

Assessment of observed general practice trainee performance:

Confirmatory factor analysis of the General Practice Registrar Competency Assessment Grid

Alison Fielding, Rachel Turner, Katie Mulquiney, Elizabeth Holliday, Jason Dizon, Jean Ball, Amanda Tapley, Rosa Canalese, Katie Fisher, Alexandria Turner, Linda Klein, Parker Magin

Background and objective

Validity of workplace-based assessments (WBA) is important, yet rare in general practice. The General Practice Registrar Competency Assessment Grid (GPR-CAG) assesses registrar in-consultation performance, as observed during external clinical teaching visits. GPR-CAG content validity and initial internal consistency is established. This study aimed to confirm the GPR-CAG factor structure.

Methods

Confirmatory factor analysis was conducted using factors and items from an exploratory factor analysis. Separate analyses were conducted for Term 1 (GPT1; 16 items) and Term 2 (GPT2; 27 items) registrar data (n=827 and n=796, respectively).

Results

Analyses supported the existing GPT1 GPR-CAG factor structure, with fit indices showing good fit of the data to the model and most factor loadings being moderate-to-high. For GPT2, a single model change was strongly supported, after which model fit was good, the factor structure coherent and most item loadings moderate-to-high.

Discussion

This study supports the GPR-CAG as a psychometrically robust, evidence-based WBA.

MEDICAL EDUCATION PROGRAMS are increasingly focused on competency-based medical education (CBME).^{1,2} Formative assessment is essential in CBME implementation.

Miller's pyramid provides a theoretical foundation for competency assessment comprising knowledge (knows), competence (knows how), performance (shows how) and action (does).³ Assessment of the 'does' level of Miller's pyramid usually requires direct workplace observation.^{4,5} Workplace-based assessment (WBA) instruments for direct observation of performance are critical to formative competency assessment.^{1,6}

Using an assessment instrument with existing validity when conducting direct observation of clinical skills is vital.⁶ However, there is a lack of psychometrically evaluated instruments for this purpose,^{7,8} particularly for general practitioners (GPs).^{7,9} A systematic review in 2009 found that only 11 instruments were implemented in general practice and, of these, three had evidence of content validity and four had evidence of an internally consistent structure.⁷ More recently, a narrative review of the literature from 2000 to 2020 also found only a small body of published work on competency-based assessments in general practice within the USA and Canada (considered the 'early adopters' of competency-based medical education in family practice).¹⁰

The General Practice Registrar Competency Assessment Grid (GPR-CAG) is a WBA instrument designed to assess GP registrars' development of clinical competencies during their training and to assist with the early identification of registrars needing assistance. The GPR-CAG is performed by an experienced GP, external to the registrar's supervisor, using direct observation of the registrar's clinical consultations during external clinical teaching visits (ECTVs). ECTVs are typically in-person and conducted over a clinic session (3–4 hours) and entail the visitor observing the registrar and providing feedback – both in-person and, subsequently, via an ECTV report (during the period reported here, incorporating the GPR-CAG items). ECTVs are performed during a registrar's community-based training term and are the most important component of WBA within Australian vocational general practice training. ECTVs, though, are resource-intensive and it is important that the methods of assessment of registrars during ECTVs are validated.

The initial development of the CAG involved generation of assessment items based on a literature review and expert opinion.¹¹ In 2017, an initial exploratory factor analysis (EFA) provided evidence for the content validity and internal consistency of GPR-CAG, supporting the use of GPR-CAG as a GP vocational training WBA instrument.¹¹ The EFA resulted in a refined and more parsimonious CAG instrument. The original 25 items applicable to registrars in Term 1 of Australian general practice training (GPT1) were refined to a four-factor, 16-item model (refer to Appendix 1, available online only). For Term 2 of training (GPT2), the original 57 items were

refined to a seven-factor, 27-item model (refer to Appendix 2, available online only). This GPR-CAG version was routinely used in formative assessments of directly observed registrar consultations from 2017.

The EFA-derived GPR-CAG requires further analysis of validity by confirmatory factor analysis (CFA). CFA is an established important step in the psychometric evaluation of assessment instruments.¹² In this study, the aim was to confirm the factor structure of the GPR-CAG using an independent GP registrar sample.

Methods

Study design

This was a cross-sectional analysis of routinely collected GPR-CAG data within a retrospective cohort of registrars, whereby CFA was undertaken on the data.

Setting

The setting for this study was GP Synergy, the former regional training organisation for New South Wales/Australian Capital Territory. At the time of this study, GP Synergy was a not-for-profit, geographically defined, education and training organisation training approximately one-third of Australia's GP registrars.

Participants

Participants were GP Synergy registrars (enrolled with both The Royal Australian College of General Practitioners [RACGP] and The Australian College of Rural and Remote Medicine [ACRRM]) undertaking their first or second (of three) training terms (ie GPT1 and GPT2) with ECTV GPR-CAG data available from 2017 to 2019. GPT1 and GPT2 GPR-CAG observations were included as these are the early training terms where formative feedback and identification of registrars experiencing difficulties is of most use.

