

Developing a Single-Item General Self-Efficacy Scale: An Initial Study

Weiwei Di (✉ nie19.dw@e.ntu.edu.sg)

Nanyang Technological University <https://orcid.org/0000-0003-1990-079X>

Youyan Nie

Nanyang Technological University

Bee Leng Chua

Nanyang Technological University

Stefanie Chye

Nanyang Technological University

Timothy Teo

Murdoch University

Research

Keywords: General self-efficacy, single item, reliability, validity

Posted Date: April 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-342642/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Purpose General self-efficacy represents a global sense of personal capability across various situations and tasks. The aims of present study were to develop and validate a single-item general self-efficacy scale which balances the practical demands and psychometric concerns.

Methods The psychometric properties of the proposed Single-Item General Self-Efficacy Scale (GSE-SI) was examined among 231 Singaporean adults. GSE-SI was selected through experts and four statistical methods. Besides, the reliability, criterion-related, construct-related validity and discriminative power of GSE-SI were examined.

Results Three reliability tests demonstrated a good reliability of GSE-SI (.594 .607 and .726, respectively, $M = .642$), comparing with other single item scales. It also showed a satisfactory criterion-related validity (i.e. correlation with a multiple-item general self-efficacy scale, $r = .795$). The construct-related validity was supported by the correlations between general self-efficacy and six relevant constructs (i.e. positive correlations with life satisfaction and positive emotions, negative correlations with negative emotions, task and perceived stress and illness symptoms). Importantly, GSE-SI and multiple-item scale showed consistent correlation patterns with relevant constructs. GSE-SI also performed similar discriminations across three respondent clusters divided based on six constructs, in accordance with the multiple-item scale.

Conclusions GSE-SI is a reliable and valid measurement of general self-efficacy and could be recommended in future research.

Introduction

Research on quality of life has gained popularity in the past few years and people's self-efficacy beliefs have been used as one of the key indicators of their quality of life in various types of research and practice (1, 2). As a widely investigated self-efficacy belief, general self-efficacy (GSE) was conceptualized as a global sense of one's competency across various situations (3). A strong sense of GSE contributes to a more positive life (4, 5) with higher achievement (6, 7), better health (1, 8) and social skills (9, 10). A great deal of research has shown that general self-efficacy was not only correlated with many indicators of life quality such as life satisfaction (11), physical health (12) and social life (5), but also widely used as an indicator to monitor and evaluate the effectiveness of intervention programmes in improving quality of life, such as stress and depression intervention (10, 13, 14).

From the literature, the existing GSE scales were all multiple-item scales (3, 15-19). Among them, three GSE scales developed by Sherer, Maddux (3), Jerusalem and Schwarzer (15) and Chen, Gully (16) are the most widely used measurements in GSE research. Nevertheless, it should be noted that the length of the scale and answering time are major concerns, especially for research with repeated measures, such as longitudinal studies or intervention programmes. Some researchers argued that shortened scales comprised of a few items could alleviate the practical issues, such as the 6-item General Self-Efficacy

Scale (17). However, the additional items, even the second or third item, might affect the willing of respondents to participate in the study, especially when the survey is online mode. In addition, single item might be more suitable if the psychometric properties of single-item measures could be adjusted to a similar level as multiple-item ones. Hence, there is a trend to develop single-item measures (17, 20).

Single-item measures have received great attention from quality of life and psychological researchers and a number of related single-item scales have been proposed. For instance, DeSalvo, Fisher (21) have developed two single-item self-rated measures to assess patients' general health. Moreover, many other general health single-item scales have been widely used (2, 22, 23). Besides, single-item life satisfaction scale (24), single-item quality of life scale (25), single-item social identification scale (26) and many others (e.g. physical activities, depressive and somatization symptoms, emotional exhaustion) have also been proposed and widely used in related research (27-29). However, as one of the most widely used constructs in psychological research, single-item GSE scale has not been developed. Moreover, the approaches to develop single-item measures were different in previous studies. There has been no unanimity among single-item research on the developing method. Therefore, the purpose of the current study is to develop and validate a single-item general self-efficacy scale, moreover, to compare and integrate different single-item developing methods.

Method Of Developing Single-item Scales

Item Selection

In prior single-item measurement studies, how to select the particular item has not received a broad consensus. Most researchers decided on the particular item according to expert judgement recommended by Rossiter (30). In addition, some statistical methods have also been suggested to select item, such as item-to-total correlations and factor analysis (31, 32). However, statistical methods have not been widely used and reported. Among the few studies involved statistical methods, different statistical methods were used in different studies (20). Thus, it is hard to compare which selection method is better. As a result, the justification for selected item is weak due to the lack of cross validations using multiple methods. Therefore, the first purpose of the current study is to select single item scale using both expert judgement and three statistical methods introduced by Loo (31), Sarstedt and Wilczynski (20) and Sarstedt, Diamantopoulos, Salzberger and Baumgartner (32).

