SPINE An International Journal for the study of the spine, Publish Ahead of Print

DOI: 10.1097/BRS.0000000000004266

The Function Assessment scale for Spinal Deformity (FASD): validity and reliability of a new clinical scale

Pieter Severijns, PhD[1,2,3], Thomas Overbergh, MS[1], Thijs Ackermans, PhD[1], Erica Beaucage-Gauvreau, PhD[1], Simon Brumagne, PhD[2], Kaat Desloovere, PhD[2,3], Lennart Scheys, PhD[1,4], Lieven Moke, PhD[1,4]

1. Institute for Orthopaedic Research and Training (IORT), Department of Development and Regeneration, Faculty of Medicine, KU Leuven, Leuven, Belgium
2. Department of Rehabilitation Sciences, KU Leuven, Leuven, Belgium
3. Clinical Motion Analysis Laboratory (CMAL), University Hospitals Leuven, Leuven, Belgium
4. Division of Orthopaedics, University Hospitals Leuven, Leuven, Belgium

Corresponding author:

Pieter Severijns

Research Orthopedie

Herestraat 49 3000 Leuven, Belgium;

Email: pieter.severijns@kuleuven.be;

The manuscript submitted does not contain information about medical device(s)/drug(s).

The IORT has received funding from KU Leuven C2 funds and Medtronic.

Relevant financial activities outside the submitted work: board membership, consultancy, grants, payments for lectures.

Mini abstract

Current evaluation methods in adult spinal deformity fail to assess functional impairments. Therefore, this study introduced a valid and reliable clinical scale, the Function Assessment scale for Spinal Deformity. By measuring function and balance, this scale can increase our insights on the impact of spinal deformity on functioning.

Abstract

Study design: Cross-sectional study

Objective: To develop and validate the Function Assessment scale for Spinal Deformity (FASD).

Summary of Background Data: Spinal malalignment impacts daily functioning. Standard evaluation of adult spinal deformity (ASD) is based on static radiography and patient-reported scores, which fail to assess functional impairments. A clinical scale, quantifying function and balance of patients with ASD, could increase our insights on the impact of ASD on functioning.

Methods: To develop the FASD, 70 ASD patients and 20 controls were measured to identify the most discriminating items of the Balance Evaluation Systems Test and Trunk Control Measurement Scale. Discussions between experts on the clinical relevance of selected items led to further item reduction. The FASD's discriminative ability was established between 43 patients and 19 controls, as well as between three deformity subgroups. For its responsiveness to treatment, ten patients were re-evaluated six months post-operatively. Concurrent validity was assessed through correlation analysis with radiographic parameters (Pelvic tilt; Sagittal vertical axis (SVA); Pelvic incidence minus lumbar lordosis (PI-LL); Coronal vertical axis) and patient-reported scores (Oswestry Disability Index; Scoliosis Research Society outcome questionnaire; Falls Efficacy Scale-International). Test-retest and interrater reliability were tested on two groups of ten patients using intraclass correlation coefficients (ICC).

Results: Patients with ASD, mainly with sagittal malalignment, scored worse compared to controls on FASD ($p<0.001$) and its subscales. No significant improvement was observed six months post-operatively ($p=0.758$). FASD correlated significantly to all patient-reported scores and to SVA and PI-LL. Reliability between sessions (ICC=0.97) and raters (ICC=0.93) was excellent. Subscales also showed good to excellent reliability, except FASD 1 on 'spinal mobility and balance' between sessions (ICC=0.71).

Conclusions: FASD proved to be a valid and reliable clinical scale for evaluation of functional impairments in ASD. Objective information on function and balance might ultimately guide physiotherapeutic treatment towards improved functioning.

Key words: Adult spinal deformity; Postural balance; Validation study; Reproducibility of results; Function Assessment scale for Spinal Deformity

Level of Evidence: 2

Introduction

Adult spinal deformity (ASD) is characterized by a three-dimensional malalignment of the spine[1]. This might cause a shift in the gravity line position within the base of support, leading to poor balance control[2] and increased risk of falling[3]. Standard evaluation in ASD consists of static radiographic imaging, to quantify spinal alignment during standing[4], and patient-reported outcome scores (PROMs) on disability and quality of life (QOL), to assess treatment outcomes and the impact of ASD on daily life[5]. A clinical scale quantifying the functional capacities and balance control of patients with ASD, which is lacking in the current evaluation methods, could increase our insights on the impact of ASD on functional deterioration, disability and increased fall risk.

In other pathologies affecting postural control, functional scales are widely accepted as diagnostic and follow-up tools in clinical practice to guide physiotherapeutic treatment and pre- and post-operative rehabilitation[6]. Recently, balance impairments in ASD were identified using the Balance Evaluation Systems Test (BESTest) and Trunk Control Measurement Scale (TCMS)[7]. Although their reliability in ASD has been established[8], the time cost associated with BESTest (27 items)[6,9], the strict focus of TCMS on trunk control during sitting[10] and the reported ceiling effects of these scales in the ASD population[8] make these scales less appropriate to use in ASD clinical practice. Diebo et al.[11] recently introduced the Dubousset Functional Test (DFT), containing four items ('Up and walking test', 'Steps test', 'Down and sitting test' and 'Dual tasking test') to measure global function in ASD during activities of daily living (ADL). However, no patient data, nor a validation of this scale is available yet[11], and the sole focus on gross motor skills might lead to a restricted profiling of the patient's functional abilities. The shortcomings of these existing scales stress the need for an ASD-specific scale that combines balance and global function, applicable in clinical practice. Reliable and valid information on function and balance could provide therapeutic cues for functional impairments. Consequently, this information might contribute to the development of evidence-based physiotherapy, including pre- and post-operative rehabilitation, which is currently lacking within ASD[12].

The aim of this study was to introduce and validate a novel ASD-specific clinical scale, namely the Function Assessment scale for Spinal Deformity (FASD), as well as test its reliability between sessions and raters.

Methods

Participants

Seventy patients with ASD and twenty controls (age >18y) were included. The study was approved by our institution's ethical commission (file-number 'S58082') and all subjects signed informed consent. Exclusion criteria for both groups were musculoskeletal lower limb disorders or neurological conditions. Patients were included from the outpatient spinal clinic in case of de novo degenerative scoliosis, progressive adolescent idiopathic scoliosis (AIS) into adulthood, hyperkyphosis or flat back deformity. Iatrogenic or post-traumatic deformity, previous spinal fusion or inability to walk 50 meters independently led to exclusion.

To control for the deformity heterogeneity, patients were subdivided into three groups based on sagittal alignment using the Scoliosis Research Society (SRS)-Schwab classification (pelvic tilt (PT), sagittal vertical axis (SVA) and pelvic incidence minus lumbar lordosis (PI-LL))[4]:

1. **ASD 1:** Decompensated sagittal malalignment (SVA>4cm with PI-LL>10° and/or PT>20°) ± coronal deformity;
2. **ASD 2:** Compensated sagittal malalignment (SVA<4cm with PI-LL>10° and/or PT>20°) ± coronal deformity;
3. **ASD 3:** Coronal malalignment (Cobb angle ≥20°) and non-pathological sagittal alignment (SVA<4cm with PI-LL<10° and PT<20°).

Thoracic kyphosis (TK), lumbar lordosis (LL) and coronal vertical axis (CVA) were also measured.

Ten patients (ASD 1: 5, ASD 2: 3, ASD 3: 2) received spinal corrective surgery for uncontrollable pain not responding to conservative care, decreased QOL and/or curve progression. To test the FASD's responsiveness to treatment, as well as the impact of multilevel fusion on function and balance, they were remeasured six months post-operatively. The upper instrumented vertebra varied from L2 to T5. The lower instrumented level was the pelvis.

Scale development

Full details on scale development are described in Appendix A (see Supplementary text, Supplemental Digital Content 1, FASD development, http://links.lww.com/BRS/B807). Briefly, the FASD was developed based on BESTest[9] and TCMS[10], selected through a narrative literature review and discussions within a team of spinal surgeons, physiotherapists and movement scientists. These scales were applied on seventy patients and twenty controls to identify discriminative items. A further qualitative appraisal by the expert panel, in terms of feasibility, safety and clinical relevance, led to further item reduction. Ultimately, one item on step ascent/descent, missing in both scales, was added and a scoring option was added to the sit-to-stand (STS) item of BESTest. Reaching distances or execution timings were adjusted (items 5,7,8,9,12&14) using a combination of receiver-operating characteristic (ROC) curves[13] and distribution histograms. Item scoring was based on BESTest and TCMS, with some modifications. Specifically, TCMS items, consisting of multiple sub-questions with separate ordinal scales, were combined into one ordinal scale each (items 1,2&3). The resulting FASD contains 14 items and three subscales: FASD 1 - spinal mobility and balance (3 items); FASD 2 - stability limits and sensory integration (5 items); FASD 3 – ADL and stability in gait (6 items). The maximum score is 56 with higher scores indicating better performance. The FASD can be found in Appendix B (see Supplementary text, Supplemental Digital Content 2, FASD, http://links.lww.com/BRS/B808).

FASD data collection

Twelve items were scored using BESTest/TCMS data. Items on step ascent/descent and STS were scored using videos, obtained during a motion analysis study which was also part of the larger ASD research project within our institution. 43/70 patients (ASD 1: 16; ASD 2: 14; ASD 3: 13) and 19/20 controls completed both the balance assessment (BESTest/TCMS) and motion analysis (step ascent/descent and STS), and were included to obtain full FASD data. (Figure 1)

Validity & reliability

Discriminative ability and responsiveness to treatment

FASD performance was compared between patients and controls, between deformity subgroups and between pre- and post-operative conditions.

Concurrent validity

Relationships within the patient group (n=43) were investigated between FASD and the ASD standard evaluation, i.e. radiography and PROMs. Radiographic parameters of interest were PT, SVA, PI-LL and CVA. Following PROMs were used: SRS outcome questionnaire (SRS-22r)[14], Oswestry Disability Index (ODI)[15] and Falls Efficacy Scale-International (FES-I)[16]. Relations between radiography and PROMs were also investigated.

