



Assessment Practices for Dietetics Trainees: A Systematic Review



Janica Jamieson, Grad Dip Diet; Claire Palermo, PhD, MNutrDiet, MPH, Grad Cert Health Prof Ed;

Margaret Hay, PhD, MAPS, Grad Dip Epi, Epi Grad Cert Health Prof Ed; Simone Gibson, PhD, Grad Dip Diet, Grad Cert Acad Pract

ARTICLE INFORMATION

Article history:

Submitted 28 May 2018

Accepted 17 September 2018

Keywords:

Assessment

Dietetics

Education

Systematic review

Trainee

Supplementary materials:

Figures 1, 2, 3, and 6 are available at www.jandonline.org.

2212-2672/Copyright © 2019 by the Academy of Nutrition and Dietetics.

<https://doi.org/10.1016/j.jand.2018.09.010>

ABSTRACT

High-quality education is essential to produce competent graduates in the field of dietetics. Assessment is a fundamental component of education and driver of learning, yet little is known about methods used to assess dietetics trainees. The objective of this review is to evaluate the practices and outcomes of methods used to assess dietetics trainees. A systematic review of the literature was undertaken. MEDLINE, the Cumulative Index to Nursing and Allied Health Literature Plus, Embase, and the Education Resources Information Center databases were searched from inception until May 31, 2017, using key terms that identified studies reporting practices for the assessment of dietetics trainees. Abstract and title screening was completed by three independent reviewers followed by full-text screening using the eligibility criteria. Quantitative and qualitative data were extracted. Study outcomes were evaluated using Miller's Pyramid, Kirkpatrick's Hierarchy, and the principles of programmatic assessment. Thirty-seven studies were identified. Assessments targeted all levels of Miller's Pyramid with the *does* level being the most prevalent ($n=23$). Most studies focussed on evaluating Level 1 (participation) ($n=16$) and Level 2b ($n=16$) (knowledge and skills) of Kirkpatrick's Hierarchy. Studies described single assessment instruments that focussed on instrument validity and reliability. Few studies considered a program of assessment or the role of expert judgment. Six themes were identified from qualitative data: (1) assessment for learning and professional development, (2) assessment requires motivated and skilled assessors, (3) trainees value authentic and global assessment, (4) assessment is evolving and context-sensitive, (5) poor assessment has negative implications, and (6) assessment evokes an emotional response. Studies focused on the development and evaluation of single quantitative-based instruments applied in isolation, with low-level outcomes sought. There is room to improve practices and design programs of assessment that combine quantitative and qualitative data for meaningful trainee feedback and credible assessment decisions. Comprehensive evaluation of assessment practices is required and must consider the contribution to improved health outcomes in all practice settings.

J Acad Nutr Diet. 2019;119(2):272-292.

DIETITIANS HAVE AN IMPORTANT ROLE IN MEETING the complex and changing health needs of the population, including the burgeoning lifestyle-related chronic disease epidemic observed globally.^{1,2} The dietetics workforce will need to expand, with more practitioners equipped with the skills and attributes

necessary to provide safe and effective health care.^{1,2} High-quality training programs are essential to deliver this increased workforce and produce competent graduates.^{1,3} Yet little is known about the practices and outcomes of methods used to assess dietitians' preparedness for the workforce.²

Assessment is a fundamental component of education and is a powerful driver of learning.^{4,5} It involves the collection and judgment of information pertaining to the performance and competence of a trainee.⁶ A wide variety of assessment methods and instruments to prepare health professionals for practice has been described.⁷ Despite the breadth of literature, challenges to assessment remain and include increased trainee numbers, work placement shortages, and credentialing requirements.^{1,3,8} In the dietetics profession, it has been observed that there are inconsistent performance expectations in assessment and a reliance on historical

The Continuing Professional Education (CPE) quiz for this article is available for free to Academy members through the MyCDRGo app (available for iOS and Android devices) and through www.jandonline.org (click on "CPE" in the menu and then "Academy Journal CPE Articles"). Log in with your Academy of Nutrition and Dietetics or Commission on Dietetic Registration username and password, click "Journal Article Quiz" on the next page, then click the "Additional Journal CPE quizzes" button to view a list of available quizzes. Non-members may take CPE quizzes by sending a request to journal@eatright.org. There is a fee of \$45 per quiz (includes quiz and copy of article) for non-member Journal CPE. CPE quizzes are valid for 1 year after the issue date in which the articles are published.

antecedent.^{9,10} Such challenges can undermine and compromise the rigor of assessment decisions.¹¹

It has been argued that best practice assessment should be based on a broad sampling of performance using different fit-for-purpose instruments applied in different contexts that include the perspectives of multiple skilled assessors. The collection and aggregation of this information should be purposeful and planned, described as a program of assessment. The use of such assessment methods allows for credible and defensible decisions.^{2,12,13} Assessment practices need to result in change beyond the trainee level, demonstrating positive outcomes for recipients of health care services.¹⁴ These concepts have recently been proposed in the dietetics profession.¹⁵

The assessment of health care professionals has traditionally relied on single instruments with demonstrated validity and reliability that use tick-boxes and ratings to identify easy-to-measure skills and attributes.^{4,12} Although this has been similarly observed in the dietetics profession,² there has been no systematic synthesis of the assessment practices used for the training of dietitians. Understanding current and historical assessment practices will highlight accomplishments and identify gaps to drive future assessment innovation and research. As such, this systematic review had the aim to identify and evaluate assessment practices and outcomes for dietetics trainees.

METHODS

Protocol

The systematic review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines,¹⁶ was registered with the International Prospective Register of Systematic Reviews on June 19, 2017, and was last updated on April 20, 2018 (42017068333). Ethics approval was not sought.

Search Strategy

The search strategy was developed using the principles of the Cochrane Review Protocol¹⁷ and was guided by a health sciences librarian. The electronic databases MEDLINE (Ovid), Cumulative Index to Nursing and Allied Health Literature Plus (EBSCO), Embase (Ovid), and the Education Resources Information Center Plus (ProQuest) were searched from their inception to May 31, 2017. The search strategy was piloted using MEDLINE (Ovid) with modifications made to the final search terms (Figure 1, available at www.jandonline.org). Publications were imported into EndNote X8¹⁸ and duplicates removed using the software function. The reference lists of included full text articles were hand searched to identify additional publications.

Eligibility Criteria

Publications were included in the case that trainees were undertaking a dietetics qualification. Studies where dietetics trainees formed part of a mixed-discipline cohort were included when the dietetics trainee data could be separated from that of other disciplines. Publications were included when they described at least one assessment and reported at least one outcome. No restrictions to date, setting, or language were applied. Primary studies of quantitative, qualitative, and mixed-method design—including published and

RESEARCH SNAPSHOT

Research Question: What are the assessment practices and outcomes for dietetics trainees?

Key Findings: Thirty-seven studies were identified in this systematic review with findings demonstrating a focus on the development of single instruments targeting the assessment of practice. No studies evaluated high-level outcomes, such as patient and organizational benefit. Instead, studies focused on trainee participation and the acquisition of knowledge and skills. Trainees placed value on authentic assessment that provided meaningful feedback.

unpublished abstracts and theses—were included to capture all relevant data (Figure 2, available at www.jandonline.org).