Reliability

Since internal consistency reliability (Cronbach's alpha) cannot be conducted for single item, some other statistical methods to examine the reliability of single item scale have been proposed. First, Wanous and Reichers (33) first introduced the method to utilize the correction for attenuation formula to estimate the minimum level of reliability of a single-item scale. In previous studies, this method has been used by some single-item researchers (26). Second, Weiss (34) proposed that factor analysis could also be used to estimate single-item reliability and a few researchers have applied it into practice (35). Furthermore, Wanous and Hudy (36) estimated the reliability of a single-item teaching effectiveness measure with two

extraction methods in factor analysis, principal axis and maximum likelihood. Their results showed that the two extraction methods were almost identical.

However, there is no broad consensus in the best way to estimate single item reliability. Moreover, it can be noted that previous studies rarely employed sufficient and various methods to evaluate reliability of single-item scales although single-item scales suffer from the doubt about potential statistical disadvantages (37). Thus, the second purpose of the current study is to evaluate and compare the reliability of the proposed single item scale based on multiple approaches. In the current study, the reliability will be estimated by both the correction for attenuation formula and factor analysis (principal axis factoring and maximum likelihood).

Criterion-Related and Construct-Related Validity

Previous single-item studies mainly focused on criterion-related and construct-related validity. For criterion-related validity, the correlation between single-item and the corresponding multiple-item scale was often used as the support (27). In the case of construct-related validity, it was mainly verified by the correlations between the target construct and the theoretically relevant variables (24). In present study, the third purpose is to evaluate the criterion-related and construct-related validity of the proposed single-item scale. Specifically, for criterion-related validity, Pearson's correlations between the multiple-item scale and the proposed single-item scale will be calculated. For construct-related validity, Pearson's correlations between GSE and six theoretically related constructs will be conducted. A large body of findings has reported significant correlations between GSE and life satisfaction, mental well-being, stress and illness symptoms (4, 5, 38). Therefore, in the present study, the scores from scales measuring life satisfaction, positive emotions, negative emotions, perceived stress, task stress and illness symptoms will be assessed as well.

Enhancing the Validity of Single-Item Scales

The use of single-item measure has been challenged because the single-item validity was often regarded as less persuasive than that of multiple-item measures. However, it should be noted that, considering the effect of measurement errors, the observed correlations of single-item measures could be biased. That is, the validity of single-item measures could be underestimated, especially the criterion-related and construct-related validity. Regarding this problem, Pedhazur (39) has proposed a method to adjust the single-item correlations. When the single-item reliability is applied into the correction for attenuation formula, the single-item correlations could be enhanced. The validity of single-item measures could be strengthened accordingly. Hence, the fourth purpose of present study is to examine the adjusted correlations of the proposed scale with relevant constructs as a more valid use of the single-item scale.

Discriminative Power

There is lack of information on the discriminative power in previous single-item scale studies even though it is also a vital indicator of the scale validity. In the few single-item studies that reported the discriminative power analysis, demographic variables (e.g. gender, age) were mainly used (40) rather than psychological variables. However, from a perspective of life quality, the discrimination across psychological variables could be more meaningful than demographics. Therefore, the fifth purpose of present study is to evaluate the discriminative power of the proposed single-item scale. Specifically, we will group respondents by cluster analysis based on six relevant psychological constructs, then a one-way ANOVA will be conducted across different respondent groups.

The Significance of the Present Study

The overall aim of the present study is to develop a single-item GSE scale by conducting various selection approaches and sufficient reliability and validity analysis. There are two major contributions of the current study. First, the current study will significantly contribute to scale development and GSE research domains by providing an alternative choice for researchers (single item vs multiple item), especially when length of the scale and time are crucial factors in a study. The proposed single-item general self-efficacy scale has the potential to provide a psychometrically sound tool to address the practical demands and measurement concerns in applied research. The proposed scale could be widely used in both academic and practical quality of life research, such as correlational studies and intervention programmes.

Second, since single-item scale reliabilities are commonly lower than those of multiple-item scales, the present study will also provide persuasive empirical evidence in support of single-item scale reliability and suggestions on how to adjust single-item scales to improve its reliability with an aim to approximate the effects produced by multiple-item scales in empirical research.

