



Developing an Instrument for Assessing Self-Efficacy in Data Mining and Analysis

Yu-Min Wang¹, Chei-Chang Chiou^{2*}, Wen-Chang Wang¹ and Chun-Jung Chen^{2,3}

¹ Department of Information Management, National Chi Nan University, Puli, Taiwan, ² Department of Accounting, National Changhua University of Education, Changhua City, Taiwan, ³ Department of Finance, National Changhua University of Education, Changhua City, Taiwan

OPEN ACCESS

Edited by:

Mu-Yen Chen,
National Taichung University
of Science and Technology, Taiwan

Reviewed by:

Chingmu Chen,
Chung Chou University of Science
and Technology, Taiwan
Wen-Tan Chang,
Guangdong University of Finance
and Economics, China

*Correspondence:

Chei-Chang Chiou
ccchiou@cc.ncue.edu.tw

Specialty section:

This article was submitted to
Human-Media Interaction,
a section of the journal
Frontiers in Psychology

Received: 06 October 2020

Accepted: 23 December 2020

Published: 15 January 2021

Citation:

Wang Y-M, Chiou C-C,
Wang W-C and Chen C-J (2021)
Developing an Instrument
for Assessing Self-Efficacy in Data
Mining and Analysis.
Front. Psychol. 11:614460.
doi: 10.3389/fpsyg.2020.614460

With the continuous progress and penetration of automated data collection technology, enterprises and organizations are facing the problem of information overload. The demand for expertise in data mining and analysis is increasing. Self-efficacy is a pivotal construct that is significantly related to willingness and ability to perform a particular task. Thus, the objective of this study is to develop an instrument for assessing self-efficacy in data mining and analysis. An initial measurement list was developed based on the skills and abilities about executing data mining and analysis, and expert recommendations. A useful sample of 103 university students completed the online survey questionnaire. A 19-item four-factor model was extracted by exploratory factor analysis. Using the partial least squares-structural equation modeling technique (PLS-SEM), the model was cross-examined. The instrument showed satisfactory reliability and validity. The proposed instrument will be of value to researchers and practitioners in evaluating an individual's abilities and readiness in executing data mining and analysis.

Keywords: self-efficacy, data mining, measurement instrument, big data, artificial intelligence

INTRODUCTION

With the penetration and advent of data storage technologies and automatic data collection techniques, the big data age is coming. Although these technologies bring rich and diverse digital data to organizations, they can also cause serious information overload. Organizations of all sizes are under pressure to extract large amounts of data and process it into useful information and knowledge. Therefore, organizations increasingly need professionals to develop and deploy data mining technologies for competitive advantage (Nemati and Barko, 2003).

Data mining is a multi-disciplinary field (Chung and Gray, 1999; Feelders et al., 2000). Successful and effective data mining requires a collaborative effort in a number of areas, including statistics, artificial intelligence, database management, data visualization, subject area expertise, data analysis expertise, and data mining algorithms (Chung and Gray, 1999; Feelders et al., 2000; Nemati and Barko, 2003). However, at present instruments to properly and accurately measure individual abilities in data mining and analysis remain lacking. This study addresses this gap in research and practice.

Self-efficacy is an important construct in social science and information management (Compeau and Higgins, 1995). It has critical influences on task success and performance (Torkzadeh and Van Dyke, 2001). The purpose of this paper is to empirically develop an instrument for assessing an individual's self-efficacy in data mining and analysis. Self-efficacy in data mining and analysis represents an individual's judgment of their capabilities and skills to use data mining techniques

for analysis and discovery in a given domain (Bandura, 1997; Wilson et al., 2007; Wang Y. Y. et al., 2019).

The remainder of this paper is organized as follows. Section “Background and Literature Review” reviews the related literature. Section “Research Methods” describes the research method and section “Results” presents the results of data analysis. Section “Application Analysis” describes the application analysis. Finally, the conclusion, implications, and research limitations are discussed in section “Conclusion and Implications.”

BACKGROUND AND LITERATURE REVIEW

Data Mining

In the past, corporate decisions were often made subjectively by decision makers, leading to errors. With the rapid development of science and technology, companies have gradually begun to use objective data to make decisions. In particular, the accumulation of data at large companies has increased rapidly and technology-assisted data analysis (e.g., data mining analysis) has gradually become an important tool for corporate decision-making. Data mining technology is an indispensable technology in the era of big data analysis. Hand et al. (2001) define data mining as the analysis of data sets (usually a large number of data sets) to identify unexpected relationships and summarize the data in novel patterns, and then provide useful information. Jain and Srivastava (2013) observed that data mining algorithms are divided into two functional types, predictive and descriptive, and eight application types, classification, estimation, forecasting, correlation analysis, sequence, time series, description, and visualization (Dunham, 2003).

Data mining technology is not only used in corporate decision-making, but is widely used in various industries. For example, in business management, Alola and Atsa'am (2020) applied data mining technology to measure the psychological capital of employees in the organization, and noted that when measuring the psychological capital of employees in recruitment interviews and promotion evaluations, data mining classification models can be useful as tools for human resource management. Zhen and Yao (2019) analyzed the lean production and technological innovation of the manufacturing industry based on the support vector machine algorithm and data mining technology. Data mining can discover novel, effective, potential, and finally understandable data patterns from a deeper level, and encode the data to predict the development trend of the enterprise. Machine learning support vector machine methods are used to analyze and model the collected data. Ding et al. (2019) indicated that the current cloud computing technology is developing rapidly, gradually integrating into IoT data mining technology and forming a new model. On this basis, the construction of an IoT data mining model based on cloud computing technology was studied. Another example is application in medicine. Zhao et al. (2020) used data mining to study the risk factors that can predict IHD during pheochromocytoma surgery, and observed that data mining techniques are increasingly being used in clinical and

medical decision-making to provide continuous support for the diagnosis, treatment, and prevention of disease. Massi et al., 2020 noted that the healthcare industry is an interesting target for fraudsters. The availability of large amounts of data makes it possible to solve this problem through the use of data mining techniques, thereby making the review process more effective. The purpose of this research was to use the hospital discharge chart in the management database to develop a new type of data mining model specifically for fraud detection between hospitals. Qian and Liu (2020) proposed data mining technology that first determined the classification of index parameters. They then used this data mining technology to establish a sports training analysis mechanism to complete the construction of the index analysis model.

Data mining technology has also been widely used in the education field and is now being used more and more widely in teaching activities (Calders and Pechenizkiy, 2012; Maldonado and Seehusen, 2018). Data mining technology can be used to analyze educational data and explore educational research issues (Campagni et al., 2015). It can be used to improve educational practices and learning materials (Romero and Ventura, 2013), and to predict student performance, group students, plan courses, discover bad student behavior, model students, and classify courses based on student preferences (Romero and Ventura, 2010; Goyal and Vohra, 2012; Maldonado and Seehusen, 2018). The main focus of educational data exploration is to help solve problems related to the learning process of students, as well as to help schools conduct adaptive curriculum planning and students conduct adaptive learning (Calders and Pechenizkiy, 2012; Maldonado and Seehusen, 2018).