Data collection

GPR-CAG data collection occurred during routine ECTVs. GP Synergy ECTV visitors undertook specific ECTV training, including in the use of the GPR-CAG instrument.

Each registrar completed a minimum of five ECTVs during their training, of which GPR-CAG data were recorded for each ECTV.

This study only included GPR-CAG data from ECTVs conducted at approximately 4 months of general practice training (full-time equivalent, FTE) for GPT1 registrars, and at approximately 10 months FTE for GPT2 registrars. These timepoints were selected as they reflect allowance of adequate training time for registrars to have developed the skills expected for that training term.

GPR-CAG data were extracted from GP Synergy's routine training database.

The datasets generated and analysed during the current study cannot be publicly shared because of ethical or privacy reasons.

Statistical methods

Participants' demographic characteristics were summarised using frequencies with percentages.

The 2017–19 dataset used for the CFA was independent of the data used in the original EFA (which used 2014–16 GPR-CAG data). The GPR-CAG used during 2017–19 encompassed the 16 EFA-established items for GPT1 (refer to Appendix 1), and the 27 EFA-established items for the GPT2 model (refer to Appendix 2).¹¹ The CFA was conducted using these established factors and items.

EFA-identified factor structures for GPT1 and GPT2 were tested by CFA using structural equation modelling (SEM), with models fitted using the R lavaan package.¹³ Given Likert-scale items with incomplete data, estimation was performed using robust full information maximum likelihood (RFIML). Compared to conventional maximum likelihood estimation assuming multivariate normality, RFIML produces stable and reliable results for ordinal data under a range of missing data conditions, particularly when a sample size is <1000.¹⁴

In the SEM, items were specified to load on factors identified in the EFA, with item cross-loadings to other factors specified to equal zero. Correlations between factors were freely estimated. Adequacy of model fit to the data was assessed using goodness-of-fit statistics. Following the recommendations of Hu and Bentler,¹⁵ we adopted a two-index approach to presenting goodness-of-fit statistics, reporting the Absolute Fit Index (standardised root mean square residual: SRMR) supplemented with an incremental fit index (root mean squared error of

approximation: RMSEA). The recommended cut-offs of 0.08 for SRMR and 0.06 for RMSEA were used,¹⁵ where lower values indicate better model fit.

A priori, we planned to fit and report parameter estimates for the same model structures reported in the EFA; we did not intend to perform post-hoc model modification because such modifications can be highly sensitive to chance variation across samples. This risk is increased further when the sample is not large, and by sequential modification, where each modification rectifies a smaller inconsistency between the model and the data.^{16,17}

For GPT1, the specified model converged without error and its fit statistics and parameter estimates were reported directly. For GPT2, the specified model converged with a warning implying model misspecification. After inspecting the factor variance-covariance matrix, modification indices (MI) and expected parameter change (EPC), a single model change was strongly supported from both statistical and theoretical perspectives. This change involved transferring item CG7 from Factor 2 to Factor 5. Further details of this modification are provided in the results section.

Reliability of the final factors was assessed using Cronbach's alpha as a measure of internal consistency.

Sample size

All registrars providing ECTV GPR-CAG data from 2017 to 2019 were included, comprising 827 registrars for GPT1 and 796 registrars for GPT2. These samples were approximately 50% higher than those used in the EFA (555 and 537). Given incomplete data, 477 and 348 complete cases were available for GPT1 and GPT2, respectively. However, estimation via RFIML used all available data for each case.

Given retrospective data collection with maximum sample sizes was determined by the collection period, post-hoc power analyses were not performed. Rather, the absolute number of participants and items were compared with previous recommendations for CFA. Based on these recommendations (briefly outlined below), the GPT1 and GPT2 sample sizes were judged as favourable for estimating the CFA.

Rules of thumb for CFA describe the study sample sizes of about 800 as 'very good' to 'excellent', and our complete case sample sizes as 'good' to 'very good'.¹⁸ Comrey suggests adequacy of $N > 200$ for models with < 40 items,¹⁹ and the present study samples well exceed the median reported $N = 200$ for SEM.²⁰

For $N:p$ ratios, at least 10 cases per indicator variable are accepted as sufficient for CFA, with our GPT1 and GPT2 samples providing ~52 and ~29 cases per item, and ~30 and 13 cases per item for complete cases. Other authors suggest that when $N > 300$, a ratio of 5–10 participants per item is adequate.²¹

Finally, Monte Carlo simulations have shown SEM models reliably converge and produce proper solutions when $N > 200$ and each latent variable has three or more indicators. Even for various adverse model structures, $N > 300$ –500 is generally adequate.^{22,23}

Cross-validation of the CFA models was not performed because of the sample size constraints. The need for further evaluation and confirmation of these models is noted.

Ethics

Approval as a quality assurance (QA) project was obtained from The University of Newcastle Human Research Ethics Committee (reference number: QA147). Registrars provided written informed consent for data routinely collected as part of the GP Synergy educational program to be used for QA purposes.

