




Cite this: DOI: 10.1039/d6rp00083e

Psychometric network analysis of gendered patterns in experimental self-efficacy among Malaysian pre-university chemistry students

Byron MC Michael Kadum *^a and Mageswary Karpudewan ^b

Gender comparisons of experimental self-efficacy (ESE) in chemistry laboratories are typically based on mean scores, which do not reveal how ESE beliefs are organised as an interconnected system. A cross-sectional sample of 655 Malaysian pre-university chemistry students (179 male; 476 female) completed a 12-item ESE measure spanning conceptual understanding, procedural complexity, laboratory hazards, and resource sufficiency. Measurement invariance was examined using multi-group CFA; configural invariance was supported, with borderline evidence for full metric invariance, so gender comparisons are interpreted cautiously as patterns under this operationalisation. Gender-specific networks were estimated using regularised Gaussian graphical models; strength centrality and permutation-based network comparison tests (NCTs) were used to evaluate gender differences, with bootstrap and case-dropping analyses assessing robustness. Both networks were domain-clustered, with strongest within-domain connections and relatively limited cross-domain links. Descriptively, conceptual understanding (item 3), “I am confident that I understand the chemical processes in the experiment,” was the most central node in the female network, whereas laboratory hazards (item 2), “I am confident of working in the laboratory without chemical spillage,” was the most central node in the male network; however, NCT indicated no differences in network structure or global strength, and no individual edges or centralities differed significantly. Overall, ESE showed similar network organisation across gender in this context. The identified clusters and cross-domain links, particularly those involving conceptual sense-making, provide hypotheses for future research on laboratory support rather than direct evidence of instructional effects.

Received 20th February 2026,
Accepted 19th May 2026

DOI: 10.1039/d6rp00083e

rsc.li/cerp

1. Introduction

Laboratory work is a defining component of chemistry education because it is where abstract chemical ideas are connected to empirical evidence through observation, measurement, and interpretation. Importantly, chemistry understanding is not built at a single level. Students are expected to coordinate what they can see in the laboratory (macroscopic changes), explain those observations in terms of particles and interactions (sub-microscopic processes), and communicate reasoning through equations, symbols, graphs, and calculations (symbolic representations). Johnstone’s (1991) triangle highlights that moving across these three levels is central to learning chemistry – and also a key source of difficulty when students struggle to connect them coherently.

1.1. Self-efficacy in chemistry education

Chemistry learning routinely requires students to coordinate what they know with what they can do in demanding contexts, including problem solving, modelling chemical phenomena, and making decisions under uncertainty during practical work. Within this landscape, self-efficacy has become a central construct for understanding why students with comparable preparation may nevertheless differ in persistence, engagement, and performance. Self-efficacy refers to learners’ beliefs about their capability to organise and execute actions needed to attain designated outcomes (Bandura, 1986, 1997). A defining feature of self-efficacy is its task- and domain-specificity (Wang and Richarde, 1988; Moreno *et al.*, 2021): confidence is not treated as a general trait but as a judgement tied to particular performances (*e.g.*, balancing redox equations, explaining molecular structure–property relations, interpreting data from a titration). This specificity matters in chemistry, where students often experience uneven confidence across different knowledge demands and practices. Consistent with broader educational research, self-efficacy is typically theorised as developing through four sources: mastery

^a Kolej Matrikulasi Labuan, Malaysian Institute of Chemistry, Malaysia.
E-mail: byrnkdm@gmail.com

^b School of Educational Studies, Universiti Sains Malaysia, 11800, Penang, Malaysia



experiences, vicarious experiences, social persuasion, and physiological/affective states (Bandura, 1997).

Across chemistry education research, self-efficacy is consistently framed as consequential because it shapes students' motivation and learning behaviours. Importantly, chemistry self-efficacy is not only an outcome of instruction but may also operate as a mechanism through which learning experiences are translated into sustained engagement and achievement in chemistry. For instance, among first-year chemistry undergraduates in New Zealand, students with higher self-efficacy tended to attribute success to effort, whereas those with lower self-efficacy were more likely to interpret difficulties as a lack of ability – an appraisal associated with discontinuing their studies (Dalgety and Coll, 2006). Consistent with this, Naibert *et al.* (2024), using data from 1485 undergraduates across six U.S. institutions, reported that stronger chemistry self-efficacy was associated with a mastery-oriented focus rather than failure avoidance, which in turn aligned with better performance on high-stakes assessments. Similar patterns are evident at the secondary level: among 370 Israeli Grade 11 students, higher chemistry self-efficacy was linked to greater willingness to engage with complex problem-solving rather than avoidance (Avargil, 2019).

Evidence in chemistry education indicates that students' chemistry self-efficacy is not static, but can shift over time in response to instructional experiences and performance-related feedback. For example, studies have tracked chemistry self-efficacy repeatedly during a semester and reported systematic within-semester change alongside achievement patterns, consistent with self-efficacy being sensitive to ongoing course experiences (*e.g.*, assessment outcomes, mastery experiences, and classroom demands) (Villafañe *et al.*, 2014, 2016). Complementary longitudinal work using self-efficacy items embedded within established chemistry attitudes/experiences instruments likewise shows that students' self-efficacy can vary across timepoints as they progress through chemistry learning experiences (Dalgety and Coll, 2006). Taken together, these findings support conceptualising chemistry self-efficacy as a course-responsive belief that may shape students' persistence and longer-term decisions about continuing in STEM pathways (Villafañe *et al.*, 2016).

1.2. Experimental self-efficacy

Building on the representational demands highlighted by Johnstone's (1991) triangle and the high-stakes nature of practical work in Malaysia's pre-university chemistry programme (Abdullah *et al.*, 2022; Matriculation Division, 2022a, 2022b), it follows that students' confidence must be considered not only in general chemistry learning but also in the specific context where macroscopic observations are generated and then translated into submicroscopic explanations and symbolic representations – namely, the laboratory. In this setting, learners are expected to act competently under authentic constraints (*e.g.*, safety, time, limited apparatus, measurement accuracy, and uncertain outcomes) while simultaneously coordinating theory with evidence. Given the task- and domain-specific nature of self-efficacy (Bandura, 1986, 1997; Wang and Richarde, 1988; Moreno *et al.*, 2021),

students may hold markedly different confidence judgements for laboratory work compared to lecture-based problem solving. Thus, experimental self-efficacy can be conceptualised as students' beliefs in their capability to plan, execute, and interpret chemistry experiments successfully within the procedural and safety demands of laboratory practice (Kolil *et al.*, 2020, 2023).

The chemistry laboratory introduces performance demands that are partly distinct from those typically emphasised in classroom learning. Beyond conceptual understanding, students must demonstrate competence in handling chemicals and glassware, following and adapting procedures, making accurate measurements, observing and recording changes systematically, troubleshooting errors, and drawing defensible conclusions from imperfect data. At the same time, these demands are affectively charged: laboratory activities can elicit anxiety related to hazards, breakage, contamination, or “getting the wrong result.” Such emotional arousal is theoretically relevant because physiological and affective states are one of the four principal sources of self-efficacy beliefs (Bandura, 1997). Consistent with this, chemistry laboratory research shows that uncertainty and anxiety can shape how confidently students approach practical tasks. For example, Kurbanoglu and Akin (2010), studying chemistry undergraduates at four Turkish universities, reported that affective responses to laboratory uncertainty were associated with students' confidence in undertaking experimental work. Similarly, Galloway and Bretz (2015), surveying 3583 undergraduates across 15 U.S. institutions, found that negative affective expectations (*e.g.*, worry and confusion) can operate as self-fulfilling prophecies, inhibiting the development of laboratory confidence. Notably, however, Wong *et al.* (2021) reported a contrasting pattern among 320 Malaysian 14-year-old secondary school students: despite limited exposure to hands-on laboratory work, students reported high practical self-efficacy. The authors further suggested that this perceived experimental self-efficacy was fully mediated by students' conceptions of learning science rather than by direct practical experience, highlighting that students' interpretive frames may also shape efficacy beliefs for what it means to learn science.

Prior literature suggests that experimental self-efficacy (ESE) is multidimensional, with students differentiating confidence across key laboratory competencies such as safe practice, apparatus handling, procedural execution, measurement/data collection, and interpretation, rather than holding a single global judgement (Uzuntiryaki-Kondakci and Capa-Aydin, 2013; Alkan, 2016). Extending this laboratory-specific framing, Kolil *et al.* (2020, 2023) empirically characterised ESE as manifesting through four dominant, laboratory-relevant facets – conceptual understanding, procedural complexity, laboratory hazards, and sufficiency of resources. This facet structure may be interpreted in relation to Bandura's (1997) four sources of self-efficacy – mastery experiences, vicarious experiences, social persuasion, and physiological/affective states. This operationalisation of ESE provides a domain-appropriate lens for representing the kinds of competence judgements Malaysian pre-university students must make to participate productively in high-stakes practical work (Abdullah *et al.*, 2022; Matriculation Division, 2022a, 2022b).



1.3. Gender differences in ESE

Evidence on gender differences in ESE is mixed and appears to depend on (i) whether self-efficacy is measured as a lab-specific construct or a broader chemistry/science belief, and (ii) whether gender is tested as a mean difference or as a moderator of relationships between self-efficacy and performance. In several studies in the chemistry context, gender differences in overall ESE levels are small or non-significant. For example, Turkish secondary school students' chemistry ESE (specifically, cognitive and psychomotor beliefs) did not differ significantly by gender, even though gender differences were detected for several laboratory perception dimensions (Merve Mustafaoglu and Alkan, 2024). Similarly, among 198 first-year chemistry majors from Kazakhstan universities, ESE showed no statistically significant gender difference, although female students displayed a slightly higher mean (Sailaubay *et al.*, 2024). A comparable “no gender difference” pattern is also reported in chemistry teacher education among 191 Indonesian pre-service teachers, where self-efficacy did not significantly differ between males and females (Wahyudiati *et al.*, 2020).

Notwithstanding these null findings, one study indicates that gender differences can emerge more clearly when ESE is assessed within a chemistry laboratory anxiety–self-efficacy framework. In a sample of 363 Turkish pre-service science teachers, female participants reported significantly higher ESE than male participants (Kirbaşlar *et al.*, 2015). This finding suggests that gender effects may be more detectable in certain teacher-education populations and measurement contexts – particularly when laboratory confidence is conceptualised alongside affective factors that shape laboratory participation (*e.g.*, anxiety in chemistry laboratory environments).