Methods

Sample and Procedures

The current sample comprises 231 adult participants above 21 years old (172 females, 59 males) enrolled in Singapore. Anonymous surveys with same contents were administered to them through either paper-based or online surveys. A convenient sampling was used to recruit 186 adults responded on paper-based questionnaires. The administration was under supervision of research team members, who were responsible to offer further explanations and ensure the confidentiality. Additionally, a snowball sampling was utilized to enrol 45 participants responded online via Qualtrics software. An independent-samples t-test was conducted and no significant difference was found between two survey approaches. English was the language used in all survey forms.

Measures

General Self-efficacy

A shortened version of NGSE developed by Chen, Gully (16) was adapted and used in the current study. It included 7 items and was labelled as NGSE-7. The coefficient alpha of the NGSE-7 was .870 in this sample. Participants responded to all items on a 5-point rating scale from 1 (*Not true of me*) to 5 (*Very true of me*).

Theoretically Relevant Variables

Life Satisfaction, Task Stress and Illness Symptoms. The 5-item Life Satisfaction (LS) scale, 9-item Task Stress (TS) scale and 4-item Illness Symptoms (IS) scale have been developed by Pettegrew and Wolf (41) to measure teacher stress. In this study, some items were slightly reworded to apply into general population (e.g., "Trying to provide a good education in an atmosphere of decreasing financial support is very stressful." into "Trying to get a good work outcome is very stressful."). The internal consistency reliabilities of the three scales were comparatively high in present sample ($\alpha = .788, .835, .732$, respectively). To maintain the consistency of whole survey form style, the responding format was simplified from original 6-point to 5-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree).

Perceived Stress. This construct was measured with the 14-item Perceived Stress (PS) scale developed by Cohen, Kamarck (42) to assess PS level of community samples with a junior high school certificate and above. Two of the items were slightly reworded to fit more general situations. The coefficient alpha for the PS scale was .819 in this sample. To align with overall rating method, in this study, items were scored on a 5-point frequency rating scale ranging from 1 (Never) to 5 (Always). The PS scale scores were obtained by summing each item score after reversing scores of seven positive items.

Positive and Negative Emotions. Both positive emotions (PEM) and negative emotions (NEM) were measured by revised scales based on the Scale of Positive and Negative Experience (SPANE) developed by Diener, Wirtz (43). In this study, two items (happy and good) retained from SPANE - P and another added two items (cheerful and excited) composed the PEM scale. The NEM scale consists of six items, including two items (angry and sad) derived from SPANE - N and four reworded items (frustrated, lonely, disrespected and miserable). The coefficient alpha of PEM and NEM scales were .898 and .812, respectively. All items were responded on the same 5-point scale from 1 (Never) to 5 (Always).

Data Analysis and Results

Selection of Single Item

The particular item was selected from NGSE-7 based on following five approaches (32): (1) expert judgement, (2) the highest item-to-total correlations (ITC), (3) the highest factor loading extracted from principal axis factoring (PAF), (4) the highest factor loading extracted from confirmatory factor analysis (CFA) and (5) the highest squared multiple correlations (SMC) in CFA.

Nine scholars with earned doctorates in psychology and with faculty appointments at a research-intensive university provided their comments on each item of the NGSE-7. Experts' views were not

consistent, thus their views had limited value in the selection of the single item. Therefore, four statistical methods were used, and the results are presented in Table 1. Four different statistical methods converged on one particular item, pointing out that item 4 captures the construct of GSE most. Only one factor was extracted in EFA, with an eigenvalue of 3.460, accounting for 49.43% of total variance and a one-factor congeneric CFA model was applied. Item 4 was reported with the highest ITC, factor loadings and SMC among all seven items. Furthermore, Cronbach's alpha of the whole scale would decrease from .870 to .844 if item 4 deleted, less than the value of any other item. Overall, the results from four statistical methods showed clear and relatively constant ranks for all items and item 4 was elected as the most representative one in NGSE-7 consistently. Therefore, based on the item selections and cross-validations across five different methods, item 4 ("*I believe I can succeed at most of any endeavour to which I set my mind*") was finally selected to be the Single-Item General Self-Efficacy Scale (GSE-SI).