Self-Efficacy

According to the theory of social cognition, perceptual self-efficacy is the key mechanism for exercising human agency within a causal structure involving the ternary causality of people, environment, and behavior (Bandura, 1986). Self-efficacy belief is an individual's belief in their ability to achieve expected results, overcome obstacles, resist adversity, self-regulate in the face of urgent circumstances, discern many competing choices and negotiate important life changes (Basili et al., 2020). Self-efficacy means an individual's confidence in their own problem solving and task completion ability (Sun and Chen, 2016; Ghazi et al., 2018). İncirkus and Nahcivan (2020) observe that self-efficacy refers to people's belief in their ability to implement an action plan, deal with challenges, and make the judgments that make a particular action successful. Mamaril et al. (2016) and Liu et al. (2020) indicated that self-efficacy is an individual's conjecture and judgment of whether they have the ability to complete a certain behavior, which can reflect the individual's belief in taking appropriate action to address environmental challenges. It contains expectations of results and expectations of effectiveness (Bandura, 1997). The former is the belief that certain actions will ensure certain results, while the latter is the belief that one can complete these actions and obtain results (Sun and Chen, 2016). Bandura and Cervone (1986) and Sullivan et al. (2006) argue that since people who are confident in a task will expect success, concentrate on thinking about how to succeed, persist

in facing difficulties, and avoid low self-efficiency tasks, self-efficacy beliefs are highly positively correlated with work and academic performance. Thus, when self-efficacy beliefs can be improved, performance improvement will occur (Dunlap, 2005; McLaughlin et al., 2008; Kuiper et al., 2010).

Many studies have explored the self-efficacy of students in academic fields and the self-efficacy of employees in practical fields. Research on employees largely explores personal self-efficacy in specific work situations (Bandura, 1986; Judge et al., 1998; Bandura and Locke, 2003). Bandura and Locke (2003) argue that self-efficacy is positively related to individual behavioral processes and results, such as perseverance in adversity, efforts to achieve high achievements, and ultimately high performance in various fields. Chae and Park (2020) indicate that expectations of personal self-efficacy determine how much task-related effort will be expended. Therefore, beliefs related to self-efficacy are the most powerful predictors of individual behavior and persistence in adversity (Bandura, 1986). Bandura (1986) and Bandura and Locke (2003) contend that when individuals have a high sense of self-efficacy, the resources they are willing to invest in tasks will increase, leading to better results. Other studies have explored the relationship between self-efficacy and entrepreneurial enthusiasm and entrepreneurial behavior (Shane et al., 2003; Murnieks et al., 2014). Shane et al. (2003) observed that self-efficacy and enthusiasm are two important factors in maintaining entrepreneurial efforts. Sun (2020) showed that self-efficacy mediates the relationship between entrepreneurial enthusiasm and entrepreneurial behavior. Researchers have also explored general self-efficacy, individuals' perception of their ability to perform in various situations, in the general workplace (Smith, 1989; Scholz et al., 2002; Chen et al., 2004). Results show that general self-efficacy is positively correlated with job performance (Beattie et al., 2016) and knowledge sharing (Srivastava et al., 2006). Chae and Park (2020) explored the relationship between an employee's general self-efficacy and task performance and knowledge-sharing. The results showed that the high general self-efficacy of key employees has a positive impact on task performance but has a negative impact on knowledge sharing.

Most studies of the self-efficacy of students agree that self-efficacy has a positive impact on learners' academic achievement and personal success (Vancouver et al., 2001; Honicke and Broadbent, 2016; Basili et al., 2020). Fernandez-Rio et al. (2017) indicated that academic self-efficacy beliefs affect the perception of ability in the self-regulation process that is beneficial to learning. Cooper (2015) demonstrated that self-efficacy can help students at risk overcome their at-risk conditions and positively impact their academic performance. Schunk (1994) and Carroll et al. (2009) demonstrated that students with higher self-efficacy beliefs can better manage their own learning and are more likely to do better academically. Klassen and Usher (2010) and Talsmaa et al. (2018) all observed that people with high self-efficacy set more difficult goals, put in more effort, persist in challenges for a longer time, and show resilience in adversity, which can improve academic achievement (Bandura, 1997). Klassen and Usher (2010) contended that self-efficacy has a key and powerful influence on academic achievement. Pajares and

Kranzler (1995) found that self-efficacy can effectively predict academic achievement. Multon et al. (1991), Richardson et al. (2012), and Honicke and Broadbent (2016) conducted a meta-analysis of self-efficacy, finding that self-efficacy is strongly correlated with academic achievement.

Many researchers have found that self-efficacy plays an important role in the process and results of individual behavior. However, since self-efficacy is a kind of behavioral cognition, a psychological scale to measure personal self-efficacy is needed. A number of different self-efficacy scales have been developed for various fields, such as self-efficacy in the medical field (Lorig et al., 1989; Incirkus and Nahcivan, 2020), general self-efficacy scales in the workplace (Chen et al., 2004), self-efficacy scale for engineering education (Mamaril et al., 2016), multi-dimensional self-efficacy scale for adolescents (Bandura, 1990), teacher research self-efficacy scale (Wester et al., 2019), teacher self-efficacy scale for student-oriented teaching (Kilday et al., 2016), college student self-efficacy scale (Khasawneh et al., 2009), and a mathematical self-efficacy energy scale (Betz and Hackett, 1983). Based on the development of education in the high-tech era, the popularization of technology-assisted teaching has led many researchers to study the role of self-efficacy when the Internet or technology is applied to teaching, and develop numerous Internet and technology-related self-efficacy scales, such as the Internet self-efficacy scale (Hsu and Chiu, 2004; Kao et al., 2011), the computer ethical self-efficacy scale (Kuo and Hsu, 2001), and the Internet ethical self-efficacy scale (Williamson et al., 2011). With the development of Internet and high technology, though big data analysis and artificial intelligence have gradually become common across various industries, data mining and artificial intelligence self-efficacy scales remain lacking. Therefore, the main purpose of this research is to develop a self-efficacy scale for data mining and analysis.

RESEARCH METHODS

Based on the prior measures and definitions of self-efficacy, this study conceptually defines "self-efficacy in data mining and analysis" as an individual's judgment of his or her ability to successfully execute data mining and analysis. The initial instrument, which consisted of 28 items, was developed based on the review of the literature on skills and abilities for executing data mining and analysis (Fayyad et al., 1996; Chung and Gray, 1999; Mitchell, 1999; Chapman et al., 2000; Feelders et al., 2000; Liao, 2008; Han et al., 2011; Tufféry, 2011; McCormick et al., 2013; Singhal and Jena, 2013; Abbott, 2014; Jian and Hsu, 2014; Xue, 2014; Marvin, 2016; Salcedo and McCormick, 2017; Struhl, 2017; Chang and Kung, 2019; Liao and Wen, 2019; Wang, 2019; Wang Y. S. et al., 2019) and expert experience. Three global items for measuring perceived overall self-efficacy were added to serve as a criterion. All items were measured using a seven-point Likert-type scale with anchors of "(1) strongly disagree, (2) disagree, (3) slightly disagree, (4) neutral, (5) slightly agree, (6) agree, and (7) strongly agree." **Table 1** shows all 31 items.

The survey methodology was adopted and empirical data for this study were collected using an Internet questionnaire survey in Taiwan. University students with data mining knowledge or experiences were qualified to participate in the survey, and were asked to fill in the questionnaire based on their experiences and self-perceptions. Every respondent in the survey was given an NT 100-dollar coupon as an incentive. The survey duration was 2 months: from April to May in 2020. This study obtained 103 useful responses. There were more females than males in the sample (51.5 and 48.5%). The proportion of college students in the sample is higher than that of graduate students (85.4 and 14.6%). The respondents had an average age of 21.6 years. On average, they took 4.03 courses and 12.57 credits in data mining.