Crucially, the literature suggests that gendered patterns in self-efficacy may be expressed less as stable mean-level gaps and more as differences in how efficacy beliefs are measured, organised, and connected to learning processes. In chemistry education, this concern is reflected in the importance of establishing measurement equivalence before interpreting gender differences, because apparent group differences are meaningful only when the instrument functions comparably across gender groups (Rocabado *et al.*, 2020). This point is reinforced by Saputra *et al.* (2024), who examined the Indonesian version of the High School Chemistry Self-Efficacy Scale through confirmatory and multigroup invariance analyses before interpreting gender- and grade-level differences. Consistent with this broader point, Sezgintürk and Sungur (2020), using a multidimensional model of science self-efficacy that included a practical-work dimension (*i.e.*, ESE), found no overall multivariate gender difference and almost identical practical-work self-efficacy means for girls and boys. However, the explanatory structure differed: the set of predictors accounted for more variance in girls' practical-work self-efficacy than in boys', implying that the drivers of laboratory-related confidence may differ by gender even when average levels are similar. In chemistry education, Villafañe *et al.* (2014) similarly showed that chemistry self-efficacy patterns may vary across time and demographic subgroups, suggesting that self-efficacy should not be interpreted only as a fixed mean-level attribute. Related science education research also suggests that gender can

shape the affective and motivational conditions surrounding self-efficacy: Kiran and Sungur (2012), for example, found no gender difference in science self-efficacy or strategy use, but reported gender differences in emotional arousal and interpersonal sources of efficacy. Collectively, these findings support the view that gender may influence how self-efficacy beliefs are organised, supported, and connected to broader learning processes, rather than producing a uniform “female lower/male higher” pattern.

Evidence that gender can condition the translation of ESE into outcomes is clearest in work using ESE scores across scientific disciplines and examining their relations with laboratory performance. For example, Kolil *et al.* (2023) reported that ESE was positively associated with laboratory scores in both groups, but the gender interaction differed in direction and magnitude across disciplinary contexts. In their chemistry group, the ESE \times Gender interaction was positive ($B = 0.213$, $\beta = 0.066$, $p = 0.027$), indicating a stronger ESE-laboratory score relationship for males than for females. In their physics/biology group, however, the interaction was negative and somewhat larger in absolute size ($B = -0.266$, $\beta = -0.090$, $p < 0.001$), indicating a weaker relationship for males and a stronger one for females. Thus, gender was not a consistent direct predictor of laboratory performance, but it did condition how strongly ESE related to laboratory scores across disciplinary contexts (Kolil *et al.*, 2023). In chemistry laboratory curriculum interventions, improvements in inquiry-related self-efficacy have also been observed for both men and women, indicating that well-designed laboratory experiences can raise experimental confidence across genders rather than amplifying gaps (Winkelmann *et al.*, 2015). Taken together, the evidence indicates that gender differences in ESE are best understood as context- and model-dependent: mean differences are often small or absent, but gender may still matter through (i) differences in belief structure and (ii) differences in how efficacy beliefs connect to performance (Sezgintürk and Sungur, 2020; Kolil *et al.*, 2023).

1.4. Psychometric network analysis

Most research on chemistry self-efficacy – including laboratory-related efficacy – has used score-based, variable-centred approaches (*e.g.*, comparing mean differences by gender, correlating total/subscale scores with achievement, or modelling paths in SEM) (Kirbaşlar *et al.*, 2015; Winkelmann *et al.*, 2015; Sezgintürk and Sungur, 2020; Wahyudiati *et al.*, 2020; Kolil *et al.*, 2023; Merve Mustafaoglu and Alkan, 2024; Sailaubay *et al.*, 2024). Although these methods are valuable for estimating overall group differences and predictive relations, they implicitly treat ESE as a single underlying attribute that can be adequately summarised by aggregate scores. This assumption is restrictive for laboratory learning, where confidence is inherently multifaceted and context-bound (*e.g.*, managing hazards, executing procedures, interpreting outcomes, coping with time/apparatus/resource constraints). In such settings, two students (or two genders) can show similar mean levels of ESE, yet hold different configurations of confidence – meaning the “same score” can reflect different belief profiles and different internal dependencies among beliefs. Although other multidimensional approaches could also move beyond a single summary score, psychometric



network analysis was selected because the present study aimed not only to represent ESE as multidimensional, but to examine how specific confidence judgements are conditionally connected within a belief system (Borsboom *et al.*, 2021). This approach complements, rather than replaces, the four-factor and measurement-invariance analyses reported here. Its added value lies in identifying tightly linked clusters, central and bridging beliefs, and possible gender differences in the organisation and overall connectivity of those item-level relations, rather than only comparing totals, subscales, or latent factors (Epskamp *et al.*, 2018; van Borkulo *et al.*, 2023). For ESE in chemistry, this is particularly relevant because students may show similar overall levels of confidence while differing in how conceptual, procedural, safety-related, and resource-related beliefs are structured and connected.

Psychometric network analysis operationalises this systems perspective by modelling ESE items as nodes and estimating edges as conditional associations (typically regularised partial correlations) between items after controlling for all others (Epskamp *et al.*, 2018; Borsboom *et al.*, 2021). Conceptually, this approach aligns with the idea that psychological attributes can emerge from direct interactions among components, rather than being explained solely by a latent common cause (Borsboom *et al.*, 2021). For ESE in chemistry, this is theoretically meaningful: beliefs about safe handling, glassware confidence, procedural fluency, and interpretive competence are likely to co-activate during real practical work and may reinforce one another through repeated mastery experiences, feedback, and affective responses in the lab. Network models can therefore reveal which beliefs form tight clusters, which beliefs bridge across domains, and which conditional links represent key “routes” through which confidence may generalise from one aspect of laboratory work to another (Epskamp *et al.*, 2018; Borsboom *et al.*, 2021). These structural insights are largely inaccessible when items are collapsed into subscale totals.

A further justification is that network models can identify central and bridging components of ESE – beliefs that are highly connected within the network or that connect otherwise distinct clusters – thereby offering a principled way to locate potential leverage points for instructional design and intervention (Epskamp *et al.*, 2018). This is particularly useful in chemistry laboratory contexts because small improvements in a central belief (*e.g.*, confidence in handling hazards or apparatus) may plausibly cascade into other connected components (*e.g.*, willingness to engage, persistence with procedures, or confidence in interpreting results). Network methodology also emphasises the need to evaluate the robustness of estimated edges and centrality indices, encouraging researchers to report accuracy/stability evidence (*e.g.*, bootstrapped confidence intervals) rather than treating network features as fixed properties (Epskamp *et al.*, 2018).

Importantly, psychometric network analysis strengthens gender-focused inquiry by moving beyond the question of whether male and female students differ in overall level to whether they differ in structure – that is, whether the organisation of efficacy beliefs differs by gender even when mean differences are mixed or small. The network comparison test (NCT) enables formal tests of group differences in (i) overall connectivity (global strength)

and (ii) network structure (patterns of conditional associations), which directly matches the theoretical possibility that gender effects operate through different belief architectures rather than through simple mean gaps (van Borkulo *et al.*, 2023). A psychometric network approach provides an empirically and theoretically grounded way to advance knowledge on ESE in chemistry by (a) mapping how specific confidence beliefs cohere under authentic laboratory demands, (b) identifying central and bridging beliefs that may be instructionally strategic, and (c) testing whether the organisation of those beliefs differs by gender – thereby offering explanations that complement, rather than duplicate, score-based comparisons (Epskamp *et al.*, 2018; Borsboom *et al.*, 2021; van Borkulo *et al.*, 2023).

In chemistry laboratories, self-efficacy is rarely isolated to a single skill. Feeling capable of working safely may shape willingness to engage with uncertain outcomes; confidence in procedural control may influence whether students persist when results deviate from expectation; and confidence in interpreting evidence may determine whether students treat unexpected data as informative rather than as failure. A network perspective is therefore useful because it models ESE as an interconnected belief system rather than as independent subscale scores. Importantly, if the belief framework differs by gender, then instructional supports may need to be targeted differently; conversely, if the framework is invariant, then universal leverage points for strengthening ESE can be prioritised.

There is currently no consensus on the exact *a priori* power requirements or minimum sample sizes for cross-sectional psychological network models, as the required N depends on the number of nodes, network sparsity, and underlying edge weights (Epskamp *et al.*, 2017, 2018; Hevey, 2018). Simulation work on Ising and Gaussian graphical models by Epskamp *et al.* (2018) indicated that the sample sizes typically used in psychological research can produce poorly recovered networks, whereas larger samples yield more accurate edge-weight estimates and more stable centrality indices. In their tutorial on network accuracy, Epskamp *et al.* (2018) varied sample size between 100 and 5000 and demonstrated that the correlation-stability (CS) coefficient for centrality increases as a function of N, recommending that CS values should not fall below 0.25 and preferably be above 0.50 before centrality differences are interpreted. In a related methodological paper, Epskamp *et al.* (2017) emphasised that network models often contain many parameters relative to typical psychological sample sizes and cautioned that results may depend heavily on modelling assumptions when N is small, whereas larger samples tend to yield more robust and accurate networks. Hevey (2018) similarly argued that larger samples produce more stable and interpretable networks and noted the limited empirical basis for conventional *a priori* power analyses for network models. Although these simulation and tutorial studies were conducted mainly in clinical and health-psychology contexts, they examined general statistical properties of regularised psychological network models for Likert-type variables rather than disorder-specific content (Epskamp *et al.*, 2017, 2018; Hevey, 2018). The Experimental Self-Efficacy (ESE) items used in this study are likewise domain-specific psychological indicators



(science/chemistry self-efficacy) measured on an ordinal response scale, so the same issues of parameter-to-sample-size ratio, sparsity, and centrality stability are applicable. In chemistry education research, large-scale survey studies typically involve a few hundred participants; accordingly, the present study drew on these domain-general methodological recommendations to justify a relatively large sample for a modest 12-node network, to consider the adequacy of the male and female subgroup N s separately, and to interpret the smaller male network more cautiously.

1.5. The present study

To this end, this study extends score-based research on ESE by testing whether the structure and connectivity of laboratory confidence differ by gender in a high-stakes pre-university chemistry context. Although psychometric network analysis is increasingly used to study complex psychological constructs, its application to ESE among Malaysian pre-university chemistry students remains limited, despite this being a high-stakes population learning within an intensive, centrally designed STEM programme. The following research questions guide the investigation:

- (a) What is the network structure of ESE for male and female students?
- (b) Which ESE items emerge as most central within each gender network?
- (c) Are the male and female ESE networks invariant in overall structure and global strength?

2. Methodology

2.1. Research design and participants

The study employed a cross-sectional research design to examine the psychological network structure of ESE among Malaysian pre-university students. A total of 655 science students aged approximately 18–19 years, from a public pre-university college in Malaysia participated. Network structures were estimated separately for male and female students to allow gender-based comparisons. Of the 655 participants, 179 were male (27.33%) and 476 were female (72.67%). In the year of data collection, the science student population at this college was 2285, with approximately 30% males and 70% females, indicating that the sample closely reflected the institution's overall gender distribution.

The study site was a government-funded pre-university college that offers a science curriculum broadly aligned with the Cambridge Advanced Level (A-Level). English was the primary medium of instruction, and all examination questions were set in English. Students admitted to this college were typically high performers in the Sijil Pelajaran Malaysia (SPM) examination, which is comparable to the British General Certificate of Education Ordinary Level (GCE O-Level). After completing the one-year programme, most students progressed to STEM-related undergraduate degrees at universities in Malaysia and abroad. These representational demands are particularly relevant in Malaysia's pre-university chemistry context, where laboratory learning is structured, compulsory, and high-stakes (Matriculation Division, 2022a, 2022b).

