



Reliability, Technical Error of Measurement and Validity of Height Measurement Using Portable Stadiometer

Azli Baharudin^{1*}, Mohamad Hasnan Ahmad¹, Balkish Mahadir Naidu¹,
Nurul Rufaidah Hamzah², Nor Azian Mohd Zaki¹, Ahmad Ali Zainuddin^{1,3}
and Noor Safiza Mohd Nor¹

¹Centre for Nutrition Epidemiology, Institute for Public Health, 50590 Jalan Bangsar, Kuala Lumpur, Malaysia

²Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Sri Iskandar, 32610 UiTM, Perak, Malaysia

³Universiti Teknologi MARA, 42300 UiTM, Bandar Puncak Alam, Selangor, Malaysia

ABSTRACT

This study sought to examine the reliability and validity of height measurements using a portable stadiometer as compared to a mechanical scale. Samples from 142 adults aged 22 to 57 were taken during data collection in November 2014. There was a high degree of reliability for the inter-examiner, intra-examiner and inter-instrument aspects with regards to mean difference, the inter correlation coefficient (ICC) and Bland-Altman Plot. For the inter-examiner aspect, the height measurement taken by the first examiner was 0.01 cm higher than that by the second examiner with an ICC of 0.999. For the intra-examiner aspect, the difference was 0.1 cm; this was higher in the first measurement compared to the second. The ICC was also 0.999. For the inter-instrument aspect, measurement taken by stadiometer was 0.61 cm higher than the measurement taken by mechanical scale and the ICC was 0.997. The Bland-Altman plot showed a distribution of differences between measurements in the inter-examiner, intra-examiner and inter-instrument aspects that were close to zero within the narrow range of $\pm 1.96SD$. The technical error of measurement (TEM), coefficient of reliability (R) and coefficient of variation (CV) for the inter-examiner, intra-examiner and inter-instrument aspects were within the acceptable limits. This study suggests that the portable stadiometer is reliable and valid for use in community surveys.

Article history:

Received: 22 October 2015

Accepted: 14 February 2017

E-mail addresses:

ps_azlibaharudin@moh.gov.my (Azli Baharudin),
mha.hasnan@gmail.com (Mohamad Hasnan Ahmad),
balkishmahadir@gmail.com (Balkish Mahadir Naidu),
nurulhamzah17@gmail.com (Nurul Rufaidah Hamzah),
norazianmz@moh.gov.my (Nor Azian Mohd Zaki),
ahmadali@moh.gov.my (Ahmad Ali Zainuddin),
safiza@moh.gov.my (Noor Safiza Mohd Nor)

*Corresponding Author

Keywords: Stadiometer, reliability, technical error of measurement, validity of height

INTRODUCTION

Anthropometry is an easy, fast and inexpensive way to assess the nutritional status of a person. Simple measurements such as height, weight,

waist circumference and hip circumference are constituents of anthropometric measurements. They can be easily measured and provide information such as body mass index (BMI) and waist-hip ratio for further health diagnosis (Ulijaszek & Kerr, 1999; Sanchez-Garcia et al., 2007).

Nevertheless, anthropometric measurement has its limitations, with the need for trained examiners and relatively high between-measurement technical errors and mechanical limitations (NYORC, 2006; Haniff et al., 2008). Surveys that involve a large number of samples require a group of skilled and trained people to do the measurement, and this can lead to measurement errors (Ulijaszek & Kerr, 1999). It was found that instrument error increased the variance of height, weight, and mean BMI generally (Biehl et al., 2013). Measurement precision and reliability refer to “the extent to which repeated measurements give the same value” while accuracy and validity mean “how close a measurement is to its ‘true’ value (Ulijaszek & Kerr, 1999; Cameron, 2002; Biehl et al., 2013).

Nowadays, numerous anthropometry instruments have been produced. Some of them have been modified for more practical use in certain places, situations or conditions. One of them is height measurement. The height measurement is estimated from a fixed rod on the wall to an adjusted and portable stadiometer. There is a concern about the reliability of stadiometers and the ensuing technical errors that may arise. As the usage of the instrument (SECA 213) for height measurement is still new in Malaysia, this study was aimed at assessing inter- and intra-examiner reliability, technical error of measurement and validity of the SECA Stadiometer 213 for measuring height in community surveys.