Data from 103 university students was tested against the proposed 28-item instrument using a two-step assessment approach. In the first stage, the exploratory factor analysis (EFA) and the criterion-related analysis was used to purify the measure, remove noise items, and acquire factor structure. In the second stage, the partial least squares-structural equation modeling (PLS-SEM) was used to assess the hierarchical component

model (HCM) based on the EFA result. Internal consistency (reliability), convergent validity, and discriminant validity were checked for the model.

RESULTS

EFA Results

Exploratory factor analysis was used to purify the measurement instrument. Before conducting the EFA, three tests were performed to check the adequacy of the survey data for EFA. First, Cronbach's α coefficient was computed to ensure the internal inconsistency of the measurement items (Churchill, 1979). The results showed that the 28-item instrument had an α coefficient of 0.97, indicating that the measure was unidimensional. Second, Bartlett's test of sphericity was used to assess the overall significance of the correlations among the measurement items (Hair et al., 1998). The results demonstrated a satisfactory suitability of the data for factor analysis ($\chi^2 = 3387.31, p < 0.001$). Third, the Kaiser-Meyer-Olkin statistic was computed for

TABLE 1 | The measurement items.

Items
Q1. I clearly understand the main applications of data mining, e.g., classification, estimation, forecasting, association, and cluster analysis
Q2. I clearly understand the procedure and main steps of data mining
Q3. I am familiar with standards for data mining and modeling
Q4. I have the ability to conduct data mining in a professional field (such as consumer behavior analysis, sales data) to discover useful information or knowledge
Q5. I have the ability to understand and interpret the outputs derived from data mining
Q6. I am familiar with at least one major programming language for data mining, such as R, Python, or Java
Q7. I think I have the programming skills required for data mining
Q8. I know how to use information retrieval methods to find useful information from a large amount of data
Q9. When I search for information, I can use keyword search accurately
Q10. I have the relevant ability of database system
Q11. I have the ability to clean, select, transform, and synthesize data
Q12. I have the ability to execute online analytical processing (OLAP)
Q13. I have the ability to use SQL (Structured Query Language)
Q14. I have the ability to build a data warehouse
Q15. I am familiar with at least one data exploration tool, such as WEKA, RapidMiner, IBM SPSS modeler, and Statistica
Q16. I have the ability to carry out pre-processing of data mining
Q17. I have the ability to execute classification analysis
Q18. I have the ability to execute cluster analysis
Q19. I have the ability to execute the feature selection
Q20. I have the ability to visualize the data
Q21. I have the relevant statistical skills required for data mining
Q22. I have the ability to execute the decision tree analysis
Q23. I have the ability to execute discriminant analysis
Q24. I have the ability to execute association analysis
Q25. I have the ability to execute sequential pattern analysis or causal analysis
Q26. I have the ability to execute time-series analysis
Q27. I have the ability to execute artificial neural networks (ANN) analysis
Q28. I have the ability to use at least one data mining technique for data analysis or discovery
G1. Overall, I think I have professional ability in data mining*
G2. Overall, I think my data mining skills capabilities meet the needs of practitioners*
G3. Overall, I think I have good and complete data mining knowledge*

*Criterion item.

checking sampling adequacy. The statistical score was 0.91 and greater than 0.50, indicating high shared-variance and relatively low uniqueness (Hair et al., 1998). These test results suggested that EFA was worth pursuing.

The principle-components analysis was used as an extraction technique and varimax method was used to rotate the factor matrix. Referring to Kaiser (1960), Sethi and King (1991), and Hair et al. (1998), four rules were applied in EFA: (1) a factor with an eigenvalue greater than 1.00 was retained; (2) an item with all factor loadings below 0.55 was removed; (3) an item with two or more factor loadings (rounding numbers) above 0.55 was dropped; and (4) an item with two or more correlation coefficients with other items greater than 0.85 was removed. **Table 2** shows the EFA results. The results show that 77.54 percent of variance is explained by four factors and 19 items are left in the instrument. These factors are labeled “Data mining techniques,” “Programming and database,” “Basic knowledge and procedure of data mining,” and “Data retrieval and statistical presentation.” The respective Cronbach’s α coefficients are 0.94, 0.91, 0.87, and 0.84. All the coefficients exceed the acceptable standard of 0.70.

The criterion-related validity was assessed by the correlation between the sum of scores on all 19 items in the instrument and the validity criterion (sum of three criterion items). The correlation was 0.78, significant at 0.001, representing satisfactory criterion-related validity.

TABLE 2 | EFA results.

Items	Factor 1	Factor 2	Factor 3	Factor 4
Q1			0.56	
Q2			0.83	
Q3			0.61	
Q4			0.78	
Q6		0.86		
Q7		0.85		
Q8				0.59
Q9				0.68
Q10		0.74		
Q12	0.67			
Q13		0.82		
Q15	0.69			
Q16	0.70			
Q20				0.78
Q21				0.69
Q22	0.84			
Q24	0.78			
Q26	0.83			
Q27	0.82			
Eigenvalue	5.40	3.57	2.97	2.79
Variance explained	28.44%	18.80%	15.65%	14.66%
Cumulative variance explained	28.44%	47.23%	62.88%	77.54%
α coefficient	0.94	0.91	0.87	0.84

Factor 1, data mining techniques; Factor 2, programming and database; Factor 3, basic knowledge and procedure of data mining; Factor 4, data retrieval and statistical presentation.

The multitrait-multimethod (MTMM) approach was used for evaluating the convergent and discriminant validity of the instrument. **Table 3** shows the correlation coefficients between items. Convergent validity is acceptable when the correlation coefficients of the same factor are significantly different from zero and large enough for further investigation (Doll and Torkzadeh, 1988). The smallest within-factor correlation coefficients are: Data mining techniques = 0.50, Programming and database = 0.60, Basic knowledge and procedure of data mining = 0.43, Data retrieval and statistical presentation = 0.54. All coefficients are significantly different from 0 ($p < 0.01$) and large enough, demonstrating the convergent validity of the measures.

The discriminant validity for each item was assessed by counting the number of times correlated more closely with items of other factors than items of its own theoretical factor (Wu and Wang, 2006). Such counts should be less than 50 percent of the comparisons. As shown in **Table 3**, there were 45 violations out of 264 comparisons, representing acceptable discriminant validity.

PLS-SEM Results

According to the two-stage HCM method suggested by Hair et al. (2017) and the rationale of EFA results, a reflective-formative measurement model was built. The repeated indicators approach was adopted for analyzing the higher-order measurement model (**Figure 1**). This model hypothesized that the four reflective first-order factors formed one second-order factor. Self-efficacy in data mining and analysis is multi-faceted and the four factors of Data mining techniques, Programming and database, Basic knowledge and procedure of data mining, and Data retrieval and statistical presentation are components of self-efficacy in data mining and analysis. Therefore, the formative type (components second-order construct) is reasonable. The 19 items are reflective indicators of these four first-order factors.

There are two parts in the measurement evaluation. First, internal consistency (rho_A), convergent validity (AVE, outer loading) and discriminant validity (HTMT) were checked for the reflective part of the model, the measurement of the four factors. Second, the convergent validity, collinearity, and significance of the path coefficients were evaluated for the formative part of the model, the four factors forming the higher-order component, self-efficacy.

