Validation of an Accelerometer-based Fall Prediction Model

Ying Liu, Stephen J. Redmond, *Senior Member, IEEE,* Tal Shany, Jane Woolgar, Michael R. Narayanan, *Member, IEEE,* Stephen R. Lord and Nigel H. Lovell, *Fellow, IEEE*

***Abstract*— Falls are a common and serious problem faced by older populations. There is a growing interest in estimating the risk of falling for older people using body-worn sensors and simple movement tasks, allowing appropriate fall prevention programs to be administered in a timely manner to the high risk population. This study investigated the capability and validity of using a waist-mounted triaxial accelerometer (TA) and a directed routine (DR) that includes three movement tasks to discriminate between fallers and non-fallers and between multiple fallers and non-multiple fallers. Data were collected from 98 subjects who were stratified into two separate groups, one for model training and the other for model validation. Logistic regression models were constructed using the TA features from the entire DR and from each single DR task, and were validated using unseen data. The best models were obtained using features from the alternate step test to classify between fallers and non-fallers with $\kappa = 0.34 - 0.41$, sensitivity $= 68\% - 71\%$ and specificity $= 63\% - 73\%$. However, the overall validation performances were poor. The study emphasizes the importance of independent validation in fall prediction studies.**

I. Introduction

Falls suffered by older people are a major public health problem facing many countries. Fall-related injuries are a major cause of hospitalization and institutionalization. It was estimated that the cost of health care related to falls by older people ($\geq$ 65 years old) in Australia was more than $600 million in 2007-08 [1]. Falls suffered by older people can also increase their fear of falling, and diminish their confidence in performing normal daily activities, which can lead to reduced engagement in normal activities and a further deterioration of their general well-being. A high risk of falling is often associated with muscle weakness, poor balance control, gait deficiency, cognitive impairment and fear of falling [2]. Most of these deficits can be reflected in the way people move. Many studies have reported that mobility dysfunction is strongly associated with a higher risk of falling [3], [4]. It is proposed that through an analysis of the movement patterns of older people, those with a higher risk of falling can be identified. Accurate fall risk screening would help facilitate the prevention of future falls by administrating appropriate intervention strategies to high risk populations.

A body-worn sensor-based system could be useful for unsupervised assessment, and for long-term fall risk monitoring, allowing a proper intervention program to be administered in time. Much research effort has been directed at the concept of using body-worn sensors for movement pattern analysis for fall risk assessment [5], [6], [7], [8]. Some studies attempted to replicate a clinical fall risk tool [5], [6], while others tried to classify between high risk and low risk classes according to their fall history or prospective falls [7], [8]. However, most sensor-based fall risk assessment studies only reported the performance of the model on training data with a small sample size, making their predictive accuracy uncertain.

Previously, a model was developed to approximate the assessment provided by a widely-used clinical fall risk assessment tool, using a waist-worn triaxial accelerometer (TA) and a directed routine (DR) with a cohort of 68 subjects [6]. However, the developed model tended to be over-fitted, with more than 30 features selected. In addition, no appropriate validation was performed. The current analysis aims to investigate the capability and validity of using these TA signals to discriminate between fallers and non-fallers and between multiple fallers and non-multiple fallers, according to their prospective 12-month fall data. In addition to these 68 subjects, data were also collected from another cohort with higher risks of falling, using the same waist-attached TA and the DR. Data from these two cohorts are stratified into training and testing groups to perform two-fold cross validation. The predictive capability and validity of TA signals from each of the three DR movement tasks is also examined.

II. Methods

A. Instrumentation

A small triaxial accelerometer (TA) device with a size of $71 \times 50 \times 18$ mm was used in both studies. The accelerometry sensors had a range of ± 1.5 G (where G $= 9.81$ ms^{-2}), and were sampled at a rate of 40 Hz per channel. Data were streamed via a Class 1 Bluetooth radio link to a connected computer in real-time. The TA was attached to the waist at the right anterior iliac area, to measure body movement during the DR assessment. The three axes of the TA were approximately aligned with the vertical (x-axis), mediolateral (y-axis) and anteroposterior (z-axis) axes of the subject's frame of reference.