The pre-university programme in Malaysia requires students to complete practical work routinely (e.g., scheduled laboratory sessions and a set number of experiments each semester), with explicit expectations related to safe practice, correct technique, accurate measurement, error awareness, and evidence-based conclusions (Abdullah *et al.*, 2022). In such settings, successful participation depends not only on conceptual knowledge but also on students' confidence to plan, execute, and interpret experimental work under real constraints. Therefore, self-efficacy becomes pivotal in shaping whether students initiate tasks, invest effort, and persist when laboratory outcomes are ambiguous or do not match expectations.

The college was selected using a convenience sampling approach, considering practical considerations such as travel distance and the feasibility of obtaining administrative approval. Malaysia has 15 government-funded pre-university colleges, each typically enrolling about 1500 to 3000 students per intake. The participating college usually admits between 1900 and 2400 science students per cohort and follows a centrally designed national STEM curriculum. Its students are competitively selected through a national merit-based matriculation process. For the science stream, applicants are typically required to obtain at least a credit in Bahasa Melayu, English, Additional Mathematics, Chemistry, and either Physics or Biology, a grade B in Mathematics, and a pass in History. Selection is then made on a meritocratic basis, considering academic merit, co-curricular achievement, and family income; the merit calculation allocates 90% to academic performance and 10% to co-curricular performance. Consequently, although the data formally came from a single institution, the context shares common curricular features with other colleges, so findings may be most transferable to similar pre-university settings; however, generalisation beyond this site requires replication.

The final overall sample of 655 pre-university students substantially exceeded the $N = 359$ used in Epskamp *et al.* (2018) work on PTSD and fell in the mid-to-upper range of sample sizes (100–5000) examined in their simulation studies. Because networks were estimated separately for males and females, the adequacy of these subgroup sizes was also considered. Epskamp *et al.* (2018) showed that regularised Gaussian graphical models can recover the general structure of sparse networks at $N \approx 100$, with edge-weight and centrality estimates becoming increasingly stable as N increases. The female subsample ($N = 476$) lay in the middle of this simulated range, whereas the male subsample ($N = 179$) was closer to the lower bound but still above $N = 100$. For a 12-node network, this corresponded to a case-to-node ratio of roughly 40:1 for females and 15:1 for males, which compares favourably with examples and recommendations in the methodological literature (Epskamp *et al.*, 2017, 2018; Hevey, 2018). Taken together, the overall $N = 655$ and the subgroup N s (179 males, 476 females) placed this study in a range where prior work suggests that regularised psychological networks can be estimated with satisfactory accuracy and stability, while still warranting somewhat more cautious interpretation for the smaller male network.



2.2. Research instrument

The psychometric instrument used in this study was the experimental self-efficacy (ESE) scale developed by Kolil *et al.* (2020, 2023) to assess students' self-efficacy for carrying out laboratory work. Although the scale was initially introduced in the context of virtual-laboratory-supported chemistry instruction, the items target core laboratory tasks and judgements that are also relevant to face-to-face chemistry laboratory work. This interpretation is consistent with prior score-based validity evidence for ESE score interpretations in chemistry laboratory classroom contexts (Kolil *et al.*, 2023) and with the content and response-process evidence gathered in the present study. The scale was developed based on Bandura's (1986) definition of self-efficacy as individuals' judgements of their capability to organise and execute actions required to achieve specific tasks. The ESE scale comprises 12 items grouped into four factors – conceptual understanding (CU), laboratory hazards (LH), procedural complexity (PC), and sufficiency of resources (SR) – with three items per factor. CU refers to how well students understand the theoretical basis of an experiment, including the underlying ideas, concepts, and principles being investigated. LH refers to potential accidents that may occur in the laboratory, such as breaking apparatus or spilling chemicals. PC captures the difficulties students experience when performing numerical calculations and following experimental procedures, including applying the required formulas, carrying out computations, and following the steps in the correct sequence. SR assesses whether students perceive that there are sufficient laboratory materials and support to complete the experiments, such as adequate glassware, instruments, and instructor guidance. Participants rated all items on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

The instrument development and administration process proceeded in four stages. First, the initial adapted ESE items were reviewed by three chemistry education subject matter experts on 5 December 2024 to gather evidence based on test content and identify wording revisions. Second, the revised pilot version was administered on 7 January 2025 to two randomly selected intact chemistry lecture groups, comprising 221 pre-university students, using a Google Form after their chemistry lecture session, with permission obtained from the respective lecturer to use approximately 10 to 15 minutes of class time. Third, response-process interviews with seven students were conducted on 17 January 2025 to identify any remaining wording difficulties. Fourth, the resulting final 12-item version was administered to the main sample from 21 to 25 April 2025 using a Google Form with eight randomly selected intact chemistry lecture groups that had not participated in the pilot phase. With permission from the respective lecturers, students completed the survey during the final 10 to 15 minutes of their chemistry lecture sessions. In both the pilot and main phases, students were asked to respond with reference to their current face-to-face chemistry laboratory experiences within the pre-university chemistry programme.

2.2.1. Pilot study. Evidence supporting the interpretation of ESE scores in this context was examined by gathering validity and reliability evidence with the study population. Specifically,

analyses were conducted to evaluate content, construct, and criterion-related aspects of validity, as well as internal consistency. In line with Lewis (2022), a pilot study was carried out to address four sources of validity evidence: test content, response processes, internal structure, and reliability. The pilot sample comprised 221 pre-university students who were not included in the main study, along with three chemistry education subject matter experts.

2.2.2. Test content validity. Evidence based on test content in the present study was gathered through consultation with three subject matter experts in chemistry education, all of whom held doctoral qualifications in STEM or chemistry education and had 20 to 22 years of experience in pre-university chemistry teaching and laboratory instruction. In addition, these experts had experience conducting chemistry education research and reviewing research instruments used in such studies. They reviewed the items to determine whether they adequately represented ESE. The experts indicated the items to be appropriate for eliciting ESE judgements about face-to-face chemistry laboratory work among pre-university students, and that the wording was comprehensible for this group. However, they recommended revising two CU items for clarity: changing “I believe I have a sound grasp of the theory behind laboratory experiments before performing experiments” to “I believe I have a good understanding of the theory behind laboratory experiments before performing experiments,” and “I am confident that I understand the underlying chemical phenomena in the experiment” to “I am confident that I understand the chemical processes in the experiment.”

2.2.3. Response process. To gather evidence for response process validity, seven pre-university students who were randomly selected from the pilot study group were interviewed. They were instructed to respond to the items from the ESE scale and comment on their understanding. Student responses revealed uncertainty about the phrase “draw relevant” in the PC item “When presented with laboratory results, I know how to interpret them and draw relevant conclusions from them”; therefore, the phrase was replaced with “make correct.” For the remaining items, including those previously revised based on feedback from the three subject matter experts in chemistry education, students agreed that the wording was clear and understandable.

2.2.4. Internal structure. The modified ESE scale was completed by 221 pre-university students from the pilot sample to examine its internal structure. To this end, a confirmatory factor analysis (CFA) was carried out. The findings supported the hypothesised four-factor structure of the ESE scores in the pilot sample. Although the chi-square statistic was significant, $\chi^2(48) = 113.99, p < 0.001$, which formally indicated misfit, this outcome was likely influenced by the well-known sensitivity of the chi-square statistic to sample sizes (Knehta *et al.*, 2019). Therefore, greater weight was placed on alternative fit indices. The comparative fit index (CFI = 0.99) and Tucker–Lewis index (TLI = 0.98) were both above the commonly recommended cut-off of 0.95, suggesting very good incremental fit (Hu and Bentler, 1999). The root mean square error of approximation (RMSEA = 0.08, 90% CI [0.06, 0.10]) pointed to an adequate, though not optimal, level of



approximate fit, while the standardised root mean square residual (SRMR = 0.04) reflected good fit (Hu and Bentler, 1999). Taken together, these indices indicated that the proposed four-factor ESE model satisfactorily represented the internal structure of the observed data.

2.2.5. Reliability. Internal consistency reliability was evaluated using McDonald's omega and Cronbach's alpha, given recommendations that ω typically provides a more accurate estimate of scale reliability than α , particularly when the assumption of tau-equivalence is unlikely to hold (Dunn *et al.*, 2014; McNeish, 2018). In line with conventional psychometric guidelines, coefficients of approximately 0.80 are considered good and values above 0.90 high or excellent for research instruments (Nunnally and Bernstein, 1994). Item–rest correlations were also inspected, as item–total correlations and related inter-item statistics are widely used indicators of internal consistency in scale development, with values below about 0.30 often treated as a criterion for item removal (Clark and Watson, 1995; Boateng *et al.*, 2018). For the 12-item ESE scale, internal consistency reliability was examined in the pilot sample ($N = 221$). The total scale showed excellent internal consistency, with McDonald's omega $\omega = 0.919$ (SE = 0.008, 95% CI [0.904, 0.935]) and Cronbach's alpha $\alpha = 0.918$ (SE = 0.014, 95% CI [0.890, 0.945]), comfortably exceeding commonly recommended thresholds. Item–rest correlations ranged from 0.51 to 0.80 (Table 1), indicating that all items contributed positively to a common ESE construct.

2.2.6. Statistical analysis procedure. All analyses were conducted using JASP version 0.95.4. Descriptive statistics (means and standard deviations) were first computed for each ESE item to summarise the response distributions. Internal consistency for the scale was then examined using Cronbach's alpha and McDonald's omega. Reporting reliability indices prior to network estimation is standard practice in network psychometrics research in psychology and education (Ji *et al.*, 2024; Zhang *et al.*, 2024).

For the network analysis, each of the 12 ESE items was treated as a node, and an edge represented the unique association between two items after statistically controlling for the remaining items. There were no missing responses across the 12 ESE items; therefore, all cases were retained for network estimation. Networks were estimated for the full sample and separately for male and female students using Gaussian graphical models (GGMs), as implemented in the Network (Frequentist) module in JASP. In this framework, the network is based on partial correlations, meaning that an edge is retained only when two items remain associated after the other items in the network have been considered (Hevey, 2018).

To avoid an overly dense network driven by trivial associations, the least absolute shrinkage and selection operator (LASSO) regularisation was applied. This procedure shrinks very small edge weights towards zero, producing a sparser and more interpretable network. The extended Bayesian information criterion (EBIC-glasso) was then used to select the most parsimonious network, with hyperparameter $\gamma = 0.5$ to balance simplicity and edge retention. Because responses were recorded on a 5-point Likert scale, items were treated as approximately continuous, following common practice in applied network psychometrics, although this choice may influence edge estimates (Hevey, 2018; Huth *et al.*, 2023). To address this, robustness analyses were conducted and small edges were interpreted cautiously (Hevey, 2018; Huth *et al.*, 2023; Diaz-Milanes *et al.*, 2024). In addition, the gender-specific networks were re-estimated using higher γ values (0.75 and 1.00) as a sensitivity check. These analyses produced identical weight matrices and network structures, indicating that the findings were robust to stricter regularisation.