Table 4 shows the PLS results and relative standards of the reflective part of the measurement model. All rho_A values for the factors exceeded the recommended value of 0.7, supporting internal consistency. The average variance extracted (AVE) values for the four factors are 0.74, 0.80, 0.72, and 0.68. All AVE values are greater than 0.5, justifying the convergent validity. As shown in **Table 4**, the outer loadings of all items are significant and above 0.7, confirming the convergent validity of this measure. Finally, the heterotrait-monotrait (HTMT) was used to assess discriminant validity. As shown in **Table 4**, all HTMT values are below the threshold value of 0.9, confirming discriminant validity (Hair et al., 2017). In sum, the reflective part of the measurement model demonstrates adequate reliability and validity.

Table 5 shows the PLS results and relative standards of the formative part of the measurement model. Three

TABLE 3 | Correlation coefficient between items.

	Q1	Q2	Q3	Q4	Q6	Q7	Q10	Q13	Q8	Q9	Q20	Q21	Q12	Q15	Q16	Q22	Q24	Q26
Q2	0.66																	
Q3	0.63	0.72																
Q4	0.50	0.67	0.57															
Q6	0.41	0.47	0.49	0.46														
Q7	0.30	0.35	0.43	0.41	0.80													
Q10	0.35	0.51	0.47	0.50	0.71	0.60												
Q13	0.43	0.42	0.57	0.35	0.80	0.70	0.76											
Q8	0.60	0.50	0.53	0.48	0.58	0.52	0.50	0.46										
Q9	0.39	0.44	0.34	0.50	0.48	0.36	0.59	0.38	0.69									
Q20	0.60	0.39	0.43	0.46	0.42	0.35	0.47	0.46	0.64	0.53								
Q21	0.48	0.24	0.37	0.21	0.27	0.20	0.32	0.31	0.51	0.43	0.63							
Q12	0.47	0.65	0.65	0.49	0.53	0.46	0.50	0.52	0.56	0.38	0.36	0.35						
Q15	0.49	0.60	0.69	0.45	0.54	0.44	0.39	0.45	0.50	0.26	0.28	0.35	0.77					
Q16	0.58	0.49	0.55	0.32	0.56	0.47	0.36	0.51	0.61	0.36	0.57	0.56	0.64	0.72				
Q22	0.48	0.53	0.63	0.29	0.38	0.31	0.35	0.46	0.56	0.36	0.41	0.56	0.68	0.60	0.65			
Q24	0.51	0.39	0.58	0.34	0.43	0.37	0.43	0.49	0.58	0.31	0.51	0.68	0.62	0.54	0.61	0.78		
Q26	0.49	0.55	0.63	0.37	0.39	0.36	0.39	0.42	0.65	0.42	0.52	0.61	0.71	0.62	0.68	0.84	0.78	
Q27	0.48	0.47	0.63	0.32	0.43	0.45	0.39	0.46	0.65	0.37	0.45	0.53	0.71	0.66	0.67	0.75	0.72	0.84

The correlation coefficients between items of the same factor will be shown in bold.

analyses were executed. First, convergent validity was evaluated. Convergent validity is the extent to which a measure correlates positively with other measures of the same construct using different indicators (Hair et al., 2017). Therefore, this study used redundancy analysis for assessing convergent validity. The redundancy analysis method is useful for analyzing a directional relationship between two sets of multivariate data (Lambert et al., 1988). We created one exogenous self-efficacy construct that are measured by 19 items and one endogenous self-efficacy construct that are first measured by three global items. Then we examine the path coefficient through which the exogenous construct influences the endogenous construct. The path coefficient is 0.82, above threshold value of 0.8, confirming convergent validity (Wong, 2019). Second, the collinearity issue was assessed. Collinearity should be evaluated in a model with multiple variables as a possible predictor-predictor redundancy phenomenon (Kock and Lynn, 2012). When two or more predictor variables in a multiple regression model are highly correlated, multicollinearity occurs, which will cause the variance inflation and increase the type I error, making some coefficients appear significant when they are not (Lombardi et al., 2017). When the variance inflation factor (VIF) is higher than the threshold value of 5.0, a potential collinearity problem can exist. As shown in Table 5, all VIF values are below 5.0, indicating no collinearity problem. Third, the significance of the path coefficients from the four factors to the high-order self-efficacy construct was examined. The path coefficients are 0.51, 0.21, 0.22, and 0.22. All path coefficients are significant.

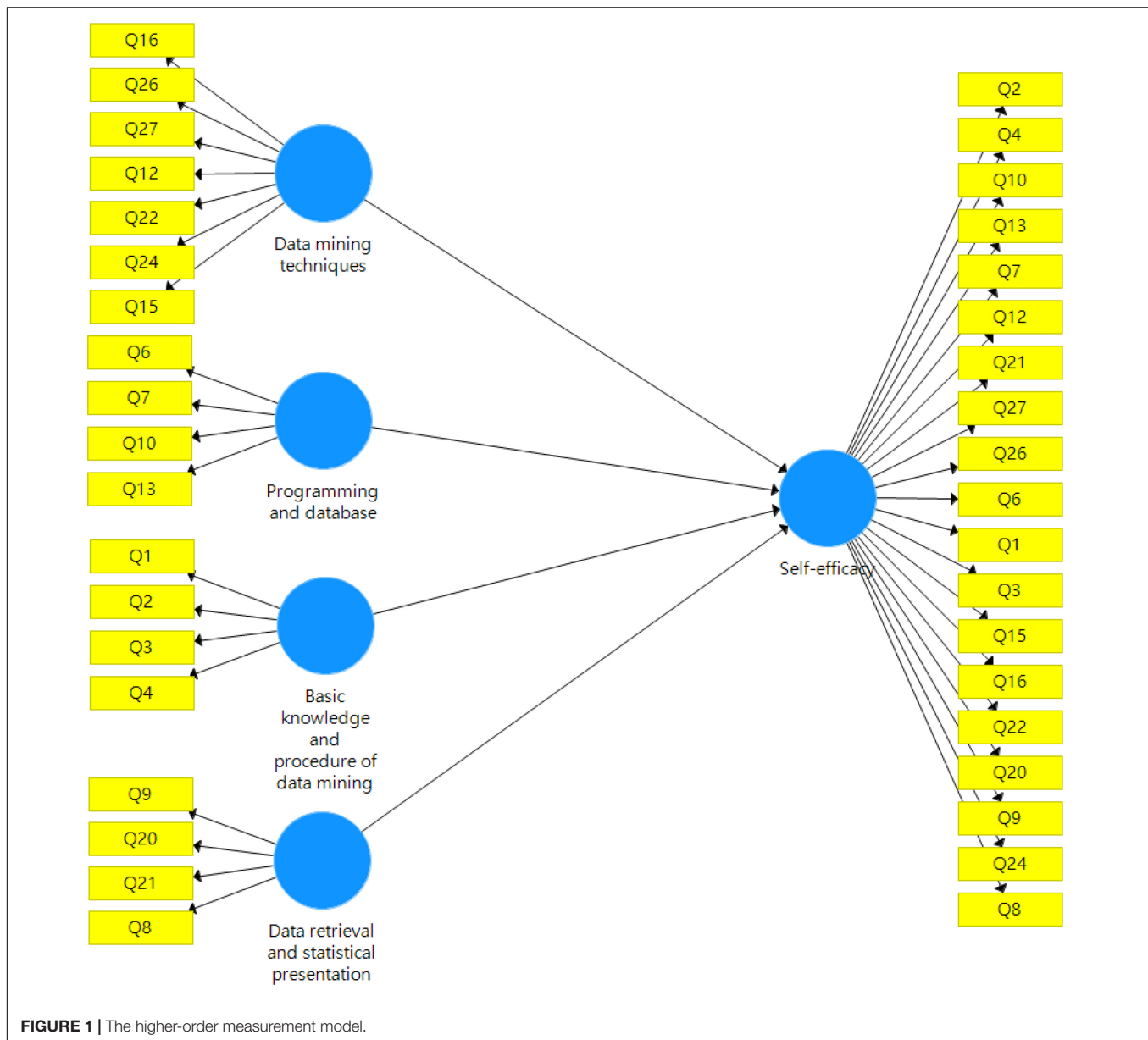
All indices and statistics in Tables 4, 5 have reached relevant assessment standards. The measurement model has satisfactory reliability and validity.