B. Directed Routine Assessment

The DR is a set of simple movement tasks which takes approximately five minutes to complete. It includes three different movement tasks, which are the Timed Up-and-Go

Y. Liu, S. J. Redmond, T. Shany, M. R. Narayanan and N. H. Lovell are with Graduate School of Biomedical Engineering, UNSW Australia. n.lovell@unsw.edu.au

J. Woolgar is with Northern Sydney Local Health District, Manly Hospital, Sydney NSW 2095, Australia.

S. R. Lord is with Neuroscience Research Australia and the School of Medical Sciences, UNSW Australia.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

Test (TUGT), Alternate Step Test (AST) and Sit-to-Stand transfer with five repetitions (STS5):

- The TUGT was performed by rising from a sitting position, walking for 3 m in a straight line, turning around, walking back to the chair, and sitting down in the chair, as quickly as possible.
- The AST was performed by standing in front of a platform (which is 19 cm high and 40 cm wide), and alternately placing the left foot on the platform and then back to the floor, repeating four times for each foot, as fast as possible.
- The STS5 was performed by doing five sit-to-stand transfers on a normal height (43 cm) chair, with arms folded in front of the chest, as quickly as possible.

C. Subjects

1) NeuRA subject cohort (2007-2008): In the previous study, 68 subjects aged from 72 to 91 years (80.00 ± 4.40 years, 69% females), were recruited from the Neuroscience Research Australia (NeuRA, formerly named Prince of Wales Medical Research Institute), Sydney, Australia [5]. The University of New South Wales (UNSW, Sydney, Australia) ethics committee approved the study. The inclusion criterion were that the participants must be able to perform the DR assessment and a clinical fall risk assessment, and must have no known major cognitive impairment. All 68 subjects of the NeuRA cohort completed all the three DR tasks and the fall diaries for the following twelve months.

2) Manly subject cohort (2012-2013): A second dataset was collected at the physiotherapy outpatients clinic at Manly Hospital , Sydney, Australia. The study was approved by the Northern Sydney Central Coast Ethics Committee. A total of 44 subjects aged from 68 to 92 years (80.61 ± 6.08 years, 75% females) were recruited. All subjects recruited were referred for physiotherapy assistance with the ultimate aim of preventing falls. The same inclusion criteria as the NeuRA study was used.

Fifteen of the 44 subjects dropped out, mainly due to health problems, before the end of the 12-month follow-up period. Moreover, many subjects could not complete all three DR tasks under the described conditions. Of the 29 subjects who provided a 12-month prospective fall dairies, 17 subjects performed the TUGT task, 27 subjects were able to perform the AST task, and 21 subjects were able to perform the STS5 task. Only 11 subjects were able to perform all three DR tasks and completed all fall diaries in the subsequent 12 months. Data from subjects who were only capable of performing one or two DR tasks are still considered useful, and were used in the modelling analysis based on a single DR task. A comparison of the participants (who provided usable data) in the two cohorts is shown in Table I.

D. Stratification

Subjects from the Manly cohort were generally frailer than those from the NeuRA cohort, for the reason that the Manly subjects were recruited to a physiotherapy service after being referred by relevant clinicians as already having a suspected risk of falling. The NeuRA study, however, involved recruitment from the community, which attracted generally healthy individuals with a lower risk of falling. It is also apparent from Table I that there is a much larger proportion of multiple fallers in the Manly cohort than in the NeuRA cohort. Neither one of the subject cohorts seems to represent the general older population. Thus, data from both cohorts were pooled together and then randomly stratified by gender, age, and fall history into two matched groups, to give training and validation groups with a range of fall risks which are more representative of the community-dwelling older population. Each of the two stratified groups were used as a training sample once, while the other group would be used for validation of the trained model (two-fold cross validation).

TABLE I

COMPARISON BETWEEN THE NEURA COHORT AND SUBJECTS WHO HAD USABLE DATA FROM THE MANLY COHORT.

	Sample size	Mean (SD) age (years)	Number of females vs. males	Previous 12 mo. falls			Prospective 12 mo. falls		
				0	1	2+	0	1	2+
NeuRA	68	80.00 (4.40)	47 vs. 21	46	13	9	42	17	9
Manly*	30	80.37 (5.97)	26 vs. 4	6	9	15	8	11	11

*Only subjects who had prospective 12-month fall diaries and were able to perform at least one DR task (TUGT, AST or STS5) were included.