Table 1 Item Selection Based on Four Statistical Methods ($n = 231$)

Items	ITC	Factor loading (PAF)	Factor loading (CFA)	SMC
1. I will be able to achieve most of the goals that I have set for myself.	.615	.670	.679	.460
2. When facing difficult tasks, I am certain that I will accomplish them.	.665	.724	.723	.522
3. In general, I think that I can obtain outcomes that are important to me.	.696	.759	.766	.587
4. I believe I can succeed at most of any endeavour to which I set my mind.	.701	.770	.779	.607
5. I will be able to successfully overcome many challenges.	.644	.696	.696	.485
6. I am confident that I can perform effectively on many different tasks.	.669	.711	.694	.481
7. Compared to other people, I can do most tasks very well.	.537	.573	.561	.315

Note. ITC: item-to-total correlation; PAF: principal axis factoring; CFA: confirmatory factor analysis; SMC: squared multiple correlation

Reliability

The results obtained from three different reliability tests (i.e. 1) correction for attenuation formula, 2) PAF in EFA and 3) maximum likelihood in EFA) were listed in Table 2. Since the two variables in current study

come from the same conceptual domain (GSE), the estimated reliability of GSE-SI was .726 based on the following formula (44):

$$r_{xy} = \sqrt{r_{xx}r_{yy}}$$

Based on the communality obtained from EFA (45), the reliability of GSE-SI was at least .644 and .607, respectively. Overall, the reliability values ranged from .594 to .726, with a mean of .642, indicating a meritorious reliability of GSE-SI.

Table 2 Reliability Based on Three Methods ($n = 231$)

	CAF	Communalities (PAF)	Communalities (ML)
I believe I can succeed at most of any endeavour to which I set my mind.	.726	.594	.607

Note. CAF: correction for attenuation formula; PAF: principal axis factoring; ML: maximum likelihood

Validity

Criterion-Related and Construct-Related Validity

Concurrent validity was assessed by calculating the Pearson's correlation between the GSE-SI and NGSE-7. The results showed a strong positive correlation of .795 at .01 level (two-tailed), indicating that the responses to the GSE-SI are highly correlated with those to the NGSE-7.

Construct-related validity were evaluated by comparing the Pearson's correlations between the two GSE measures and six theoretically related variables. As Table 3a shown, the patterns of correlations with other six measures were exactly same across two scales ($p < .01$, two-tailed). Positive emotions were positively correlated with both GSE-SI and NGSE-7. Similarly, positive correlations between life satisfaction and GSE-SI and NGSE-7 were obtained with comparatively lower but still significant coefficients. Negative emotions and illness symptoms were negatively correlated with both two GSE measures with limited difference. Besides, perceived stress showed strongest negative correlations with two GSE measures and task stress showed the same correlation pattern with slightly lower coefficients.

To further test the construct validity of GSE-SI, the variance reductions were calculated based on the correlations for two GSE measures. When for the GSE-SI were subtracted from those for the NGSE-7, the average reduction was .055, ranging from .022 to .115 (Table 3b). Therefore, single-item and multiple-item GSE scales did not show a significant difference in correlations with other variables.

Adjusting with Reliability: More Valid Use

To gain more confidence, the correlation coefficients between GSE-SI and six theoretically relevant variables were corrected with its reliability based on the correction for attenuation formula (39) with three estimated and the mean reliability of GSE-SI applied (Table 3a). Based on the corrected correlations, the variance reductions decreased in all four correction scenarios, comparing to uncorrected results (Table 3b). Specifically, the average variance reductions calculated by three estimated reliabilities were .013, -.010 and -.008, respectively. Notably, when applying the mean reliability, the reduction in the average variance was only -.001. In addition to the means, standard deviations based on four reliabilities were almost the same (.025) and lower than original values. Overall, the slight differences in correlations indicated that the GSE-SI and the NGSE-7 shared a very similar correlation with the six theoretically relevant constructs.

Table 3a Correlations with Related Variables: Single-item versus Multiple-item General Self-efficacy Scale

	/		Corrected /			
	GSE-SI	NGSE-7				
Positive emotions	.256**/.066	.368**/.135	.317/.100	.351/.123	.347/.120	.337/.114
Life satisfaction	.192**/.037	.311**/.097	.254/.065	.281/.079	.278/.077	.270/.073
Negative emotions	-.179**/.032	-.245**/.060	-.233/.054	-.258/.066	-.255/.065	-.248/.061
Task stress	-.290**/.084	-.342**/.117	-.372/.138	-.412/.170	-.407/.166	-.396/.157
Perceived stress	-.343**/.118	-.483**/.233	-.445/.198	-.492/.242	-.486/.237	-.473/.224
Illness symptoms	-.188**/.035	-.239**/.057	-.258/.067	-.285/.081	-.282/.080	-.274/.075

* $p < .01$.

The reliability estimated by correction for attenuation formula (.726) was applied.

The reliability estimated by principal axis factoring (.594) was applied.

The reliability estimated by maximum likelihood (.607) was applied.

The mean reliability (.642) was applied.

