5	4	3	7	8	10	10
Group total				Low = 9		Mid = 18			High = 20	

Table 2. Classification group differences in the field test expressed as a factor of the Smallest Detectable Difference (SDD, see Appendix I).

	SDD 95%	Low - Mid	Low - High	Mid - High
Forward speed avg. (m/s)	0.038	***6.2***	***6.5***	0.3
Forward speed best (m/s)	0.046	***5.2***	***6.2***	1.0
Forward acceleration avg. (m/s^2)	0.085	***5.3***	***6.0***	0.6
Rotational speed curve avg. (°/s)	3.409	***2.0***	***2.0***	0.0
Rotational speed turn best (°/s)	12.065	***1.5***	***1.3***	0.2
Rotational acceleration avg. ($°/s^2$)	18.740	***5.5***	***5.5***	0.0

Notes: Factors of SDDs over 1 are marked bold

Table 3. Classification group statistics in the match and field test data.

	Match				Field Test			
	ANOVA	Bonferroni post-hoc			ANOVA	Bonferroni post-hoc		
		Low - High	Low - Mid	Mid - High		Low - High	Low - Mid	Mid - High
Forward speed avg. (m/s)	***0.000***	***0.000***	***0.021***	0.214	***0.000***	***0.000***	***0.000***	1.000
Forward speed best (m/s)	***0.000***	***0.000***	0.993	***0.003***	***0.000***	***0.000***	***0.003***	1.000
Forward acceleration avg. (m/s^2)	***0.001***	***0.001***	0.139	0.105	***0.003***	***0.003***	***0.010***	1.000
Rotational speed curve avg. (°/s)	***0.002***	***0.004***	***0.007***	1.000	***0.009***	***0.012***	***0.016***	1.000
Rotational speed turn best (°/s)	***0.003***	***0.004***	***0.013***	1.000	0.068	0.146	0.078	1.000
Rotational acceleration avg. ($°/s^2$)	***0.006***	***0.005***	0.115	0.443	***0.002***	***0.003***	***0.004***	1.000

Notes: Significance levels are shown, with all levels $p<0.05$ marked bold. Result description is based on adjacent class groups, that is between low-mid and between mid-high. Differences between the low and high classified athletes are obvious and not used in further interpretation of results.

Table 4. Pearson correlation and mean differences between match and field test performance (n=12);

	Pearson correlation	Mean diff.	p value T-Test
Forward speed avg. (m/s)	***0.735***	0.42	**0.000**
Forward speed best (m/s)	***0.756***	-0.19	**0.001**
Forward acceleration avg. (m/s^2)	***0.702***	0.92	**0.000**
Rotational speed curve avg. (^{o}/s)	***0.721***	1.70	0.221
Rotational speed turn best (^{o}/s)	0.616	0.60	0.936
Rotational acceleration avg. (^{o}/s^2)	***0.745***	64.0	**0.002**

Notes: all Pearson correlations were significant ($p<0.05$), >0.7 marked bold; if match performance exceeds test outcomes, a negative value is shown in the mean difference; significance levels <0.05 in the T-test are marked bold.

Table 5. The TwoStep clustering method applied to the dataset of the 47 athletes measured in the field test revealed two clusters (A & B). The table shows the distribution of athlete's classification over the two clusters. cluster performance characteristics and their differences.

Class	Cluster A	Cluster B	mean diff	Factor SDD 95%	p value T-Test
Low	8	1			
Mid	4	14			
High	2	18			
Total	14	33			
Forward speed avg. (m/s)	1.87	2.13	0.26	6.83	**0.000**
Forward speed best (m/s)	2.60	2.90	0.30	6.51	**0.000**
Forward acceleration avg. (m/s^2)	1.97	2.60	0.63	7.37	**0.000**
Rotational speed curve avg. (^{o}/s^2)	64.5	71.9	7.4	2.16	**0.000**
Rotational speed turn best (^{o}/s^2)	193.9	213.9	20.0	1.66	**0.001**
Rotational acceleration avg. (^{o}/s^2)	307.3	404.7	97.4	5.20	**0.000**

Notes: If optimized for group size (most athletes per class in each cluster), there is a clear split (dashed line) between the low and mid/high classification groups. The lower part of the table shows the wheelchair mobility performance outcomes per cluster and their difference, also expressed as a factor of the SDD 95% (Appendix I).

Table 6. ICC, SEM and SDD 95% of wheelchair mobility performance outcomes measured twice in the standardized field test.

	ICC	SD	SEM	SDD 95%
Forward speed avg. (m/s)	**0.947**	0.059	0.014	0.038
Forward speed best (m/s)	**0.947**	0.072	0.016	0.046
Forward acceleration avg. (m/s^2)	**0.950**	0.138	0.031	0.085
Rotational speed curve avg. (^{o}/s)	0.870	3.41	1.23	3.41
Rotational speed turn best (^{o}/s)	0.837	10.78	4.35	12.07
Rotational acceleration avg. (^{o}/s^2)	**0.944**	28.57	6.76	18.74

Figure 1. Measurement setup, with inertial sensors on wheels and frame and measurements during a match. (Photograph by www.frankvanhollebeke.be).