Study Selection

Study selection was coordinated using Covidence.¹⁹ Three independent reviewers (J. J. and S. G. or C. P.) completed title and abstract screening. Disagreements were resolved by reviewers coming together and reaching a consensus.¹⁶ All studies had an English language abstract available. No studies published in a language other than English met the eligibility criteria. Full texts were obtained for studies meeting the criteria and three reviewers (J. J. and S. G. or C. P.) independently screened for inclusion using the eligibility criteria. Disagreements were resolved through discussion. Reviewers did not screen their own publications. Duplicate articles not detected by Endnote X8¹⁸ were manually removed during full-text screening.

Data Extraction

Microsoft Excel²⁰ was used to develop a data extraction worksheet. Data extraction included publication year, location, funding sources, study aim and design, study population, type and purpose of assessment, and results. The worksheet was piloted with three studies with subsequent minor revisions. Qualitative data were extracted verbatim from the results, findings, or discussion section of publications, and abstracts were checked for additional findings.²¹ Data included quotations and themes to position the review within the original publication and participant experience.²² Quantitative data were extracted verbatim. One reviewer (J. J.) was responsible for extracting all data, which was verified by a secondary reviewer (S. G. or C. P.). Discrepancies were resolved through discussion.

Critical Appraisal

Publications were evaluated for methodologic quality and education outcomes. The Medical Education Research Study Quality Instrument (MERSQI)²³ was used to assess the methodologic quality of all publications with a score ranging between 4.5 and 18. MERSQI is widely used to evaluate observational, experimental, or quasiexperimental medical education research.²⁴⁻²⁷ The methodologic quality of each study is considered in the interpretation of findings. The Critical Appraisal Skills Programme Qualitative Checklist²⁸ was used to evaluate qualitative studies in addition to MERSQI. Two independent reviewers (J. J. and S. G. or C. P.)

assessed each publication using the critical appraisal tools. Discrepancies were resolved through discussion.

Outcomes

Miller's Pyramid provided the framework for the evaluation of study outcomes. Miller's Pyramid is a 5-level framework conceptualizing the development and assessment of competence.²⁹ The five levels are *knows*, *knows how*, *shows how*, *does*, and *is*. *Knows* is the base of competence and *knows how* is the awareness of how to apply knowledge. *Shows how* is the application of knowledge to demonstrate problem solving and reasoning ability and *does* is the application within the practice setting. *Is* represents professional identity.^{29,30} Study outcomes were classified using Kirkpatrick's Hierarchy, which evaluates the impact of medical education research³¹⁻³³ using four levels: Level 1 (participation), Level 2a (attitudes and perceptions) and Level 2b (knowledge and skills), Level 3 (behavioral change), and Level 4a (organization practice) and 4b (patient benefits).³¹

Assessment instrument reliability and validity coefficients reported in the studies were interpreted as poor (<0.5), moderate (0.5 to 0.75), good (0.75 to 0.90), and excellent (>0.90).³⁴ Validity was classified as either demonstrated or not demonstrated. The researchers developed a framework to evaluate each study using the principles of programmatic assessment (Figure 3, available at www.jandonline.org). Programmatic assessment encapsulates best practice in health education and recognizes that assessment decisions should be informed by broad sampling using different instruments, contexts, and assessors as part of a program.¹³ Each study was evaluated by two independent reviewers (J. J. and S. G. or C. P.). Discrepancies were resolved through discussion.

Data Analysis

A segregated approach was used for data analysis whereby qualitative and quantitative synthesis was separated until aggregation for interpretation in the discussion.³⁵ Qualitative and quantitative data in mixed-method studies were similarly separated for synthesis and then aggregated for interpretation.

A thematic analysis of qualitative data was undertaken independently by two reviewers (J. J. and S. G.). Independently, the reviewers applied inductive coding to extracted qualitative data whereby labels were assigned to segments of text. The reviewers then independently identified themes by grouping codes together with the aim of capturing the relational meaning. The reviewers then came together to discuss the analysis. Using an iterative approach, the initial themes identified by each independent reviewer were revised until a consensus on the key themes was achieved. The last step aimed to generate concepts and understanding related to the systematic review question.³⁶

Due to the heterogeneity of the studies a descriptive analytical approach was applied to the quantitative data rather than a statistical or meta-analysis. The purpose of the descriptive analysis was to explore the types and quality of assessment for dietetics trainees and identify their impact. Statistical significant was taken at the level of $P < 0.05$ for all studies.³⁷

RESULTS

Literature Search

A total of 2,840 publications were identified after the removal of duplicates. Application of the inclusion criteria provided 37 publications for data synthesis (Figure 4).

Assessment Description

Study Characteristics. Forty-eight different methods of assessments were identified of which five were repeated in multiple articles,³⁸⁻⁴³ leaving 43 individual assessments. Six studies reported more than one assessment method.^{39-41,44-46} One study⁴⁷ evaluated multiple assessment instruments used by 14 different dietetics departments within the United Kingdom, for the purpose of the analysis they have been considered as one assessment method.

Twenty-one studies were from the United States,^{38-43,46,48-61} seven from the United Kingdom,^{47,62-67} six from Australia,^{44,45,68-71} two from Canada,^{72,73} and one from Malaysia.⁷⁴ Studies were published between 1965 and 2016 with the highest frequency of publications observed in 2010 (n=6)^{47,52,62,64,69,70} followed by 2015,^{44,46,50} 1980,^{49,60,72} and 1976 (n=3 each).^{38,43,56}

Types of Assessments. Performance evaluation instruments (n=22),^{38-41,43,44,46,47,49,51,52,54-56,58,59,61,63,65,67,72,74} objective structured clinical examinations (OSCEs) (n=7),^{44,45,50,53,64,66,70} tests (n=4),^{39,40,48} examinations (n=3),^{41,42,57,60} portfolios (n=3),^{62,69,73} written essay or report (n=2),^{45,68} quizzes (n=1),⁷¹ and a presentation (n=1)⁴⁵ were reported. Performance evaluation instruments were published between 1965 and 2016. OSCEs were published after 2004 and portfolios after 2007. Tests were reported between 1965 and 1978 and examinations between 1980 and 2002, including the USA National Registration Examination for Dietitians.