Table 3b Variance Reduction in Related Variables: Single-item versus Multiple-item General Self-efficacy Scale

	Variance reduction compared with NGSE-7				
Positive emotions	.069	.035	.012	.015	.021
Life satisfaction	.060	.032	.018	.020	.024
Negative emotions	.028	.006	-.006	-.005	-.001
Task stress	.033	-.021	-.053	-.049	-.040
Perceived stress	.115	.035	-.009	-.004	.009
Illness symptoms	.022	-.010	-.024	-.023	-.018
<i>M</i>	.055	.013	-.010	-.008	-.001
<i>SD</i>	.035	.025	.026	.025	.025

The reliability estimated by correction for attenuation formula (.726) was applied.

The reliability estimated by principal axis factoring (.594) was applied.

The reliability estimated by maximum likelihood (.607) was applied.

The mean reliability (.642) was applied.

Discriminative Power

Two steps were conducted to classify the participants based on six theoretically relevant constructs. First, a hierarchical cluster analysis was utilized to explore the optimal number of clusters. In this step, Ward's method and the squared Euclidean distance were opted, meanwhile, participants were allocated into different clusters. The results showed that the first huge gap in the coefficient values from the agglomeration schedule indicated that the significant clusters occurred at stage 228, suggesting three clusters as the optimum solution. Second, a k-means cluster analysis was conducted using the means of three previous clusters as the initial cluster centres. Subsequently, the participants were relocated into the final three clusters. Importantly, a cross validation procedure was repeated five times to confirm the validity of this 3-cluster solution. The sample was randomly divided into two parts (115 and 116 participants, respectively). The similar two steps were conducted on two parts except that the initial centres for two parts in k-means cluster analysis both used the cluster means based on one part.

Subsequently, the Cohen's κ values were calculated to test the consistency between new and original clusters. The mean Cohen's κ was .824 across five times cross validations, verifying the stable 3-cluster solution (46). Descriptive statistics and a one-way ANOVA was used to investigate the differences across clusters in six related constructs. The results were presented in Table 4. Homogeneity of variance could be assumed only in illness symptoms, in which the Bonferroni post hoc tests were conducted. Where homogeneity could not be assumed, the Brown-Forsythe F ratio was reported and Games-Howell post hoc tests were used. The ANOVA results revealed significant differences between three clusters across all six relevant constructs ($p < .001$). A post-hoc test showed that all comparisons were significant ($p < .01$). Thus, based on the score patterns, the three clusters were named as negative (high on four negative constructs, low on two positive constructs), balanced (no obvious tendency) and positive (high on two positive constructs, low on four negative constructs).

A one-way ANOVA was conducted to assess the discriminative power of the GSE-SI across three clusters. As expected, GSE measured by two GSE scales both increased from negative group to positive group. As Table 5 shown, both GSE-SI and NGSE-7 had significantly discriminated between three clusters in an extremely similar pattern ($F(2, 228) = 9.127$ and 21.122 , respectively, $p < .001$).

Table 4 Discriminations on Related Variables across Clusters

	Negative ($n = 60$)		Balanced ($n = 115$)		Positive ($n = 56$)		F ratio	p -Value
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Positive emotions	2.858	.632	3.387	.557	4.013	.579	54.686	< .001
Life satisfaction	3.250	.590	3.784	.529	4.329	.471	59.531	< .001
Negative emotions	2.883	.586	2.216	.467	1.804	.490	64.083	< .001
Task stress	3.643	.476	3.024	.427	2.264	.487	125.662	< .001
Perceived stress	3.258	.283	2.727	.208	2.305	.260	201.278	< .001
Illness symptoms	3.121	.812	2.554	.605	1.786	.581	59.732	< .001

Table 5

Discriminations across Clusters: Single-item versus Multiple-item General Self-efficacy Scale.

	Negative (<i>n</i> = 60)		Balanced (<i>n</i> = 115)		Positive (<i>n</i> = 56)		F ratio	<i>p</i> -Value
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
GSE-SI	3.870	.853	4.170	.652	4.430	.657	9.127	< .001
NGSE-7	3.700	.594	3.993	.453	4.289	.424	21.122	< .001

Discussion

The present study developed a single-item scale of general self-efficacy based on various approaches, furthermore, the reliability and validity of proposed scale were assessed.

Item selection

In single-item scale research, there is no broad consensus for the best method to select the single item and expert judgement were often used. However, the current study found that experts could only provide information with limited value. Thus, researchers may need to consider some other approaches in item selection to help decision making, such as statistical methods. Sarstedt, Diamantopoulos (32) have concluded that experts can hardly be expected to select the best item due to the instability. Furthermore, they noted that the item selections were of high consistency among all statistical methods. The current findings were consistent with previous study in that four statistical selection processes converged to one particular item (32). In summary, decisions on single item selection should be based on multiple approaches and more comprehensive information such as the consistency of results obtained from different methods.