METHODOLOGY

This cross-sectional study was conducted among adults in a selected government clinic. Data collection was undertaken in November 2014. A total of 142 adults were recruited using convenient sampling. Respondents with no deformities affecting height were recruited. Adults with mental and terminal illness were excluded from this study. The sample size was estimated based on Walter et al. (1998), with two replicates per subject; an expected reliability coefficient (r) of at least 0.8 ($H1: \rho=0.8$), reliability of 0.7 ($H0: \rho=0.7$) or higher were required to be minimally acceptable, $\alpha=0.05$ and $\beta=0.2$ (corresponds to 80% power); this required a total number of 130 subjects. Using a 10% drop-out rate accounted for by poor response, the final target sample size was 142.

The height of respondents was obtained by using two instruments i.e. the portable SECA Stadiometer 213 as a ‘test’ instrument and the SECA Medical 703 Digital Column Scale as a ‘reference’ instrument, which is standardised and used widely in health facilities. A good stadiometer is able to read to 0.1 cm or 1/8th of an inch; it is stable and has a horizontal headpiece that can be brought into contact with the most superior part of the head.

Data collection was carried out in November 2014. Two trained examiners with a background in public health nursing conducted the measurement of height for each respondent. Not being a part of the study team, the two examiners were not aware of the study’s objectives. Height was measured to the nearest 0.1 cm from the subject head-to-toe in an upright position with five points of the subject’s body touching the wall.

On the day of data collection, each examiner measured the respondents' height individually. The first examiner measured the respondent's height, followed by the second examiner using the same instrument, the SECA Stadiometer 213. After that, the respondent was measured again, but this time using the SECA Medical 703 Digital Column Scale. The examiners were requested not to reveal their previous readings. The data capture form was designed in such a way that all the recordings of previous readings were obscured immediately after each recording; this was to minimise recall bias. A measurement form was provided to the examiners for their easier recoding and data management later.

Statistical analysis was reported using:

1. Absolute mean difference
 - Generally, absolute mean difference is the measure of a statistical dispersion difference of two random variables. In this paper, it is explained by the dispersion difference in inter-examiner, intra-examiner and inter-instrument aspects of height measurement using the SECA Stadiometer 213 and the SECA Medical 703 Digital Column Scale. It is the base analysis to verify the difference or similarity between two readings (Haniff et al., 2008).
2. Correlation coefficient (r)
 - This is commonly used to show the relationship (similarities) between two readings. Intra-class correlation (ICC) will be used for this purpose. The value of the reliability coefficient ranged from 0 to 1, where $ICC < 0$ indicated "no reliability", ≥ 0 but < 0.2 "slight reliability", 0.2 to < 0.4 "fair reliability", 0.4 to < 0.6 "moderate reliability", 0.6 to < 0.8 "substantial reliability" and 1 "almost perfect reliability" (MUSC, 2006; Geeta et al., 2009).
3. Bland-Altman plot (Bland, 1986)
 - This is a method to compare two measurement techniques or methods by plotting the differences between the two methods against the averages of the two methods. It is set as a model to show the spread of differences in the readings, the mean difference and the upper and lower limits of agreement for both inter- and intra-examiner reliabilities (Geeta et al., 2009).
4. Technical error of measurement (TEM)
 - This is an important analysis to represent the measurement quality and control dimension. It is a typical way to indicate the error margin in anthropometry that has been adopted by the International Society for the Advancement of Kinanthropometry for the accreditation of anthropometrists. It is indeed the standard deviation between repeated measures. The TEM index allows anthropometrists to verify the degree of accuracy when performing and repeating anthropometrical measurements (intra-examiner) and when comparing their measurement with measurements from other anthropometrists (inter-examiner). A relative TEM value less than 0.20 indicates that the measurements taken are acceptable (Geeta et al., 2009).

To determine the precise height measurement, four differences are commonly used, which are Technical Error of Measurement (TEM), the relative of Technical Error Measurement (rTEM), the Coefficient of Reliability (R) and the Coefficient of Validity (CV) (Ulijaszek & Kerr, 1999).

$$\text{Absolute TEM} = \sqrt{\frac{\sum d_i^2}{2n}} \quad (1)$$

where T_0 = Absolute TEM

d = difference between 1st and 2nd reading,

n = number of respondents,

i = the number of readings

From the equation (1) it was transformed into T_1 to obtain the error expressed as a percentage corresponding to the total average of the variable to be analysed. To compare TEM across anthropometric measurement or study occasions, we converted the absolute TEM (α) to %TEM (β). The equation (2) of β is shown:

$$\text{relative TEM (\% TEM)} = \frac{T_0}{\mu} \times 100 \quad (2)$$

where T_1 = relative TEM

μ = Average of measurement

Acceptable T_1 levels were 5% or less for skinfolds and 1% for other anthropometrical measure (Gore & Gore, 2000). The lower the T_1 obtained, the better the precision to perform the measurement.