E. Signal Processing

The accelerometry signals from the DR assessment tasks were first segmented into smaller sections demarking the start, end, and other fiducial events in the signals, before feature extraction. A set of automatic event detection algorithms developed by Redmond *et al.* were employed for signal segmentation. Refer to Narayanan *et al.* [5] for information on the event markers used in this study, and Redmond *et al.* [9] for details of the automatic segmentation algorithms.

Following segmentation, a total of 123 features were extracted from the accelerometry signals generated during the DR assessment. There were 51 temporal and energy-related feature, as described in the study of Narayanan *et al.* [5]; seventy-two additional features were then extracted from the frequency spectra of the $x-$, $y-$ and $z-$axis acceleration components and an acceleration magnitude signal, as described by Liu *et al.* in [6]. In addition to the TA based features, age and gender were also considered for fall prediction. Table II shows a summary of the 125 features.

F. Fall Prediction Model Training and Validation

A logistic regression model was employed to classify between low fall risk subjects and high fall risk subjects, according to their prospective 12-month fall diaries. The study examined the fall risk categorization in two ways: one classifies between non-fallers and fallers, and the other is between non-multiple fallers and multiple fallers. A forward stepwise feature selection method was employed to obtain an nearly optimal subset of features that would fit the training data well. The feasibility of using accelerometry features from all three DR tasks for fall prediction, and using features from each single movement task for fall prediction,

TABLE II

SUMMARY OF THE 125 CANDIDATE FEATURES (ADAPTED FROM [10]).

Feature no.	Feature name
1	TUGT total time duration (log transformed)
2-5	TUGT time intervals for standing, walking 3 m, turning, walking back, sitting (log transformed)
7	TUGT f_{step}
8, 9	TUGT RMS of high-pass filtered SVM and TUGT SMA
10 - 13	TUGT first 6 harm. freq. ratio of SVM, x_{BA}, y_{BA}, z_{BA}
14 - 17	TUGT fund., 2nd, 3rd, 4th harm. magnitude ratio of SVM
19 - 22	TUGT fund., 2nd, 3rd, 4th harm. magnitude ratio of x_{BA}
24 - 27	TUGT fund., 2nd, 3rd, 4th harm. magnitude ratio of y_{BA}
29 - 32	TUGT fund., 2nd, 3rd, 4th harm. magnitude ratio of z_{BA}
18, 23, 28, 33	TUGT even to odd harm. magnitude ratio of SVM, x_{BA}, y_{BA}, z_{BA}
34 - 42	AST total time duration and time intervals for each stepping movement (log transformed)
43, 44	Standard deviation and normalized SD of AST time differences (log transformed)
45, 46	AST dissimilarity of leading foot steps and trailing foot steps
47	AST dissimilarity of leading/trailing step pairs
48, 49	AST RMS of high-pass filtered SVM and AST SMA
50, 51	AST SMA of weakest cycle and strongest cycle
52	AST max. - min. cycle SMA
53	AST SMA ratio between strongest and weakest cycle
54	AST SMA variance per cycle
55, 56	AST SMA of leading foot cycles and lagging foot cycles
57	AST SMA ratio of leading/trailing leg energy
58, 59	AST SMA variance for leading foot cycles and trailing foot cycles
60 - 63	AST first 5 harm. freq. ratio of the SVM, x_{BA}, y_{BA}, z_{BA}
64 - 67	AST fund., 2nd, 3rd, 4th harm. magnitude ratio of SVM
69 - 72	AST fund., 2nd, 3rd, 4th harm. magnitude ratio of x_{BA}
74 - 77	AST fund., 2nd, 3rd, 4th harm. magnitude ratio of y_{BA}
79 - 82	AST fund., 2nd, 3rd, 4th harm. magnitude ratio of z_{BA}
68, 73, 78, 83	AST even to odd harm. magnitude ratio of SVM, x_{BA}, y_{BA}, z_{BA}
84 - 89	STS5 total time duration and time intervals for each STS movement (log transformed)
90, 91	Standard deviation and normalized SD of STS time differences (log transformed)
92	STS5 dissimilarity of sit-to-stand cycles
93, 94	STS5 RMS of high-pass filtered SVM and STS5 SMA
95, 96	STS5 SMA of the weakest and the strongest cycle
97	STS5 max. - min. cycle SMA
98	STS5 SMA ratio between strongest and weakest cycles
99	STS5 SMA variance per cycle
100 - 103	STS5 first 4 harmonics frequency ratio of SVM, x_{BA}, y_{BA}, z_{BA}
104 - 107	STS5 fund., 2nd, 3rd, 4th harm. magnitude ratio of SVM
109 - 112	STS5 fund., 2nd, 3th, 4th harm. magnitude ratio of x_{BA}
114 - 117	STS5 fund., 2nd, 3rd, 4th harm. magnitude ratio of y_{BA}
119 - 122	STS5 fund., 2nd, 3rd, 4th harm. magnitude ratio of z_{BA}
108, 113, 118, 123	STS5 even to odd harm. magnitude ratio of SVM, x_{BA}, y_{BA}, z_{BA}
124-125	Age, Gender