Reliability

Several methods have been proposed in this study to estimate single-item reliability. However, we did not find evidence to suggest the best. Therefore, the mean reliability can be considered as the reliability of single-item scales when differences appear across several estimation methods. Despite the notion that single-item scale reliabilities may be inferior to those of multiple-item scales, the reliability of GSE-SI in present study is shown to be satisfactory when compared to multiple-item GSE scales (.70-.90, .75-.91, .85-.90) (3, 15, 16). Comparing with other empirically-supported single-item scales, Postmes, Haslam (26) found the average estimate reliability of single-item scales in a meta-analysis to be .51, ranging from .14 to .68 (e.g., single-item perceived stress scale (47) and single-item burnout scale (48)), revealing a remarkable reliability of GSE-SI proposed in this study (mean reliability = .642). Moreover, the “true” reliability of GSE-SI is probably higher since both methods produce the minimum estimate of reliability. Thus, results of the present study indicate that GSE-SI has a meritorious reliability.

Criterion-Related and Construct-Related Validity

In terms of criterion-related validity of single-item measures, the correlations between homogeneous single-item measures and multiple-item scales were commonly at a modest level (27, 49). Therefore, based on the present findings, criterion-related validity of the proposed GSE-SI can be considered to be quite high. Regarding construct-related validity, it was often supported by the correlations between single-item measures with relevant constructs. In the current study, GSE-SI also showed similar and consistent correlation patterns as NGSE-7 with six relevant constructs. However, it can be noted that the values of correlations with theoretically relevant variables of GSE-SI were slightly lower than those of NGSE-7.

The Importance to Use Reliability to Adjust Correlation

To gain more valid use of single-item measures, some researchers have suggested that single item correlations could be corrected by its reliability using the correction for attenuation formula (39), a method used by Cheung and Lucas (24) to support the concurrent validity of their single-item measure. In present study, the correlations with relevant constructs of GSE-SI were comparable to the correlations of the well-established multiple-item scale after corrected with reliability. Moreover, the corrected variance differences and standard deviations between two GSE measures were also calculated. In the four correcting contexts, the extremely low variance reductions and stable standard deviations all suggested that GSE-SI had correlated with other related constructs as strongly and consistently as the well-established NGSE-7, providing strong and sufficient support for the construct-related validity of GSE-SI. Therefore, when single item is used in correlational study, adjusted corrections using reliability are recommended to achieve similar correlation strength comparing with multiple items.

Discriminative Power

In previous studies, discriminative power was mainly verified across demographic variables such as gender and age (40). Therefore, the analysis methods to identify respondent groups were often simple. In the present study, the evaluation of the discriminative power of the proposed GSE-SI mainly focused on the respondents' different respond patterns. Moreover, three respondent groups in this study were divided based on five times cross-validated cluster analysis rather than simply demographic characteristics in previous. Overall, the significant differences across three groups and similar discrimination patterns between two GSE measures supported that GSE-SI differentiate respondents as powerful as the well-established multiple-item NGSE-7.

Limitations and Future Directions

This study conducted various analysis methods and sufficient cross-validations throughout the development of GSE-SI and several limitations were noted which future research could address. First, as one of the first attempts to develop the single-item GSE-SI, this study had employed convenient sampling with a limited sample size. To further establish the norm of GSE-SI, future research could expand the sample size aimed at achieving representativeness for empirical research. Second, the current sample

was only recruited in Singapore. In spite of the useful results in this study, further cross-culture validations are required in the future, with a view to establish validity and reliability for use of the GSE-SI into other cultures or contexts.

Conclusions

Contrary to the impression that single-item scales are inferior to multiple-item scales, the proposed Single-item General Self-Efficacy Scale in this study showed satisfactory reliability, criterion-related validity, construct-related validity and discriminative power compared to the well-established multiple-item scale, NGSE-7. Thus, current findings support the feasibility of using single-item measures in future general self-efficacy research. Furthermore, since the single-item measures reduce much cost to both researchers and participants, using a single-item scale of general self-efficacy should be encouraged and be recommended as a reliable and valid measure in future research.

Declarations

Funding

National Institute of Education (NIE) Academic research fund.

Conflicts of interest/Competing interests

No.

Availability of data and material

Available by request.

Authors' contributions

First author, Weiwei Di, analyzed data and wrote the manuscript. Second author, Youyan Nie, designed the research, co-analyze the data and co-write the manuscript. Third co-authors, Bee Leng Chua, Stefanie Chye, Timothy Teo participated in the research design and editing the manuscript. All authors read and approved the final manuscript.