From TEM, the coefficient of reliability (R) was very useful for comparing relative reliability of different measurements. Equation (3) is shown:

$$\text{Coefficient of reliability, R} = 1 - \left(\frac{(T_0)^2}{\sigma^2} \right) \quad (3)$$

If the range of R is close to 0 then the reading of the anthropometrical measure is considered as not reliable but if the range is close to one it is considered as being completely reliable. If the R values are greater than 0.95, this indicates that the measurement is sufficiently precise (Ulijaszek & Kerr, 1999).

Finally, the coefficient of variance (CV) was calculated to express sample variability relative to the mean of the sample, with the following formula. Equation (3) is shown:

$$\text{Coefficient of variability, CVs} = \frac{\sigma}{\mu} \times 100 \quad (4)$$

where σ = standard deviation,

The CVs provide a general view of the performance of the method. If the CVs are less than or equal to 5%, they generally imply good method performance, while CVs greater than or equal to 10% imply that the method did not perform well (Zady, 2006). In comparing methods, the percentage of the CVs is a good indicator (Bland, 2006).

RESULTS

Sample Characteristics

A sample of 142 adults aged 22 to 57 years were involved during the one-month data collection in November 2014. Among them, 44 (31%) were men and 98 (69%) were women. Mean height was 167.48 ± 5.71 cm and 156.38 ± 6.60 cm for men and women, respectively and it was not significantly different by gender.

Reliability

There were three ways in which the inter-examiner, intra-examiner and inter-instrument aspects were examined in this study. The first was by looking at the absolute mean difference. Absolute mean difference for the inter-examiner, intra-examiner and inter-instrument aspects were 0.01, 0.10 and 0.61, respectively as illustrated in Table 1. Although mean differences in the inter-instrument aspect was high, by independent t-test, the measurements were determined as being not significantly different between the stadiometer and mechanical scale.

The second method was by determination of the correlation coefficient. Results of the correlation coefficient of analysis of the inter-examiner, intra-examiner and inter-instrument aspects used the intra-class coefficient (ICC) as given in Table 1. All the ICCs were nearly in perfect agreement, which means a strong correlation between the two readings. This indicated a high degree of reliability between the two examiners, within examiners and between the two instruments.

Table 1
Absolute Mean Difference and Intra-Class Coefficient (ICC) of Height Measurement

Pair comparison	Summary statistic			Absolute mean diff. (1) – (2)	P value	ICC
	N	Mean (SD)	Median (Min, Max)			
Inter-examiner	142			0.01	0.989	0.999
Examiner#1 (1)		159.60 (7.85)	158.65 (141.75, 179.65)			
Examiner#2 (2)		159.59 (7.81)	158.58 (141.65, 179.60)			
Intra-examiner	142			0.10	0.915	0.999
1 st measurement (1)		159.55 (7.85)	158.60 (141.80, 179.60)			
2 nd measurement (2)		159.65 (7.85)	158.80 (141.70, 179.90)			
Inter-instrument	142			0.61	0.511	0.998
Stadiometer (1)		159.60 (7.85)	158.65 (141.75, 179.65)			
Stadiometer (2)		158.99 (7.92)	158.35 (141.95, 179.50)			

P value was obtained by independent t-test

Validity

The third method used was the Bland-Altman method. Figure 1 shows that the measurements taken by the first examiner were consistent with those taken by the second examiner, with an average difference of 0.012 cm, upper limit at 0.591 cm and lower limit at -0.566 cm. Figure 2 shows that the measurements taken by the first examiner at first reading were consistent with those of the second reading, with an average difference of -0.099 cm, upper limit at 0.436 cm and lower limit at -0.636 cm. Figure 3 shows that the measurements taken by mechanical scale were consistent with those taken by stadiometer with an average difference of 0.616 cm, upper limit at 1.775 cm and lower limit at -0.543 cm.

Inter-examiner. The Bland and Altman plot in Figure 1 shows that the differences between examiner 1 and examiner 2 were consistent for the measurements in which the data were collected. However, it was noted that there were many points outside the upper control limit of 0.51 cm and the lower control limit of -0.566 cm.