Networks were visualised using a force-directed (Fruchterman–Reingold) layout, with thicker edges indicating stronger partial correlations. Node centrality was quantified using the indices available in JASP – strength, closeness, betweenness,

Table 1 Frequentist individual item reliability statistics of the modified ESE scale

Item	Item description	Item-rest correlation		
		Estimate	Lower 95% CI	Upper 95% CI
CU1	I believe I have a good understanding of the theory behind laboratory experiments before performing experiments.	0.637	0.551	0.709
CU2	Experimental concepts become clearer to me as I perform the experiment.	0.625	0.538	0.700
CU3	I am confident that I understand the chemical processes in the experiment.	0.696	0.621	0.758
LH1	I can usually handle the glass apparatus in the laboratory on my own without any fear of breakage and injury.	0.505	0.400	0.598
LH2	I am confident of working in the laboratory without chemical spillage.	0.585	0.491	0.665
LH3	I am always alert in the laboratory and have minimal accidents.	0.539	0.438	0.626
PC1	After an experiment, I have no difficulty figuring out how my calculation procedures and errors affected my results.	0.713	0.641	0.772
PC2	When presented with laboratory results, I know how to interpret them and make correct conclusions from them.	0.795	0.741	0.839
PC3	I do not struggle with processing information in background articles and relating them to my own laboratory procedures and results.	0.763	0.702	0.813
SR1	I find it easy to complete the exercise in the laboratory even though there is limited personal participation in performing experiments.	0.652	0.569	0.722
SR2	It is easy for me to understand theory and concepts properly in spite of limited availability of physical instruments.	0.728	0.659	0.784
SR3	I do not find it challenging to understand an experiment even if there is only one try due to limited availability of chemicals.	0.725	0.655	0.782

CU = conceptual understanding; LH = laboratory hazards; PC = procedural complexity; SR = sufficiency of resources.



and strength/expected influence – to characterise how strongly and how efficiently each node is connected to the rest of the network (Hevey, 2018; Huth *et al.*, 2023; Diaz-Milanes *et al.*, 2024). Higher centrality values were interpreted as indicating items that may play a more influential role in the ESE network.

The accuracy and stability of the estimated networks were evaluated using non-parametric bootstrapping procedures, following established guidelines for psychological network analysis (Epskamp *et al.*, 2018; Hevey, 2018; Huth *et al.*, 2023). Bootstrapped 95% confidence intervals were computed for each edge weight to assess estimation precision; narrower intervals were taken to indicate more reliable edges. Stability of centrality indices (especially node strength and expected influence) was assessed *via* case-dropping subset bootstrapping and summarised using the correlation-stability (CS) coefficient, which reflects the maximum proportion of cases that can be dropped while retaining a high correlation ($r \geq 0.70$) between the original and subset-based centrality estimates (Epskamp *et al.*, 2018; Hevey, 2018). In line with current recommendations, CS values of at least 0.25 were regarded as acceptable, with values of 0.50 or higher considered preferable for drawing substantive conclusions about central nodes (Epskamp *et al.*, 2018).

In addition to standard centrality indices, bridge expected influence (BEI) was computed to quantify which ESE items connect different domains. Clusters were pre-specified according to the four ESE domains (CU, LH, PC, SR). One-step BEI was operationalised as the sum of a node's edge weights linking it to nodes outside its own domain (Jones *et al.*, 2021). Because all retained edges in the estimated networks were positive, BEI yields the same rank ordering as bridge strength, but retains an interpretation aligned with expected influence. To evaluate whether the estimated networks recover the hypothesised four-domain organisation, community detection was performed using the Louvain algorithm and modularity (Q) was computed for each gender-specific network (Blondel *et al.*, 2008). Bridge indices and modularity results were used descriptively to strengthen the interpretation of cross-domain connectivity patterns.

Before comparing networks by gender, the measurement invariance of the four-factor ESE model was evaluated using multi-group CFA. The configural model showed acceptable fit (CFI = 0.924, RMSEA = 0.059, SRMR = 0.056), supporting configural invariance across males and females. Imposing equality constraints on factor loadings (metric invariance) resulted in a small reduction in model fit (CFI = 0.910; Δ CFI = -0.014; Δ RMSEA = 0.001; Δ SRMR = 0.016), although the χ^2 difference test was non-significant ($\Delta\chi^2(12) = 17.921$, $p = 0.118$). Adding intercept constraints (scalar invariance) did not further degrade fit relative to the metric model (CFI = 0.918; Δ CFI = 0.008; Δ RMSEA = -0.005; Δ SRMR = 0.000; $\Delta\chi^2(8) = 4.525$, $p = 0.807$). Collectively, these results support configural invariance, indicating that the same four-factor ESE configuration was recovered across males and females. However, support for full metric invariance was only borderline, meaning that exact equality of item-factor loadings across gender was not clearly established. Because metric invariance concerns whether the strength of item-factor relations is similar across groups, incomplete loading

equivalence limits stronger claims that individual items are functioning in the same way for males and females (Gregorich, 2006; Rocabado *et al.*, 2020). Accordingly, the gender-based network comparisons in the present study are interpreted cautiously as descriptive comparisons of broad network organisation under the current operationalisation, rather than as definitive evidence that specific item-level contrasts reflect substantively different gender effects in ESE (Sass, 2011; Rocabado *et al.*, 2020). In other words, the findings are interpreted mainly in terms of whether the overall arrangement of ESE relations appears broadly similar across groups, while recognising that stronger between-group claims would require stronger evidence of invariance.

To investigate gender differences in the structure of ESE, independent-group network comparison tests were conducted on the male and female networks. Permutation-based tests were used to examine (a) global strength invariance (*i.e.*, whether the sum of absolute edge weights differed between groups) and (b) network structure invariance (*i.e.*, whether the overall pattern of edge weights differed), mirroring procedures used in recent applications comparing symptom networks across gender (Zhang *et al.*, 2024). Statistical significance was evaluated at $\alpha = 0.05$, and differences in global strength and selected edge weights were reported to aid interpretation of any observed group differences (Hevey, 2018; Zhang *et al.*, 2024).

3. Results

3.1. Descriptive information of data

Table 2 presents the descriptive statistics for the 12 ESE items for the overall sample and by gender. Across all participants ($N = 655$), mean item scores ranged from 3.33 for PC1 (SD = 0.85) to 4.37 for LH3 (SD = 0.76), with standard deviations ranging from approximately 0.76 to 0.95, indicating generally moderate to high levels of ESE and comparable variability across items. Descriptively, items in the LH and CU clusters tended to show higher means, while items in the PC and SR clusters were slightly lower but still above the scale midpoint. Gender-stratified means followed a very similar pattern: female ($N = 476$) and male ($N = 179$) students showed broadly comparable levels of ESE, with absolute mean differences typically below 0.15 scale points. Males reported slightly higher scores on most items, particularly SR1 (3.45 *vs.* 3.31) and SR2 (3.48 *vs.* 3.33), whereas females showed marginally higher means on CU2 (4.27 *vs.* 4.17) and LH3 (4.40 *vs.* 4.28). Overall, the absence of extreme floor or ceiling effects and the similarity of item distributions across genders support the inclusion of all 12 items as nodes in the subsequent psychological network models of ESE (Hevey, 2018).

3.2. ESE psychological network analysis

3.2.1. Network edge structure and density. Fig. 1 shows the gender-specific ESE networks estimated from regularised partial correlations. Each network comprised 12 items, yielding 66 possible undirected edges based on the formula $n(n - 1)/2$. The female network retained 43 non-zero edges (density = 0.652;



Table 2 Mean and standard deviation scores for all ESE nodes (items)

	Overall (N = 655)	Female (N = 476)	Male (N = 179)
CU1	3.508 ± 0.820	3.481 ± 0.809	3.581 ± 0.847
CU2	4.241 ± 0.766	4.267 ± 0.761	4.173 ± 0.778
CU3	3.562 ± 0.814	3.534 ± 0.795	3.637 ± 0.859
LH1	3.921 ± 0.945	3.901 ± 0.936	3.972 ± 0.968
LH2	3.829 ± 0.901	3.813 ± 0.925	3.872 ± 0.835
LH3	4.366 ± 0.760	4.399 ± 0.760	4.279 ± 0.757
PC1	3.333 ± 0.847	3.309 ± 0.836	3.397 ± 0.877
PC2	3.440 ± 0.803	3.437 ± 0.772	3.447 ± 0.881
PC3	3.412 ± 0.845	3.403 ± 0.827	3.436 ± 0.893
SR1	3.350 ± 0.952	3.311 ± 0.945	3.453 ± 0.967
SR2	3.366 ± 0.871	3.326 ± 0.857	3.475 ± 0.901
SR3	3.337 ± 0.937	3.313 ± 0.916	3.402 ± 0.992

sparsity = 0.348), while the male network retained 42 non-zero edges (density = 0.636; sparsity = 0.364), indicating highly comparable global connectivity across gender, with only a marginally denser network among females. All retained edges were positive, suggesting that ESE components tend to co-vary in the same direction when controlling for the remaining items.

In terms of edge structure, both networks were characterised by stronger within-construct than between-construct connectivity, producing clear clustering across the four ESE domains (CU, LH, PC, and SR). Supporting this pattern, the female network retained all 12 within-construct edges, whereas the male network retained 11 of 12 (with one within-PC edge shrunk to zero, *i.e.*, between PC1 and PC3). Moreover, within-construct edges were substantially larger on average than cross-construct edges (female: $M_{\text{within}} \approx 0.278$ vs. $M_{\text{between}} \approx 0.068$; male: $M_{\text{within}} \approx 0.291$ vs. $M_{\text{between}} \approx 0.070$), indicating a modular organisation with strong internal coherence and comparatively weaker bridging ties. Practically, this means that students' confidence beliefs tended to reinforce one another most strongly within the same ESE domain, whereas cross-domain links were present but comparatively limited. Although the female network was marginally denser overall, this should not be interpreted as materially greater between-domain connectivity, because cross-domain edge weights were very similar

across gender and the later network comparison test indicated no significant difference in global structure.

3.2.2. Key connections between ESE items. In practical terms, the edge weights reported in the current study are regularised partial correlations, meaning that they reflect the unique association between two confidence judgements after the remaining ESE items in the network have been considered (Hevey, 2018; Epskamp *et al.*, 2018). Thus, stronger edges indicate item pairs that remain closely linked even after the rest of the ESE system has been considered, whereas weaker or absent edges suggest either looser coupling or relations that may be expressed more indirectly through neighbouring nodes (Hevey, 2018; Epskamp *et al.*, 2018). The following results, therefore, focus on the strongest and most interpretable connections in each gender network and consider how these links fit into the broader ESE structure.