Results

Registrar demographic characteristics are shown in Table 1 (available online only).

GPT1 competencies

For the GPT1 GPR-CAG competencies, the CFA model converged; fits indices met the prespecified thresholds, indicating good fit of the model to the data, with SRMR=0.055 (< 0.08) and RMSEA=0.058 (< 0.06).

The magnitude of factor loadings ranged from 0.36 to 0.82, with 75% of loadings being ≥ 0.5 . All loadings were statistically significant with $P < 0.001$ (refer to Table 2; available online only) for factor loadings with standard errors, P values and error variances, and Figure 1 (available online only) for

the path diagram). Compared to the EFA, there was broad similarity in the relative loading of individual items on the factors in the CFA. For example, items with the highest loadings on individual factors were reasonably consistent between the EFA and the CFA. There was moderate correlation among the factors, with polychoric correlation coefficients ranging from 0.53 to 0.77.

For additional detail, Appendices 3 and 4 (available online only) show pairwise item (polychoric) correlations and factor correlations, respectively.

The overall Cronbach alpha was 0.856 and the Factor Cronbach alpha ranged from 0.61 to 0.75 (Table 2). Refer to Appendix 5 (available online only) for further internal consistency data.

GPT2 competencies

For GPT2, the model converged with a warning that the covariance matrix was not positive definite. The resulting factor variance-covariance matrix showed high covariance between Factor 2 and Factor 5. A single model modification was made, based on the largest absolute expected parameter change (EPC) accompanied by a significant modification index (MI), as previously recommended.^{24,25} This change (EPC=0.35) suggested cross-loading of item CG7 ('negotiates agenda taking patient and doctor's needs into account') on Factor 5, in addition to, or instead of, Factor 2 (as in the original model). GP experts in vocational training also strongly supported transferring item CG7 to Factor 5, from a theoretical perspective.

In a refitted model allowing CG7 to cross-load on Factor 2 and Factor 5, loading of CG7 on Factor 2 was reduced to approximately 0 and was non-significant. The model was thus refitted with item CG7 loading only on Factor 5. This was the only modification implemented.

Final model fit indices for GPT2 indicated good fit of the model to the data, with SRMR=0.054 (< 0.08) and RMSEA=0.05 (< 0.06).

The magnitude of factor loadings ranged from 0.38 to 0.79, with 85% of loadings being ≥ 0.5 . All loadings were statistically significant with $P < 0.001$ (refer to Table 3; available online only) for factor loadings with standard errors, P values and error variances,

and Figure 2 (available online only) for the path diagram). For GPT2, there was also broad similarity in the items with highest loading on each factor in the CFA. There was moderate correlation among the factors, with polychoric correlation coefficients ranging from 0.47 to 0.81.

For additional detail, Appendices 6 and 7 (available online only) show pairwise item (polychoric) correlations and factor correlations, respectively.

The overall Cronbach alpha was 0.907, and Factor Cronbach alpha ranged from 0.58 to 0.81 (Table 3). Refer to Appendix 8 (available online only) for further internal consistency data.

Discussion

Main findings

This study provides further evidence of the validity of GPR-CAG to support the use of GPR-CAG as a GP vocational training WBA instrument. Overall, the CFA showed that the previously established items and factors for both GPT1 and GPT2 GPR-CAGs were good fits. Clinically/educationally congruent improvements in factor loadings were identified by moving one item within the GPT-2 GPR-CAG.

Implications for educational practice and further research

ECTVs are mandated intensive requirements in Australian general practice vocational training. It is vital that assessments occurring within ECTVs are robust. We have confirmed the GPR-CAG, which has been successfully implemented as a WBA during ECTVs, to be psychometrically robust. The GPR-CAG has further been proven to have predictive use for registrar performance in RACGP fellowship examinations.²⁶ This highlights the use of GPR-CAG assessment within ECTVs for enabling implementation of early, proactive and tailored support strategies for registrars identified as at-risk of poor performance. Thus, the GPR-CAG provides a robust and evidence-based WBA for inclusion in assessment frameworks in GP vocational training.

Limitations

Although ECTV visitors received the same comprehensive training in GPR-CAG

completion, inter-rater reliability and test-re-test reliability have not been evaluated. Also, patient data (demographics or nature and complexity of presentations) were not available, which would assist in considering external validity of the findings.

An inherent limitation of the GPR-CAG is that not all included competencies will be observable in any one three-hour ECTV.

Also, the ECTVs on which these analyses were based were conducted prior to Australian general practice training having transitioned from Regional Training Organisations to the RACGP and ACRRM. This has resulted in structural changes to the conduct of ECTVs and the feedback to registrars. These will influence the implementation of the GPR-CAG, but not the validity of the GPR-CAG factor structure.

Conclusion

This study provides further evidence for the validity of the GPR-CAG as a WBA instrument within vocational general practice training.