Intra-examiner. The Bland and Altman plot in Figure 2 shows that the differences between measurement 1 and measurement 2 were almost consistent for the measurements. However, it was noted that there were more points outside the upper control limit of 0.436 cm and lower control limit of -0.636 cm.

Inter-instrument. The Bland and Altman plot in Figure 3 shows that the differences between the Stadiometer and the Seca Mechanical Scale measurements were consistent for the measurements compared to when using the intra-examiner and inter-examiner aspects. Only several points were outside the lower control limit of -0.543 cm.

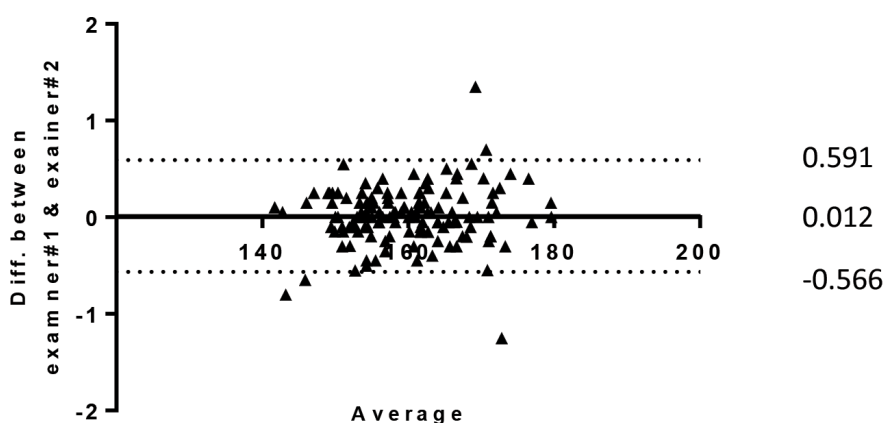


Figure 1. Bland-Altman plot difference vs average of first and second examiner

Height Measurement Reliability Using Portable Stadiometer

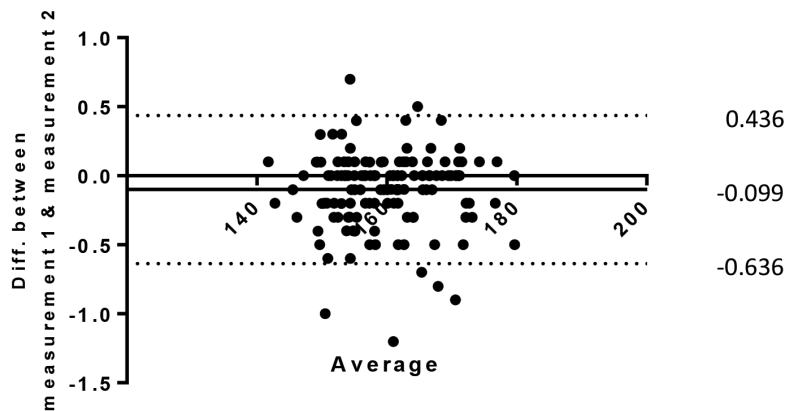


Figure 2. Bland-Altman plot difference vs average of first and second measurement

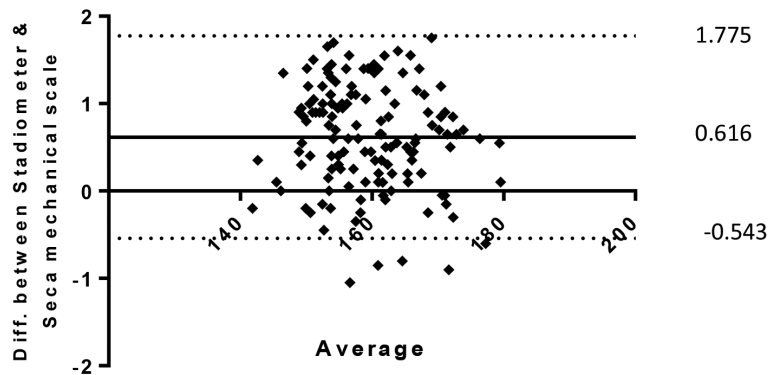


Figure 3. Bland-Altman plot difference vs average stadiometer and seca mechanical scale

Technical Error Measurement (TEM)

The results for the TEM are tabulated in Table 2. The relative TEMs (T_1) for the inter- and intra-examiner aspects were 0.13% and 0.13%, respectively, while for the inter-instrument aspect, it was 0.38%. The coefficient of reliability, (R), for inter, intra-instrument and inter-instrument aspects were 99.92%, 99.93% and 99.41%, respectively. The coefficients of variability (CVs) for the inter- and intra-examiner and inter-instrument aspects were 4.895%, 4.908% and 4.943%, respectively.