In the female network (see SI, Appendix 1: Table A), the strongest and most interpretable retained edges were concentrated within the same ESE domains, particularly laboratory hazards and sufficiency of resources. Specifically, the strongest conditional associations were observed among laboratory hazards items - LH2-LH3 (0.408) and LH1-LH2 (0.407) - followed closely by strong cohesion within sufficiency of resources (SR1-SR2 = 0.404; SR2-SR3 = 0.373). Within-domain connectivity was also evident for conceptual understanding (CU1-CU3 = 0.348; CU2-CU3 = 0.315) and procedural complexity (PC2-PC3 = 0.318; PC1-PC2 = 0.260). Substantively, this pattern suggests that safety-related judgements such as avoiding spillage, handling apparatus safely, and remaining alert in the laboratory tended to cohere closely, as did confidence judgements related to coping with resource constraints. Likewise, the conceptual items clustered around sense-making during experiments, indicating that confidence in understanding theory, gaining clarity during experimentation, and understanding chemical processes tended to reinforce one another.

A comparable pattern emerged in the male network, as presented in Fig. 1 and Appendix 2: Table B (see SI), although the single strongest retained edge occurred within procedural

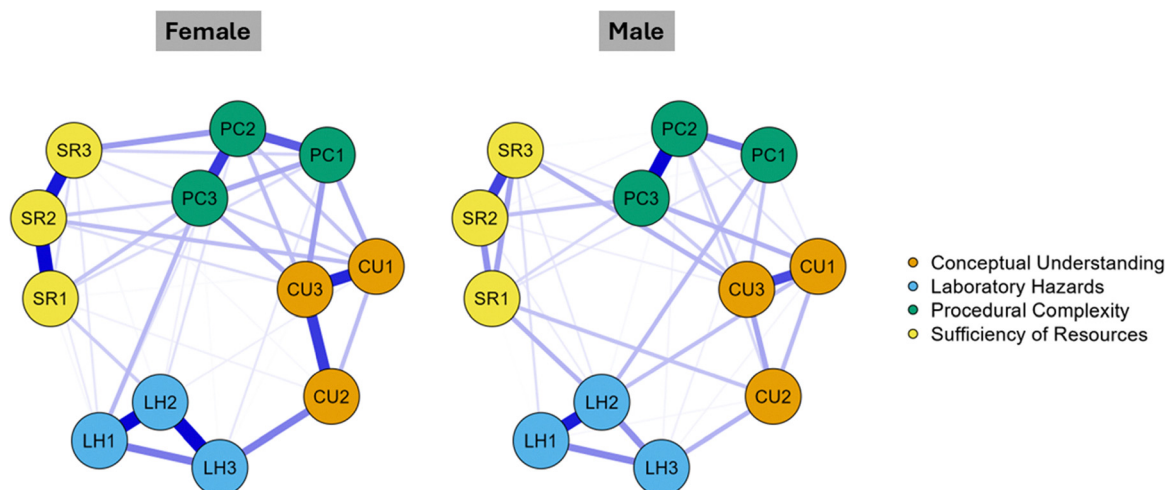


Fig. 1 Gender-based comparison of network edge structure in Malaysian pre-university chemistry students.



complexity, where PC2–PC3 (0.522) was the largest connection in the entire network. Strong within-construct links were also observed for laboratory hazards (LH1–LH2 = 0.464; LH1–LH3 = 0.243; LH2–LH3 = 0.223), indicating that confidence in safe handling, avoiding spillage, and remaining alert tended to form a tightly coupled safety-technique cluster. Sufficiency of resources also showed coherent within-domain clustering (SR2–SR3 = 0.386; SR1–SR2 = 0.225; SR1–SR3 = 0.188), suggesting that confidence under limited participation, limited instruments, and one-try-only conditions was experienced as a connected set of resource-related judgements. Conceptual understanding remained internally cohesive as well, led by CU1–CU3 (0.368) and CU2–CU3 (0.180), indicating that confidence in understanding theory, gaining conceptual clarity during experimentation, and understanding chemical processes was organised as a sense-making cluster. The procedural complexity pattern is especially informative because PC1 and PC3 were both linked to PC2, but not directly to each other. This suggests that these two forms of procedural confidence may converge on a shared interpretive core rather than forming a single undifferentiated procedural block. Specifically, PC1 concerns understanding how calculation procedures and errors affected results, whereas PC3 concerns relating background information to one's own procedures and results; both judgements plausibly converge on PC2, which focuses on interpreting laboratory results and making correct conclusions. Once this shared interpretive component is considered, the direct PC1–PC3 relation may no longer remain. In this sense, PC2 appears to function as a local interpretive hub within the procedural complexity domain.

Beyond these within-construct clusters, a smaller number of cross-construct “bridge” edges helped connect the ESE domains (Fig. 1). In the female network, the largest cross-construct connection was CU2–LH3 (0.204), followed by bridges linking procedural complexity with resources (PC2–SR3 = 0.160) and conceptual understanding with procedural complexity (CU3–PC1 = 0.153; CU1–PC1 = 0.135; CU3–PC3 = 0.126). In the male network, the most prominent bridges included CU1–PC3 (0.155), CU2–LH3 (0.154), CU3–SR3 (0.152), and links involving hazards with resources/procedures (LH2–SR1 = 0.151; LH2–PC1 = 0.139). Although these cross-domain edges were smaller than the strongest within-domain edges, they remain important because they indicate where confidence in one domain may connect to confidence in another. Overall, these patterns show that while both gender networks are primarily organised around strong within-domain coherence, the retained cross-domain edges consistently connect conceptual understanding with the other ESE components, particularly procedural complexity and laboratory hazards.

3.2.3. Cluster structure and bridge nodes across ESE domains.

To evaluate whether the estimated networks recover the hypothesised four-domain organisation (CU, LH, PC, SR), community detection using the Louvain algorithm was performed on each gender network. In both the female and male networks, the detected communities matched the four intended ESE domains exactly, indicating clear domain clustering. Consistent with this, modularity values were moderate-to-high (female $Q = 0.364$;

male $Q = 0.345$), supporting a modular organisation characterised by strong within-domain coherence and comparatively weaker cross-domain links.

Bridge expected influence (BEI) was then used to quantify which items most strongly connect domains. As shown in Table 3, in the female network, bridge connectivity was most prominent for procedural complexity items, with PC3 (BEI = 0.590) emerging as the strongest bridge node, followed by PC1 (0.456) and PC2 (0.448). These bridge effects reflected cross-domain links connecting procedural processing with conceptual understanding and constraint-related confidence (e.g., PC3–CU3 = 0.126; PC3–SR1 = 0.095; PC3–SR2 = 0.092; PC3–LH1 = 0.112). In the male network (Table 3), bridge connectivity was most prominent for CU3 (0.521) and LH2 (0.514), followed by CU1 (0.481) and PC3 (0.471). Notably, LH2 showed cross-domain links with safety/resource and procedural nodes (e.g., LH2–SR1 = 0.151; LH2–PC1 = 0.139; LH2–CU1 = 0.130), whereas CU3 primarily linked conceptual understanding with procedural and resource-related confidence (e.g., CU3–SR3 = 0.152; CU3–PC1 = 0.117; CU3–PC2 = 0.114). These bridge patterns help explain why some item pairs are not directly connected despite belonging to the same broader ESE system: confidence judgements may relate to one another indirectly through local hubs or cross-domain connectors rather than through a single direct edge.

Because the network comparison test indicated no statistically significant gender differences in global structure or strength, these bridge patterns are interpreted descriptively as potential connective “routes” within each gender network rather than as evidence of robust between-group differences.

3.2.4. Centrality measure. Centrality indices for the female and male networks are presented in Fig. 2 and Appendix 3: Table C (see SI). Centrality values are presented as standardised scores (mean = 0); therefore, negative values indicate below-average connectivity relative to other nodes, not “negative importance”. Given that the estimated networks contained only positive edges, expected influence followed the same pattern as strength; therefore, the following interpretations focus on strength as the primary index of node importance. In practical terms, a more central node may be understood as a confidence belief that is more tightly linked to multiple other parts of the ESE system, and therefore may represent a more strategically connected target for instructional support.

In the female network, CU3 was the most central item, exhibiting the highest strength ($S = 1.430$), alongside comparatively high betweenness ($B = 1.992$) and closeness ($C = 1.614$). The next highest-strength nodes were SR2 ($S = 1.105$), PC3 ($S = 0.971$), and

Table 3 Bridge expected influence by gender

Rank	Female node	BEI	Male node	BEI
1	PC3	0.590	CU3	0.521
2	PC1	0.456	LH2	0.514
3	PC2	0.448	CU1	0.481
4	CU1	0.446	PC3	0.471
5	CU3	0.444	SR1	0.412

Higher = more cross-domain connectivity.



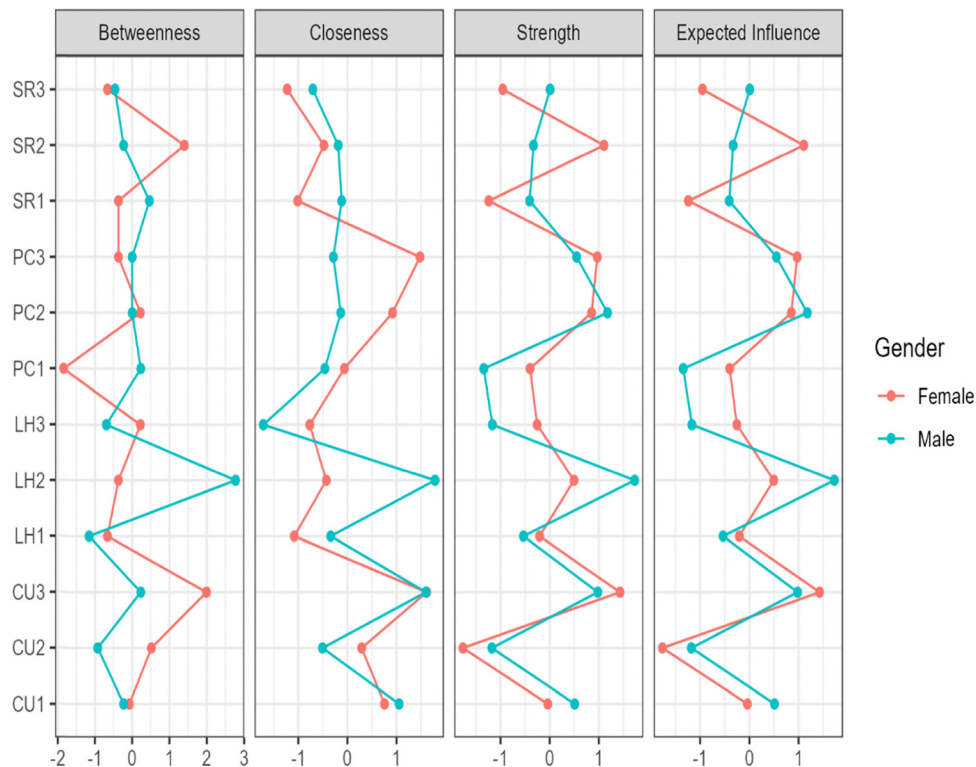


Fig. 2 Comparison of node centrality in female and male ESE networks.