Reliability analysis

FASD's test-retest reliability was assessed on ten subjects using BESTest/TCMS data, measured twice by the same rater. For step ascent/descent and STS, videos from repeated motion analyses were used (see 2.3). Repeated measurements occurred within two weeks to avoid alterations in clinical presentation, however never on the same day to avoid learning effects or investigator bias. Since the same rater performed both assessments, the resulting data represents a mixture of intrarater and test-retest reliability[13]. Interrater reliability was assessed on ten different patients. Three raters scored simultaneously but independently, while the instructions were provided by the same rater. Videos on step ascent/descent and STS were rated separately. (Figure 1)

Statistical analysis

Due to non-normality of a large part of the data, verified by the Shapiro-Wilk test, all statistical analyses were carried out using nonparametric methods. Differences in FASD performance, demographics, radiographic parameters and PROMs between patients and controls, as well as between controls and pre- and post-operative conditions, were assessed with an independent Mann-Whitney U test. Differences between deformity subgroups were analyzed using a Kruskal-Wallis test with post-hoc Bonferroni correction for multiple testing. Pre- to post-operative changes were evaluated using a related Wilcoxon signed rank test. Concurrent validity was analyzed through Spearman rank correlation coefficients (r_s).

Reliability between sessions and raters on scale- and subscale-level was determined by intraclass correlation coefficients (ICC) with a two-way random effects model for absolute agreement for single measurement (ICC(2,1)) (<0.50: poor; 0.50-0.75: moderate; 0.75-0.90: good; >0.90 excellent[13]). The standard error of measurement (SEM) and smallest detectable difference (SDD) were calculated according to following formulas: SEM=SDx√(1-ICC) and SDD=SEMx1.96√2.[13] Cronbach's alpha was used to define internal consistency (>0.70: strong)[13]. Reliability of items was assessed through free-marginal multi-rater kappa, allowing unrestricted distribution of cases into categories[17,18] (<0.41: poor; 0.41-0.60: moderate; 0.61-0.80: good; >0.81: excellent[19]). Percentages of agreement, as well as 95% confidence intervals (CI) for ICC and kappa scores were calculated. Bland-Altman plots were used to determine systematic bias between sessions and raters[20].

The significance level was set at $p<0.05$. Statistics were performed through SPSS 26 (IBM Corp. Armonk, NY, USA) and Randolph's online kappa calculator[17].

Results

Participants

The ASD group for FASD analysis, did not differ from controls on demographics, except for height. The ASD group, as well as all deformity subgroups, scored significantly worse on PROMs, except for ASD 3 on FES-I. Post-operatively, SRS-22r significantly improved, in contrast with ODI and FES-I. (Table 1)

Discriminative ability and responsiveness to treatment

FASD differed significantly between patients and controls (ASD: 45; Control: 52; p<0.001), as did all subscales. The FASD total (p<0.001) and subscale scores (FASD 1: p<0.001; FASD 2: p=0.006; FASD 3: p<0.001) discriminated between subgroups (i.e. ASD 1 vs controls; ASD 2 vs controls; ASD 1 vs ASD 3). (Table 2)

Surgical patients scored, both pre- and post-operatively, worse on FASD (Pre: 38.5; Post: 34.5; Control: 52; p<0.001) and all subscales compared to controls. Performance on FASD 1 decreased post-operatively (Pre: 10.0; Post: 8.0; p=0.020). (Table 2)

Concurrent validity

Total FASD correlated significantly with all PROMs, as well as with SVA and PI-LL. FASD 1 only showed significant correlations with SVA and FES-I. FASD 2 and 3 significantly correlated with all PROMS and radiographic parameters, except for PT. (Table 3)

Reliability analysis

Total FASD showed excellent reliability between sessions (ICC=0.97, p<0.001) and raters (ICC=0.93, p<0.001), as did all subscales, except for FASD 1 between sessions (ICC=0.71; p=0.012). SEM's and SDD's were low for FASD (Test-retest: SEM=1.15, SDD=3.18; Interrater: SEM=1.00, SDD=2.76) and its subscales. Strong internal consistency was observed for FASD (0.90), FASD 2 (0.84) and FASD 3 (0.82). (Table 4)

On item-level, moderate to excellent kappa scores between sessions were found for all items, except for "5b. Standing on foam–eyes closed" (κ=0.20, 40% agreement) and "7a. One leg stance–Left" (κ=0.33, 50% agreement). Interrater reliability on item-level was good to excellent. (Table 5)

Bland-Altman plots revealed no systematic bias between sessions or raters. (Figure 2)

Discussion

This study introduced the FASD and proved its valid and reliable use in the ASD population. This novel scale on balance and general function could improve our insights on the impact of ASD on daily functioning.

Discriminative ability of the FASD and its subscales was established through the observed significant differences between patients and controls. Subgroup analysis showed that only sagittally deformed patients (ASD 1&2) differed from controls, in contrast with scoliosis patients with non-pathological sagittal alignment (ASD 3). These results suggest that sagittal malalignment and its compensations are associated with decreased function and balance, complementing their previously established association with pain and disability[21]. Responsiveness to treatment was assessed by remeasuring ten patients six months after spinal corrective surgery. However, since no significant differences were found post-operatively, responsiveness to treatment of the FASD could not be established. Only FASD 1 showed decreased post-operative performance, possibly due to decreased spinal mobility after fusion. These findings are in line with Laratta et al., reporting no significant improvement on the Berg Balance Scale six months after spinal fusion[22]. Also, the fact that surgical patients showed no significant improvements on the functional PROMs (ODI and FES-I), supports the assumption that functional abilities remained indeed decreased after surgical deformity correction. Moreover, the minimal clinical important difference (MCID) of 10.1 on ODI[23] was not reached post-operatively. Therefore, this lack of post-operative functional improvement cannot be attributed entirely to a decreased responsiveness to treatment of the FASD. Given the clinical importance of persisting functional impairments from pre- to post-operatively, the impact of multilevel fusion on function and balance should be further investigated in larger surgical samples.

Due to the relative novelty of functional assessment in ASD, defining FASD's concurrent validity is challenging. Radiographic analysis and PROMs form the cornerstone of clinical decision-making, and were therefore correlated to FASD. Interestingly, FASD related more to PROMs than radiographic parameters. This corresponds to previous findings on BESTest in ASD, which showed to be a better predictor of QOL than radiographic parameters[7], and supports the current shift within ASD from static alignment measurement towards a more comprehensive assessment, including dynamic function[25]. Significant correlations with FES-I and ODI suggest that FASD measures aspects of function relevant to the patient's own beliefs of his/her balance, safety and disability. In accordance, previous literature showed that the BESTest[9] significantly correlated with the Activities-specific Balance Confidence Scale, which similarly to the FES-I measures a person's confidence to keep balance during ADL.

Reliability between sessions and raters showed to be good to excellent for total score and subscales, except for FASD 1 between sessions. This is in line with the original reliability of TCMS in cerebral palsy[10] and BESTest in a heterogeneous population with and without balance disorders[9]. The FASD's reliability showed to be better or equal to that of TCMS and BESTest in ASD[8], with higher or comparable ICC's (Test-retest: ICC_{FASD}=0.97, $ICC_{BESTest}$=0.90, ICC_{TCMS}=0.88; Interrater: ICC_{FASD}=0.93, $ICC_{BESTest}$=0.94, ICC_{TCMS}=0.76) and smaller SEM's (Test-retest: SEM_{FASD}=1.15, $SEM_{BESTest}$=2.19, SEM_{TCMS}=1.66; Interrater: SEM_{FASD}=1.00, $SEM_{BESTest}$=2.32, SEM_{TCMS}=2.35) and SDD's (Test-retest: SDD_{FASD}=3.18, $SDD_{BESTest}$=8.30, SDD_{TCMS}=4.61; Interrater: SDD_{FASD}=2.76, $SDD_{BESTest}$=6.43, SDD_{TCMS}=6.52)[8]. The lower test-retest reliability of FASD 1 (ICC=0.71) was still higher compared to that of TCMS' 'dynamic reaching' subscale in ASD (ICC=0.27), on which FASD 1 was based[8]. This suggests that pooling the subquestions of the selected TCMS items into three FASD items, with one ordinal scale each (item 1,2&3), improved reliability. Small differences in sitting position could have decreased these item's reliability between sessions[8], since the zero position depends on the subject's most upright position, possibly influenced by fatigue[26]. Challenging items, such as standing on a foam with eyes closed (item 5b) and one leg stance (item 7), showed decreased reliability between sessions, but not between raters, suggesting greater influence of day-to-day performance. Besides improved reliability, also the reported ceiling effects of BESTest and TCMS in ASD[7,8] were addressed by FASD, illustrated in Appendix C (see Supplementary text, Supplemental Digital Content 3, histograms on data distribution of FASD, BESTest and TCMS, http://links.lww.com/BRS/B809). FASD showed an increased variability in total scores and a significantly decreased mean score compared to TCMS and BESTest.

Limitations

A first limitation is the small sample size of the surgical group. Future studies on larger samples are needed to prove FASD's ability to guide (physiotherapeutic) treatment. Second, balance scales are vulnerable to the assessor's subjective interpretation. However, good to excellent interrater reliability and its simplicity, in terms of material and time cost, make FASD clinically more interesting to measure function than advanced tools, such as motion analysis. Third, despite the small sample and the use of retrospective data, reliability showed to be comparable to other functional scales. Fourth, this study did not control for the presence of spinal stenosis, which might impact functional performance[36]. Finally, free-marginal kappa, used to assess reliability on item-level, does not take the size of the rating difference into account. Consequently, one- and two-point differences between raters or sessions resulted in the same reliability scores. However, on an ordinal scale, one-point differences can be considered to be less incorrect.