APPLICATION ANALYSIS

Through rigorous empirical analysis, this study has developed a reliable and valid instrument for measuring an individual's self-efficacy in data mining and analysis. This section presents the application analysis of the instrument from three perspectives. First, the correlation between education and self-efficacy in data mining and analysis is assessed. Second, measurement invariance from the gender perspective is evaluated. Finally, the norms of this instrument are developed.

The Correlation Between Education and Self-Efficacy in Data Mining and Analysis

This study found that there is a significant positive correlation between total self-efficacy level and credits taken by university students in data mining and analysis related courses. The correlation coefficient is 0.41, significant at 0.001. This relationship is significant and positive. The regression analysis is also tested. The independent variable is credits taken by university students in data mining and analysis related courses, and the dependent variable is total self-efficacy level. The results are $\beta = 0.41$, $T = 4.57$, and significance level < 0.001 . These



findings support the effectiveness of university education in the data mining and analysis domain.

Measurement Invariance

Measure invariance is also called measurement equivalence (Wong, 2019). It refers to the degree of a measure retains the measurement properties across observations and contexts (Mangos and Johnston, 2008). Measure invariance should be checked prior to executing multi-group analysis in the future study. This study assessed the measurement invariance from the gender perspective. Referring to Hair et al. (2017) and Wong (2019), three steps were applied: (1) Configural invariance is developed using the same path model, data treatment, and analysis algorithm. (2) Compositional invariance is evaluated by

comparing path coefficients. (3) Composite means and variances are assessed if compositional invariance exists.

For analysis, we split the sample into two groups based on gender. The male group has 53 responses and the female group has 50 responses. First, the same two PLS path models for these two groups were developed. The analysis parameters and algorithm were set the same for configural invariance. Then path coefficients were estimated and compared for examining compositional invariance. The modified two independent-sample *t*-test of Keil et al. (2000) was used to compare whether the path coefficients between male and female samples are significantly different. The results are shown in **Table 6**. One relationship (Data mining techniques → Self-efficacy) was found to have different path coefficients. This implies that males and females have different perceptions about the influence of data mining

techniques on self-efficacy. Compositional variance in measuring data mining techniques may exist across gender.

Norms

The composite scores were computed by summing the 19- item scores. However, a raw composite score on a measurement instrument may be not sufficiently informative (Churchill, 1979). A better way of assessing an individual's self-efficacy is to compare the individual score with norms – the total distribution of the scores achieved by other people. The tentative norm of the self-efficacy instrument was presented in Table 7. These statistics offer a frame of reference and comparison for potential instrument users. The instrument users can use the norms as the benchmark for evaluating relative abilities and scores against others.

CONCLUSION AND IMPLICATIONS

Most data-mining studies focus on development of innovative algorithms, comparisons of different algorithms, and application analysis. However, relatively few studies evaluate individuals' capabilities and talents in data mining. This study is a pioneering effort to develop and validate an instrument for assessing an individual's self-efficacy in data mining and analysis. The measure items are developed based on relevant data-mining literature and practical experiences. The instrument is purified and validated empirically. Finally, nineteen items are exclusively used to assess an individual's self-efficacy in data mining and analysis. The results reveal that self-efficacy in data mining and analysis is a higher-order construct composed of four dimensions: Data mining techniques, Programming and database, Basic knowledge and procedure of data mining, and Data retrieval and statistical presentation. The results enhance our understanding of the nature and dimensionality of self-efficacy in data mining and analysis. The research findings have several implications for practitioners and researchers.

First, the instrument developed in this study can be used as an assessment and diagnosis tool. Students and practitioners can use this instrument to assess their abilities in data mining and analysis

TABLE 4 | PLS results: The reflective part.

Tests	Factor 1	Factor 2	Factor 3	Factor 4
rho_A	0.94	0.92	0.88	0.86
All coefficients are above the minimum standard of 0.7				
AVE	0.74	0.80	0.72	0.68
All AVEs are above the minimum standard of 0.5				
Outer loading	0.81–0.92	0.86–0.93	0.79–0.90	0.77–0.88
All loadings are above the minimum standard of 0.7				
HTMT	0.61–0.76	0.61–0.64	0.64–0.76	0.65–0.75
All HTMT indexes are below the maximum threshold of 0.9				

Factor 1, data mining techniques; Factor 2, programming and database; Factor 3, basic knowledge and procedure of data mining; Factor 4, data retrieval and statistical presentation.

TABLE 5 | PLS results: The formative part.

Tests	Results
Convergent validity (redundancy analysis)	Path coefficient = 0.82 The path coefficient (HOC → criterion) is above the minimum standard of 0.8
Collinearity	VIF = 2.47, 1.73, 2.25, 2.16 All VIFs are below the maximum threshold of 5.0
Significance of path coefficients	Path coefficients = 0.51, 0.21, 0.22, 0.22 All path coefficients (LOC → HOC) are significant at 0.001 level

TABLE 6 | Comparisons of path coefficients by gender.

Paths	Male		Female		P-value
	β	SD	β	SD	
Data mining techniques → Self-efficacy	0.54	0.04	0.43	0.02	0.03
Programming and database → Self-efficacy	0.22	0.04	0.23	0.02	0.86
Basic knowledge and procedure of data mining → Self-efficacy	0.21	0.03	0.25	0.02	0.27
Data retrieval and statistical presentation → Self-efficacy	0.24	0.02	0.21	0.02	0.16

TABLE 7 | Percentile scores for the instrument.

Percentile	Composite score				
	Total	Factor 1	Factor 2	Factor 3	Factor 4
10	45.40	10.80	8.00	9.40	12.00
20	57.00	14.00	11.00	12.00	13.00
30	61.20	18.00	15.00	14.00	15.00
40	68.20	21.00	16.00	15.00	16.00
50	74.00	23.00	19.00	16.00	17.00
60	77.80	27.40	20.00	17.00	19.00
70	87.60	29.00	21.00	19.00	19.00
80	94.00	31.20	23.00	20.00	20.20
90	99.60	35.00	24.00	21.00	22.60

Factor 1, data mining techniques; Factor 2, programming and database; Factor 3, basic knowledge and procedure of data mining; Factor 4, data retrieval and statistical presentation.

and take action to address weaknesses. Enterprises can use this instrument to assess employee abilities. When enterprises recruit data-mining professionals, they can design exam questions using the four dimensions. Instructors in universities can refer to the items, dimensions, and relative influences of these dimensions in designing data-mining programs and allocating course credits.

Second, this study finds that “data mining techniques” have the highest influence on self-efficacy (β = 0.51) among the four factors. This implies that “data mining techniques” are the requisite capabilities that individuals need to effectively perform data mining and analysis. When individuals have mastery of

data mining techniques, they have the knowledge and abilities to handle decision tree, association, time-series, and artificial neural network analysis, and the pre-processing of data mining. These are indispensable and fundamental capabilities.