Most performance evaluation instruments were reported in the United States (n=16)^{38-41,43,46,49,51,52,54-56,58,59,61} followed by the United Kingdom (n=4),^{47,63,65,67} and Australia,⁴⁴ Canada,⁷² and Malaysia (n=1 each).⁷⁴ Australia published the most OSCEs (n=3)^{44,45,70} followed by the United Kingdom,^{64,66} and the United States^{50,53} (n=2 respectively) and all tests and examinations were from the United States (n=11).^{39-42,48,57,60} Portfolios were reported by Australia,⁶⁹ Canada,⁷³ and the United Kingdom⁶² (n=1 each). All written tasks, quizzes, and presentations were from Australia.^{45,68,71}

Assessment Settings. Assessments (n=48) occurred within both the work-integrated learning setting, such as internships or placements (n=25),^{38-43,45,47,49,51,53,55-58,61-63,65,67,69,72,74} and the university setting (n=27).^{39-41,44,46,48,50,52,54,55,59,60,64,66,68-71,73,74} The university setting included coursework (n=25)^{39-41,44,46,48,52,54,55,59,60,64,66,68-71,73} and clinics (n=2).^{50,74} The most common work-integrated learning setting was the hospital (n=20)^{38-40,42,43,47,49,51,53,55-58,61,62,65,67,69,72,74} with one study occurring in a preventative health setting.⁴⁵ Four studies applied the assessment to both a work-integrated learning and a university setting.^{39,55,69,74}

Types of Assessors. The use of university-based assessors (n=22)^{38-40,44-46,50,52,54-57,59,60,64-66,68-71,73} and work-based

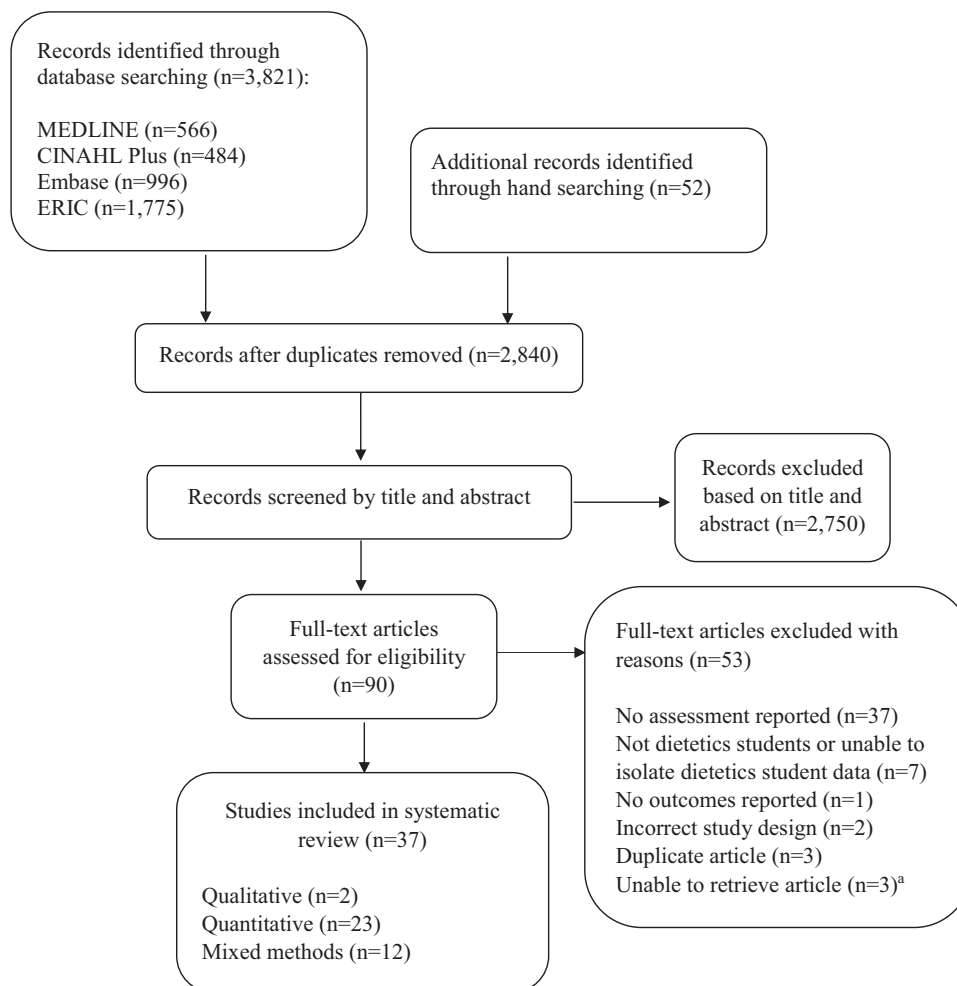


Figure 4. Assessment practices and outcomes for dietetics trainees: Literature search and selection process. ^aArticles could not be sourced by a health science librarian. CINAHL—the Cumulative Index to Nursing and Allied Health Literature Plus. ERIC—Education Resources Information Center.

assessors (WBA) (n=19)^{39,40,43-45,49,52,53,55,56,58,61,62,64,65,69,72-74} was reported, followed by trainee self-assessment (n=9),^{49,54-56,62,63,68,69,72} professional organization (n=2),^{41,42} peer (n=2),^{55,68} and simulation patients (n=1).⁵² An assessor was not specified in three studies.^{47,48,51} For those studies reporting an assessor, 65% (n=24) had one type of assessor^{38,39,41-43,46-48,50,51,53,57-61,63,65-67,70-72,74} and 35% (n=13)^{40,44,45,49,52,54-56,62,64,68,69,73} used different types of assessors with the most common combination being WBAs and university-based assessors.

Knowledge and Skills Assessed. Assessments (n=43) evaluated clinical competence (n=21),^{38,40,43,47,49-51,53,55-58,61,62,64-66,69-72,74} counseling and verbal communication (n=8),^{44,46,52,54,63,67} preventative health (n=3),⁴⁵ critical thinking (n=2),^{39,40,59} written communication (n=1),^{39,40} personality attributes (n=1),^{39,40} and a combination of skills (n=7).^{40-42,48,60,68,73} Clinical competence was assessed between 1969 and 2016 with performance evaluation instruments (n=13)^{38,40,43,47,49,51,55,56,58,61,65,72,74} being the most common method, followed by OSCEs (n=5),^{50,53,64,66,70} portfolios (n=2),^{52,69} an examination (n=1),⁵⁷ tests (n=1),⁴⁰

and quizzes (n=1).⁷¹ Counseling and verbal communication appraisals were reported after 1984 and were assessed using either a performance evaluation instrument (n=7)^{44,46,52,54,63,67} or an OSCE (n=1).⁴⁴ Studies are summarized in Figure 5.

Critical Appraisal

The mean MERSQI score was 9.5 ± 3.0 (range=5 to 14.5). Common study weaknesses were single group design (78%),^{38,39,42,43,45,47-49,51-56,58-60,62-66,68-74} single institutions (57%),^{38,39,41-43,45,49,52-56,59,60,64,66,68,69,71,73,74} low or no reported response rate (59%),^{38-40,43,44,48,49,52,53,55-60,62,63,66,70,72-74} reliance on student participation (51%),^{38,43,47,50,53,55,58-60,62-64,66,68-73} not reporting internal structure (62%),^{38-41,43,44,48-50,53,55,57-59,61,63-66,68-70,72} content (53%),^{39,40,42,49,53,55,57,62-64,66,68-72} or relationship to other variables (59%)^{38,43,44,47,49-51,53,55,56,58,59,61-64,66,68,69,71,72,74} for assessments, and outcomes that focused on satisfaction, attitudes, perceptions, opinions, and general facts (54%).^{38,41,43,47,50,51,53-56,58,59,62,64,66,68-72} The appropriateness^{41,42,44-51,54,57,59,61,62,64,67-71,73,74} and complexity^{40-42,44-52,54,56,57,61,62,64,65,67,70,74} of