Future research perspectives

Although this study showed the reliability and validity of the FASD to measure function and balance, the underlying mechanisms of functional impairment in patients with ASD remain unclear. Poor to moderate correlations between FASD and radiographic parameters, as well as impaired FASD performance after surgical deformity correction, indicate that the skeletal deformity only partly explains functional impairment. Consequently, future research is required to further reveal the underlying mechanisms of decreased FASD performance. As observed in related pathologies, such as AIS[27,28] and LBP[29], further research should explore proprioceptive disorders. Function and balance assessment, combined with measurements of central and peripheral proprioceptive signal processing[30,31], could increase our insights on the complex interaction between ASD and postural control. Previous research in ASD on muscle structures reported decreased trunk[32] and lower limb[33] strength, sarcopenia[34] and increased fatty infiltration in spinopelvic muscles[35]. These findings on muscle structures should be linked to functional FASD performance. Eventually, these insights could result in treatments targeting the measured functional impairments to improve the patient's daily functioning.

Conclusions

This study introduced the FASD, a novel scale to measure function and balance in the ASD population. The FASD showed to be reliable between sessions and raters. Its discriminative ability and its relation to radiographic parameters and PROMs support the validity of the FASD to be used in the ASD population. Future research should investigate the causes of balance impairments in ASD, ultimately leading to treatments directly targeting these impairments.

Acknowledgements

We thank Kristel Van de Loock and Rowie De Buysscher for subject recruitment and follow-up, as well as Tereza Doškářová, Boris De Koninck and Viktor Weuts for their contribution to the data collection.

Key points

- The Function Assessment scale for Spinal Deformity, as well as its subscales, showed to be reliable between sessions and raters.
- Good discriminative ability and significant correlations with radiographic parameters and patient-reported disability, risk of falling and quality of life, support the validity of the scale in the adult spinal deformity population.
- Sagittal malalignment relates more to decreased function and balance compared to purely coronal malalignment.
- Six months after surgical correction of the deformity, performance on the Function Assessment scale for Spinal Deformity was not significantly improved.

References

1. Ailon T, Smith JS, Shaffrey CI, et al. Degenerative spinal deformity. *Neurosurgery* 2015;77:S75–91.

2. Le Huec JC, Saddiki R, Franke J, et al. Equilibrium of the human body and the gravity line: the basics. *Eur Spine J* 2011;20:1–6.

3. Godzik J, Frames CW, Smith Hussain V, et al. Postural stability and dynamic balance in adult spinal deformity: Prospective pilot study. *World Neurosurg*. Epub ahead of print 2020. DOI: 10.1016/j.wneu.2020.06.010.

4. Schwab F, Ungar B, Blondel B, et al. SRS-Schwab Adult Spinal Deformity Classification: A Validation Study. *Spine (Phila Pa 1976)* 2012;37:1077–82.

5. Smith JS, Shaffrey CI, Fu K-MG, et al. Clinical and radiographic evaluation of the adult spinal deformity patient. *Neurosurg Clin N Am* 2013;24:143–56.

6. Mancini M, Horak FB. The relevance of clinical balance assessment tools to differentiate balance deficits. *Eur J Phys Rehabil Med* 2010;46:239–48.

7. Moke L, Severijns P, Schelfaut S, et al. Performance on Balance Evaluation Systems Test (BESTest) impacts health-related quality of life in Adult Spinal Deformity Patients. *Spine (Phila Pa 1976)* 2017;1.

8. Severijns P, Overbergh T, Scheys L, et al. Reliability of the balance evaluation systems test and trunk control measurement scale in adult spinal deformity. *PLoS One* 2019;14:e0221489.

9. Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. *Phys Ther* 2009;89:484–98.

10. Heyrman L, Molenaers G, Desloovere K, et al. A clinical tool to measure trunk control

in children with cerebral palsy: The Trunk Control Measurement Scale. *Res Dev Disabil* 2011;32:2624–35.

11. Diebo BG, Challier V, Shah N V., et al. The Dubousset Functional Test is a Novel Assessment of Physical Function and Balance. *Clin Orthop Relat Res*. Epub ahead of print 2019. DOI: 10.1097/CORR.0000000000000820.

12. Glassman SD, Carreon LY, Shaffrey CI, et al. The costs and benefits of nonoperative management for adult scoliosis. *Spine (Phila Pa 1976)* 2010;35:578–82.

13. Portney LG, Watkins MP. *Foundations of Clinical Research: Application to Practice*. 2009.

14. Bridwell KH, Cats-Baril W, Harrast J, et al. The validity of the SRS-22 instrument in an adult spinal deformity population compared with the Oswestry and SF-12: A study of response distribution, concurrent validity, internal consistency, and reliability. *Spine (Phila Pa 1976)*. Epub ahead of print 2005. DOI: 10.1097/01.brs.0000153393.82368.6b.

15. Fairbank JC, Pynsent PB. The Oswestry Disability Index. *Spine (Phila Pa 1976)* 2000;25:2940–52; discussion 2952.

16. Tinetti ME, Richman D, Powell L. Falls efficacy as a measure of fear of falling. *Journals Gerontol*. Epub ahead of print 1990. DOI: 10.1093/geronj/45.6.P239.

17. Randolph J. Online Kappa Calculator [Computer software].

18. Wongpakaran N, Wongpakaran T, Wedding D, et al. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: A study conducted with personality disorder samples. *BMC Med Res Methodol*. Epub ahead of print 2013. DOI: 10.1186/1471-2288-13-61.

19. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics*. Epub ahead of print 1977. DOI: 10.2307/2529310.

20. Martin Bland J, Altman D. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *Lancet* 1986;1:307–10.

21. Glassman SD, Bridwell K, Dimar JR, et al. The impact of positive sagittal balance in adult spinal deformity. *Spine (Phila Pa 1976)* 2005;30:2024–9.

22. Laratta JL, Glassman SD, Atanda AA, et al. The Berg balance scale for assessing dynamic stability and balance in the adult spinal deformity (ASD) population. *J Spine Surg*. Epub ahead of print 2019. DOI: 10.21037/jss.2019.09.15.

23. Carreon LY, Bratcher KR, Canan CE, et al. Differentiating minimum clinically important difference for primary and revision lumbar fusion surgeries: Clinical article. *J Neurosurg Spine*. Epub ahead of print 2013. DOI: 10.3171/2012.10.SPINE12727.

24. Crawford CH, Glassman SD, Bridwell KH, et al. The minimum clinically important difference in SRS-22R total score, appearance, activity and pain domains after surgical treatment of adult spinal deformity. *Spine (Phila Pa 1976)*. Epub ahead of print 2015. DOI: 10.1097/BRS.0000000000000761.

25. Diebo BG, Shah N V., Pivec R, et al. From static spinal alignment to dynamic body balance: Utilizing motion analysis in spinal deformity surgery. *JBJS Rev* 2018;6:e3.

26. Bae J, Theologis AA, Jang J-S, et al. Impact of Fatigue on Maintenance of Upright Posture: Dynamic Assessment of Sagittal Spinal Deformity Parameters after Walking 10 Minutes. *Spine (Phila Pa 1976)*. Epub ahead of print September 2016. DOI: 10.1097/BRS.0000000000001898.

27. Simoneau M, Richer N, Mercier P, et al. Sensory deprivation and balance control in idiopathic scoliosis adolescent. *Exp Brain Res*. Epub ahead of print 2006. DOI: 10.1007/s00221-005-0246-0.

28. Le Berre M, Guyot MA, Agnani O, et al. Clinical balance tests, proprioceptive system and adolescent idiopathic scoliosis. *Eur Spine J*. Epub ahead of print 2017. DOI: 10.1007/s00586-016-4802-z.

29. Brumagne S, Janssens L, Knapen S, et al. Persons with recurrent low back pain exhibit a rigid postural control strategy. *Eur Spine J* 2008;17:1177–84.

30. Brumagne S, Janssens L, Knapen S, et al. Persons with recurrent low back pain exhibit a rigid postural control strategy. *Eur Spine J* 2008;17:1177–84.

31. Pijnenburg M, Caeyenberghs K, Janssens L, et al. Microstructural integrity of the superior cerebellar peduncle is associated with an impaired proprioceptive weighting capacity in individuals with non-specific low back pain. *PLoS One*. Epub ahead of print 2014. DOI: 10.1371/journal.pone.0100666.

32. Kobayashi T, Chiba H, Jimbo S, et al. Clinical, physical, and radiographic analyses of lumbar degenerative kyphosis and spondylolisthesis among community-based cohort. *Eur Spine J*. Epub ahead of print 2016. DOI: 10.1007/s00586-016-4615-0.

33. Cerpa M, Lenke LG, Fehlings MG, et al. Evolution and Advancement of Adult Spinal Deformity Research and Clinical Care: An Overview of the Scoli-RISK-1 Study. *Glob Spine J*. Epub ahead of print 2019. DOI: 10.1177/2192568219828729.

34. Ohyama S, Hoshino M, Terai H, et al. Sarcopenia is related to spinal sagittal

imbalance in patients with spinopelvic mismatch. *Eur Spine J*. Epub ahead of print 2019. DOI: 10.1007/s00586-019-06066-2.

35. Ferrero E, Skalli W, Lafage V, et al. Relationships between radiographic parameters and spinopelvic muscles in adult spinal deformity patients. *Eur Spine J*. Epub ahead of print 2020. DOI: 10.1007/s00586-019-06243-3.

36. Arbit E, Pannullo S. Lumbar stenosis: A clinical review. *Clin Orthop Relat Res*. Epub ahead of print 2001. DOI: 10.1097/00003086-200103000-00016.

Figure 1. Overview of the data collection. ASD: Adult Spinal Deformity; BESTest: Balance Evaluation Systems Test; TCMS: Trunk Control Measurement; FASD: Function Assessment scale for Spinal Deformity; STS: Sit-to-stand.