Authors

Alison Fielding BND (Hons), PhD, Research Manager, GP Training Research, The Royal Australian College of General Practitioners, Newcastle, NSW; Conjoint Lecturer, School of Medicine and Public Health, The University of Newcastle, Newcastle, NSW

Rachel Turner BSc (Hons), MBBS, FRACGP, DCH, General Practitioner, Charlestown Square Medical Centre, Lake Macquarie, NSW

Katie Mulquoney BND (Hons), MClInEpid (Distinc), Senior Research Assistant, The Royal Australian College of General Practitioners, Newcastle, NSW
Elizabeth Holliday PhD, BSc (Hons), MSc, Professor of Biostatistics, School of Medicine and Public Health, The University of Newcastle, Newcastle, NSW

Jason Dizon MMedStats, BMath, BCompSc, Statistician, Data Sciences, Hunter Medical Research Institute, Newcastle, NSW

Jean Ball BMath, Grad Dip Med Stats, Statistician, Data Sciences, Hunter Medical Research Institute, Newcastle, NSW

Amanda Tapley M Med Stat, B Biomed Sci, Senior Research Officer and Statistician, The Royal Australian College of General Practitioners, Newcastle, NSW

Rosa Canalese MBBS, Dip Paed, FRACGP, MPH, Grad Cert Uni Teach, Senior Medical Adviser, Avant Mutual Group Limited, Sydney, NSW

Katie Fisher BMed/MD, FRACGP, Conjoint Lecturer, School of Medicine and Public Health, The University of Newcastle, Newcastle, NSW

Alexandria Turner BBiotech (Hons), PhD, Senior Research Assistant, The Royal Australian College of General Practitioners, Newcastle, NSW

Linda Klein BSc (Hons), MSc, PhD, Conjoint Senior Lecturer, School of Medicine and Public Health, The University of Newcastle, Newcastle, NSW;

Manager, GP Training Research, The Royal Australian College of General Practitioners, Newcastle, NSW
Parker Magin PhD, FRACGP, Conjoint Professor, School of Medicine and Public Health, The University of Newcastle, Newcastle, NSW; Adjunct Professor, School of Population Health, University of New South Wales, Sydney, NSW; Senior Manager, GP Training Research, The Royal Australian College of General Practitioners, Newcastle, NSW

Competing interests: None.

Funding: This project was internally funded by GP Synergy. GP Synergy was funded by the Australian Government Department of Health.

Provenance and peer review: Not commissioned, externally peer reviewed.

AI declaration: The authors confirm that there was no use of artificial intelligence (AI)-assisted technology for assisting in the writing or editing of the manuscript and no images were manipulated using AI.

Correspondence to:

Parker.magin@newcastle.edu.au

Acknowledgements

The authors wish to acknowledge Dr Chris Starling, Regional Head of Education, Hunter Manning and Central Coast, and Dr Vanessa Moran, Director of Education and Training NSW and ACT, at GP Synergy for providing medical education expertise regarding suggested model modifications.