TABLE III

COMPARISON BETWEEN THE TWO STRATIFIED GROUPS.

	Sample size	Mean (SD) age (years)	Number of females vs. males	Previous 12 mo. falls 0	1	2+	Prospective 12 mo. falls[3] 0	1	2+
DR Group I[1]	39	80.03 (4.59)	28 vs. 11	24	9	6	22	14	3
DR Group II	40	79.75 (4.47)	29 vs. 11	24	7	9	22	10	9
TUGT Group I[2]	43	79.56 (4.66)	31 vs. 12	27	8	8	24	13	6
TUGT Group II	42	80.02 (4.20)	30 vs. 12	23	10	9	22	13	7
AST Group I[2]	47	80.34 (4.67)	35 vs. 12	25	12	10	22	15	10
AST Group II	48	80.15 (4.64)	35 vs. 13	26	9	13	27	11	10
STS5 Group I[2]	45	80.40 (4.78)	33 vs. 12	26	9	10	23	15	7
STS5 Group II	44	80.39 (4.80)	33 vs. 11	23	11	10	22	10	12

[1] The DR Group I & II were stratified using subjects from both cohorts who have usable accelerometry data for all three DR tasks.
[2] The TUGT group I & II were stratified using subjects from both cohorts who have usable accelerometry data for TUGT, similar to AST group I & II, and STS5 group I & II.
[3] The prospective 12-month fall data were not used in the stratification process.

were all investigated. Age and gender were included in all investigated feature pool before stepwise selection.

III. RESULTS

Table III shows a comparison of demographics for the two stratified groups using the pooled assessment data from both cohorts, when using accelerometry signals from all three DR tasks, and when using accelerometry signals from single DR tasks. Sample size increases when using accelerometry data from single DR task, as there were fewer subjects who were able to perform all three DR tasks. Table III shows the sample size, mean age, and female to male ratio of the stratified groups. The numbers of subjects in each faller class (non-faller, single faller and multiple faller) according to their fall history and according to prospective fall diaries are also listed in Table III.

Table IV shows the performance of the logistic regression models in classification between non-multiple fallers and multiple fallers (2+ falls). The accuracy in classification of the validation sample is shown in the table, as well as the selected features after stepwise feature selection. Table IV also lists the sensitivity and specificity of the classification performance during model validation, as well as Cohen's kappa coefficient, as a measure of the agreement between the estimated classes and the real faller categories [11]. The performance of the logistic regression models in classification between non-fallers and fallers is shown in Table V.

TABLE IV

THE PERFORMANCE OF THE LOGISTIC REGRESSION MODEL IN CLASSIFICATION BETWEEN NON-MULTIPLE FALLERS AND MULTIPLE FALLERS.

Train	Test	Selected Features*	Accuracy	Sensitivity	Specificity	κ
DR Group I	DR Group II	{10,14,46}	73%	11%	90%	0.02
DR Group II	DR Group I	{21,101,102,116}	82%	0%	89%	-0.10
TUGT Group I	TUGT Group II	{4,18}	83%	14%	97%	0.16
TUGT Group II	TUGT Group I	{1,10,15,21}	79%	50%	84%	0.28
AST Group I	AST Group II	{35}	73%	0%	92%	-0.11
AST Group II	AST Group I	{}	79%	0%	100%	0.00
STS5 Group I	STS5 Group II	{99,100,110,117,118}	73%	22%	86%	0.09
STS5 Group II	STS5 Group I	{95,102,121}	77%	20%	94%	0.18

*Feature no. refers to Table II.