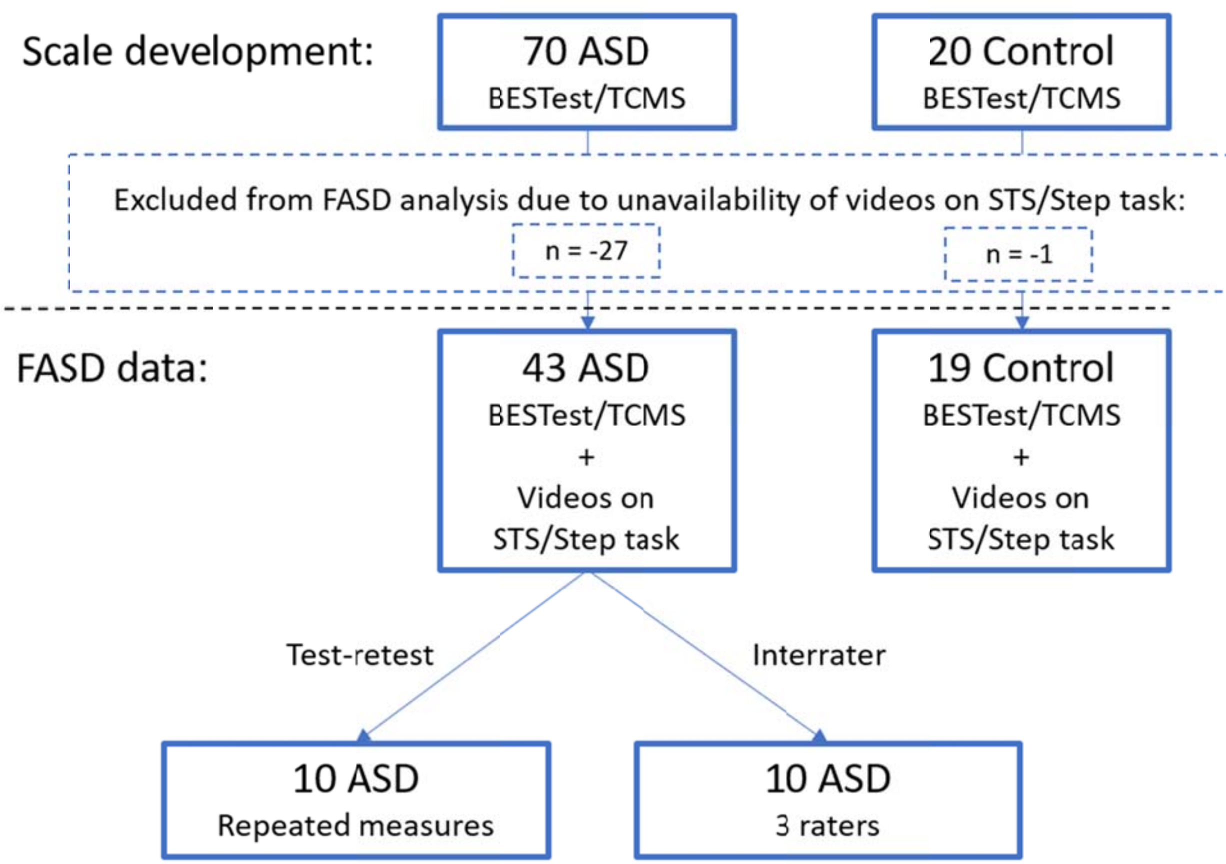


Figure 2. Bland-Altman plots for test-retest and interrater reliability of FASD.

On the horizontal axis the mean FASD score between sessions/raters is displayed. On the vertical axis the difference in scores between sessions/raters is plotted. The solid thin line represents the zero axis. The bold black line represents the mean difference. The dotted lines indicate the limits of agreement (1.96 x standard deviation). MD = Mean difference ± standard deviation; FASD: Function Assessment scale for Spinal Deformity

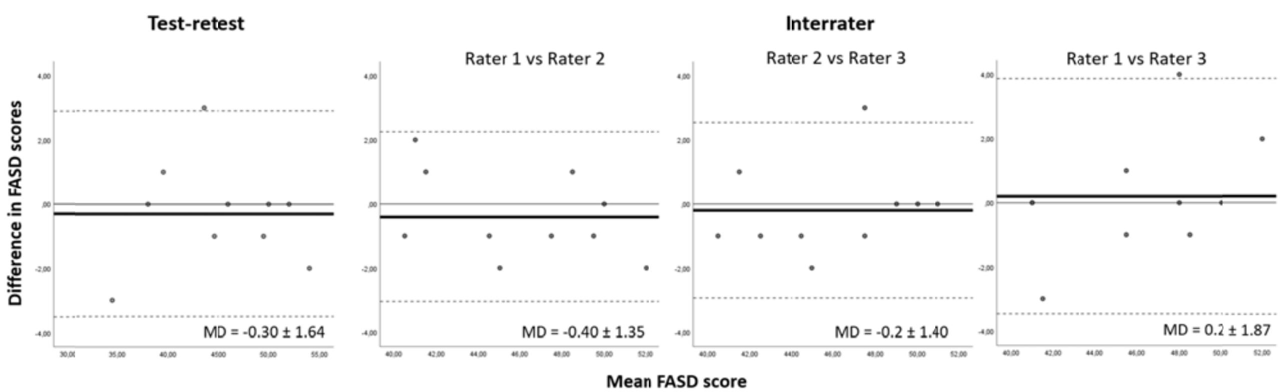


Table 1. Sample size, demographics, radiographic parameters and patient-reported outcome scores (PROMS) of FASD subgroup.

	A. Control vs ASD			**B. Pre- vs post-op**		*p -value*		
	Control (n=19)	ASD total (n=43)	*p-value*	Pre-op (n=10)	Post-op (n=10)	*Pre vs Post*	*Pre vs Control*	*Post vs Control*
A. Demographics								
Age (year)	65.0 (15.0)	61.0 (14.0)	0.708	64.2 (8.0)	64.2 (6.5)	0.180	0.839	0.982
Height (cm)	168.0 (8.5)	161.0 (9.0)	**0.022**	161.0 (10.8)	161.3 (10.4)	0.180	**0.014**	0.014
Weight (kg)	66.4 (15.0)	63.4 (14.4)	0.737	60.3 (22.8)	58.5 (22.7)	0.655	0.542	0.456
BMI (kg/m^2)	23.7 (3.4)	24.2 (5.2)	0.303	23.7 (7.9)	23.7 (7.9)	0.655	0.875	0.875
Gender (F/M)	13F/6M	36F/7M	0.178	8F/2M	8F/2M	1.000	0.789	0.789
B. Radiographic parameters								
PT (°)	20.7 (12.6)	19.7 (12.4)	0.951	17.4 (18.6)	17.5 (19.3)	0.959	0.573	0.330
SVA (mm)	5.7 (25.4)	28.2 (48.1)	**<0.001**	40.0 (125.5)	24.0 (56.1)	0.092	**0.001**	0.126
PI–LL (°)	-2.6 (13.3)	8.4 (23.3)	**0.022**	12.8 (31.4)	-2.35 (25.7)	**0.007**	**0.007**	0.701
Coronal (D/T/L/N)	19N	16D/23L/4N	**<0.001**	3D/5L/2N	10N	**<0.001**	**<0.001**	1.000
TK (°)	45.7 (13.8)	43.9 (22.7)	0.234	35.5 (29.1)	51.3 (25.1)	**0.017**	**0.019**	0.604
LL (°)	57.0 (18.4)	41.1 (32.3)	**0.001**	24.6 (24.2)	48.8 (14.5)	**0.007**	**<0.001**	**0.016**
CVA (mm)	8.3 (13.6)	17.6 (19.9)	**0.003**	25.1 (47.6)	15.6 (28.5)	0.285	**0.003**	**0.027**
C. PROMs								

SRS-22 (1-5)	4.6 (0.4)	3.3 (1.0)	**<0.001**	2.6 (1.2)	3.5 (0.6)	**0.008**	**<0.001**	**<0.001**
ODI (%)	0.0 (6.0)	32.0 (24.2)	**<0.001**	47.0 (31.0)	36.9 (27.3)	0.058	**<0.001**	**<0.001**
FES-I (16-64)	17.0 (3.0)	24.5 (17.0)	**0.001**	31.0 (24.0)	25.0 (15.0)	0.090	**0.002**	**<0.001**

C. Between group analysis

	Control (n=19)	ASD 1 (n=16)	ASD 2 (n=14)	ASD 3 (n=13)	*p-value*	*Post-hoc differences between groups* C vs 1	C Vs 2	C Vs 3	1 Vs 2	1 Vs 3	2 Vs 3
A. Demographics											
Age (year)	65.0 (15.0)	64.2 (14.5)	61.5 (11.8)	59.0 (16.0)	0.836						
Height (cm)	168.0 (8.5)	160.3 (8.0)	162.8 (11.3)	160.5 (12.8)	0.096						
Weight (kg)	66.4 (15.0)	63.8 (10.7)	64.7 (19.0)	59.6 (17.7)	0.953						
BMI (kg/m^2)	23.7 (3.4)	24.3 (6.0)	23.4 (6.2)	24.0 (4.6)	0.460						
Gender (F/M)	13F/6M	14F/2M	10F/4M	12F/1M	0.289						
B. Radiographic parameters											
PT (°)	20.7 (12.6)	23.7 (17.3)	25.5 (10.6)	17.2 (5.7)	**0.012**						0.006
SVA (mm	5.7 (25.4)	81.1 (65.8)	17.4 (28.8)	16.1 (11.4)	**<0.001**	<0.001			0.001	<0.001	

)											
PI–LL (°)	-2.6 (13.3)	16.4 (29.0)	14.1 (13.8)	-3.3 (6.6)	**<0.001**	0.017	0.024			0.004	0.006
Coronal (D/T/L/N)	19N	4D/10L/2N	6D/6L/2N	6D/7L	**<0.001**	<0.001	<0.001	<0.001			
TK (°)	45.7 (13.8)	44.2 (22.0)	43.0 (32.7)	43.5 (19.3)	0.699						
LL (°)	57.0 (18.4)	29.7 (21.2)	37.3 (33.1)	57.3 (11.1)	**<0.001**	<0.001	0.011			0.003	
CVA (mm)	8.3 (13.6)	19.4 (30.0)	16.9 (21.6)	13.8 (16.3)	**0.008**	0.008					
C. PROMs											
SRS-22 (1-5)	4.6 (0.4)	2.9 (1.3)	3.1 (1.3)	3.5 (0.4)	**<0.001**	<0.001	<0.001	<0.001			
ODI (%)	0.0 (6.0)	32.7 (35.1)	27.0 (27.0)	30.0 (19.0)	**<0.001**	<0.001	0.001	<0.001			
FES-I (16-64)	17.0 (3.0)	27.5 (21.0)	27.0 (18.5)	22.0 (9.0)	**0.004**	0.007	0.025				

Medians and interquartile ranges are reported; Significance level: $p < 0.05$.