References

- Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach* 2010;32(8):676–82. doi: 10.3109/0142159X.2010.500704.
- lobst WF, Sherbino J, Cate OT, et al. Competency-based medical education in postgraduate medical education. *Med Teach* 2010;32(8):651–56. doi: 10.3109/0142159X.2010.500709.
- Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990;65(9) Suppl:S63–67. doi: 10.1097/00001888-199009000-00045.
- Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet* 2001;357(9260):945–49. doi: 10.1016/S0140-6736(00)04221-5.
- Kogan JR, Holmboe E. Realizing the promise and importance of performance-based assessment. *Teach Learn Med* 2013;25 Suppl 1:S68–74. doi: 10.1080/10401334.2013.842912.
- Kogan JR, Hatala R, Hauer KE, Holmboe E. Guidelines: The do's, don'ts and don't knows of direct observation of clinical skills in medical education. *Perspect Med Educ* 2017;6(5):286–305. doi: 10.1007/S40037-017-0376-7.
- Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: A systematic review. *JAMA* 2009;302(12):1316–26. doi: 10.1001/jama.2009.1365.
- Wetzell AP. Factor analysis methods and validity evidence: A review of instrument development across the medical education continuum. *Acad Med* 2012;87(8):1060–69. doi: 10.1097/ACM.0b013e31825d305d.
- Pelgrim EA, Kramer AW, Mokkink HG, van den Elsen L, Grol RP, van der Vleuten CP. In-training assessment using direct observation of single-patient encounters: A literature review. *Adv Health Sci Educ Theory Pract* 2011;16(1):131–42. doi: 10.1007/s10459-010-9235-6.
- Danilovich N, Kitto S, Price DW, Campbell C, Hodgson A, Hendry P. Implementing competency-based medical education in family medicine: A narrative review of current trends in assessment. *Fam Med* 2021;53(1):9–22. doi: 10.22454/FamMed.2021.453158.
- Fielding A, Mulquoney K, Canalese R, et al. A general practice workplace-based assessment instrument: Content and construct validity. *Med Teach* 2020;42(2):204–12. doi: 10.1080/0142159X.2019.1670336.
- Mueller RO, Hancock GR. Factor analysis and latent structure, confirmatory. In: Smelser NJ, Baltes PB, editors. *International encyclopedia of the social & behavioral sciences*. 1st edn. Elsevier, 2001; p. 5239–44. doi: 10.1016/B0-08-043076-7/00426-5.
- Rosseel Y. lavaan: An R package for structural equation modeling. *J Stat Softw* 2012;48(2):1–36. doi: 10.18637/jss.v048.i02.
- Jia F, Wu W. Evaluating methods for handling missing ordinal data in structural equation modeling. *Behav Res Methods* 2019;51(5):2337–55. doi: 10.3758/s13428-018-1187-4.
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling* 1999;6(1):1–55. doi: 10.1080/10705519909540118.
- MacCallum RC, Roznowski M, Necowitz LB. Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychol Bull* 1992;111(3):490–504. doi: 10.1037/0033-2909.111.3.490.
- Hutchinson SR. The stability of post hoc model modifications in confirmatory factor analysis models. *J Exp Educ* 1998;66(4):361–80. doi: 10.1080/00220979809601406.
- Comrey AL, Lee, HB. A first course in factor analysis. 2nd edn. Psychology Press, 1992.
- Comrey AL. Factor-analytic methods of scale development in personality and clinical psychology. *J Consult Clin Psychol* 1988;56(5):754–61. doi: 10.1037/0022-006X.56.5.754.
- MacCallum RC, Austin JT. Applications of structural equation modeling in psychological research. *Annu Rev Psychol* 2000;51(1):201–26. doi: 10.1146/annurev.psych.51.1.201.
- DeVellis RF. *Scale development: Theory and applications*. 4th edn. Sage, 2017.
- MacCallum RC, Widaman KF, Zhang S, Hong S. Sample size in factor analysis. *Psychol Methods* 1999;4(1):84–99. doi: 10.1037/1082-989X.4.1.84.
- Myers ND, Ahn S, Jin Y. Sample size and power estimates for a confirmatory factor analytic model in exercise and sport: A Monte Carlo approach. *Res Q Exerc Sport* 2011;82(3):412–23. doi: 10.1080/02701367.2011.10599773.
- Kaplan D. Model modification in covariance structure analysis: Application of the expected parameter change statistic. *Multivariate Behav Res* 1989;24(3):285–305. doi: 10.1207/s15327906mbr2403_2.
- Luijben T, Boomsma A. Statistical guidance for model modification in covariance structure analysis. In: Edwards D, Raun N, editors. *COMPSTAT 1988: Proceedings in computational statistics*. Physica, 1988; p. 335–40. doi: 10.1007/978-3-642-46900-8_46.
- Fisher K, Fielding A, Ralston A, et al. Exam prediction and the General Practice Registrar Competency Assessment Grid (GPR-CAG). *Educ Prim Care* 2023;34(5-6):268–76. doi: 10.1080/14739879.2023.2269884.

Table 1. Registrar demographics

Variable	GPT1	GPT2
Total (n)	827	796
Gender		
Female	483 (58.4)	472 (59.3)
Male	344 (41.6)	323 (40.6)
Age (years) (mean±SD)	33.5 (±6.9)	34.1 (±6.9)
Country of medical graduation		
Australia	626 (75.8)	614 (77.1)
International	200 (24.2)	182 (22.9)
Pathway		
General	523 (63.2)	504 (63.3)
Rural	304 (36.8)	292 (36.7)
Full-time/Part-time status		
Full time	691 (83.6)	627 (78.8)
Part time	136 (16.4)	169 (21.2)

Data are presented as n (%) unless otherwise stated.

GPT1, Term 1 of Australian general practice training; GPT2, Term 2 of Australian general practice training; SD, standard deviation.

Table 2. GPT1: Factor loadings and internal consistency of identified factors (standardised Cronbach alphas) in the confirmatory factor analysis^A