TABLE V

THE PERFORMANCE OF THE LOGISTIC REGRESSION MODEL IN CLASSIFICATION BETWEEN NON-FALLERS AND FALLERS.

Train	Test	Selected Features*	Accuracy	Sensitivity	Specificity	κ
DR Group I	DR Group II	{11,44,51,57,90,121}	63%	58%	67%	0.25
DR Group II	DR Group I	{5,10,13,36,100,102}	62%	47%	73%	0.20
TUGT Group I	TUGT Group II	{2,4}	52%	25%	77%	0.02
TUGT Group II	TUGT Group I	{8,10,33}	47%	53%	42%	-0.06
AST Group I	AST Group II	{38,42,47,73}	67%	71%	63%	0.34
AST Group II	AST Group I	{35,42,52}	70%	68%	73%	0.41
STS5 Group I	STS5 Group II	{109}	56%	55%	56%	0.11
STS5 Group II	STS5 Group I	{86,123}	47%	54%	60%	0.14

*Feature no. refers to Table II.

IV. DISCUSSION AND CONCLUSION

Firstly, this study has investigated the ability of stepwise logistic regression modelling to classify between non-multiple fallers and multiple fallers, when using accelerometry data from three different DR tasks. It must be noted that the multiple faller categories are heavily unbalanced in all the cases, whichever DR tasks were used, as there are only 20 out of all 98 subjects (drawn from both cohorts) in the multiple faller category (see Table III). Using accelerometry data from all three DR tasks, the selected models obtained a poor validation performance when testing on the other subject group (judged from sensitivity, specificity and κ derived from estimated classes and the real faller classes). The model trained on DR Group I obtained a validation performance with sensitivity $= 11\%$, specificity $= 90\%$ and $\kappa = 0.02$, while the model trained on DR Group II had a similar validation performance (sensitivity $= 0\%$, specificity $= 89\%$ and $\kappa = -0.10$). Using only accelerometry data from the TUGT, the selected logistic regression model had the

best validation performance in classification between non-multiple fallers and multiple fallers. The model trained on TUGT Group I had a validation performance of sensitivity $= 14\%$, specificity $= 97\%$ and $\kappa = 0.16$, and the model trained on TUGT Group II gave sensitivity $= 5\%$, specificity $= 84\%$ and $\kappa = 0.28$ in validation. Both AST and STS5 features showed poor validation performance. Age and gender were included in the feature pool, but were not selected in any logistic regression models.

Secondly, the logistic regression model was also investigated in classification between non-fallers and fallers; these two fall risk classes are much more balanced (48 of the 98 subjects are in the faller category). From Table V, when using accelerometry data from all three DR tasks, the model performed relatively better than the performance in classifying between non-multiple fallers and multiple fallers. The model trained on DR Group I had a validation sensitivity $= 58\%$, specificity $= 67\%$ and $\kappa = 0.25$, while the model trained on DR Group II achieved sensitivity $= 47\%$, specificity $= 0.73$ and $\kappa = 0.20$ in validation. Interestingly, the model using AST data alone obtained the best performance, with sensitivity $= 71\%$, specificity $= 63\%$ and $\kappa = 0.34$ for the model trained on AST Group I, and sensitivity $= 68\%$, specificity $= 73\%$ and $\kappa = 0.41$ for the model trained on AST Group II. One AST feature was selected in both models, which is the log-transformed time interval for the last stepping movement. The TUGT data and STS5 data, when used in the stepwise logistic regression model separately, both had a poor performance in validation. Again, age and gender were not selected in any models.

It is possible that the trained models are over-fitted to the available training data. The large feature dimensionality and the small sample size might contribute to the over-fitting problem, as there will be an increased chance of discovering a false relationship between some feature and the target in the small training sample, especially when all 123 features from the three DR tasks were included, with only less than 40 subjects used for model training; although the feature number was reduced significantly after feature selection. The same dimensionality problem can be seen from the previous study by the author [6], where an extremely over-optimized model was obtained with over 30 features selected for only 68 training sample, and no independent validation performed.