FASD: Function Assessment scale for Spinal Deformity; ASD 1: Decompensated sagittal malalignment group; ASD 2: Compensated sagittal malalignment group; ASD 3: Coronal malalignment group with non-pathological sagittal alignment; BMI: Body Mass Index; F: Female; M: Male; PT: Pelvic tilt; SVA: Sagittal vertical axis; PI: Pelvic incidence; LL: Lumbar lordosis; Coronal: SRS-Schwab Coronal classification; D: Double; T: Thoracic; L: Lumbar; N: No major coronal deformity; TK: Thoracic kyphosis; CVA: Coronal vertical axis; PROMs: Patient-reported outcome scores; SRS-22: Scoliosis Research Society Outcomes Questionnaire; ODI: Oswestry Disability Index; FES-I: Fall efficacy scale.

Table 2. Between group differences on the Function Assessment scale for Spinal Deformity (FASD)

	A. Controls vs ASD			**B. Pre- vs post-op**		*p-value*		
	Control (n=19)	ASD total (n=43)	*p-value*	Pre-op (n=10)	Post-op (n=10)	*Pre vs Post*	*Pre vs Control*	*Post vs Control*
FASD total	52.0 (5.0)	45.0 (12.0)	**<0.001**	38.5 (10.3)	34.5 (8.5)	0.758	**<0.001**	**<0.001**
A FASD 1	13.0 (2.0)	11.0 (3.0)	**<0.001**	10.0 (3.3)	8.0 (4.5)	**0.020**	**0.001**	**<0.001**
B FASD 2	18.0 (4.0)	17.0 (8.0)	**0.044**	13.5 (7.3)	13.5 (4.5)	0.354	**0.002**	**0.001**
C FASD 3	20.0 (2.0)	17.0 (4.0)	**<0.001**	15.0 (3.5)	14.0 (2.3)	0.766	**<0.001**	**<0.001**

	A. Between group analysis					*Post-hoc differences between groups*					
	Control (n=19)	ASD 1 (n=16)	ASD 2 (n=14)	ASD 3 (n=13)	*p-value*	C vs 1	C Vs 2	C Vs 3	1 Vs 2	1 Vs 3	2 Vs 3
FASD total	52.0 (5.0)	39.0 (16.5)	42.4 (11.8)	49.0 (4.0)	**<0.001**	<0.001	0.027			0.020	
A FASD 1	13.0 (2.0)	10.0 (2.0)	12.0 (3.3)	12.0 (3.0)	**<0.001**	<0.001					
B FASD 2	18.0 (4.0)	14.0 (11.3)	14.5 (7.3)	19.0 (3.0)	**0.006**	0.022				0.026	
C FASD 3	20.0 (2.0)	14.5 (4.8)	18.0 (4.0)	19.0 (3.5)	**<0.001**	<0.001	0.039			0.009	

Medians (and interquartile ranges) are reported. Significance level: $p < 0.05$.

FASD 1: Spinal mobility and balance; FASD 2: Stability limits and sensory integration; FASD 3: ADL and stability in gait; Significance level: $p<0.05$. C: Control; 1: ASD 1; 2: ASD 2; 3: ASD 3.

Table 3. Concurrent validity by correlation analysis between FASD, radiographic parameters and HRQOL

A. FASD vs Rx	PT		SVA		PI-LL		CVA	
	r_s	p	r_s	p	r_s	p	r_s	p
FASD	-0.17	0.280	-0.60	**<0.001**	-0.42	**0.005**	-0.30	0.051
Subscales:								
• FASD 1	0.05	0.735	-0.38	**0.012**	-0.13	0.391	-0.15	0.923
• FASD 2	-0.29	0.062	-0.52	**<0.001**	-0.46	**0.002**	-0.32	**0.038**
• FASD 3	-0.16	0.302	-0.64	**<0.001**	-0.37	**0.016**	-0.45	**0.002**

B. FASD vs HRQOL	SRS-22r		ODI		FES-I	
	r_s	p	r_s	p	r_s	p
FASD	0.42	**0.005**	-0.58	**<0.001**	-0.59	**<0.001**
Subscales:						
• FASD 1	0.21	0.167	-0.25	0.053	-0.37	**0.018**
• FASD 2	0.36	**0.019**	-0.50	**<0.001**	-0.52	**0.001**
• FASD 3	0.38	**0.012**	-0.58	**<0.001**	-0.44	**0.004**

C. Rx vs HRQOL	SRS-22r		ODI		FES-I	
	r_s	p	r_s	p	r_s	p
PT	0.07	0.638	-0.07	0.338	0.13	0.419
SVA	-0.11	0.478	0.21	0.090	0.26	0.100
PI-LL	-0.21	0.177	0.22	0.081	0.28	0.085
CVA	-0.25	0.111	0.28	0.073	0.21	0.187

Spearman correlation coefficients are reported. Significance level: $p < 0.05$.

FASD: Function Assessment scale for Spinal Deformity; FASD 1: Spinal mobility and balance; FASD 2: Stability limits and sensory integration; FASD 3: Activities of daily living (ADL) and stability in gait; PT: Pelvic tilt; SVA: Sagittal vertical axis; PI-LL: Pelvic incidence minus Lumbar lordosis: CVA: Coronal Vertical Axis; SRS-22r: Scoliosis Research Society outcome questionnaire; ODI: Oswestry Disability Index; FES-I: Fall efficacy scale. r_s: Spearman correlation coefficient; p: p-value.

Table 4. Measures for test-retest and interrater reliability and internal consistency of FASD

	Internal consistency	Test-retest (n = 10)					Interrater (n = 10)				
	Cronbach's α	ICC	*p-value*	95% CI	SEM	SDD	ICC	*p-value*	95% CI	SEM	SDD
FASD (0-56)	0.90	0.97	**<0.001**	0.88-0.99	1.15	3.18	0.93	**<0.001**	0.81-0.98	1.00	2.76
A. Spinal mobility and balance (0-14)	0.68	0.71	**0.012**	0.12-0.93	0.75	2.08	0.91	**<0.001**	0.77-0.98	0.57	1.57
B. Stability limits and sensory integration (0-21)	0.84	0.87	**<0.001**	0.56-0.97	1.19	3.31	0.93	**<0.001**	0.81-0.98	0.51	1.42
C. ADL and stability in gait (0-21)	0.82	0.94	**<0.001**	0.78-0.99	0.60	1.66	0.98	**<0.001**	0.94-1.00	0.38	1.06

FASD: Function Assessment scale for Spinal Deformity; ADL: Activities of daily living; CI: Confidence interval; SEM: Standard error of measurement; SDD: Smallest detectable difference; Significance level: $p < 0.05$

Table 5. Test-retest and interrater reliability of FASD on item-level

	Test-retest (n = 10)			Interrater (n = 10)		
	Free marginal κ	95% CI	%	Free marginal κ	95% CI	%
A. Spinal mobility and balance						
1. a. Side bending – Left	0.60	0.20-1.00	70	0.73	0.47-1.00	80
b. Side bending – Right	0.60	0.20-1.00	70	0.73	0.47-1.00	80
2. a. Side reaching – Left	0.55	0.10-1.00	70	1.00	1.00	100
b. Side reaching – Right	1.00	1.00	100	0.90	0.70-1.00	93.33
3. a. Cross reaching / spinal rotation – Left	0.80	0.31-1.00	80	0.90	0.70-1.00	93.33
b. Cross reaching / spinal rotation – Right	0.25	-0.24-0.74	50	0.80	0.54-1.00	86.67
B. Stability limits and sensory integration						
4. Toe rise	0.73	0.38-1.00	80	0.91	0.74-1.00	93.33
5. a. Standing on foam – eyes open	1.00	1.00	100	0.91	0.74-1.00	93.33
b. Standing on foam – eyes closed	0.20	-0.23-0.63	40	0.91	0.74-1.00	93.33
6. Forward reactive response	0.60	0.20-1.00	70	0.73	0.47-1.00	80
7. a. One leg stance – Left	0.33	-0.10-0.77	50	0.64	0.36-0.93	73.33
b. One leg stance – Right	0.47	0.04-0.89	60	1.00	1.00	100
8. Forward reaching	0.47	0.04-0.89	60	0.91	0.74-1.00	93.33
C. ADL and stability in gait						
9. Sit on floor and stand up	0.60	0.20-	70	1.00	1.00	100

		1.00				
10. Sit-to-stand (chair)	0.73	0.38-1.00	80	1.00	1.00	100
11. a. Step ascent	0.87	0.61-1.00	90	1.00	1.00	100
b. Step descent	0.87	0.61-1.00	90	0.82	0.59-1.00	86.67
12. Overground walking	0.73	0.38-1.00	80	1.00	1.00	100
13. Walking with head turns	0.60	0.20-1.00	70	0.91	0.74-1.00	93.33
14. TUG-test	0.87	0.61-1.00	90	0.91	0.74-1.00	93.33

FASD: Function Assessment scale for Spinal Deformity; ADL: Activities of daily living; TUG-test: Timed Get-Up-and-Go-Test; K: Kappa; CI: Confidence interval; %: Percentage of agreement.