Table 2
Technical Error of Measurement (TEM) & Coefficient of Variation (CVs)

	Inter-examiner (Examiner 1 & Examiner 2)	Intra-examiner (1 st measurement & 2 nd measurement)	Inter-instrument (Stadiometer & Mechanical Scale)
Σ (Deviation) ²	12.31	11.95	102.94
Absolute TEM, T_0	0.21	0.21	0.60
VAV	159.60	159.60	159.30
Relative TEM %, T_r	0.13	0.13	0.38
Coefficient of reliability (R)	0.99929	0.999314	0.994155
Coefficient of variations (CVs)	4.895	4.908	4.943

DISCUSSION

Anthropometric measurements have different types of error. This study looked, in particular, at imprecision as one of the components of reliability. Imprecision is the variability of repeated measurements due to intra- examiner, inter-examiner and inter-instrument measurement differences (Ulijaszek, 1999) (Table 3). There are a few indices that are often used to assess intra- and inter-examiner and inter-instrument aspect variability. These include technical error of measurement (TEM), coefficient variation (CV), coefficient of reliability (R), intra-class correlation coefficient (ICC) and the Bland and Altman plot. There are only a handful of published articles on measurement errors. There is one landmark review article (Ulijaszek, 1999) that captured two anthropometric measurements among which were the Stadiometer and Mechanical Scale. This study examined the reliability of the inter-examiner, intra-examiner and inter-instrument aspects. The ICC is an estimate of the proportion of the combined variance for the true biological value for any anthropometric measure and for the measurement errors associated with it. We found that the ICC values were close to 1; 0.999 for the inter-examiner and 0.999 for the intra-examiner and 0.997 for the inter-instrument aspects. This indicates high variability between repeated measures on the same subject, which means the accuracy of the stadiometer varied between different examiners.

Table 3
Correlation of Coefficient of Inter- and Intra-Examiner and Inter-Instrument Measurement

Variables	N	Inter-examiner		N	Intra-examiner		N	Inter-instrument	
		ICC	Coefficient Variation		ICC	Coefficient Variation		ICC	Coefficient Variation
Height	137	0.999	5.15%	137	0.999	5.07%	137	0.997	5.07%

A number of methods of measuring inconsistency are available but the preferred method involves calculation of the intra- and inter-examiner and intra-instrument aspects. TEM gives information on the error margin of a trait and therefore is an accuracy index. Determination of acceptable levels of measurement errors is not straightforward and relates to the variable being studied as well as the height of the subjects (Ulijaszek & Kerr, 1999). However, Ulijaszek and Kerr (1999) have suggested acceptable intra- and inter-examiner limits for a range of anthropometric measure including height, weight and skinfold site (Ulijaszek & Kerr, 1999). The results of anthropometric measurements, TEM (T_0), rTEM (T_I), R and CVs values are presented in Table 1.

Changes in the mean, correlation coefficient and Bland and Altman plots revealed a high degree of inter-examiner reliability. The p values of the change in the mean showed no statistical significance. The Bland and Altman plot shown in the three graphs are valid because all the points are close to each other. This makes for high validity and was also found, by Bland and Altman, that the two examiners were consistent with an average of 0.012 cm and an upper limit of 0.591 cm to lower limit of -0.566 cm. However, the reading of an upper limit and lower limit was slightly bigger in difference. For intra-examiner reliability, the average is -0.99 cm with an upper limit of 0.436 cm to lower limit of -0.636 cm. It was found that the difference between mean upper limit and lower limit was more precise compared to those of the inter-examiner aspect. The reading was more valid because the same person took the reading. In addition, the mean inter-instrument average was 0.616 cm and upper limit 1.775 cm to lower limit -0.543 cm. This suggested that the examiner must have been poorly trained at handling the instrument. The ICC for both intra- and inter-examiner was almost perfect: 0.9900 and 0.9990, respectively. For inter-instrument was 0.954. All findings showed that the reliability of the three experiments was absolutely perfect.