Author(s) Year Country	Assessment type	Assessment purpose and description	Assessment setting	Assessors	Miller's Pyramid ^a	Kirkpatrick's Hierarchy ^b
Brennan and Lennie ⁶² 2010 United Kingdom	Portfolio	Assess clinical competence during Placement B and/or C Individual portfolio, content derived from trainee activities and tasks	Workplace (National Health Service)	Trainee (self) WBA ^c	Does	Level 3
Chambers and Hubbard ⁴⁸ 1978 United States	Test	Assess competence for Plan IV Minimum Academic Competencies (American Dietetic Association) Structure: 1. Food 2. Nutrition 3. Nutrition in disease 4. Management	University coursework	NR ^d	Knows	Level 1
Cochran and Spears ⁴⁹ 1980 United States	Performance evaluation instrument	<i>Modified Ingalsbe & Spears Critical Incident</i> Assess performance during hospital-based placement Critical incidents recorded and reviewed on 3 occasions during placement. Instrument structure: 1. Planning and organizing 2. Gathering and evaluating data 3. Applying scientific principles to clinical dietetics 4. Observing, reporting, and documenting 5. Relating to instructors, managers, employees, peers, and clientele 6. Adaptability to new and stressful situations 7. Using learning opportunities 8. Using creativity 9. Acceptance of professional responsibility 10. Judgment regarding professional values	Workplace (health service)	Trainee (self) WBA	Does	Level 2b
Daniels and Magarey ⁶⁸ 2000 Australia	Written task	Develop feedback and critical thinking. Trainees complete essay and receive and provide peer feedback. Trainees submit essay with an appraisal of peer feedback quality	University coursework	Trainee (self) Peer University	Knows	Level 1

(continued on next page)

Figure 5. Assessment practices and outcomes for dietetics trainees: Description of studies (1965-2016).

Author(s) Year Country	Assessment type	Assessment purpose and description	Assessment setting	Assessors	Miller's Pyramid ^a	Kirkpatrick's Hierarchy ^b
Earl ⁶³ 1984 United Kingdom	Performance evaluation instrument	Assess communication skills (group presentations). Instrument completed on multiple occasions. Structure: background, environmental conditions, audience information, presentation aids, self-rating of communication skills, trainee assessment of audience perception, and areas for improvement	Workplace (various)	Trainee (self)	Does	Level 2b
Farahat and colleagues ⁵⁰ 2015 United States	OSCE ^e	Assess preparedness for clinical placement. OSCE repeated in 3 consecutive tutorials. Structure: 15-min orientation and 7 stations 1. 15-min chart review 2. 20-min simulated patient encounter 3. 25-min charting 4. 10-min health care professional interaction 5. 25-min article reading and questions 6. 20-min video observation of a dietitian client consultation 7. 40-min debrief. Trainee viewed their encounter (station 2) later and answered self-reflection questions	University clinic (Clinical Skills Education Centre)	University	Shows	Level 2a
Fiedler and colleagues ⁵¹ 1981 United States	Performance evaluation instrument	<i>The Counseling Checklist, Indirect Patient Care</i> Assess performance during hospital placement. Structure: 5 categories each with criteria describing the range of performance. 1. Verbal communication 2. Nonverbal communication 3. Interpersonal relationships 4. Organization 5. Application of knowledge	Workplace (hospital)	NR	Does	Level 2b
Gibson and Davidson ⁴⁴ 2015 Australia	OSCE	Assess communication skills. OSCE repeated on 2 occasions. Structure: 4 stations (20-min each) 1. Patient referral and planning 2. Simulated patient interview 3. Dietary assessment 4. Debrief	University coursework	University WBA	Shows	Level 1
	Performance evaluation instrument	<i>Abbreviated Calgary-Cambridge</i> instrument adapted to include a global rating scale				

(continued on next page)

Figure 5. (continued) Assessment practices and outcomes for dietetics trainees: Description of studies (1965-2016).

Author(s) Year Country	Assessment type	Assessment purpose and description	Assessment setting	Assessors	Miller's Pyramid ^a	Kirkpatrick's Hierarchy ^b
Hawker and colleagues ⁷⁰ 2010 Australia	OSCE	Assess clinical skills. OSCE completed once. Structure: 3 stations, total 60-min 1. Diet history from a simulated patient 2. Qualitative and semiquantitative diet history assessment 3. Anthropometric measurements and estimation of food portions	University coursework	University	Shows	Level 2b
Henry and Smith ⁵² 2010 United States	Performance evaluation instrument	<i>Feedback and Counseling using Simulation (FOCUS)</i> instrument Assess nutrition counseling skills. Trainees participated in 2 simulation counseling sessions and were assessed using Feedback on Counseling using Simulation tool. Structure: 1. Information-gathering critical points (yes/no). 2. Counseling behaviors (acceptance, genuineness, empathy, communication factors, and counseling process) (6-point scale). 3. Global rating (3-point scale)	University coursework	Simulated patients WBA University	Shows	Level 1
Hipskind and colleagues ⁵³ 2013 United States	OSCE	Develop malnutrition diagnosis skills. Three stations with simulated patients	Workplace	WBA	Shows	Level 1
Horacek and colleagues ⁵³ 2007 United States	Performance evaluation instrument	<i>Modified Dietitians Interviewing Rating Scale (DIRS)</i> Assess counseling skills. Twenty questions classified into 6 categories. 1. Comprehensiveness of the sessions. 2. Organization. 3. Transitional statements. 4. Questioning skills. 5. Rapport. 6. Plan/education. Criteria for accurate and competent skill. Scoring on a scale of 1 (inappropriate/inaccurate/incompetent display of skills) to 5 (accurate/appropriate/competence display of skills)	University coursework	Trainee (self) University	Does	Level 1

(continued on next page)

Figure 5. (continued) Assessment practices and outcomes for dietetics trainees: Description of studies (1965-2016).

Author(s) Year Country	Assessment type	Assessment purpose and description	Assessment setting	Assessors	Miller's Pyramid ^a	Kirkpatrick's Hierarchy ^b
Ingalsbe and Spears ⁵⁵ 1979 United States	Performance evaluation instrument	<i>Clinical Performance Evaluation (CPE)</i> Assess performance during placement. Instrument used to document multiple observations that are compiled and assessed. Structure: 1. Planning and organizing. 2. Observing, reporting, and document. 3. Applying scientific principles to foodservice management. 4. Checking. 5. Relating to instructors, managers, employees, peers, and clientele. 6. Adapting to new or stressful environments. 7. Using learning opportunities. 8. Using creativity. 9. Accepting professional responsibility. 10. Judging professional values. Instrument based on the Critical Incident Technique	University coursework Workplace (hospital, university, schools)	University WBA Trainee (self) Peer	Does	Level 1
Isenring ⁷¹ 2014 Australia	Quiz	Trainee self-assessment of clinical knowledge. Ten online multiple-choice questions with automated feedback	University coursework	University	Knows how	Level 1
Johnson and Hurley ⁵⁶ 1976 United States	Performance evaluation instrument	Assess performance during placement. Structure: 1. Self-education. 2. Organization. 3. Communication. 4. Gathering, recording, and assessing data. 5. Planning nutrition care. 6. Providing nutrition care. 7. Monitoring foodservice. 8. Team utilization and participation. Five-point criteria using Bloom's taxonomy	Workplace (clinical)	WBA University Trainee (self)	Does	Level 2b
Karupaiah and colleagues ⁷⁴ 2016 Malaysia	Performance evaluation instrument	Assess clinical competence. Completed at weeks 1, 6, 9, and 12 (final week). Six components with subcategories based on the Nutrition Care Process, documentation, and personal development. 1. Nutrition assessment. 2. Nutrition diagnosis. 3. Nutrition intervention. 4. Nutrition monitoring and	University clinic and workplace (inpatient and outpatient)	WBA	Does	Level 2b

(continued on next page)

Figure 5. (continued) Assessment practices and outcomes for dietetics trainees: Description of studies (1965-2016).