PC2 ($S = 0.852$), indicating that these items contributed most strongly to overall connectivity. Conversely, CU2 showed the lowest strength ($S = -1.764$), followed by SR1 ($S = -1.237$) and SR3 ($S = -0.953$).

In the male network, LH2 emerged as the strongest hub, with the highest strength ($S = 1.729$), and it also showed the highest betweenness ($B = 2.773$) and closeness ($C = 1.784$). Other high-strength nodes were PC2 ($S = 1.177$) and CU3 ($S = 0.980$), followed by PC3 ($S = 0.550$) and CU1 ($S = 0.511$). The lowest strength values were observed for PC1 ($S = -1.340$), CU2 ($S = -1.177$), and LH3 ($S = -1.165$).

Descriptively, CU3 appeared among the more central nodes in the female network, whereas LH2 appeared among the more central nodes in the male network; however, these differences were not statistically supported in the network comparison test and are therefore interpreted as exploratory patterns. Overall, centrality patterns were broadly similar across gender, with CU3 and LH2 emerging as prominent nodes in the female and male networks, respectively; however, group differences in centrality were not statistically significant in the network comparison test.

3.2.5. Network comparison test. Independent-group Gaussian network comparison tests were first used to assess whether the ESE networks were invariant across gender. The analysis provided no evidence of differences in the overall network structure (edge weights) between male and female students ($M = 0.118$, $P = 0.851$). Likewise, the test of global strength invariance showed no significant difference in overall connectivity, with network strength values of 3.855 for males and 3.684 for females, indicating comparable levels of network connectivity

across the two groups ($S = 0.171$, $P = 0.699$). Additional permutation tests showed that neither the strength of individual edges nor the centrality indices differed significantly between female and male students. No edge displayed a P -value below 0.05, and all comparisons of centrality strength were likewise non-significant (all $P > 0.05$). Overall, these results point to a lack of substantial gender differences in the network configuration and overall connectivity of ESE items among Malaysian pre-university chemistry students.

3.2.6. Network robustness test. Network robustness was evaluated by estimating (a) edge-weight accuracy using non-parametric bootstrapped 95% confidence intervals (CIs), and (b) centrality stability using a case-dropping bootstrap procedure. For both female and male networks (see SI, Appendix 4: Fig. A and Appendix 5: Fig. B, respectively), bootstrapped CIs indicated that the strongest edges were estimated with greater precision, whereas many smaller edges showed wide and overlapping CIs, limiting fine-grained comparisons among weaker connections. Case-dropping results suggested that strength centrality was the most robust index (particularly in the female network), while closeness and betweenness displayed substantially lower stability and therefore were not emphasised in interpretation. Given known stability limitations of closeness and betweenness in estimated psychological networks, interpretation focused on strength (and expected influence where relevant), supported by bootstrap accuracy and case-dropping stability analyses. Accordingly, the practical interpretations in the present study focus primarily on the stronger and more stable edges, whereas finer distinctions among weaker connections are treated more cautiously.



4. Discussion

In both gender groups, the ESE networks showed a coherent, domain-aligned structure, with stronger within-domain connections (among items within conceptual understanding, laboratory hazards, procedural complexity, and sufficiency of resources) and fewer, smaller cross-domain bridges. All retained edges were positive, indicating that higher confidence in one practical-work capability tended to co-occur with higher confidence in related capabilities after conditioning on other items. This pattern is consistent with how psychometric networks typically represent mutually reinforcing belief systems in educational contexts (Hevey, 2018; Borsboom *et al.*, 2021).

A key contribution of the network approach in this study is not simply that items cluster into the four expected domains – an outcome already consistent with prior evidence supporting a four-factor structure – but that the network quantifies which specific items connect domains and therefore may represent plausible leverage points for broadening students' ESE. Cluster detection recovered the four intended ESE domains exactly in both gender networks (female $Q = 0.364$; male $Q = 0.345$), indicating coherent domain organisation. However, BEI highlighted a small subset of items with disproportionately high cross-domain connectivity, providing hypotheses for future intervention studies for instruction than domain-level scores alone. Against this structural backdrop, node centrality was examined to identify the most interconnected beliefs within each gender network.

Centrality estimates (strength) indicated that CU3 (“I am confident that I understand the chemical processes in the experiment”) was the most central node in the female network, whereas LH2 (“I am confident of working in the laboratory without chemical spillage”) was the most central node in the male network. Descriptively, this suggests that conceptual process understanding may function as a key organising belief among females, while spillage-related safety/technique confidence may be a salient hub among males. However, these contrasts should be interpreted as exploratory because centrality is sensitive to sampling variability and should not be over-interpreted without robust stability evidence (Epskamp *et al.*, 2018).

Permutation-based network comparison provided no evidence of gender differences in overall network structure or global strength, indicating that the conditional association framework of ESE appears broadly similar across male and female students in this Malaysian pre-university sample. Given the borderline evidence for full metric invariance, this result is interpreted as no evidence of network differences under the current operationalisation, rather than as definitive evidence that the underlying construct is fully invariant across gender. This framing reinforces the value of formal network comparison and measurement evaluation, rather than relying on visual inspection of apparent network differences (van Borkulo *et al.*, 2023).

The dominance of within-domain edges across both networks suggests that students hold tightly coupled confidence beliefs within each domain – understanding chemical processes (CU), managing hazards (LH), navigating procedural demands (PC), and coping with resource constraints (SR). This

pattern supports the conceptualisation of ESE as task- and context-specific judgements about capability in authentic laboratory conditions, consistent with self-efficacy theory (Bandura, 1986, 1997). From a network psychometrics standpoint, such clustering is expected when items reflect strongly related components of a broader competence perception (Hevey, 2018; Borsboom *et al.*, 2021).

Although fewer and generally smaller, cross-domain connections were theoretically meaningful and were concentrated in a limited set of bridge items. In the female network, BEI was highest for PC3 (0.590), followed by PC1 (0.456) and PC2 (0.448), indicating that procedural complexity beliefs – especially confidence in connecting preparatory information to procedures and outcomes – were most strongly positioned to link domains (*e.g.*, PC3–CU3; PC3–LH1; PC3–SR1/SR2). In the male network, BEI was highest for CU3 (0.521) and LH2 (0.514), suggesting that conceptual process understanding and spillage-related technique/safety confidence were the most prominent cross-domain connectors (*e.g.*, CU3–SR3; CU3–PC1/PC2; LH2–SR1; LH2–PC1; LH2–CU1). Taken together, these bridge patterns reinforce the practical reality of chemistry laboratory work: students must align conceptual reasoning with procedural execution while managing hazards and constraints (Villafañe *et al.*, 2016; Avargil, 2019).

The emergence of CU3 (female) and LH2 (male) as prominent hubs is consistent with the idea that different belief elements can “anchor” a student's broader confidence system. However, centrality should be interpreted cautiously because (i) it does not imply causality, and (ii) some centrality indices (especially closeness and betweenness) are less stable in estimated psychological networks, motivating the emphasis on strength and expected influence supported by bootstrap accuracy and case-dropping procedures (Epskamp *et al.*, 2018; Hevey, 2018). Accordingly, these node-level contrasts are best treated as hypothesis-generating.

In addition to CU3, SR2 also appeared relatively central in the female network. Descriptively, this suggests that confidence in understanding chemistry under resource-constrained laboratory conditions may have been more strongly integrated into the broader ESE system for female students. In practical terms, this may indicate that for female students, resource-related confidence was tied not only to perceived adequacy of materials, but also to their broader sense of conceptual functioning in the laboratory. However, this pattern should be interpreted cautiously because the network comparison test did not indicate statistically significant gender differences in centrality.

The present results align with the wider literature showing that gender differences in ESE are not consistently observed and may depend on population, learning design, and how laboratory confidence is conceptualised and measured (Winkelmann *et al.*, 2015; Wahyudiati *et al.*, 2020). At the same time, studies have reported gender differences under particular conditions – for example, when laboratory confidence is examined alongside affective variables such as laboratory anxiety, or in specific teacher-education contexts (Kırbaşlar *et al.*, 2015). The current findings, therefore, extend prior work in two ways: (i) they show small gender differences at the mean level, and (ii) they demonstrate that



gender similarity also appears at the level of belief architecture, with comparable within-domain organisation and bridging patterns, consistent with the modularity and BEI results reported above (Borsboom *et al.*, 2021; van Borkulo *et al.*, 2023).

The lack of detectable gender differences in network structure or global strength may reflect several plausible mechanisms:

(a) Shared laboratory ecology and standardised expectations.

In a high-stakes pre-university setting, students often experience similar laboratory routines, assessment demands, and structured practical formats, which may reduce the formation of divergent confidence pathways by gender (Abdullah *et al.*, 2022; Matriculation Division, 2022a, 2022b).

(b) Task-specific measurement attenuates broad gender generalisations. Because ESE is measured as confidence in clearly defined practical-work demands (*e.g.*, understanding processes, avoiding spillage, interpreting results under constraints), gender differences that sometimes appear in broader “science confidence” measures may be less pronounced here (Bandura, 1997; Villafañe *et al.*, 2016).

(c) Borderline metric invariance encourages cautious inference. Although multi-group CFA supported configural invariance and showed borderline evidence for metric invariance, this reinforces the interpretation of gender comparisons primarily as overall patterning rather than as strong claims about measurement functioning differences.

Within social cognitive theory, self-efficacy refers to judgements of capability to organise and execute actions needed to attain goals, and these judgements are shaped through learning experiences in specific contexts (Bandura, 1986, 1997). The current network findings complement this perspective by illustrating that ESE is not merely “high or low,” but a system of mutually connected capability beliefs across conceptual, procedural, safety, and constraint-related facets of laboratory work. In network terms, clusters reflect tightly linked belief components, while bridging edges suggest how certain beliefs (notably conceptual understanding) may coordinate confidence across domains (Hevey, 2018; Borsboom *et al.*, 2021). This provides a theoretically grounded basis for thinking about laboratory support: targeted strengthening of strategically connected beliefs may help reinforce the broader ESE system.

The present findings suggest several tentative pedagogical considerations for chemistry laboratory teaching and student support; however, these should be interpreted cautiously because the current network analysis is cross-sectional and does not directly test instructional effects:

(a) Support conceptual sense-making during experiments. The recurrent cross-domain involvement of conceptual-understanding items suggests that helping students connect observations, procedures, and underlying chemical processes may be a reasonable focus for laboratory support (Villafañe *et al.*, 2016; Avargil, 2019). For instance, prompts that explicitly link procedure → observation → chemical process → conclusion may be worth examining in future instructional designs, although the present findings do not establish causal effects.

(b) Treat spillage-related confidence as an exploratory area for targeted support. The prominence of LH2 in the male

network suggests that confidence related to avoiding chemical spillage may be one potentially connected aspect of laboratory self-efficacy (Townes *et al.*, 2015; Kolil *et al.*, 2020). However, because node-level contrasts were not statistically significant in the network comparison test, this pattern is best treated as exploratory rather than as evidence of a confirmed instructional entry point.