Third, this study also finds that the other three factors have significant and similar influences (β coefficients are between 0.21 and 0.22). This finding supports the claim that data mining is a multi-disciplinary field (Chung and Gray, 1999; Feelders et al., 2000). Since executing data mining requires cross-domain knowledge and skills, individuals should possess more than basic data mining techniques. If they want to successfully execute data mining projects and obtain correct outcomes, expertise such as programming and database use, basic knowledge and procedure of data mining, and data retrieval and statistical presentation, should be possessed.

Fourth, this study finds that education and self-efficacy are positively correlated. This implies that the higher the number of credits related to data mining, the higher the self-efficacy. This not only supports the effectiveness of university education, but also encourages students who want to have the abilities in data mining and analysis to take more relevant courses.

Finally, measure variance in the “data mining techniques” dimension may exist across genders. This issue should be re-verified with more samples. If measure variance remains, researchers should address gender difference in the influence of data mining techniques on self-efficacy.

This research has several limitations. First, this research only takes students as the survey object for analysis. However, data mining and analysis are applied in practical domains. It is thus possible that people who work in practical applications of data mining technology will have different self-efficacy. In the future, people working in practical applications of data mining should be surveyed for further analysis. Second, the sample size of the research is not large and the sample does not include students of diverse backgrounds. Future research should expand coverage to students from different backgrounds and compare the differences among them in self-efficacy of data mining and analysis.

REFERENCES

- Abbott, D. (2014). *Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst*. Hoboken, NJ: John Wiley & Sons.
- Alola, U. V., and Atsa'am, D. D. (2020). Measuring employees' psychological capital using data mining approach. *J. Public Affairs* 20:e2050. doi: 10.1002/pa.2050
- Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Mahwah, NJ: Prentice Hall.
- Bandura, A. (1990). *Multidimensional Scales of Perceived Academic Efficacy*. Stanford, CA: Stanford University.
- Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York, NY: Freeman and Company.
- Bandura, A., and Cervone, D. (1986). Differential engagement of self reactive influences in cognitive motivations. *Organ. Behav. Hum. Decis.* 38, 92–113. doi: 10.1016/0749-5978(86)90028-2
- Bandura, A., and Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *J. Appl. Psychol.* 88, 87–99. doi: 10.1037/0021-9010.88.1.87
- Basili, E., Gomez, P. M., Paba, B. C., Gerbino, M., Thartori, E., Lunetti, C., et al. (2020). Multidimensional scales of perceived self-efficacy(MSPSE):

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because when collecting the survey data, we had a promise to the respondents that the response contents would not be disclosed and be given to the third parties. Requests to access the datasets should be directed to corresponding author.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the Local Legislation and Institutional Requirements. Written informed consent from the participants was not required to participate in this study in accordance with the National Legislation and the Institutional Requirements. However, consent was implied via completion of the questionnaire.

AUTHOR CONTRIBUTIONS

Y-MW contributed to the research topic, data collection, statistical analysis, developing implications, and writing. C-CC took charge in literature review, writing the manuscript, and responsible for correspondence. W-CW developed the instrument and designed the questionnaire. C-JC contributed to data collection and practical implications. All authors contributed to the article and approved the submitted version.

FUNDING

The authors would like to thank the Ministry of Science and Technology, Taiwan, for financially supporting this research (Grant No. MOST 108-2511-H-001-MY3).

measurement invariance across Italian and Colombian adolescents. *PLoS One* 15:e0227756. doi: 10.1371/journal.pone.0227756

- Beattie, S., Woodman, T., Fakehy, M., and Dempsey, C. (2016). The role of performance feedback on the self-efficacy–performance relationship. *Sport Exerc. Perform. Psychol.* 5, 1–13. doi: 10.1037/spy0000051
- Betz, N. E., and Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *J. Voc. Behav.* 23, 329–345. doi: 10.1016/0001-8791(83)90046-5
- Calders, T., and Pechenizkiy, M. (2012). Introduction to the special section on educational data mining. *ACM SIGKDD Explor. Newslett.* 13, 3–6. doi: 10.1145/2207243.2207245
- Campagni, R., Merlini, D., Sprugnoli, R., and Verri, M. C. (2015). Data mining models for student careers. *Expert Syst. Applic.* 42, 5508–5521. doi: 10.1016/j.eswa.2015.02.052
- Carroll, A., Houghton, S., Wood, R., Unsworth, K., Hattie, J., Gordon, L., et al. (2009). Self-efficacy and academic achievement in Australian high school students: the mediating effects of academic aspirations and delinquency. *J. Adolesc.* 32, 797–817. doi: 10.1016/j.adolescence.2008.10.009
- Chae, H., and Park, J. (2020). Interactive effects of employee and coworker general self-efficacy on job performance and knowledge sharing. *Soc. Behav. Pers. Intern. J.* 48, 1–11. doi: 10.2224/sbp.9527