Ethics approval

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Nanyang Technological University (NTU) IRB.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Consent for publication

N.A.

References

1. Nützel A, Dahlhaus A, Fuchs A, Gensichen J, König H-H, Riedel-Heller S, et al. Self-rated health in multimorbid older general practice patients: a cross-sectional study in Germany. *BMC Family practice*. 2014;15(1):1.
2. Cullati S, Bochatay N, Rossier C, Guessous I, Burton-Jeangros C, Courvoisier DS. Does the single-item self-rated health measure the same thing across different wordings? Construct validity study. *Quality of Life Research: Care: an International Journal of Quality of Life Aspects of Treatment*; 2020.
3. Sherer M, Maddux JE, Mercandante B, Prentice-Dunn S, Jacobs B, Rogers RW. The self-efficacy scale: Construction and validation. *Psychological reports*. 1982;51(2):663–71.
4. Azizli N, Atkinson BE, Baughman HM, Giammarco EA. Relationships between general self-efficacy, planning for the future, and life satisfaction. *Personality Individ Differ*. 2015;82:58–60.
5. Luszczynska A, Gutiérrez-Doña B, Schwarzer R. General self-efficacy in various domains of human functioning: Evidence from five countries. *International journal of Psychology*. 2005;40(2):80–9.
6. Schwarzer R. *Self-efficacy: Thought control of action*: Taylor & Francis; 2014.
7. Feldman DB, Kubota M. Hope, self-efficacy, optimism, and academic achievement: Distinguishing constructs and levels of specificity in predicting college grade-point average. *Learning Individual Differences*. 2015;37:210–6.
8. Wu A, Tang CS-kk, Kwok T. Self-efficacy, health locus of control, and psychological distress in elderly Chinese women with chronic illnesses. *Aging Ment Health*. 2004;8(1):21–8.
9. Bandura A. 199 *Self-efficacy: The exercise of control*. NY: H FreemanNew York; 1997.
10. Pössel P, Baldus C, Horn AB, Groen G, Hautzinger M. Influence of general self-efficacy on the effects of a school-based universal primary prevention program of depressive symptoms in adolescents: a randomized and controlled follow-up study. *J Child Psychol Psychiatry*. 2005;46(9):982–94.
11. Capri B, Ozkendir OM, Ozkurt B, Karakus F. General self-efficacy beliefs, life satisfaction and burnout of university students. *Procedia-Social Behavioral Sciences*. 2012;47:968–73.
12. Haugland T, Wahl AK, Hofoss D, DeVon HA. Association between general self-efficacy, social support, cancer-related stress and physical health-related quality of life: a path model study in patients with neuroendocrine tumors. *Health Qual Life Outcomes*. 2016;14(1):11.
13. Blackburn L, Owens GP. The effect of self efficacy and meaning in life on posttraumatic stress disorder and depression severity among veterans. *Journal of clinical psychology*. 2015;71(3):219–28.

14. Feldstain A, Lebel S, Chasen M. An interdisciplinary palliative rehabilitation intervention bolstering general self-efficacy to attenuate symptoms of depression in patients living with advanced cancer. *Support Care Cancer*. 2016;24(1):109–17.
15. Jerusalem M, Schwarzer R. Generalized self-efficacy scale. *Measures in health psychology: A user's portfolio Causal and control beliefs*. 1995:35–7.
16. Chen G, Gully SM, Eden D. Validation of a new general self-efficacy scale. *Organizational research methods*. 2001;4(1):62–83.
17. Romppel M, Herrmann-Lingen C, Wachter R, Edelmann F, Dungen H-D, Pieske B, et al. A short form of the General Self-Efficacy Scale (GSE-6): Development, psychometric properties and validity in an intercultural non-clinical sample and a sample of patients at risk for heart failure. *GMS Psycho-Social-Medicine*. 2013;10.
18. Bosscher RJ, Smit JH, Kempen GI. Algemene competentieverwachtingen bij ouderen: Een onderzoek naar de psychometrische kenmerken van de Algemene Competentieschaal (ALCOS). *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden*. 1997.
19. Löve J, Moore CD, Hensing G. Validation of the Swedish translation of the general self-efficacy scale. *Qual Life Res*. 2012;21(7):1249–53.
20. Sarstedt M, Wilczynski P. More for less? A comparison of single-item and multi-item measures. *Die Betriebswirtschaft*. 2009;69(2):211.
21. DeSalvo KB, Fisher WP, Tran K, Bloser N, Merrill W, Peabody J. Assessing measurement properties of two single-item general health measures. *Qual Life Res*. 2006;15(2):191–201.
22. Jenkinson C, Peto V, Coulter A. Measuring change over time: a comparison of results from a global single item of health status and the multi-dimensional SF-36 health status survey questionnaire in patients presenting with menorrhagia. *Qual Life Res*. 1994;3(5):317–21.
23. Turner-Bowker DM, Saris-Baglama RN, DeRosa MA. Single-item electronic administration of the SF-36v2 Health Survey. *Qual Life Res*. 2013;22(3):485–90.
24. Cheung F, Lucas RE. Assessing the validity of single-item life satisfaction measures: Results from three large samples. *Quality of Life research*. 2014;23(10):2809–18.
25. Yohannes AM, Dodd M, Morris J, Webb K. Reliability and validity of a single item measure of quality of life scale for adult patients with cystic fibrosis. *Health Qual Life Outcomes*. 2011;9(1):1–8.
26. Postmes T, Haslam SA, Jans L. A single-item measure of social identification: Reliability, validity, and utility. *British journal of social psychology*. 2013;52(4):597–617.
27. Milton K, Bull F, Bauman A. Reliability and validity testing of a single-item physical activity measure. *Br J Sports Med*. 2011;45(3):203–8.
28. Hart DL, Werneke MW, George SZ, Deutscher D. Single-item screens identified patients with elevated levels of depressive and somatization symptoms in outpatient physical therapy. *Qual Life Res*. 2012;21(2):257–68.