Author(s) Year Country	Assessment type	Assessment purpose and description	Assessment setting	Assessors	Miller's Pyramid ^a	Kirkpatrick's Hierarchy ^b
		evaluation. 5. Documentation. 6. Personal development (personality and appearance, discipline, soft skills, and confidence). Ratings use a 5-point scale (1=poor to 5=excellent)				
Lake ⁷² 1980 Canada	Performance evaluation instrument	Competency profile for entry-level nutrition/dietetics practice Information not provided on the instrument	Workplace	WBA Trainee (self)	Does	Level 1
Lambert and colleagues ⁶⁴ 2010 United Kingdom	OSCE	Assess clinical skills. OSCE completed once. Structure: Six 10-min stations 1. Initiation of a consultation, assessment of body mass index and feedback to a third part. 2. Obtain dietary information from simulated client. 3. Interpretation of medical notes. 4. Manipulation of diet for the simulated client. 5. Calculation and planning of an enteral feed regimen. 6. Identification of portion sizes and carbohydrate content of specific food	University coursework	University staff WBA	Shows	Level 1
Lennie and Juwah ⁴⁷ 2010 United Kingdom	Performance evaluation instruments	Assess clinical competence. Multiple instruments used for UK Placements B and C. Classified into unwritten assessments (observations of performance with or without witness statements, presentations) and written assessments (case study, project, and reflection)	Workplace	NR	Does	Level 2b
Litchfield and colleagues ⁵⁷ 2002 United States	Examination	Assess clinical judgment and competence (critical thinking and problem solving). Key-feature examination. Each examination included 2 nutrition problems that entry-level dietitians are expected to manage. Each nutrition problem had 3-4 key-features. Examinations administered before and after participation in online nutrition modules (nutrition support, pediatric nutrition, and renal nutrition)	Workplace	University	Knows how	Level 3

*(continued on next page)***Figure 5.** (continued) Assessment practices and outcomes for dietetics trainees: Description of studies (1965-2016).

Author(s) Year Country	Assessment type	Assessment purpose and description	Assessment setting	Assessors	Miller's Pyramid ^a	Kirkpatrick's Hierarchy ^b
Lordly ⁷³ 2007 Canada	Portfolio	Recognition of prior learning for internship (community, administration, and clinical) credit. Trainee assembles portfolio of prior learning (eg, resume, learning narrative, reference letters, work samples, competency development profiles, and evaluation) with feedback from internship coordinator. Assessed using the Dietitians of Canada entry-level competencies	University	University staff WBA	Does	Level 3
Novascone ⁵⁸ 1985 United States	Performance evaluation instrument	Scale for <i>Evaluation Performance in Clinical Dietetics D/BARS</i> ^f . Assess clinical competence. Eight dimension statements 1. Obtaining anthropometric, biochemical, and diagnostic data. 2. Interviewing client/significant other. 3. Analyzing dietary, anthropometric, biochemical, and diagnostic data. 4. Developing client-centred nutrition care plan. 5. Counseling client/significant other. 6. Evaluating the nutrition care plan. 7. Communicating nutrition intervention. 8. Managing the delivery of nutrition care. Each statement has behavioral anchors scaled from 1-7, not applicable and "insufficient opportunity to observe"	Workplace (clinical)	WBA	Does	Level 1

(continued on next page)

Figure 5. (continued) Assessment practices and outcomes for dietetics trainees: Description of studies (1965-2016).

Author(s) Year Country	Assessment type	Assessment purpose and description	Assessment setting	Assessors	Miller's Pyramid ^a	Kirkpatrick's Hierarchy ^b
Olive and colleagues ⁴¹ 1985 United States	Examination	<i>USA National Registration Examination</i> Assess competence. Structure: 1. Normal nutrition. 2. Clinical nutrition. 3. Community nutrition. 4. Management. 5. Foodservice operations. 6. Food science	Workplace	Professional organization	Knows how	Level 2b
	Performance evaluation instrument	Structure: 1. Clinical. 2. Community. 3. Total clinical. 4. Management. 5. Foodservice operations. 6. Total administrative. Performance measured using a 4-point scale (inadequate, somewhat adequate, adequate, and outstanding)			Does	
Palermo and colleagues ⁴⁵ 2016 Australia	Written task	Assess preventative health competence. Proposal and report for a group project at a placement site. Involving plan, implement, and evaluate addressing a public health issue. Criterion-referenced assessment criteria	Workplace (preventative health)	University WBA	Knows	Level 1
	Presentation	Presentation: Group project at a placement site. Involving plan, implement, and evaluate addressing a public health issue. Criterion-referenced assessment criteria			Does	
	OSCE	Case scenario with questions regarding decisions and justification, completed individually by the trainee			Knows how	
Pender and de Looy ⁶⁶ 2004 United Kingdom	OSCE	Assess clinical skills. OSCE completed once. Four stations with 5-min reading and 10-min to complete the task. 1. Discriminatory skills (analyze, prioritize, and discriminate between information in a patient referral letter). 2. Communication skills (diet history and lifestyle assessment with a simulated patient). 3. Communication skills (feedback to a WBA). 4. Interpretation and food knowledge skills	University	University	Shows	Level 2b

(continued on next page)

Figure 5. (continued) Assessment practices and outcomes for dietetics trainees: Description of studies (1965-2016).