- Chang, Y. T., and Kung, L. (2019). *Principles and Techniques of Data Mining*. Taichung: Wunan Books.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., et al. (2000). *CRISP-DM 1.0: Step-by-Step Data Mining Guide*. Chicago: SPSS Inc.
- Chen, G., Gully, S. M., and Eden, D. (2004). General self-efficacy and self-esteem: toward theoretical and empirical distinction between correlated self-evaluations. *J. Organ. Behav.* 25, 375–395. doi: 10.1002/job.251
- Chung, H. M., and Gray, P. (1999). Data mining. *J. Manag. Inform. Syst.* 16, 11–16. doi: 10.1080/07421222.1999.11518231
- Churchill, G. A. Jr. (1979). A paradigm for developing better measures of marketing constructs. *J. Mark. Res.* 16, 64–73. doi: 10.1177/002224377901600110
- Compeau, D. R., and Higgins, C. A. (1995). Computer self-efficacy: development of a measure and initial test. *MIS Q.* 19, 189–211. doi: 10.2307/249688
- Cooper, C. L. (2015). *Students at Risk: The Impacts of Self-Efficacy and Risk Factors on Academic Achievement*. Doctoral thesis, University of Texas at Arlington, Arlington, TX.
- Ding, B., Chen, W., and Huang, Y. (2019). The construction of internet data mining model based on cloud computing. *J. Intellig. Fuzzy Syst.* 37, 3275–3283. doi: 10.3233/JIFS-179129
- Doll, W. J., and Torkzadeh, G. (1988). The measurement of end-user computing satisfaction. *MIS Q.* 12, 259–274. doi: 10.2307/248851
- Dunham, M. (2003). *Data Mining-Introductory and Advanced Topics*. London: Pearson Education.
- Dunlap, J. C. (2005). Problem based learning and self-efficacy: how a capstone course prepares students for a profession. *Educ. Technol. Res. Dev.* 53, 65–85. doi: 10.1007/BF02504858
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magaz.* 17, 37–37. doi: 10.1609/aimag.v17i3.1230
- Feelders, A., Daniels, H., and Holsheimer, M. (2000). Methodological and practical aspects of data mining. *Inform. Manag.* 37, 271–281. doi: 10.1016/S0378-7206(99)00051-8
- Fernandez-Rio, J., Cecchini, J. A., Me'ndez-Gimenez, A., Mendez-Alonso, D., and Prieto, J. A. (2017). Self-regulation, cooperative learning, and academic self-efficacy: interactions to prevent school failure. *Front. Psychol.* 8:22. doi: 10.3389/fpsyg.2017.00022
- Ghazi, C., Nyland, J., Whaley, R., Rogers, T., Wera, J., and Henzman, C. (2018). Social cognitive or learning theory use to improve self-efficacy in musculoskeletal rehabilitation: a systematic review and meta-analysis. *Physiother. Theory Pract.* 34, 495–504. doi: 10.1080/09593985.2017.1422204
- Goyal, M., and Vohra, R. (2012). Applications of data mining in higher education. *Intern. J. Comput. Sci.* 9, 113–120.
- Hair, J. F., Anderson, R. T., Tatham, R. L., and Black, W. C. (1998). *Multivariate Data Analysis*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., and Sarstedt, M. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Thousand Oaks, CA: Sage Publication Inc.
- Han, J., Pei, J., and Kamber, M. (2011). *Data Mining: Concepts and Techniques*. Amsterdam: Elsevier.
- Hand, D., Mannila, H., and Smyth, P. (2001). *Principles of Data Mining*. Cambridge: MIT Press.
- Honick, T., and Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: a systematic review. *Educ. Res. Rev.* 17, 63–84. doi: 10.1016/j.edurev.2015.11.002
- Hsu, M.-H., and Chiu, C.-M. (2004). Internet self-efficacy and electronic service acceptance. *Decis. Support Syst.* 38, 369–381. doi: 10.1016/j.dss.2003.08.001
- İncirkus, K., and Nahcivan, N. (2020). Validity and reliability study of the Turkish version of the self-efficacy for managing chronic disease 6-item scale. *Turk. J. Med. Sci.* 50, 1254–1261. doi: 10.3906/sag-1910-13
- Jain, N., and Srivastava, V. (2013). Data mining techniques: a survey paper. *Intern. J. Eng. Technol.* 2, 116–119. doi: 10.15623/ijret.2013.0211019
- Jian, Z. F., and Hsu, C. Y. (2014). *Data Mining & Big Data Analysis*. New Taipei: Future Career Publishing Co.
- Judge, T. A., Erez, A., and Bono, J. E. (1998). The power of being positive: the relation between positive self-concept and job performance. *Hum. Perform.* 11, 167–187. doi: 10.1080/08959285.1998.9668030
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educ. Psychol. Measur.* 20, 141–151. doi: 10.1177/001316446002000116
- Kao, C.-P., Wu, Y.-T., and Tsai, C.-C. (2011). Elementary school teachers' motivation toward web-based professional development, and the relationship with Internet self-efficacy and belief about web-based learning. *Teach. Teach. Educ.* 27, 406–415. doi: 10.1016/j.tate.2010.09.010
- Keil, M., Tan, B. C., Wei, K. K., Saarinen, T., Tuunainen, V., and Wassenaar, A. (2000). A cross-cultural study on escalation of commitment behavior in software projects. *MIS Q.* 24, 299–325. doi: 10.2307/3250940
- Khasawneh, A. S., Jawarneh, M., Al-Sheshani, A., Iyadat, W., and Al-Shudaifat, S. (2009). Construct validation of an Arabic version of the college students' self-efficacy scale for use in Jordan. *Intern. J. Appl. Educ. Stud.* 6, 56–70.
- Kilday, J. E., Lenser, M. L., and Miller, A. D. (2016). Considering students in teachers' self-efficacy: examination of a scale for student-oriented teaching. *Teach. Teach. Educ.* 56, 61–71. doi: 10.1016/j.tate.2016.01.025
- Klassen, R. M., and Usher, E. L. (2010). Self-efficacy in educational settings: recent research and emerging directions. *Adv. Motiv. Achiev.* 16, 1–33. doi: 10.1108/S0749-74232010000016A004
- Kock, N., and Lynn, G. S. (2012). Lateral collinearity and misleading results in variance-based SEM: an illustration and recommendations. *J. Assoc. Inform. Syst.* 13, 546–580. doi: 10.17705/1jais.00302
- Kuiper, R. A., Murdock, N., and Grant, N. (2010). Thinking strategies of baccalaureate nursing students prompted by self-regulated learning strategies. *J. Nurs. Educ.* 49, 429–436. doi: 10.3928/01484834-20100430-01
- Kuo, F. Y., and Hsu, M. H. (2001). Development and validation of ethical computer self-efficacy measure: the case of softlifting. *J. Bus. Ethics* 32, 299–315. doi: 10.1023/A:1010715504824
- Lambert, Z. V., Wildt, A. R., and Durand, R. M. (1988). Redundancy analysis: an alternative to canonical correlation and multivariate multiple regression in exploring intersert associations. *Psychol. Bull.* 104:282. doi: 10.1037/0033-2909.104.2.282
- Liao, S. X. (2008). *Knowledge Management*. Taipei: Yeh Yeh Book Gallery.
- Liao, S. X., and Wen, Z. H. (2019). *Data Mining: Artificial Intelligence and Machine Learning Development*. Taipei: DrMaster Press Co.
- Liu, Q., Mo, L., Huang, X., Yu, L., and Liu, Y. (2020). The effects of self-efficacy and social support on behavior problems in 8–18 years old children with malignant tumors. *PLoS One* 15:e0236648. doi: 10.1371/journal.pone.0236648
- Lombardi, S., Santini, G., Marchetti, G. M., and Focardi, S. (2017). Generalized structural equations improve sexual-selection analyses. *PLoS One* 12:e0181305. doi: 10.1371/journal.pone.0181305
- Lorig, K., Chastain, R. L., Ung, E., Shoor, S., and Holman, H. R. (1989). Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. *Arthrit. Rheum.* 32, 37–44. doi: 10.1002/anr.1780320107
- Maldonado, E., and Seehusen, V. (2018). Data mining student choices: a new approach to business curriculum planning. *J. Educ. Bus.* 93, 196–203. doi: 10.1080/08832323.2018.1450212
- Mamaril, N. A., Usher, E. L., Li, C. R., Economy, D. R., and Kennedy, M. S. (2016). Measuring undergraduate students' engineering self-efficacy: a validation study. *J. Eng. Educ.* 105, 366–395. doi: 10.1002/jee.20121
- Mangos, P. M., and Johnston, J. H. (2008). "Performance measurement issues and guidelines for adaptive, simulation-based training," in *Human Factors in Simulation and Training*, eds D. A. Vincenzi, J. A. Wise, M. Mouloua, and P. A. Hancock (Boca Raton, FL: CRC Press), 301–320. doi: 10.1201/9781420072846.ch16
- Marvin, L. (2016). *Decision Trees and Applications with IBM SPSS Modeler*. Scotts Valley, CA: CreateSpace Independent Publishing Platform.
- Massi, M. C., Ieva, F., and Lettieri, E. (2020). Data mining application to healthcare fraud detection: a two-step unsupervised clustering method for outlier detection with administrative databases. *BMC Med. Inform. Decis. Mak.* 20:160. doi: 10.1186/s12911-020-01143-9
- McCormick, K., Abbott, D., Brown, M. S., Khabaza, T., and Mutchler, S. R. (2013). *IBM SPSS Modeler Cookbook*. Birmingham: Packt Publishing.
- McLaughlin, K., Moutray, M., and Muldoon, O. T. (2008). The role of personality and self efficacy in the selection and retention of successful nursing students: a longitudinal study. *J. Adv. Nurs.* 61, 211–221. doi: 10.1111/j.1365-2648.2007.04492.x
- Mitchell, T. M. (1999). Machine learning and data mining. *Commun. ACM* 42, 30–36. doi: 10.1145/319382.319388