All three relative TEM values were within the acceptable limit. Our findings of TEM values for the inter- and intra- examiner was 0.13% and for inter-instrument it was 0.38%. The inter-instrument reading was higher compared to that for inter- and intra- examiner, which was found to be due to different instruments. Comparison for intra- and inter-examiner and inter-instrument aspects showed that the inter-correlation coefficient gave a precise reading, which is close to 1. From the findings of R, the Coefficient of Reliability, all three measurements reported less than 5% error due to human measurement. These indicated close to perfect intra- and inter-examiner reliability for both measurements and inter-instrument. The CVs for three experiments were below 5% error. This shows that variability was low in this sample, which was the hypothesis. However, this study can conclude that the measurement of inter-instruments were less reliable than those for the inter- and intra-examiner aspects. On the part of validity, the intra- and inter-examiner readings were 4.895% and 4.908%, respectively. For inter-instrument, it was 4.943%. Our findings showed that the most valid measurement was that of the intra-examiner compared to the others.

The reporting of examiner measurement error was common, and it is this issue that should receive adequate attention in future studies using anthropometric measures as the main study outcome. For further research, observer or examiner measurement error should be minimised through close attention to every aspect of the data collection process, such as equipment calibration and the training of research personnel.

CONCLUSION

Anthropometric measurement errors are unavoidable and should be minimised through close attention to every aspect of the data collection process including selection of examiners. The inter-examiner, intra-examiner and inter-instrument correlation coefficients for height measurement were almost perfect. The R and CV value were also within acceptable range. Therefore, it was proven that the portable stadiometer had high degrees of reliability for measuring height and was valid to be interchanged with the SECA Mechanical scale.

ACKNOWLEDGEMENT

The authors would like to thank the Director General of Health, the Ministry of Health Malaysia, for permission to publish this paper. We are grateful to the subjects for participating in this survey.

REFERENCES

- Biehl, A., Hovengen, R., Meyer, H. E., Hjelmesæth, J., Meisfjord, J., Grøholt, E. K. ... & Strand, B. H. (2013). Impact of instrument error on the estimated prevalence of overweight and obesity in population-based surveys. *BMC Public Health*, 13(1), 1–6.
- Bland, M. J., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, 327(8476), 307-310.
- Cameron, N. (2002). *Human growth and development*. United States of America, USA: Academic Press.
- Geeta, A., Jamaiyah, H., Safiza, M. N., Khor, G. L., Kee, C. C., Ahmad, A. Z., ... & Faudzi, A. (2009). Reliability, technical error of measurements and validity of instruments for nutritional status assessment of adults in Malaysia. *Singapore Medical Journal*, 50(10), 1013.
- Gore, C. J., & Australian Sports Commission (Eds.). (2000). *Physiological tests for elite athletes*. United Kingdom, UK: Human Kinetics.
- Haniff, J., Appannah, G., Nor, M., Safiza, N., Wong, N. F., Kee, C. C., ... & Abd Talib, R. (2008). Reliability and technical error of calf circumference and mid-half arm span measurements for nutritional status assessment of elderly persons in Malaysia. *Malaysian Journal of Nutrition*, 14(2), 137–150.
- MUSC. (2006). *Medical University of South Carolina*. Retrieved September 20, 2006 from www.musc.edu/dc/icrebm/index.html
- NYORC. (2006). *New York Obesity Research Center: Body composition unit*. Retrieved September 16, 2006 from www.nyorc.org/anthropometry.html
- Norton, K., & Olds, T. (1996). *Anthropometrica: A textbook of body measurement for sports and health courses*. Australia: UNSW press.
- Sánchez-García, S., García-Peña, C., Duque-López, M. X., Juárez-Cedillo, T., Cortés-Núñez, A. R., & Reyes-Beaman, S. (2007). Anthropometric measures and nutritional status in a healthy elderly population. *BMC Public Health*, 7(2), 1-9.

Height Measurement Reliability Using Portable Stadiometer

- Ulijaszek, S. J., & Kerr, D. A. (1999). Anthropometric measurement error and the assessment of nutritional status. *British Journal of Nutrition*, 82(03), 165–177.
- Walter, S. D., Eliasziw, M., & Donner, A. (1998). Sample size and optimal designs for reliability studies. *Statistics in Medicine*, 17(1), 101–110.
- Zady, M. F. (n.d.). *Z-4: Mean, standard deviation, and coefficient of variation*. Madison: Westgard QC. Retrieved March 6, 2014, from www.westgard.com/lesson34.htm#coefficient