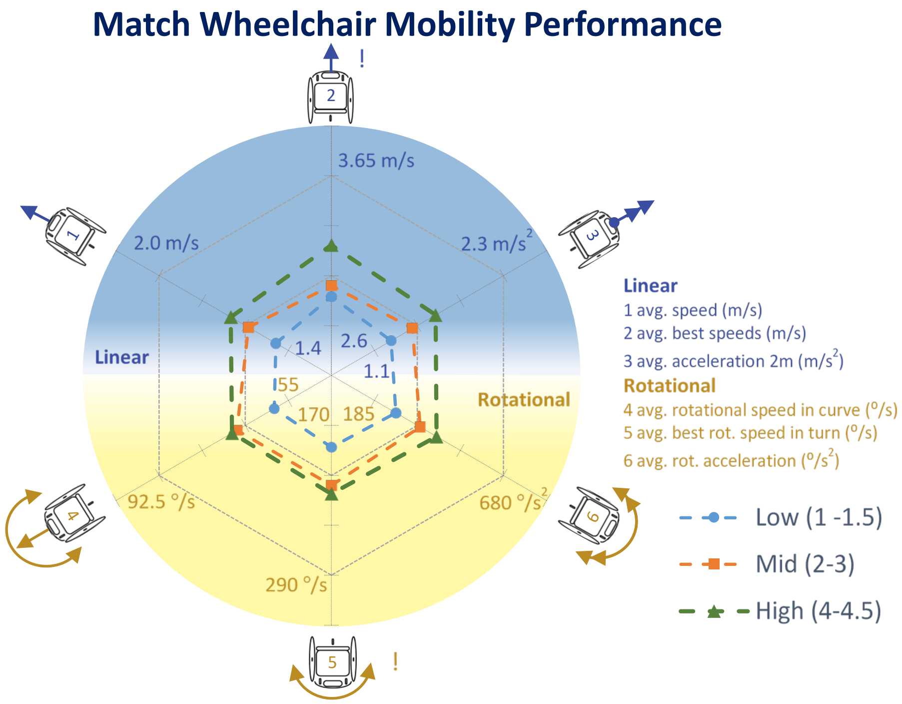

Figure 2. Wheelchair mobility performance in a match for three classification groups, adapted from van der Slikke et al., 2016.

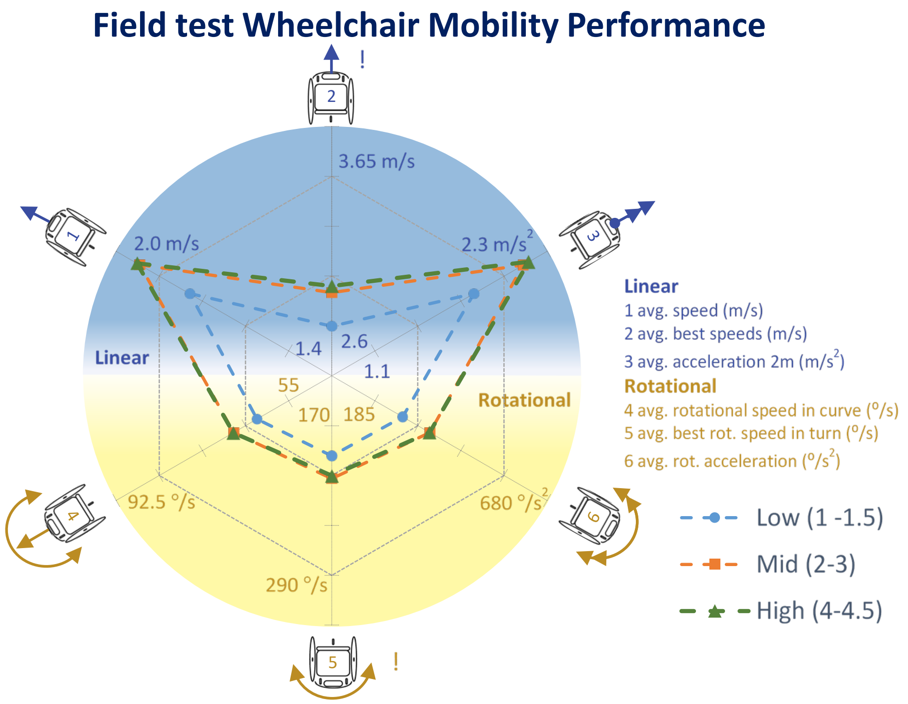

Figure 3. Best possible wheelchair mobility performance as measured in the field test for three classification groups.