Author(s) Year Country	Assessment type	Assessment purpose and description	Assessment setting	Assessors	Miller's Pyramid ^a	Kirkpatrick's Hierarchy ^b
		(knowledge of therapeutic food and application)				
Pender and de Looy ⁶⁵ 2004 United Kingdom	Performance evaluation instrument	Assess clinical competence. Instrument completed on multiple occasions. Structure: 1. Written skills (letters, record cards, case notes/nursing notes, discharge summaries, reports, and individualized diet sheets). 2. Conduct of interview, level of firm information, time management, client specificity, recording information, and complexity). 3. Dietary assessment/history taking (preparation, choice of model, interview technique, evaluation, documentation, and reflection). 4. Oral/presentation skills (pace, volume, vocabulary, eye contact, time management, use of audio-visual aids, enthusiasm for the topic, confident manner, identification of main issues, summary of content, structure, and evidence of collaboration). Skills are measured on a visual analog scale from 0-100 mm (optimal skill) with the midpoint (50 mm) marked (adequate skill). Text descriptions provided at 0 and 100 mm	Workplace (hospital)	WBA	Does	Level 3
Pope and Gines ⁴² 1986 United States	Examination	<i>USA National Registration Examination.</i> Assess competence. Structure NR	Workplace	Professional organization	Knows how	Level 3
Schumacher ⁵⁹ 2014 United States	Performance evaluation instrument	<i>California Critical Thinking Disposition Inventory (CCTDI)</i> Assess critical thinking. Seventy-five items with a 6-point scale to give a total score and 7 subscales in the areas of dispositions of truth seeking, open-mindedness, inquisitiveness,	University	University	Is	Level 2b
<i>(continued on next page)</i>						

Figure 5. (continued) Assessment practices and outcomes for dietetics trainees: Description of studies (1965-2016).

Author(s) Year Country	Assessment type	Assessment purpose and description	Assessment setting	Assessors	Miller's Pyramid ^a	Kirkpatrick's Hierarchy ^b
		analyticity, systematicity, confidence, and maturity				
Schwartz and colleagues ⁴⁶ 2015 United States	Performance evaluation instruments	Assess counseling and communication skills <i>Shortened Calgary Cambridge Observation Guide (CCOG)</i> Twenty-eight items (eg, initiating the session, gathering information, providing structure to the consultation, building a relationship, and closing the session) scored 0 (skill not achieved), 1 (skill needed improvement), or 2 (skill carried out effectively) <i>Behaviour Change Counselling Index (BECCI)</i> Eleven items (eg, patient-centered counseling, encouraged the patient to talk about behavior change, elicit patient feedback, demonstrate empathy, respect patient choices, and exchange ideas) scored 0 (poor), 1 (barely passing), 2 (fair), 3 (good), and 4 (excellent)	University coursework	University	Does	Level 2b
Shanklin and Beach ⁶⁰ 1980 United States	Examination	Assess knowledge. Recall and application questions (distribution of question type dependent on stage within the program with later stages exclusively application questions)	University coursework	University	Knows how	Level 1
Tower and Vosburgh ⁴³ 1976 United States	Performance evaluation instrument	Assess clinical competence. Structure: 1. Verbal communication. 2. Interpersonal relations. 3. Organization of work. 4. Application of knowledge. Each section is divided into behaviors that are rated using a 5-point scale (maximal effectiveness to absence of behavior). Not applicable, not observed option and space for comment. Checklist of behaviors rated as presence or absent	Workplace (hospital)	WBA	Does	Level 2b

(continued on next page)

Figure 5. (continued) Assessment practices and outcomes for dietetics trainees: Description of studies (1965-2016).

Author(s) Year Country	Assessment type	Assessment purpose and description	Assessment setting	Assessors	Miller's Pyramid ^a	Kirkpatrick's Hierarchy ^b
Turner and colleagues ⁶¹ 2000 United States	Performance evaluation instrument	Assess clinical competence. Structure: 1. Obtaining anthropometric biochemical and diagnostic data. 2. Interviewing client/significant other. 3. Analyzing dietary, anthropometric, biochemical, and diagnostic data. 4. Developing client-centered nutrition care plan. 5. Counseling client/significant other. 6. Evaluating nutrition care plan and nutrition intervention. 7. Communicating nutrition intervention. 8. Managing delivery of nutrition care. Each category has behavior/skill statements. Rating based on a 5-point scale (unacceptable to outstanding) with each point having a definition	Workplace	WBA	Does	Level 1
Volders and colleagues ⁶⁹ 2010 Australia	Portfolio	Assess and monitor clinical competence. Individual portfolio compiled throughout WIL ⁹ by trainee. Structure: 1. Introduction section (learning objectives, key policy points, assessment requirements, and reading lists). 2. Subsections for each placement block (objectives checklist, peer and WBA feedback forms, guides, goal setting, self-reflection, self-directed learning activities, and patient log). 3. Appendixes (educational resources and marking guides)	Workplace University coursework	Trainee (self) WBA University	Does	Level 1
Vosburgh and colleagues ³⁸ 1976 United States	Performance evaluation instrument	Instrument described by Tower and Vosburgh ⁴³ 1976	Workplace (hospital)	University	Does	Level 1

*(continued on next page)***Figure 5.** (continued) Assessment practices and outcomes for dietetics trainees: Description of studies (1965-2016).

Author(s) Year Country	Assessment type	Assessment purpose and description	Assessment setting	Assessors	Miller's Pyramid ^a	Kirkpatrick's Hierarchy ^b
Wenberg and Ingersoll ³⁹ 1965 United States	Test	<i>Watson-Glasser Critical Thinking Appraisal (WGCTA)</i> Measure critical thinking and includes problems, statements, arguments, interpretation of data	University	University	Is	Level 3
		<i>Cooperative English Tests (CET)</i> English language ability in reading and written expression			Knows	
		<i>Opinion, Attitudes, and Interest Survey (OAIS)</i> Three categories: 1. Academic promise scale (achiever personality, intellectual quality, and creative personality). 2. Psychological scales (social adjustment, emotional adjustment, and masculine orientation). 3. Educational-vocational interest scales (business interest, humanities interest, social science interest, physical science interest, and biological science interest)			Is	
	Performance evaluation instrument	Evaluation Form A. Behaviors (work performance, organization of work, work with other, communication, application of knowledge, personal appearance, personal development, management, and teaching ability) using a rating scale (NR)	University Workplace		Does	
Wenberg and colleagues ⁴⁰ 1969 United States	Tests	<i>Watson-Glasser Critical Thinking Appraisal, Cooperative English Tests, Opinion, Attitudes, and Interest Survey</i> described by Wenberg and Ingersoll 1965 ³⁹	Workplace	University WBA	Refer to Wenberg and Ingersoll 1965 ³⁹	Level 2b
	Performance evaluation instrument	External criterion. 3 areas: 1. Desire to achieve. 2. With people. 3. Intellectual. Includes assessor comments				

(continued on next page)

Figure 5. (continued) Assessment practices and outcomes for dietetics trainees: Description of studies (1965-2016).

Author(s) Year Country	Assessment type	Assessment purpose and description	Assessment setting	Assessors	Miller's Pyramid ^a	Kirkpatrick's Hierarchy ^b
Whitehead and colleagues ⁶⁷ 2013 United Kingdom	Performance evaluation instrument	DIET-COMMS Assess verbal communication skills. Structure: 20 items scored on 3-point scale (0 not done or NA; 1 partly achieved or attempted; 2 full achieved) with a subjective global assessment (clear pass, borderline, fail)	Workplace	WBA	Does	Level 2b

^aMiller's Pyramid levels: knows, knows how, shows, does, and is.
^bKirkpatrick's Hierarchy: Level 1=participation, Level 2a=modification of attitudes or perceptions, Level 2b=modification of knowledge or skills, Level 3=behavior change, Level 4a=change in organizational practice, Level 4b=benefit to patient or client.
^cWBA=work-based assessor.
^dNR=not reported.
^eOSCE=Objective Structured Clinical Examination.
^fD/BARS=clinical dietetics behaviorally anchored scales
^gWIL=work-integrated learning.