29. West CP, Dyrbye LN, Satele DV, Sloan JA, Shanafelt TD. Concurrent validity of single-item measures of emotional exhaustion and depersonalization in burnout assessment. *J Gen Intern Med.* 2012;27(11):1445–52.
30. Rossiter JR. The C-OAR-SE procedure for scale development in marketing. *International journal of research in marketing.* 2002;19(4):305–35.
31. Loo R. A caveat on using single-item versus multiple-item scales. *Journal of managerial psychology.* 2002.
32. Sarstedt M, Diamantopoulos A, Salzberger T, Baumgartner P. Selecting single items to measure doubly concrete constructs: A cautionary tale. *J Bus Res.* 2016;69(8):3159–67.
33. Wanous JP, Reichers AE. Estimating the reliability of a single-item measure. *Psychol Rep.* 1996;78(2):631–4.
34. Weiss DJ. Multivariate procedures. *Handbook of industrial and organizational psychology.* 1983:327–62.
35. Arvey RD, Landon TE, Nutting SM, Maxwell SE. Development of physical ability tests for police officers: a construct validation approach. *J Appl Psychol.* 1992;77(6):996.
36. Wanous JP, Hudy MJ. Single-item reliability: A replication and extension. *Organizational Research Methods.* 2001;4(4):361–75.
37. Hoepfner BB, Kelly JF, Urbanoski KA, Slaymaker V. Comparative utility of a single-item versus multiple-item measure of self-efficacy in predicting relapse among young adults. *Journal of substance abuse treatment.* 2011;41(3):305–12.
38. Andersson LM, Moore CD, Hensing G, Krantz G, Staland-Nyman C. General self-efficacy and its relationship to self-reported mental illness and barriers to care: A general population study. *Commun Ment Health J.* 2014;50(6):721–8.
39. Pedhazur EJ. *Multiple regression in behavioral research: explanation and prediction.* [Belmont. Ca.]: Wadsworth; 1997.
40. Elo A-L, Leppänen A, Jahkola A. Validity of a single-item measure of stress symptoms. *Scandinavian journal of work, environment & health.* 2003:444–51.
41. Pettegrew LS, Wolf GE. Validating measures of teacher stress. *Am Educ Res J.* 1982;19(3):373–96.
42. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *Journal of health and social behavior.* 1983:385–96.
43. Diener E, Wirtz D, Biswas-Diener R, Tov W, Kim-Prieto C, Choi D-w, et al. *New measures of well-being. Assessing well-being: Springer; 2009.* pp. 247–66.
44. Nunnally JC. *Psychometric Theory: 2d Ed: McGraw-Hill; 1978.*
45. Harman HH. *Modern factor analysis: University of Chicago press; 1976.*
46. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *biometrics.* 1977:159–74.

47. Littman AJ, White E, Satia JA, Bowen DJ, Kristal AR. Reliability and validity of 2 single-item measures of psychosocial stress. *Epidemiology*. 2006:398–403.
48. Rohland BM, Kruse GR, Rohrer JE. Validation of a single-item measure of burnout against the Maslach Burnout Inventory among physicians. *Stress Health: Journal of the International Society for the Investigation of Stress*. 2004;20(2):75–9.
49. Konrath S, Meier BP, Bushman BJ. Development and validation of the single item narcissism scale (SINS). *PLOS one*. 2014;9(8):e103469.