(c) Examine interpretation support under constrained conditions. Bridge connections involving procedural complexity and resource-related confidence suggest that students’ confidence in interpreting laboratory work may be linked to how they experience limited time, limited equipment, or limited opportunities to repeat experiments. Accordingly, scaffolds that support interpretation and planning under such constraints may warrant further examination in future intervention studies (Townes *et al.*, 2015; Kolil *et al.*, 2020).

Several limitations should frame interpretation. First, the networks are cross-sectional; edges represent conditional associations and do not imply causal direction (Hevey, 2018; Borsboom *et al.*, 2021). Second, small edges showed wide and overlapping confidence intervals, so fine-grained ranking of weaker connections should be avoided (Epskamp *et al.*, 2018). Third, the male subsample was smaller, which may reduce sensitivity for subtle group differences. Finally, because multi-group CFA showed borderline evidence for full metric invariance, gender comparisons should be interpreted cautiously as reflecting group patterning under this operationalisation; replication with explicit partial invariance testing and/or item-level non-invariance modelling would strengthen inference. Future research could (i) use longitudinal or intervention designs to examine whether strengthening conceptual sense-making changes downstream procedural/safety/resource confidence, (ii) compare networks across institutions with different laboratory infrastructures, and (iii) test moderators more proximal than gender (*e.g.*, prior laboratory exposure, laboratory anxiety, or perceived instructional support), consistent with prior evidence that gender effects can become clearer in affect-linked measurement contexts (Kırbaşlar *et al.*, 2015; Sezginürk and Sungur, 2020).

5. Conclusions

This study set out to extend understanding of ESE in chemistry by moving beyond mean-level gender comparisons and examining how ESE beliefs are organised and connected as a system. Using psychometric network analysis, ESE was modelled as a set of interrelated confidence judgements spanning conceptual understanding, procedural complexity, laboratory hazards, and sufficiency of resources, and the invariance of this belief framework across gender was tested within a Malaysian pre-university chemistry context.

Across both gender groups, the estimated networks showed a coherent, domain-aligned structure: the most prominent connections occurred within each ESE domain, while fewer and smaller cross-domain links connected domains. In practical terms, this pattern suggests that students’ confidence in



laboratory work is not a single undifferentiated belief, but a clustered system where capability judgements cohere most strongly around related demands (e.g., procedural demands with procedural demands; hazard-related demands with hazard-related demands). At the same time, the networks retained several interpretable bridge connections, with conceptual understanding – especially confidence in explaining chemical processes – appearing repeatedly as a linking component that connects to procedural and safety-related confidence, indicating a plausible coordinating role for sense-making within the overall ESE system.

At the node level, strength centrality suggested that CU3 (“I am confident that I understand the chemical processes in the experiment”) was most central in the female network, whereas LH2 (“I am confident of working in the laboratory without chemical spillage”) was most central in the male network. However, these contrasts should be treated as exploratory rather than definitive gender differences. The network comparison tests indicated no evidence of gender differences in overall network structure or global strength, and neither individual edges nor centrality indices differed significantly between groups. Moreover, robustness analyses reinforced a cautious interpretive stance: stronger edges were estimated more precisely, whereas many smaller edges had wide and overlapping confidence intervals, and closeness/betweenness showed lower stability, motivating an interpretation that prioritises strength (and expected influence where relevant).

Collectively, these findings contribute to knowledge in three main ways. First, they show that for Malaysian pre-university chemistry students, gender similarity is evident not only in overall patterns but also in the architecture of ESE – the way confidence components cohere within domains and connect across domains – providing the type of insight that network psychometrics is designed to offer. Second, the findings foreground conceptual sense-making as a potentially strategic component for supporting broader laboratory confidence, given its recurring bridging role across domains. Third, the results offer a practical reframing of ESE: rather than treating ESE as simply “high” or “low,” the network perspective supports viewing chemistry practical-work self-efficacy as a mutually reinforcing system of capability beliefs, suggesting that targeted support aimed at strategically connected beliefs may have wider benefits across the ESE network.

The study also suggests a plausible area for support for chemistry laboratory teaching and student support. Instructional designs that prioritise linking observation to explanation may strengthen confidence beyond conceptual understanding alone, particularly when prompts explicitly connect procedure → observation → chemical process → conclusion. In addition, the prominence of spillage-related confidence (LH2) in the male network suggests that technique-focused supports – such as structured practice for safe transfer/pouring and feedback on handling – may function as an accessible “micro-skill” entry point for strengthening broader laboratory confidence. Finally, bridges involving procedural complexity and resources indicate that confidence in interpreting results may be intertwined with perceived constraints; therefore, scaffolds for interpretation and

planning routines may be especially valuable in contexts where time, instruments, or opportunities to repeat experiments are limited.

In conclusion, this study provides evidence that ESE among Malaysian pre-university chemistry students is organised as a structured, interconnected system of laboratory confidence beliefs, and that this system appears broadly similar across gender in both overall connectivity and structure. These patterns may inform future longitudinal or intervention studies by identifying confidence components that appear structurally connected within the ESE network; however, they should not be interpreted as direct evidence that targeting these components would causally strengthen broader laboratory confidence.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data that support the findings of this study are not publicly available due to restrictions set by the participating institution.

Supplementary information (SI) contains additional gender-specific Experimental Self-Efficacy (ESE) network analysis outputs, including edge-weight matrices for the female and male networks, item-level centrality indices for both networks, and robustness plots showing bootstrapped edge-weight confidence intervals and centrality stability curves. See DOI: <https://doi.org/10.1039/d6rp00083e>.

Ethical considerations

Informed consent was obtained from all participants. This study received ethical approval from the participating institution, which operates under the jurisdiction of the Ministry of Education, under reference number KML.100-7/5/1 Jld.3(20).

Acknowledgements

This research received no external funding. Sincere thanks are extended to the pre-university chemistry students who participated in this study. Support from the relevant institutional and administrative parties in facilitating recruitment, data collection, and study coordination is gratefully acknowledged. Appreciation is also extended to colleagues who assisted with study logistics and to the developers of the open-source statistical software used in the network analyses.

References

- Abdullah, Z., Yahya, R., Ismail, N., Wan Aziz, W. Z. A., Kadum, B. M. C. M., Ahmad, Z., Musa, F. N. H., Ismail, F., Ali, S., Zulkeply, N. F., Wan Salim, W. S., Rosali, N. A., Yang Abd Talib, A. F. and Ahmad Badri, A. I. (ed.), (2022), *Chemistry laboratory manual semester I and II SK015 and SK025*



- thirteenth edition*, Matriculation Division Ministry of Education Malaysia.
- Alkan, F., (2016), Development of chemistry laboratory self-efficacy beliefs scale, *J. Balt. Sci. Educ.*, **15**(3), 350–359, <https://www.cceol.com/search/article-detail?id=974623>.
- Avargil, S., (2019), Learning chemistry: self-efficacy, chemical understanding, and graphing skills, *J. Sci. Educ. Technol.*, **28**(4), 285–298, DOI: [10.1007/s10956-018-9765-x](https://doi.org/10.1007/s10956-018-9765-x).
- Bandura, A., (1986), *Social foundations of thought and action: a social cognitive theory*, Prentice-Hall.
- Bandura, A., (1997), *Self-efficacy: the exercise of control*, Freeman.
- Blondel, V. D., Guillaume, J. L., Lambiotte, R. and Lefebvre, E., (2008), Fast unfolding of communities in large networks, *J. Stat. Mech.: Theory Exp.*, **2008**(10), P10008, DOI: [10.1088/1742-5468/2008/10/P10008](https://doi.org/10.1088/1742-5468/2008/10/P10008).
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R. and Young, S. L., (2018), Best practices for developing and validating scales for health, social, and behavioral research: a primer, *Front. Public Health*, **6**, 149, DOI: [10.3389/fpubh.2018.00149](https://doi.org/10.3389/fpubh.2018.00149).
- Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., Robinaugh, D. J., Perugini, M., Dalege, J., Costantini, G., Isvoranu, A.-M., Wysocki, A. C., van Borkulo, C. D., van Bork, R. and Waldorp, L. J., (2021), Network analysis of multivariate data in psychological science, *Nat. Rev. Methods Primers*, **1**(1), 58, DOI: [10.1038/s43586-021-00055-w](https://doi.org/10.1038/s43586-021-00055-w).
- Clark, L. A. and Watson, D., (1995), Constructing validity: basic issues in objective scale development, *Psychol. Assess.*, **7**(3), 309–319, <https://psycnet.apa.org/doi/10.1037/1040-3590.7.3.309>.
- Dalgety, J. and Coll, R. K., (2006), Exploring first-year science students' chemistry self-efficacy, *Int. J. Sci. Math. Educ.*, **4**(1), 97–116, DOI: [10.1007/s10763-005-1080-3](https://doi.org/10.1007/s10763-005-1080-3).
- Diaz-Milanes, D., Salado, V., Santín Vilariño, C., Andrés-Villas, M. and Pérez-Moreno, P. J., (2024), A network analysis study on the structure and gender invariance of the satisfaction with life scale among Spanish university students, *Healthcare*, **12**(2), 237, DOI: [10.3390/healthcare12020237](https://doi.org/10.3390/healthcare12020237).
- Dunn, T. J., Baguley, T. and Brunsden, V., (2014), From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation, *Br. J. Psychol.*, **105**(3), 399–412, DOI: [10.1111/bjop.12046](https://doi.org/10.1111/bjop.12046).
- Epskamp, S., Kruis, J. and Marsman, M., (2017), Estimating psychopathological networks: be careful what you wish for, *PLoS One*, **12**(6), e0179891, DOI: [10.1371/journal.pone.0179891](https://doi.org/10.1371/journal.pone.0179891).
- Epskamp, S., Borsboom, D. and Fried, E. I., (2018), Estimating psychological networks and their accuracy: a tutorial paper, *Behav. Res. Methods*, **50**(1), 195–212, DOI: [10.3758/s13428-017-0862-1](https://doi.org/10.3758/s13428-017-0862-1).
- Galloway K. R. and Bretz S. L., (2015), Measuring meaningful learning in the undergraduate chemistry laboratory: a national, cross-sectional study, *J. Chem. Educ.*, **92**(12), 2006–2018, DOI: [10.1021/acs.jchemed.5b00538](https://doi.org/10.1021/acs.jchemed.5b00538).
- Gregorich, S. E., (2006), Do self-report instruments allow meaningful comparisons across diverse population groups?: testing measurement invariance using the confirmatory factor analysis framework, *Med. Care*, **44**(11), S78–S94, DOI: [10.1097/01.mlr.0000245454.12228.8f](https://doi.org/10.1097/01.mlr.0000245454.12228.8f).
- Hevey, D., (2018), Network analysis: a brief overview and tutorial, *Health Psychol. Behav. Med.*, **6**(1), 301–328, DOI: [10.1080/21642850.2018.1521283](https://doi.org/10.1080/21642850.2018.1521283).
- Hu, L. T. and Bentler, P. M., (1999), Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives, *Struct. Equ. Model.: Multidiscip. J.*, **6**(1), 1–55, DOI: [10.1080/10705519909540118](https://doi.org/10.1080/10705519909540118).
- Huth, K. B., de Ron, J., Goudriaan, A. E., Luigjes, J., Mohammedi, R., van Holst, R. J., Wagenmakers, E. and Marsman, M., (2023), Bayesian analysis of cross-sectional networks: a tutorial in R and JASP, *Adv. Methods Pract. Psychol. Sci.*, **6**(4), 25152459231193334, DOI: [10.1177/25152459231193334](https://doi.org/10.1177/25152459231193334).
- Ji, F., Sun, H., Barnhart, W. R., Cui, T., Cui, S., Zhang, J. and He, J., (2024), Psychometric network analysis of the Intuitive Eating Scale-2 in Chinese general adults, *J. Clin. Psychol.*, **80**(5), 1098–1114, DOI: [10.1002/jclp.23657](https://doi.org/10.1002/jclp.23657).
- Johnstone, A. H., (1991), Why is science difficult to learn? Things are seldom what they seem, *J. Comput. Assist. Learn.*, **7**(2), 75–83, DOI: [10.1111/j.1365-2729.1991.tb00230.x](https://doi.org/10.1111/j.1365-2729.1991.tb00230.x).
- Jones, P. J., Ma, R. and McNally, R. J., (2021), Bridge centrality: a network approach to understanding comorbidity, *Multivar. Behav. Res.*, **56**(2), 353–367, DOI: [10.1080/00273171.2019.1614898](https://doi.org/10.1080/00273171.2019.1614898).
- Kiran, D. and Sungur, S., (2012), Middle school students' science self-efficacy and its sources: examination of gender difference, *J. Sci. Educ. Technol.*, **21**(5), 619–630, DOI: [10.1007/s10956-011-9351-y](https://doi.org/10.1007/s10956-011-9351-y).
- Kırbaşlar, F. G., Veyisoğlu, A. and Özsoy-Güneş, Z., (2015), Investigating the relationships between pre-service science teachers' self-efficacy in laboratory and anxiety towards chemistry laboratory, *Procedia - Soc. Behav. Sci.*, **174**, 43–50, DOI: [10.1016/j.sbspro.2015.01.624](https://doi.org/10.1016/j.sbspro.2015.01.624).
- Knekta, E., Runyon, C. and Eddy, S., (2019), One size doesn't fit all: using factor analysis to gather validity evidence when using surveys in your research, *CBE—Life Sci. Educ.*, **18**(1), rm1, DOI: [10.1187/cbe.18-04-0064](https://doi.org/10.1187/cbe.18-04-0064).
- Kolil, V. K., Muthupalani, S. and Achuthan, K., (2020), Virtual experimental platforms in chemistry laboratory education and its impact on experimental self-efficacy, *Int. J. Educ. Technol. High. Educ.*, **17**(1), 30, DOI: [10.1186/s41239-020-00204-3](https://doi.org/10.1186/s41239-020-00204-3).
- Kolil, V. K., Parvathy, S. U. and Achuthan, K., (2023), Confirmatory and validation studies on experimental self-efficacy scale with applications to multiple scientific disciplines, *Front. Psychol.*, **14**, 1154310, DOI: [10.3389/fpsyg.2023.1154310](https://doi.org/10.3389/fpsyg.2023.1154310).
- Kurbanoglu N. İ. and Akin A., (2010), The relationships between university students' chemistry laboratory anxiety, attitudes, and self-efficacy beliefs, *Aust. J. Teach. Educ.*, **35**(8), 48–59, DOI: [10.14221/ajte.2010v35n8.4](https://doi.org/10.14221/ajte.2010v35n8.4).
- Lewis, S. E., (2022), Considerations on validity for studies using quantitative data in chemistry education research and practice, *Chem. Educ. Res. Pract.*, **23**(4), 764–767, DOI: [10.1039/D2RP90009B](https://doi.org/10.1039/D2RP90009B).
- Matriculation Division, (2022a), Curriculum specifications chemistry (SK015) [Fact sheet], Malaysia Ministry of Education.