Figure 5. (continued) Assessment practices and outcomes for dietetics trainees: Description of studies (1965-2016).

data analysis was appropriate for most studies (62% and 59%, respectively) (Figure 6, available at www.jandonline.org). For the three qualitative studies evaluated using the Critical Appraisal Skills Programme, the main weaknesses related to insufficient rigor in the reporting of study design and data analysis.^{38,72,73}

Assessment Outcomes

Miller's Pyramid. Assessments addressed all five levels of Miller's Pyramid. Four assessments were at the *knows* level,^{39,40,45,48,68} five at the *knows how* level,^{41,42,45,57,60,71} eight at the *shows* level,^{44,50,52,53,64,66,70} 23 at the *does* level,^{38-41,43,45-47,49,51,54-56,58,61-63,65,67,69,72-74} and three assessments were at the *is* level.^{39,40,59} The *does* level occurred between 1965 and 2016 and the *shows* level was observed after 2004. The *is*, *knows how*, and *knows* levels occurred throughout the date range. The *is* level was assessed using tests that profile personality and socialization (n=4)^{39,40} and a performance evaluation instrument (n=1).⁵⁹ The *does* level was assessed using performance evaluation instruments (n=20),^{38-41,43,46,47,49,51,54-56,58,61,63,65,67,72,74} portfolios (n=3),^{62,69,73} and a presentation (n=1).⁴⁵ The *shows* level was assessed using OSCEs (n=6)^{44,50,53,64,66,70} and performance evaluation instruments (n=2).^{44,52} The *knows how* level was assessed using examinations (n=4),^{41,42,57,60} quizzes (n=1),⁷¹ and an OSCE (n=1).⁴⁵ The *knows* level was assessed using tests (n=3)^{39,40,48} and written methods (n=2).^{45,68}

Kirkpatrick's Hierarchy. The majority of assessments (n=16) focused on Level 1 (participation).^{38,44,45,48,52-55,58,60,61,64,68,69,71,72} These studies reported satisfaction and experience. The assessment in one study was classified as Level 2a (attitudes and perceptions)⁵⁰ and 16 were Level 2b (knowledge and skills).^{40,41,43,46,47,49,51,56,57,59,63,65-67,70,74} Level 2b studies reported changes in trainee performance determined by either an assessor or the trainee. Four studies were classified as Level 3 (behavior change).^{39,42,62,73} These studies reported a willingness by trainees to apply their learning in future practice. No assessments achieved a Level 4a (organization practice) or Level 4b (patient benefits).

Programmatic Assessment. Thirty-six studies were classified using the principles of programmatic assessment framework (Figure 3, available at www.jandonline.org). The framework could not be applied to one study because it reported a range of performance evaluation instruments that were used across 14 different dietetics departments.⁴⁷ The majority of studies (72%) described a single assessment method and did not articulate how the assessment fit within a program.^{39-43,46,48,49,51-53,55-57,59-61,63-68,71,72,74} Instrument validity and reliability were reported in 58% of studies.^{38,41,43-46,48,50,51,53,54,56,58,60,61,64,65,67,70,73,74} Of the studies assessing at the *does* and *is* levels of Miller's Pyramid (n=23), 22% considered the validity of the assessor, which typically involved training.^{38,45,46,54,58,65,67,73} Assessment was considered on a stakes-based continuum in 22% of studies^{38,40,45,62,69-71,73} and in 53% of studies assessment was aligned with curriculum objectives and provided meaningful feedback.^{38,44-46,49,50,52-54,60,61,67-69,71-73} The role of expert judgment in assessment was considered in 36% of studies.^{38,43-45,54,58,64-67,69,70,73}

Qualitative Analysis. Qualitative data from 14 studies^{38,43,50,51,56,58,62,66-69,71-73} were analyzed to produce six themes.

Theme 1: Assessment for learning and professional development. Assessment was shown to provide valuable learning and feedback, supporting the professional and personal development of trainees. Summative assessment had a diagnostic role in identifying deficits in knowledge and ability. Formative assessment was highly valued by trainees and provided the opportunity to practice skills and make mistakes in a safe environment that facilitated development and learning. The provision of feedback was a key function of assessment and was essential for trainee development. Feedback was either external, such as from a WBA, or internal through trainee self-reflection. Trainees placed significant value on constructive evaluation that was provided using various methods such as written feedback, debriefing, peer feedback, and self-reflection. Assessment functioned to develop both professional competencies and personal attribute, including confidence, self-esteem, professional identity, and a commitment to lifelong learning. However, the assessment of personal attributes was recognized as posing a greater challenge than that of professional competence.

Theme 2: Assessment requires motivated and skilled assessors. The credibility and trustworthiness of assessment relied on the ability and attributes of the assessors. All assessors required and sought training and guidance on assessment instruments and processes, as well as performance expectations. Assessors needed to be motivated and engaged with the assessment. They required experience to build confidence and familiarity that alleviated anxiety and concern. The combination of the right person, with training and experience, brought clarity to assessment, which resulted in consistent and appropriate application.

Theme 3: Trainees value authentic and global assessment. Authentic global assessment with sufficient observation of trainee performance using a variety of methods was desired, if not always achieved. Assessment that involved genuine and meaningful practice experiences with real clients was highly valued by trainees because it provided insight into how knowledge and skills are applied to the workplace. Current assessment practices focused on single moments of trainee performance and gave a fragmented view of competence. It was recognized that combining single assessment moments could achieve the desired global and holistic view of a trainee's ability. This was supported by the need for frequent and multiple observations to sufficiently capture performance. Documentation of trainee performance relied heavily on quantitative criterion-based methods although the role of qualitative methods was recognized. Scales were a commonly reported means to document trainee performance. However, scales have limited ability to provide a global picture. Trainees with a borderline performance posed a unique challenge to assessment and their management was not well defined.

Theme 4: Assessment is evolving and context-sensitive. Assessment is an evolving process that required evaluation,

ideally with input from the assessors. Assessors were well placed to identify issues and could contribute to discussions allowing for a consensus regarding modifications. Assessment was not one size fits all, it required tailoring to the setting to which it was applied. As such, assessment needed to be flexible and responsive to the varied needs of different users while maintaining alignment with the curriculum. Contextual idiosyncrasies restricted the intended implementation of assessment creating issues, providing further evidence for the need to evaluate assessment practices.

Theme 5: Poor assessment has negative implications. Poor assessment practices had potential to undermine the experience of the assessor and trainee, and jeopardize the quality of assessment decisions. Efficient and effective assessment design was paramount. Assessment should be practical, organized, and time-efficient. Paperwork that is repetitive and overly detailed hindered efficient assessment. Appropriate descriptors and ratings were important. Incomplete or missing ratings negatively influenced the assessor's ability to describe trainee performance. The cost of poor assessment was described in time and money; institutional reputation, perception and liability; and trainee emotional, learning, and financial investment. Acceptance of assessment by the profession was a key consideration. Poor assessment influenced the quality of assessment decisions and the experience of the users, leading to difficult and inconsistent assessment of performance, which negatively influenced trainee development.