Factor ^A	Item ^B	Standardised solution			Error variance (SE)	P value
		Mean (SD)	Loading (SE)	P value		
Factor 1: Consultation techniques subserving patient-centeredness caring ($\alpha=0.751$)	CG10 – Listens attentively – uses appropriate listening skills and silence	3.13 (0.49)	0.69 (0.03)	<0.001	0.53 (0.04)	<0.001
	CG15 – Uses concise easily understood questions and comments, avoids jargon	3.02 (0.50)	0.49 (0.04)	<0.001	0.76 (0.04)	<0.001
	CG23 – Demonstrates appropriate non-verbal behaviour – eye contact, posture, position and movement, vocal cues, rate volume, tone	3.17 (0.46)	0.56 (0.03)	<0.001	0.69 (0.04)	<0.001
	CG5 – Listens attentively to patient's opening, without interrupting	3.13 (0.46)	0.66 (0.03)	<0.001	0.56 (0.04)	<0.001
	CG6 – Confirms list and screens for further problems	2.96 (0.49)	0.44 (0.04)	<0.001	0.80 (0.04)	<0.001
	CG9 – Uses open and closed techniques appropriately	3.02 (0.46)	0.65 (0.03)	<0.001	0.58 (0.04)	<0.001
Factor 2: Skills in formulating and articulating coherent hypotheses and management plans ($\alpha=0.730$)	CG36 – Appropriate hypotheses are articulated and problems defined	3.00 (0.41)	0.63 (0.04)	<0.001	0.61 (0.05)	<0.001
	CG39 – Clearly outlines the plan of management for each defined problem. Explains expected outcomes and influence on ongoing management of the problem	3.02 (0.46)	0.73 (0.03)	<0.001	0.47 (0.05)	<0.001
	CG40 – Provides clear information on investigations, procedures and explains process for results	3.04 (0.42)	0.74 (0.03)	<0.001	0.45 (0.05)	<0.001
	CG48 – Prescribes medications and treatments as appropriate to diagnostic conclusions, adopts a quality use of medications framework, prescribes safely	3.01 (0.37)	0.50 (0.05)	<0.001	0.75 (0.05)	<0.001
Factor 3: Attention to basic-level clinical professional behaviours and responsibilities ($\alpha=0.609$)	CG2 – Introduces self and clarifies role	3.10 (0.34)	0.64 (0.04)	<0.001	0.58 (0.06)	<0.001
	CG3 – Shows interest and respect, attends to patient's physical comfort	3.20 (0.43)	0.69 (0.04)	<0.001	0.52 (0.05)	<0.001
	CG51 – Medical records accurate and contemporaneous – all key and relevant information for the consultation recorded. Management plan and follow-up arrangements clearly recorded	3.09 (0.40)	0.41 (0.05)	<0.001	0.83 (0.04)	<0.001
	CG54 – Writes clear referral letters stating reason for referral, expected outcomes and provides all necessary pertinent information to facilitate the referral	3.04 (0.34)	0.36 (0.07)	<0.001	0.87 (0.05)	<0.001
Factor 4: Proficiency in physical examination skills ($\alpha=0.661$)	CG34 – Performs an appropriate physical examination – is mindful of patient comfort and privacy throughout the physical examination	3.02 (0.44)	0.82 (0.04)	<0.001	0.33 (0.06)	<0.001
	CG35 – Physical examination is accurate and clinical signs correctly elicited	2.97 (0.44)	0.61 (0.04)	<0.001	0.63 (0.05)	<0.001

^AOverall Cronbach $\alpha=0.856$.

^BCG, competency grid (note: CG numbers refer to item numbers in the original 57-item competency grid – prior to the original exploratory factor analysis).

GPT1, Term 1 of Australian general practice training; SD, standard deviation; SE, standard error.

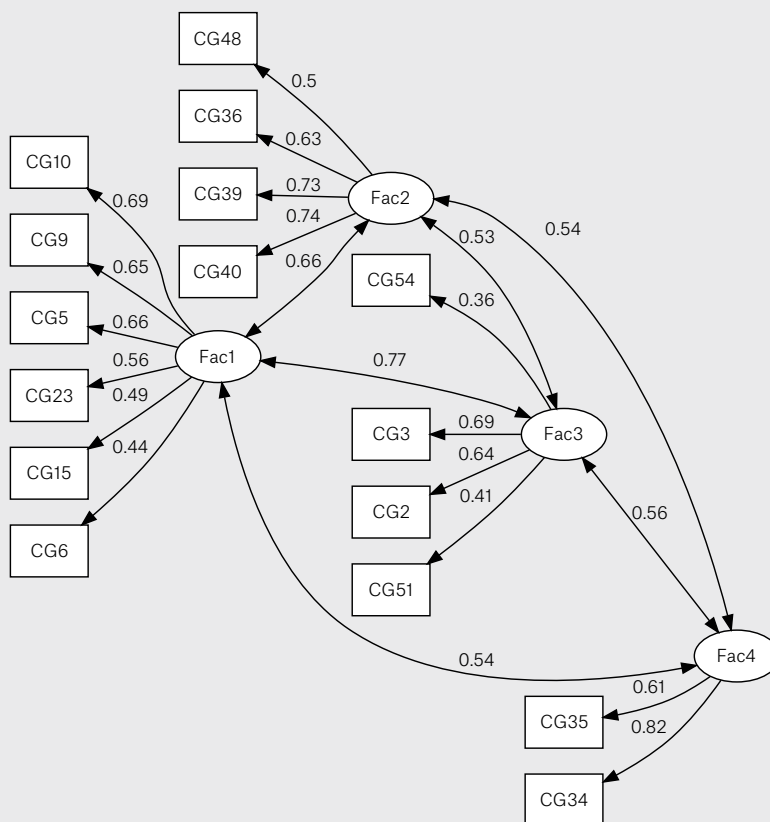


Figure 1. GPT1 path diagram.

CG, competency grid; GPT1, Term 1 of Australian general practice training; Fac, factor.