Appendix A. Development of the Function Assessment scale for Spinal Deformity (FASD)

A.1 Narrative literature review on function and balance assessment scales

A.2 Selection of function and balance assessment scales

A.3 Data collection on selected scales in ASD and control sample

A.4 Item selection and adjustments based on collected data and expert opinion

A.5 Function Assessment scale for Spinal Deformity – finalization

A.1 Narrative literature review on function and balance assessment scales

First, a narrative literature review of function and balance scales (published before January 2016) was performed, using the 'Pubmed' database (search terms: 'balance scale', 'balance assessment', 'dynamic balance', 'function scale', 'functional evaluation', 'physiotherapy'). This resulted in the identification of eight scales to measure balance and function, namely the Berg Balance scale (BBS)[1], the Trunk Impairment scale (TIS)[2], the Trunk Control Measurement Scale (TCMS)[3], the Physical Performance Test (PPT)[4], the Fullerton Advanced Balance Scale (FAB)[5] and the Balance Evaluation Systems Test (BESTest)[6], as well as its shorter versions, the Mini BESTest[7] and the Brief BESTest[8].

A.2 Selection of function and balance assessment scales

To decide which scales could potentially form the base of the new ASD-specific scale, first all subscales and items of the eight selected scales were screened and compared by two physiotherapists (PS and TD). Based on this screening the scales were categorized according to their scope.

Table A1. Scope of the selected function and balance assessment scales.

Gross motor function and balance	Trunk control and balance during sitting	Gross and fine motor function
BBS (13 items; 0-56)	TIS (17 items; 0-23)	PPT (9 items; 0-36)
FAB (10 items; 0-40)	TCMS (15 items; 0-58)	
BESTest (27 items; 0-100%)		
Mini BESTest (14; 0-28)		
Brief BESTest (6 items; 0-24)		

After this first screening, the PPT was eliminated from the selection because of its partial focus on fine motor function of the upper body (such as writing and eating) and its limited amount of items on gross motor function and balance[4].

The remaining scales for potential use in ASD were further studied and discussed by a team of experts within our institute in the field of ASD, including two surgeons, three physiotherapists and two movement scientists.

The TIS and TCMS were particularly of interest to apply in ASD because of their focus on the trunk. Both scales were developed to assess trunk control during sitting in neurological disorders, more specifically for patients after stroke (TIS)[2] and children with cerebral palsy (TCMS)[3]. A thorough investigation of both scales showed that the TCMS was more appropriate to use in ASD for two reasons. First, the TCMS evaluates three-planar motions within and beyond the base of support[3], in contrast with the TIS evaluating only biplanar motions within the base of support[2]. However, ASD is characterized by a three-dimensional deformity of the spine and therefore information on trunk control in all three planes seems warranted. Second, foot support is not allowed in the TCMS, which limits lower limb compensations[3], in contrast with the TIS, in which patients are tested while sitting with the feet supported[2]. These lower limb compensations during sitting might lead to ceiling effects in patients with ASD, and therefore the TCMS was chosen over the TIS.

The BBS and FAB showed a marked overlap with the BESTest. Respectively, 8 out of 14 for BBS and 8 out of 10 items for FAB were similar to items of the BESTest[1,5]. Three items of the BBS showed also similarity to the TCMS. (Table A2) In addition, one of the items of the FAB, namely 'two footed jump', is contraindicated in patients with osteoporosis, which is known to be highly prevalent in elderly with ASD[9]. For these reasons, the BBS and FAB were not retained. The Mini BESTest and Brief BESTest are directly derived from the BESTest, and consequently all items of these two scales are incorporated in the original BESTest[6–8]. Therefore, it was decided to select the BESTest, because of its extensiveness, covering a large spectrum of function and balance, and its possibility to identify balance disorders within different balance systems (biomechanical constraints, stability limits/verticality, anticipatory postural adjustments, postural responses, sensory orientation and stability in gait)[6].

Table A2. Overlap between BBS/FAB and BESTest/TCMS

Berg Balance Scale (BBS)	Overlap with	Fullerton Advanced Balance scale (FAB)	Overlap with
1. Sitting to standing	BESTest	1. Stand with feet together and eyes closed	BESTest
2. Standing unsupported	BESTest	2. Reach forward	BESTest
3. Sitting unsupported	TCMS	3. Turn 360 degrees	BESTest
4. Standing to sitting		4. Step up onto and over a bench	BESTest
5. Transfers (sit)		5. Tandem walk	
6. Standing with eyes closed	BESTest	6. Stand on one leg	BESTest
7. Standing with feet together	BESTest	7. Stand on foam with eyes closed	BESTest
8. Reaching forward with outstretched arm	BESTest/TCMS	8. Two-footed jump	
9. Retrieving object from floor		9. Walk with head turns	BESTest
10. Turning to look behind	TCMS	10. Reactive postural response	BESTest
11. Turning 360 degrees	BESTest		
12. Placing alternate foot on stool	BESTest		
13. Standing with one foot in front			
14. Standing on one foot	BESTest		

In conclusion, the BESTest[6] and the TCMS[3] were selected to apply on subjects with ASD and consequently form the base for an ASD-specific function and balance assessment scale.

A.3 Data collection on selected scales in ASD and control sample

A.3.1 Participants

Seventy patients with ASD were included an compared to a group of 20 healthy controls. (Table A3) Inclusion and exclusion criteria can be found in the paper manuscript.

The patient group was subdivided into three groups with different sagittal spinal alignment, according to the SRS-Schwab Classification:

1. **ASD 1 (n=29):** Decompensated sagittal malalignment (SVA>4cm with PI-LL>10° and/or PT>20°) ± coronal deformity;
2. **ASD 2 (n=24):** Compensated sagittal malalignment (SVA<4cm with PI-LL>10° and/or PT>20°) ± coronal deformity;
3. **ASD 3 (n=17):** Scoliosis (Cobb angle of ≥20°) and non-pathological sagittal alignment (SVA<4cm with PI-LL<10° and PT<20°).

Table A3. Demographics, radiographic parameters, PROMS of total group for FASD development

	Control (n=20)	ASD (n=70)	*p-value*
A. Demographics			
Age (year)	64.5 (15)	64.0 (14.0)	0.617
Height (cm)	167.5 (7.8)	160.5 (10.4)	**0.006**
Weight (kg)	66.9 (14.1)	65.0 (13.8)	0.965
BMI (kg/m^2)	23.9 (3.4)	25.2 (5.7)	0.051
Gender (F/M)	14/6	60/10	0.107
B. Radiographic parameters			
PT (°)	19.5 (11.9)	21.9 (14.2)	0.262
SVA (mm)	6.05 (25.0)	34.1 (50.2)	**<0.001**
PI minus LL (°)	-2.7 (13.0)	13.4 (27.4)	**0.001**
Coronal (D/T/L/N)	23D/39L/8N	20N	**<0.001**
TK (°)	45.7 (13.8)	41.0 (24.7)	**0.022**
LL (°)	57.0 (18.4)	41.7 (30.7)	**<0.001**
C. Patient-reported outcome scores			
SRS-22r (1-5)	4.6 (0.4)	3.2 (1.0)	**<0.001**
ODI (%)	0.0 (5.9)	32.7 (30.5)	**<0.001**

Medians and interquartile ranges are reported; Significance level: $p < 0.05$. FASD: Function Assessment scale for Spinal Deformity; BMI: Body Mass Index; F: Female; M: Male; PT: Pelvic tilt; SVA: Sagittal vertical axis; PI: Pelvic incidence; LL: Lumbar lordosis; Coronal: SRS-Schwab Coronal classification; D: Double; T: Thoracic; L: Lumbar; N: No Major Coronal Deformity; TK: Thoracic kyphosis; SRS-22: Scoliosis Research Society Outcomes Questionnaire; ODI: Oswestry Disability Index.

A.3.2 Data collection and processing

All subjects were evaluated on both the TCMS and the BESTest. All assessments were performed by the same physiotherapist (PS). To identify which items had the ability to discriminate between patients with ASD and controls a Mann-Whitney U test was performed. Subgroups with different deformity types within the patient group were compared using a Friedman test. All statistical analyses were performed with SPSS 26 (IBM Corp. Armonk, NY, USA) and significance level was set at $p<0.05$.

A.3.3 Discriminant ability of items of TCMS and BESTest

Table A4. Discriminant ability of TCMS items between ASD and controls and between ASD subgroups.

TCMS	ASD vs Control	Between groups
A. Static sitting balance		
1. Upright sitting		
2. Lift both arms		
3. Legs crossed by therapist		
4. Legs crossed by patient	X	
5. Leg abduction	X	X
B. Selective movement control (sitting)		
6. Forward lean		X
7. Backward lean		
8. Lateral lean (sidebending)	X	
9. Pelvic lift	X	
10. Upper trunk rotation	X	
11. Lower trunk rotation	X	X
12. Forward and backward pelvic shuffle	X	
C. Dynamic reaching (sitting)		
13. Forward reaching		
14. Lateral reach	X	X
15. Lateral cross-over reach	X	

Significant differences between groups are indicated with an X. Significance level: $p < 0.05$.

Table A5. Discriminant ability of BESTest items between ASD and controls and between ASD subgroups.

BESTest	ASD vs Control	Between groups
A. Biomechanical constraints		
1. Base of support		
2. COM alignment	X	X
3. Ankle strength and range of motion	X	X
4. Hip/trunk lateral strength (standing)	X	X
5. Sit on floor and stand up		X
B. Stability limits/Verticality		
6. Lateral lean during sitting and verticality	X	
7. Functional lean forward (standing reach)	X	
8. Functional lean lateral (standing reach)	X	
C. Transitions-Anticipatory postural adjustments		
9. Sit to stand		
10. Rise to toes	X	
11. Stand on one leg	X	X
12. Alternate stair touching		
13. Standing arm raise	X	

D. Reactive postural responses		
14. In place response – forward	X	
15. In place response – backward		
16. Compensatory stepping correction – forward		
17. Compensatory stepping correction – backward	X	
18. Compensatory stepping correction – lateral	X	
E. Sensory orientation		
19. Sensory integration (firm surface – foam / eyes open – eyes closed)	X	X
20. Incline standing - eyes closed	X	X
F. Stability in gait		
21. Walk on level surface	X	
22. Change in gait speed	X	
23. Walk with head turns	X	X
24. Walk with pivot turns	X	X
25. Step over obstacle	X	X
26. Timed "get up and go" test	X	
27. Timed "get up and go" test with dual task	X	

Significant differences between groups are indicated with an X. Significance level: $p < 0.05$.