- Matriculation Division, (2022b), Curriculum specifications chemistry (SK025) [Fact sheet], Malaysia Ministry of Education.
- McNeish, D., (2018), Thanks coefficient alpha, we'll take it from here, *Psychol. Methods*, **23**(3), 412, DOI: [10.1037/met0000144](https://doi.org/10.1037/met0000144).
- Merve Mustafaoglu, F. and Alkan, F., (2024), An analysis of high school students' laboratory self-efficacy beliefs and perceptions of laboratory applications, *Periódico Tchê Química*, **21**(48), 16–31, DOI: [10.52571/ptq.v21.n48.2024_02_fatma_pgs_16_48.pdf](https://doi.org/10.52571/ptq.v21.n48.2024_02_fatma_pgs_16_48.pdf).
- Moreno, C., Pham, D. and Ye, L., (2021), Chemistry self-efficacy in lower-division chemistry courses: changes after a semester of instruction and gaps still remain between student groups, *Chem. Educ. Res. Pract.*, **22**(3), 772–785, DOI: [10.1039/D0RP00345J](https://doi.org/10.1039/D0RP00345J).
- Naibert, N., Mooring, S. R. and Barbera, J., (2024), Investigating the relations between students' chemistry mindset, self-efficacy, and goal orientation in general and organic chemistry lecture courses, *J. Chem. Educ.*, **101**(2), 270–282, DOI: [10.1021/acs.jchemed.3c00929](https://doi.org/10.1021/acs.jchemed.3c00929).
- Nunnally, J. C. and Bernstein, I. H., (1994), *Psychometric theory*, 3rd edn, McGraw-Hill.
- Rocabado, G. A., Komperda, R., Lewis, J. E. and Barbera, J., (2020), Addressing diversity and inclusion through group comparisons: a primer on measurement invariance testing, *Chem. Educ. Res. Pract.*, **21**(3), 969–988, DOI: [10.1039/D0RP00025F](https://doi.org/10.1039/D0RP00025F).
- Sailaubay, A., Myrzakhmetova, N., Tunçel, M. and Kishibayev, K., (2024), Analysis of the relationships between university students' laboratory self-efficacy beliefs, science process skills and achievement in chemistry courses, *J. Curric. Stud. Res.*, **6**(2), 36–51, DOI: [10.46303/jcsr.2024.1](https://doi.org/10.46303/jcsr.2024.1).
- Saputra, A., Tania, L., Sunyono, S., Ibrahim, N. H. and Surif, J., (2024), A confirmatory and multigroup invariance analysis of the Indonesian version of the High School Chemistry Self-Efficacy Scale: gender and grade level overview, *J. Chem. Educ.*, **101**(8), 3013–3026, DOI: [10.1021/acs.jchemed.3c01332](https://doi.org/10.1021/acs.jchemed.3c01332).
- Sass, D. A., (2011), Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework, *J. Psychoeduc. Assess.*, **29**(4), 347–363, DOI: [10.1177/0734282911406661](https://doi.org/10.1177/0734282911406661).
- Sezgintürk, M. and Sungur, S., (2020), A multidimensional investigation of students' science self-efficacy: the role of gender, *İlköğretim Online-Elementary Education Online*, **19**(1), 208–218, DOI: [10.17051/ilkonline.2020.653660](https://doi.org/10.17051/ilkonline.2020.653660).
- Towns, M., Harwood, C. J., Robertshaw, M. B., Fish, J. and O'Shea, K., (2015), The digital pipetting badge: a method to improve student hands-on laboratory skills, *J. Chem. Educ.*, **92**(12), 2038–2044, DOI: [10.1021/acs.jchemed.5b00464](https://doi.org/10.1021/acs.jchemed.5b00464).
- Uzuntiryaki-Kondakci, E. and Capa-Aydin, Y., (2013), Predicting critical thinking skills of university students through meta-cognitive self-regulation skills and chemistry self-efficacy, *Educ. Sci.: Theory Pract.*, **13**(1), 666–670. <https://files.eric.ed.gov/fulltext/EJ1016667.pdf>.
- van Borkulo, C. D., van Bork, R., Boschloo, L., Kossakowski, J. J., Tio, P., Schoevers, R. A., Borsboom, D. and Waldorp, L. J., (2023), Comparing network structures on three aspects: a permutation test, *Psychol. Methods*, **28**(6), 1273, DOI: [10.1037/met0000476](https://doi.org/10.1037/met0000476).
- Villafañe, S. M., Garcia, C. A. and Lewis, J. E., (2014), Exploring diverse students' trends in chemistry self-efficacy throughout a semester of college-level preparatory chemistry, *Chem. Educ. Res. Pract.*, **15**(2), 114–127, DOI: [10.1039/C3RP00141E](https://doi.org/10.1039/C3RP00141E).
- Villafañe, S. M., Xu, X. and Raker, J. R., (2016), Self-efficacy and academic performance in first-semester organic chemistry: testing a model of reciprocal causation, *Chem. Educ. Res. Pract.*, **17**(4), 973–984, DOI: [10.1039/C6RP00119J](https://doi.org/10.1039/C6RP00119J).
- Wahyudiati, D., Rohaeti, E., Wiyarsi, A. and Sumardi, L., (2020), Attitudes toward chemistry, self-efficacy, and learning experiences of pre-service chemistry teachers: grade level and gender differences, *Int. J. Instr.*, **13**(1), 235–254, DOI: [10.29333/iji.2020.13116a](https://doi.org/10.29333/iji.2020.13116a).
- Wang, A. Y. and Richarde, R. S., (1988), Global versus task-specific measures of self-efficacy, *Psychol. Rec.*, **38**(4), 533–541, DOI: [10.1007/BF03395045](https://doi.org/10.1007/BF03395045).
- Winkelmann, K., Baloga, M., Marcinkowski, T., Giannoulis, C., Anquandah, G. and Cohen, P., (2015), Improving students' inquiry skills and self-efficacy through research-inspired modules in the general chemistry laboratory, *J. Chem. Educ.*, **92**(2), 247–255, DOI: [10.1021/ed500218d](https://doi.org/10.1021/ed500218d).
- Wong, S. Y., Liang, J. C. and Tsai, C. C., (2021), Uncovering Malaysian secondary school students' academic hardiness in science, conceptions of learning science, and science learning self-efficacy: a structural equation modelling analysis, *Res. Sci. Educ.*, **51**(suppl 2), 537–564, DOI: [10.1007/s11165-019-09908-7](https://doi.org/10.1007/s11165-019-09908-7).
- Zhang, Z., Qiu, A., Zhang, X., Zhao, Y., Yuan, L., Yi, J., Zhang, Q., Liu, H., Lin, R. and Zhang, X., (2024), Gender differences in the mental symptom network of high school students in Shanghai, China: a network analysis, *BMC Public Health*, **24**(1), 2719, DOI: [10.1186/s12889-024-20130-7](https://doi.org/10.1186/s12889-024-20130-7).