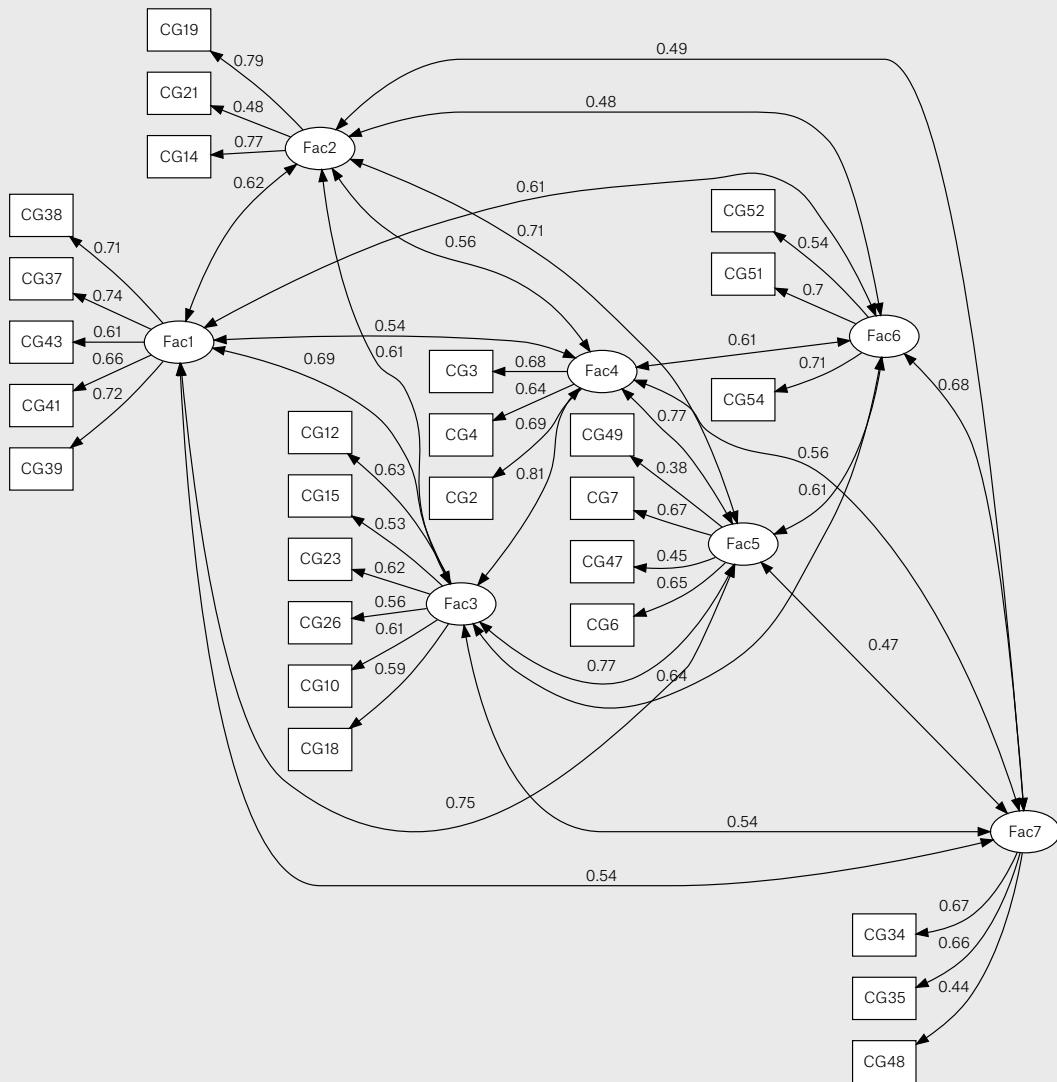


Figure 2. GPT2 path diagram.

CG, competency grid; GPT2, Term 2 of Australian general practice training; Fac, factor.

Table 3. GPT2: Factor loadings and internal consistency of identified factors (standardised Cronbach alphas) in the confirmatory factor analysis^A