A.4 Item selection based on collected data and expert opinion

The analysis of the discriminant ability of the individual items of both scales served as a first guideline for in- or exclusion of an item in the ASD-specific balance scale. The results (Table A4 and A5) showed that 9/15 items of TCMS and 21/27 items of BESTest were able to discriminate patients with ASD from controls.

At this time in the process, the expert group was consulted for the second time to evaluate the qualitative aspects of the specific items, such as the feasibility, safety and clinical relevance. Also, if items measured similar aspects of balance and function, only one item was retained (e.g. 'incline standing with eyes closed' vs 'foam standing with eyes closed' in BESTest are both items to test balance while occluding vision and changing proprioceptive information from the ankle joint).

This quantitative (discriminant ability) and qualitative analysis resulted in a first selection of three items of the TCMS and eight items of the BESTest. (Table A.6)

Table A6. First selection of items and categorization.

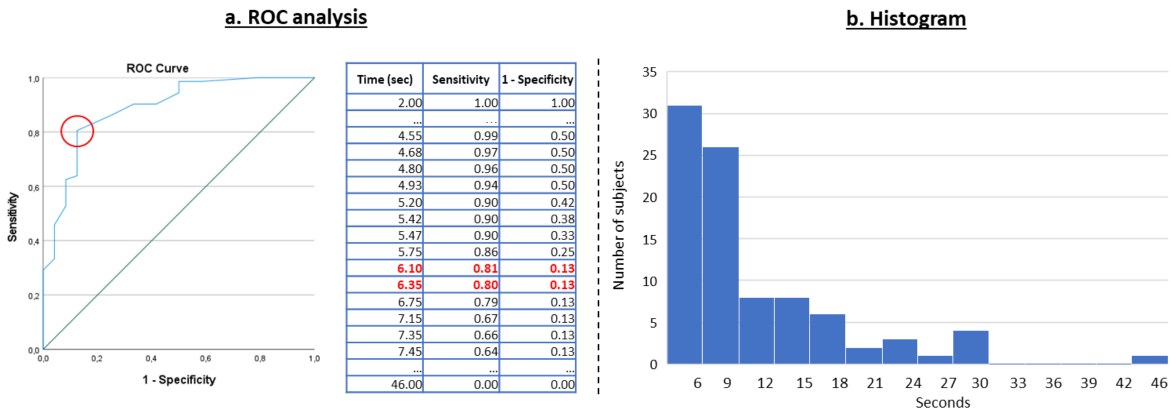
TCMS	BESTest
Spinal balance and mobility: 1. Lateral lean (sidebending sit) 2. Lateral reach (sit) 3. Lateral cross-over reach (sit)	Stability (limits) and sensory integration: 1. Functional lean forward (standing reach) 2. Rise to toes 3. Stand on one leg 4. In place response – forward 5. Sensory integration – foam standing Stability in gait: 6. Walk on level surface 7. Walking with head turns 8. Timed "get up and go" test

Although this first selection covered different aspects of function and balance, as indicated by the categories in table A6 ('Spinal balance and mobility'; 'Stability (limits) and sensory integration'; 'Stability in gait'), functional items on activities of daily living (ADL) were still lacking. An item on stair negotiation, which is considered an important ADL task determining an individual's independence, was not available in the BESTest. The 'alternate stair touching' item of BESTest, which most closely resembled stair negotiation, did however not discriminate between ASD and control. Therefor a step ascent/descent item, including two consecutive steps, was added to the selection.

Promising ADL items from the BESTest, such as 'sit to stand' and 'sit on the floor and stand up' could not discriminate between ASD and controls. Because of their functional character, the fact that multiple aspects of functionality are addressed by these items (e.g. strength, mobility, balance,…) and also other scales, such as the BBS and more recently the Dubousset Functional Test[10], included these items, we believed 'sit to stand' and 'sit on the floor and stand up' were important items to include in the FASD.

To increase the discriminant ability of these two items, the scoring options were adjusted based on the collected data. For 'sit to stand', a scoring option was added, to make a distinction between subjects that raised from a chair without hand support in one smooth movement and subjects that could also stand up without hand support but with some difficulty or hesitation. For the 'sit on the floor and stand up' item, the main reason for the observed ceiling effect, was the lack of a time aspect in the scoring options. Therefore, subjects that could independently sit down and stand back up without the use of a chair, scored all maximum, regardless of the time they needed to perform the task. Since these timings were collected for all subjects, these could be used to identify ASD-specific cut-off timings which could be incorporated in the scoring options.

The ASD-specific cut-off timings were identified using a combination of ROC-curves[11] and histograms for data distribution. With ROC-curves, the cut-off score was determined with the highest sensitivity (true positive rate), but the lowest 1-specificity (false positive rate). Histograms of data distribution were then used to check whether to round up or down towards the nearest half centimeter/second. Figure A1 gives an overview of these analyses for 'sit on the floor and stand up'.



Time (sec)	Sensitivity	1 - Specificity
2.00	1.00	1.00
...	...	...
4.55	0.99	0.50
4.68	0.97	0.50
4.80	0.96	0.50
4.93	0.94	0.50
5.20	0.90	0.42
5.42	0.90	0.38
5.47	0.90	0.33
5.75	0.86	0.25
6.10	**0.81**	**0.13**
6.35	**0.80**	**0.13**
6.75	0.79	0.13
7.15	0.67	0.13
7.35	0.66	0.13
7.45	0.64	0.13
...	...	...
46.00	0.00	0.00

Figure A1. ROC and histogram analysis to identify cut-off timing for 'sit on the floor and stand up'

In fig A1.a the ROC analysis revealed that the timing with the highest specificity and lowest 1-specificity was between 6 and 6.5 seconds. A histogram confirmed that approximately 30 subjects (1/3 of the total population) performed the task within 6 seconds, which was accepted to be a reasonable cut-off score.

To decrease the risk of ceiling effects in the total FASD, this exercise to adjust cut-off scores in the scoring options was performed for all items that included time or distance targets. (items 5,7,8,9,12&14; Appendix B)

Eventually, this resulted in a final selection of items from TCMS and BESTest, adjusted to the performances of the ASD population, complemented with one item on step ascent/descent. (Table A7).

Table A7. Final selection of items and categorization.

TCMS	BESTest
Spinal balance and mobility:	Stability (limits) and sensory integration:
1. Lateral lean (sidebending - sit)	1. Functional lean forward (standing reach)
2. Lateral reach (sit)	2. Rise to toes
3. Lateral cross-over reach (sit)	3. Stand on one leg
	4. In place response – forward
	5. Sensory integration – foam standing
	ADL and stability in gait:
	6. Walk on level surface
	7. Walking with head turns
	8. Timed "get up and go" test
	9. Sit to stand
	10. Sit on floor and stand up
	Added
	ADL and stability in gait:
	1. Step ascent/descent

A.5 Function Assessment scale for Spinal Deformity – finalization

The resulting FASD (Appendix B) contains 14 items divided in three subscales (Table A7): FASD 1 - spinal mobility and balance (3 items); FASD 2 - stability limits and sensory integration (5 items); FASD 3 - ADL and stability in gait (6 items).

All items in the FASD were scored on a three- or four-point ordinal scale and were administered bilaterally in case of clinical relevance. The total score of the scale was 56 points, with a higher score indicating better performance. Scoring for the FASD was based on the BESTest and TCMS, with some modifications. Specifically, TCMS items, consisting of two or three subquestions with separate ordinal scales, were combined into one ordinal scale each (items 1, 2 & 3).

The final version of the FASD, including instructions for assessment, can be found in Appendix B.

References

[1] Berg K, Wood-Dauphinee S, Williams JI, Gayton D. Measuring balance in the elderly: Preliminary development of an instrument. Physiother Canada 1989. https://doi.org/10.3138/ptc.41.6.304.

[2] Verheyden G, Nieuwboer A, Mertin J, Preger R, Kiekens C, De Weerdt W. The Trunk Impairment Scale: A new tool to measure motor impairment of the trunk after stroke. Clin Rehabil 2004. https://doi.org/10.1191/0269215504cr733oa.

[3] Heyrman L, Molenaers G, Desloovere K, Verheyden G, De Cat J, Monbaliu E, et al. A clinical tool to measure trunk control in children with cerebral palsy: the Trunk Control Measurement Scale. Res Dev Disabil 2011;32:2624–35. https://doi.org/http://dx.doi.org/10.1016/j.ridd.2011.06.012.

[4] Siu AL. An Objective Measure of Physical Function of Elderly Outpatients: The Physical Performance Test. J Am Geriatr Soc 1990. https://doi.org/10.1111/j.1532-5415.1990.tb01373.x.

[5] Rose DJ, Lucchese N, Wiersma LD. Development of a Multidimensional Balance Scale for Use With Functionally Independent Older Adults. Arch Phys Med Rehabil 2006. https://doi.org/10.1016/j.apmr.2006.07.263.

[6] Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. Phys Ther 2009;89:484–98. https://doi.org/10.2522/ptj.20080071.

[7] Franchignoni F, Horak F, Godi M, Nardone A, Giordano A. Using psychometric techniques to improve the balance evaluation systems test: The mini-bestest. J Rehabil Med 2010. https://doi.org/10.2340/16501977-0537.

[8] Padgett PK, Jacobs J V., Kasser SL. Is the Bestest at its Best? A Suggested Brief version based on Interrater reliability, validity, internal consistency, and theoretical construct. Phys Ther 2012. https://doi.org/10.2522/ptj.20120056.

[9] Fehlings MG, Tetreault L, Nater A, Choma T, Harrop J, Mroz T, et al. The aging of the global population: The changing epidemiology of disease and spinal disorders. Neurosurgery 2015. https://doi.org/10.1227/NEU.0000000000000953.