- Multon, K. D., Brown, S. D., and Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: a meta-analytic investigation. *J. Counsel. Psychol.* 38, 30–38. doi: 10.1037/0022-0167.38.1.30
- Murnieks, C. Y., Mosakowski, E., and Cardon, M. S. (2014). Pathways of passion: identity centrality, passion, and behavior among entrepreneurs. *J. Manag.* 40, 1583–1606. doi: 10.1177/0149206311433855
- Nemati, H. R., and Barko, C. D. (2003). Key factors for achieving organizational data-mining success. *Industr. Manag. Data Syst.* 103, 282–292. doi: 10.1108/02635570310470692
- Pajares, F., and Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemp. Educ. Psychol.* 20, 426–443. doi: 10.1006/ceps.1995.1029
- Qian, L., and Liu, J. (2020). Application of data mining technology and wireless network sensing technology in sports training index analysis. *EURASIP J. Wire. Commun. Netw.* 121, 1–17. doi: 10.1186/s13638-020-01735-z
- Richardson, M., Abraham, C., and Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychol. Bull.* 138, 353–387. doi: 10.1037/a0026838
- Romero, C., and Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Trans. Syst. Man Cybernet. Part C* 40, 601–618. doi: 10.1109/TSMCC.2010.2053532
- Romero, C., and Ventura, S. (2013). Data mining in education. *WIREs Data Min. Knowl. Discov.* 3, 12–27. doi: 10.1002/widm.1075
- Salcedo, J., and McCormick, K. (2017). *IBM SPSS Modeler Essentials: Effective Techniques for Building Powerful Data Mining and Predictive Analytics Solutions*. Birmingham: Packt Publishing.
- Scholz, U., Doña, B. G., Sud, S., and Schwarzer, R. (2002). Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *Eur. J. Psychol. Assess.* 18, 242–251. doi: 10.1027/1015-5759.18.3.242
- Schunk, D. H. (1994). "Self-regulation of self-efficacy and attributions in academic settings," in *Self-regulation of Learning and Performance: Issues and Educational Applications*, eds D. H. Schunk and B. J. Zimmerman (Hillsdale, NJ: Erlbaum), 75–99.
- Sethi, V., and King, W. R. (1991). Construct measurement in information systems research: an illustration in strategic systems. *Decis. Sci.* 22, 455–472. doi: 10.1111/j.1540-5915.1991.tb01274.x
- Shane, S., Locke, E. A., and Collins, C. J. (2003). Entrepreneurial motivation. *Hum. Resour. Manag. Rev.* 13, 257–279. doi: 10.1016/S1053-4822(03)00017-2
- Singhal, S., and Jena, M. (2013). A study on WEKA tool for data preprocessing, classification and clustering. *Intern. J. Innov. Technol. Explor. Eng.* 2, 250–253.
- Smith, R. E. (1989). Effects of coping skills training on generalized self-efficacy and locus of control. *J. Pers. Soc. Psychol.* 56, 228–233. doi: 10.1037/0022-3514.56.2.228
- Srivastava, A., Bartol, K. M., and Locke, E. A. (2006). Empowering leadership in management teams: effects on knowledge sharing, efficacy, and performance. *Acad. Manag. J.* 49, 1239–1251. doi: 10.5465/amj.2006.23478718
- Struhl, S. (2017). *Artificial Intelligence Marketing and Predicting Consumer Choice: An Overview of Tools and Techniques*. London: Kogan Page Publishers.
- Sullivan, B. A., O'Connor, M. O., and Burtis, E. R. (2006). Negotiator confidence: the impact of self-efficacy on tactics and outcomes. *J. Exper. Soc. Psychol.* 42, 567–581. doi: 10.1016/j.jesp.2005.09.006
- Sun, J. C.-Y., and Chen, A. Y.-Z. (2016). Effects of integrating dynamic concept maps with interactive response system on elementary school students' motivation and learning outcome: the case of anti-phishing education. *Comput. Educ.* 102, 117–127. doi: 10.1016/j.compedu.2016.08.002
- Sun, X. (2020). Self-efficacy mediates the relationship between entrepreneurial passion and entrepreneurial behavior among master of business administration students. *Soc. Behav. Pers. Intern. J.* 48, 1–8. doi: 10.2224/sbp.9293
- Talsmaa, K., Schüza, B., Schwarzer, R., and Norrisa, K. (2018). I believe, therefore I achieve (and vice versa): a meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learn. Individ. Differ.* 61, 136–150. doi: 10.1016/j.lindif.2017.11.015
- Torkzadeh, G., and Van Dyke, T. P. (2001). Development and validation of an Internet self-efficacy scale. *Behav. Inform. Technol.* 20, 275–280. doi: 10.1080/01449290110050293
- Tufféry, S. (2011). *Data Mining and Statistics for Decision Making*. New York, NY: John Wiley & Sons.
- Vancouver, J. B., Thompson, C. M., and Williams, A. A. (2001). The changing signs in the relationships among self-efficacy, personal goals, and performance. *J. Appl. Psychol.* 86, 605–620. doi: 10.1037/0021-9010.86.4.605
- Wang, Y. M. (2019). *Measuring the Programming Self-Efficacy*. Working paper, National Chi Nan University, Taiwan.
- Wang, Y. S., Tseng, T. H., Wang, Y. M., and Chu, C. W. (2019). Development and validation of an internet entrepreneurial self-efficacy scale. *Internet Res.* 30, 653–675. doi: 10.1108/INTR-07-2018-0294
- Wang, Y. Y., Wang, Y. S., Lin, H. H., and Tsai, T. H. (2019). Developing and validating a model for assessing paid mobile learning app success. *Interact. Learn. Environ.* 27, 458–477. doi: 10.1080/10494820.2018.1484773
- Wester, K. L., Gonzalez, L., Borders, L. D., and Ackerman, T. (2019). Initial development of the faculty research self-efficacy scale (FaRSES): evidence of reliability and validity. *J. Profess.* 10, 78–99.
- Williamson, S., Clow, K. E., Walker, B. C., and Ellis, T. S. (2011). Ethical issues in the age of the internet: a study of students' perceptions using the multidimensional ethics scale. *J. Internet Commer.* 10, 128–143. doi: 10.1080/15332861.2011.571992
- Wilson, F., Kickul, J., and Marlino, D. (2007). Gender, entrepreneurial self-efficacy, and entrepreneurial career intentions: implications for entrepreneurship education. *Entrepreneursh. Theory Pract.* 31, 387–406. doi: 10.1111/j.1540-6520.2007.00179.x
- Wong, K. K. K. (2019). *Mastering Partial Least Squares Structural Equation Modeling (PLS-SEM) with Smartpls in 38 Hours*. Bloomington, IN: iUniverse.
- Wu, J. H., and Wang, Y. M. (2006). Measuring ERP success: the ultimate users' view. *Intern. J. Operat. Product. Manag.* 26, 882–903. doi: 10.1108/01443570610678657
- Xue, W. (2014). *Data Mining Based on SPSS Modeler*. Beijing: China Renmin University Press.
- Zhao, Y., Fang, L., Cui, L., and Bai, S. (2020). Application of data mining for predicting hemodynamics instability during pheochromocytoma surgery. *BMC Med. Inform. Decis. Mak.* 20:165. doi: 10.1186/s12911-020-01180-4
- Zhen, Z., and Yao, Y. (2019). Lean production and technological innovation in manufacturing industry based on SVM algorithms and data mining technology. *J. Intellig. Fuzzy Syst.* 37, 6377–6388. doi: 10.3233/JIFS-179217

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Wang, Chiou, Wang and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.