Factor ^A	Item ^B	Standardised solution				
		Mean (SD)	Loading (SE)	P value	Error variance (SE)	P value
Factor 1: Patient-centredness, sharing ($\alpha=0.810$)	CG37 – Clearly explains diagnostic conclusions, justification and check patients understanding	3.03 (0.44)	0.74 (0.03)	<0.001	0.46 (0.04)	<0.001
	CG38 – Appropriately explains causation-seriousness, expected duration, short- and long-term consequences	2.98 (0.44)	0.71 (0.03)	<0.001	0.50 (0.04)	<0.001
	CG39 – Clearly outlines the plan of management for each defined problem. Explains expected outcomes and influence on ongoing management of the problem	3.05 (0.43)	0.72 (0.03)	<0.001	0.48 (0.04)	<0.001
	CG41 – Relates investigations and procedures to the management plan, explains value and purpose	3.04 (0.37)	0.66 (0.04)	<0.001	0.56 (0.05)	<0.001
	CG43 – Discusses possible options for management if appropriate and relevant and elicits patients' viewpoint	3.04 (0.42)	0.61 (0.04)	<0.001	0.63 (0.05)	<0.001
Factor 2: Structural aspects of history-taking ($\alpha=0.680$)	CG14 – Periodically summarise to verify own understanding, invites patient to correct interpretation	3.05 (0.48)	0.77 (0.03)	<0.001	0.41 (0.05)	<0.001
	CG19 – Summarises periodically to confirm understanding before moving on	3.02 (0.46)	0.79 (0.03)	<0.001	0.38 (0.05)	<0.001
	CG21 – Structures interview in a logical sequence	3.04 (0.51)	0.48 (0.05)	<0.001	0.77 (0.05)	<0.001
Factor 3: Higher-level caring, patient-centredness ($\alpha=0.754$)	CG10 – Listens attentively – uses appropriate listening skills and silence	3.14 (0.47)	0.61 (0.04)	<0.001	0.63 (0.05)	<0.001
	CG12 – Picks up verbal and non-verbal cues, appropriately addresses and acknowledges	3.01 (0.49)	0.63 (0.03)	<0.001	0.60 (0.04)	<0.001
	CG15 – Uses concise easily understood questions and comments, avoids jargon	3.05 (0.49)	0.53 (0.03)	<0.001	0.71 (0.04)	<0.001
	CG18 – Encourages the patient to express their feelings	3.01 (0.48)	0.59 (0.04)	<0.001	0.65 (0.05)	<0.001
	CG23 – Demonstrates appropriate non-verbal behaviour – eye contact, posture, position and movement, vocal cues, rate volume, tone	3.17 (0.45)	0.62 (0.03)	<0.001	0.61 (0.04)	<0.001
	CG26 – Accepts legitimacy of patient's views and is non-judgemental	3.19 (0.43)	0.56 (0.04)	<0.001	0.68 (0.04)	<0.001

Table continued on the next page

Table 3. GPT2: Factor loadings and internal consistency of identified factors (standardised Cronbach alphas) in the confirmatory factor analysis^A (cont'd)

Factor ^A	Item ^B	Standardised solution				
		Mean (SD)	Loading (SE)	P value	Error variance (SE)	P value
Factor 4: Minimum-required performance in patient-centred caring ($\alpha=0.703$)	CG2 – Introduces self and clarifies role	3.12 (0.34)	0.69 (0.04)	<0.001	0.52 (0.06)	<0.001
	CG3 – Shows interest and respect, attends to patient's physical comfort	3.23 (0.45)	0.68 (0.05)	<0.001	0.54 (0.07)	<0.001
	CG4 – Identifies patient's problems or issues	3.11 (0.41)	0.64 (0.05)	<0.001	0.60 (0.06)	<0.001
Factor 5: Holistic pro-active approach to patient presentations ($\alpha=0.603$)	CG47 – Appropriately implements health promotion and identifies opportunities of behaviour change if applicable	2.97 (0.47)	0.45 (0.05)	<0.001	0.79 (0.04)	<0.001
	CG49 – Discusses non-medication options as appropriate and relevant to clinical context	3.01 (0.34)	0.38 (0.06)	<0.001	0.85 (0.04)	<0.001
	CG6 – Confirms list and screens for further problems	3.05 (0.49)	0.65 (0.04)	<0.001	0.57 (0.05)	<0.001
	CG7 ^C – Negotiates agenda, taking patient and doctor's needs into account	3.01 (0.51)	0.67 (0.03)	<0.001	0.56 (0.05)	<0.001
Factor 6: Attention to minimum standards of professional communication ($\alpha=0.709$)	CG51 – Medical records accurate and contemporaneous – all key and relevant information for the consultation recorded. Management plan and follow-up arrangements clearly recorded	3.07 (0.39)	0.70 (0.05)	<0.001	0.50 (0.07)	<0.001
	CG52 – Patient medical summary information and relevant family and preventative health information recorded and regularly updated as required	2.98 (0.41)	0.54 (0.06)	<0.001	0.70 (0.07)	<0.001
	CG54 – Writes clear referral letters stating reason for referral, expected outcomes and provides all necessary pertinent information to facilitate the referral	3.04 (0.34)	0.71 (0.06)	<0.001	0.50 (0.08)	<0.001
Factor 7: High level but structured clinical tasks ($\alpha=0.576$)	CG34 – Performs an appropriate physical examination – is mindful of patient comfort and privacy throughout the physical examination	3.03 (0.37)	0.67 (0.05)	<0.001	0.55 (0.07)	<0.001
	CG35 – Physical examination is accurate and clinical signs correctly elicited	2.99 (0.39)	0.66 (0.05)	<0.001	0.57 (0.07)	<0.001
	CG48 – Prescribes medications and treatments as appropriate to diagnostic conclusions, adopts a quality use of medications framework, prescribes safely	3.03 (0.30)	0.44 (0.07)	<0.001	0.80 (0.06)	<0.001

^AOverall Cronbach $\alpha=0.907$.

^BCG, competency grid (note: CG numbers refer to item numbers in the original 57-item competency grid – prior to the original exploratory factor analysis)

^CItem moved from Factor 2 to Factor 5.

GPT2, Term 2 of Australian general practice training.