[10] Diebo BG, Challier V, Shah N V., Kim D, Murray DP, Kelly JJ, et al. The Dubousset Functional Test is a Novel Assessment of Physical Function and Balance. Clin Orthop Relat Res 2019. https://doi.org/10.1097/CORR.0000000000000820.

[11] Portney LG, Watkins MP. Foundations of Clinical Research: Application to Practice. 2009.

Appendix B. FASD

Function Assessment scale for Spinal Deformity (FASD)

Total score: /56

Material:
- Treatment table
- Foam pad
- Measuring tape
- Chair with armrests
- Stair step with two consecutive steps of normal height (18-20 cm)
- Stop watch

A. Spinal balance and mobility	Total: /14	
1. Side bending	Left	Right
The patient sits with an upright trunk on the treatment table with the hands on the upper legs and is instructed to touch the table with his/her elbow at the level of the femoral head.		
- Patient touches the table with elbow without lifting the pelvis and clear trunk shortening	3	3
- Patient touches the table with elbow with pelvic lift and small trunk shortening	2	2
- Patient touches table with elbow with pelvic lift and straight trunk	1	1
- Patient is unable to touch table with elbow	0	0
2. Side reaching	Left	Right
The patient sits with an upright trunk on the treatment table with his/her arm raised sideways and is instructed to touch the therapist's hand. (distance: one forearm length measured from the homolateral arm when raised sideways)		
- Patient can touch therapist's hand in one smooth movement and returns without falling or hand support	2	2
- Patient reaches therapist's hand slowly and with a lot of effort	1	1
- Patient cannot reach target or needs hand support when returning back to starting position	0	0
3. Cross reaching and spinal rotation (distance: half the forearm length)	Left	Right
The patient sits with an upright trunk on the treatment table and is instructed to reach to the opposite side and touch the hand of the therapist. (distance: half the forearm length measured from the heterolateral arm when raised sideways)		
- Patient can touch therapist's hand in one smooth movement and returns without falling or hand support	2	2
- Patient reaches therapist's hand slowly and with a lot of effort	1	1

- Patient cannot reach target or needs hand support when returning back to starting position	0	0
B. Stability (limits) and sensory integration	**Total: /21**	
4. Toe rise The patient is instructed to rise on the toes as high as possible for 3 seconds. - Stable and maximal height for 3 seconds - Unstable and submaximal height for 3 seconds - Less than 3 seconds - Not possible	 3 2 1 0	
5. Standing on foam	Time 'eyes open':	Time 'eyes closed':	Eyes open	Eyes closed
The patient is instructed to stand in the middle of a foam pad with the feet close to each other for 30 seconds. - Patient stands stable for 30 seconds - Patient stands unstable for 30 seconds (twitching/arm movements) - Patient stands for less than 30 seconds - Patient stands for less than 10 seconds	 3 2 1 0	 3 2 1 0
6. Forward reactive response Patient stands straight while the therapist gives some backward pressure on the patient's shoulders. The patient withstands the pressure without leaning forward, until the therapist releases the pressure. The patient tries to keep balance without taking a step. - Patient recovers stable (no large arm/trunk motion) without a step - Patient recovers unstable (arm swing/trunk motion) without a step - Patient needs a step but recovers independently (second realignment step is allowed) - Patient needs multiple steps or needs assistance to prevent a fall	 3 2 1 0	
7. One leg stance	Left:	Right:	Left	Right
Patient is instructed to stand as long as possible on one leg. The test ends when the patient reaches 30 seconds. - Patient stands stable for 30 seconds - Patient stands unstable (trunk/arm/leg motion) for 30 seconds - Patient stands for < 30 seconds - Patient stands for < 10 seconds	 3 2 1 0	 3 2 1 0

8. Forward reaching during stance Distance: Patient is instructed to reach forward as far as possible without lifting the heels and return back to starting position. - Patient reaches ≥ 30 cm - Patient reaches target between 25 and 30 cm - Patient reaches < 25 cm - Patient cannot lean forward without falling	 3 2 1 0	
9. ADL and stability in gait	**Total: /21**	
10. Sit on floor and stand up Patient is instructed to sit on the floor with the legs straight and stand up again. A chair can be used as support if necessary. - Patient sits and stands up independently in ≤ 6 seconds - Patient sits and stands up independently in more than 6 seconds - Patient needs support from chair to sit and/or to stand up - Patient cannot sit on floor or stand up without support of therapist	Time: 3 2 1 0	
11. Sit-to-stand The patient sits on a chair and tries to stand up without hand support if possible and without moving his/her feet. - Patient stands up without use of hands in one smooth movement - Patient stands up without use of hands but with hesitation, large trunk flexion or large arm movement - Patient uses hand support to stand up from chair - Patient cannot stand up independently even with hand support	 3 2 1 0	
12. Step ascent (a.) and descent (b.) Patient is instructed to ascend a stair step with two consecutive steps, turn around and descend. - Patient ascends/descends independently with single foot support, in one smooth movement and no large arm movements - Patient ascends/descends independently with single foot support but slowly, with hesitation before taking the step OR with signs of imbalance (large arm movements, wide base, excessive trunk motion, midline cross-over,…) - Patient uses double foot support while ascending/descending - Patient is unable to ascend/descend without external support	Ascent 3 2 1 0	Descent 3 2 1 0
13. Overground walking 6 m The patient is instructed to walk 6m at his/her own preferred speed.	Time:	

- Patient walks independently with normal speed (≤5.5 seconds) and no signs of imbalance - Patient walks independently with slower speed (>5.5 seconds) and no signs of imbalance - Patient walks with signs of imbalance (wide base, excessive trunk motion, inconsistent step length, midline cross-over,…) - Patient cannot walk 6 m	3 2 1 0
13. Walking with head turns Patient is instructed to walk 6m and turn his/her head to the left/right according to the therapist's instructions. - Patient walks without imbalance or changes in gait speed - Patient slows down (>10% slower than 12) while turning the head but with good balance - Patient cannot perform full head turns or shows signs of imbalance (midline cross-over or deviation) - Patient cannot perform head turns or stops walking	3 2 1 0
14. Timed Get-Up-and-Go-Test Patient is instructed to stand up from a chair, walk 3 meters, turn around and sit down again as fast as possible without running. - Patient performs test within 7.5 seconds with good balance - Patient performs test slowly (> 7.5 seconds) with good balance - Patient performs test within 7.5 seconds with signs of imbalance - Patient performs test slowly (> 7.5 seconds) with signs of imbalance	Time: 3 2 1 0
Comments:	

Appendix C. Data distribution of FASD compared to BESTest/TCMS

One of the reasons to develop an ASD-specific scale for function and balance, was the presence of ceiling effects of BESTest and TCMS when applied in the ASD population[1,2]. Selecting the discriminative items of both scales and adjusting the scoring options of selected items of BESTest and TCMS based on an extensive data collection, as described in Appendix A, aimed for a reduction of these ceiling effects.

Histograms of FASD, BESTest and TCMS were provided to compare the distribution of the total scores (expressed as percentages) across the total group (n=43). Differences between total scores on FASD, BESTest and TCMS were analyzed with the Friedman's test. Results can be found in figure C1.

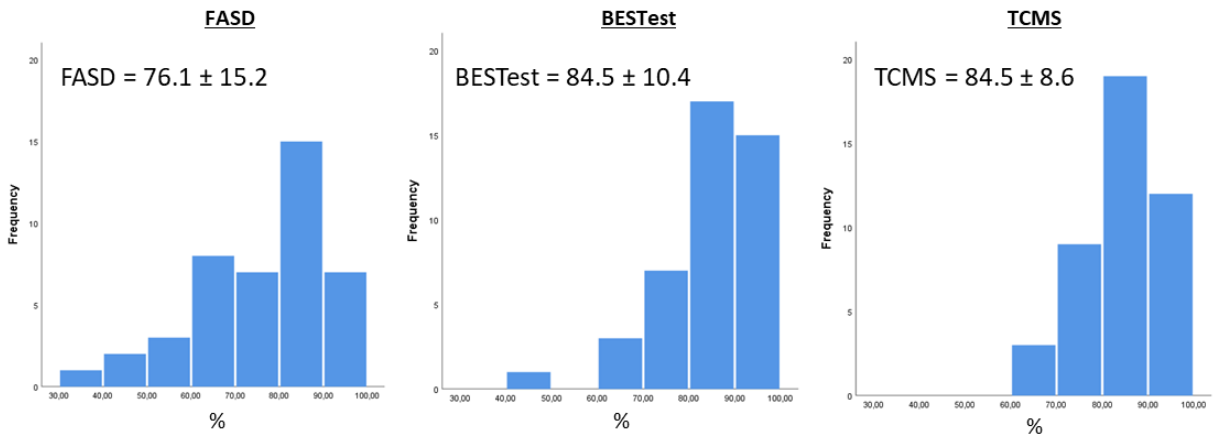


Figure C1. Histograms with distribution of, respectively, FASD, BESTest and TCMS data.
On the horizontal axis the scoring on FASD, BESTest and TCMS is displayed in percentages. On the vertical axis the frequencies of a rating can be found. Mean and standard deviation are reported. FASD: Function and balance scale for Adult Spinal Deformity; BESTest: Balance Evaluation Systems Test; TCMS: Trunk Control Measurement Scale.

Both the increased variability of scores on the FASD within the same sample, as indicated by an increased standard deviation and a larger distribution of scores on the x-axis of the histogram, and the significantly lower total score on FASD compared to TCMS and BESTest ($p<0.001$) suggest a reduction in ceiling effect.

References

1. Laratta JL, Glassman SD, Atanda AA, et al. The Berg balance scale for assessing dynamic stability and balance in the adult spinal deformity (ASD) population. *J Spine Surg*. Epub ahead of print 2019. DOI: 10.21037/jss.2019.09.15.

2. Severijns P, Overbergh T, Scheys L, et al. Reliability of the balance evaluation systems test and trunk control measurement scale in adult spinal deformity. *PLoS One* 2019;14:e0221489.