The lack of common selected features between models trained on the two stratified groups, and their subsequent poor validation performance, also indicates that the trained models do not accurately fit the true data distribution for the general population. The number of training examples (≤ 48) is less than half of the total number of the pooled sample, and will hardly be representative of the general older population. Besides, the entire sample size is also relatively small(≤ 98). While it is pooled from a both generally healthy subject cohort and a frailer cohort, it may still not represent the general population in this age group.

It is also noted that the reliability of the extracted features was not examined. A test-retest reliability investigation may also help in feature dimensionality reduction. Features which are not reliable between task repeats, or are sensitive to minor variations in device placement, should be discarded before model training.

Many researchers have also used body-worn sensors with certain movement tasks for prospective fall prediction. Greene *et al.* reported a mean accuracy of 79.69% with cross validation, in classification between fallers and non-fallers, using gait parameters extracted from two shank-attached sensors during TUGT [7]. However, it seems that features were selected based on entire data set before cross-validation, thus the reported performance does not reflect the general performance on an independent data. Doi *et al.* reported a good training performance (specificity $= 84.2\%$, sensitivity $= 68.8\%$) using a stepwise logistic regression model [8]. However, the stepwise regression was performed on features that were significantly correlated with falling in bivariate analysis, thus the trained model might be over-optimistic. Moreover, no validation performance was reported. Most of the previous studies did not have proper validation using data that are independent from the training process, and thus their expected predictive performance on unseen data is unknown. A proper independent validation is required to estimate the model's future performance when applied to the general population.

REFERENCES

[1] C. Bradley, *Hospitalisations due to falls by older people, Australia 2007-08*, ser. Series no. 61, Cat. no. INJCAT173, 2012.

[2] S. R. Lord, C. Sherrington, and H. B. Menz, *Falls in older people: Risk factors and strategies for prevention*. Cambridge: Cambridge University Press, 2001.

[3] L. A. Lipsitz, P. V. Jonsson, M. M. Kelley, and J. S. Koestner, "Causes and correlates of recurrent falls in ambulatory frail elderly," *Journals of Gerontology*, vol. 46, no. 4, pp. M114–M122, July 1991.

[4] A. Tiedemann, H. Shimada, C. Sherrington, S. Murray, and S. Lord, "The comparative ability of eight functional mobility tests for predicting falls in community-dwelling older people," *Age and Ageing*, vol. 37, pp. 430–435, 2008.

[5] M. R. Narayanan, S. J. Redmond, M. E. Scalzi, S. R. Lord, B. G. Celler, and N. H. Lovell, "Longitudinal falls-risk estimation using triaxial accelerometry," *IEEE Transactions on Biomedical Engineering*, vol. 57, pp. 534–541, 2010.

[6] Y. Liu, S. J. Redmond, N. Wang, F. Blumenkron, M. R. Narayanan, and N. H. Lovell, "Spectral analysis of accelerometry signals from a directed-routine for falls-risk estimation," *IEEE Transactions on Biomedical Engineering*, vol. 58, pp. 2308–2315, 2011.

[7] B. R. Greene, E. P. Doheny, C. Walsh, C. Cunningham, L. Crosby, and R. A. Kenny, "Evaluation of falls risk in community-dwelling older adults using body-worn sensors," *Gerontology*, vol. 58, no. 5, pp. 472–80, 2012.

[8] T. Doi, S. Hirata, R. Ono, K. Tsutsumimoto, S. Misu, and H. Ando, "The harmonic ratio of trunk acceleration predicts falling among older people: results of a 1-year prospective study," *Journal of Neuroengineering and Rehabilitation*, vol. 10, 2013.

[9] S. J. Redmond, M. E. Scalzi, M. R. Narayanan, S. R. Lord, S. Cerutti, and N. H. Lovell, "Automatic segmentation of triaxial accelerometry signals for falls risk estimation," in *32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Buenos Aires, Argentina, 2010, pp. 2234–2237.

[10] Y. Liu, S. J. Redmond, M. R. Narayanan, and N. H. Lovell, "Classification between non-multiple fallers and multiple fallers using a triaxial accelerometry-based system," in *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, September 2011, pp. 1499 –1502.

[11] J. L. Fleiss, J. Cohen, and B. S. Everitt, "Large sample standard errors of kappa and weighted kappa," *Psychological Bulletin*, vol. 72, pp. 323–327, 1969.