Theme 6: Assessment evokes an emotional response. Assessment elicited an emotional response. WBAs experience confusion and frustration. Trainees reported anxiety, competitiveness, stress, and concern regarding their performance. This was especially true for those performing below standards. However, well-designed assessment was able to overcome the stress and frustration of assessment to deliver an enjoyable and engaging experience with trainees feeling comfortable and highly satisfied with reduced anxiety.

Quantitative Analysis. Twenty-eight studies reported quantitative outcomes, these data were analysed and classified into four categories. These are trainee assessment scores (n=22),^{39,40,44-46,49,50,53,54,57,59-66,68,70,71,74} evaluation of user (assessor and trainee) satisfaction and experience (n=11),^{47,50,53,55,62,64,66,68-71} relationship between assessment scores and trainee variables (n=6),^{40-42,57,60,61} and comparison of assessment scores awarded by an assessor and by the trainee (n=3).^{41,49,54} Several studies had outcomes in more than one category (n=15).^{40,41,49,50,53,54,57,60-62,64,66,68,70,71}

Trainee assessment scores were reported as grades or attributes, with most trainees achieving the standard.^{39,44,46,52,60,62,66,71} Several studies reported a change in assessment scores with the majority of trainees demonstrating an improvement over time.^{44,49,50,53,54,61,63,65,68,70,74} High levels of trainee satisfaction were observed^{47,50,53,55,64,66,68-71} with only one study reporting neutral satisfaction.⁶² Varying levels of correlation were demonstrated among trainee assessment scores and trainee variables such as grade point averages, scores in the USA National Registration Examination for

Dietitians, American College Test, and timing of internships.^{40,42,57,60,61} Agreement was observed between the assessor and trainee score in two studies,^{41,49} whereas a third reported disagreement whereby trainees awarded themselves a higher grade.⁵⁴ Nineteen studies reported the reliability and validity of assessments (Figure 6, available at www.jandonline.org).^{38,41,43,44,48,50-52,54-56,58,60,61,64-67,74}

DISCUSSION

The aim of this systematic review was to identify and evaluate the existing research on assessment practices and outcomes for trainees within the field of dietetics. The large number of studies retrieved indicates a sustained and increasing investment in trainee assessment within the profession. To date, studies have primarily focussed on the development of single assessment instruments and sought educational outcomes at the lower levels of Kirkpatrick's Hierarchy. Many studies focus on assessment at the *does* level of Miller's Pyramid, which aligns with the value trainees place on authentic assessment that is rich in feedback.

The findings of this systematic review indicate an emphasis on the development of individual valid and reliable assessment instruments for use within the university and work-integrated learning setting. This trend is also observed within the medical profession where much of the literature has focused on the psychometric properties of single assessment instruments.⁴ Recently, a paradigm shift has occurred recognizing that single assessment instruments are flawed.^{14,15} Single instruments are unable to capture a phenomenon as complex and context-sensitive as competence, cannot measure trainee development over time, and remain prone to assessor bias despite attempts to achieve validity and reliability.^{11,14,15}

The pursuit of valid and reliable assessment instruments observed in the present study is not unexpected. Historically, attainment of psychometric properties has been considered a means to eliminate the bias and subjectivity of assessment.^{11,14} However, it is now recognized that assessment instruments perform only as well as the individuals who use them.¹⁴ Achievement of validity and reliability is unable to overcome assessor subjectivity that infiltrates interpretations of trainee performance and is reflected in assessments.^{9,10,12} Assessment is subjective, particularly at the *does* level of Miller's Pyramid where assessment circumstances cannot be standardized¹¹ and this is where much of the assessment research in the field of dietetics has occurred. Assessor training, group discussion and judgment, and consensus building all provide avenues for exploring subjectivity and making defensible and reliable assessment decisions.^{12,75} Although this study found that engaged, skilled, and motivated individuals were important for credible assessment, no studies reported specific methods to manage subjectivity.

Dietetics assessors have been reported to be uncomfortable with subjectivity and prefer objective means.¹⁰ This was observed in our review in that the majority of assessment methods employed ratings and scores. Similarly, the medical profession has traditionally focussed on quantitative assessment methods such as tick-boxes and ratings.⁷⁶ Such methods reduce competence into discrete and isolated skills and attributes, which can trivialize and compromise

assessment.¹² Moreover, quantitative assessment methods lack the rich feedback trainees require for improvement.⁷⁶ Qualitative or narrative assessment is able to provide this detailed and meaningful information and has demonstrated rigor.¹⁵ Few studies in our systematic review described the role of narrative, which was generally limited to open-text feedback following quantitative assessment despite the value placed on qualitative narratives by trainees.

This systematic review identified frustrations with current assessment practices in the field of dietetics. Assessors identified that instruments were at times missing necessary criteria to describe trainee ability, that trainee performance was fragmented rather than holistically captured, and instruments were not always fit for purpose when applied in different settings. These findings align with identified challenges in assessment within the literature.¹⁵ Best practice assessments, such as programmatic assessment, seek to overcome these issues by purposefully combining high quality quantitative and qualitative data that are collected over time by multiple skilled assessors.¹² Combining multiple items from different perspectives overcomes the inherent limitations and biases of using single assessment instruments and sole assessors and enables holistic and credible assessment decisions.^{11,12} Such an approach is feasible and has been described in recent medical and dietetics publications that were beyond the scope of this review.^{12,15,77,78} Recent dietetics publications^{15,77} seek to apply best practice to address the challenges of assessment and position the profession as early adopters for change, alongside the medical profession.

Evaluation of teaching and assessment practices has shifted from educational outcomes, whereby trainees express satisfaction with assessment and demonstrate the necessary skills for practice, to the outcomes of health care recipients.¹² Despite this change, there is little evidence that current health professional training and assessment practices produce this ultimate goal.^{11,79} As observed within medicine,¹² dietetics has focused on the lower levels of Kirkpatrick's Hierarchy; that is, the easy-to-measure education and satisfaction outcomes. Evaluation at the higher levels of Kirkpatrick's Hierarchy is challenging because influences may take decades to unfold and are influenced by multiple and varying individual and contextual factors.⁸⁰ This includes the influence of professional regulatory bodies on training and assessment requirements^{2,14}; the diverse settings in which dietitians practice; and the communities they serve, including individuals, populations, institutions, and systems. These challenges and influences necessitate evaluation methods that can capture complex initiatives that occur in dynamic settings over time and involve multiple stakeholders, and where outcomes may be attributed to multiple factors.⁸⁰ Contribution Analysis⁸⁰ and Outcome Evidence⁸¹ are promising evaluation methods. However, there are no published examples of application to the evaluation of health professional training and assessment initiatives. Adopting innovative evaluation methods to determine the influence of dietetics training and assessment practices on health care recipients is essential to positively contribute to improved health outcomes.⁷⁹

Much of the dietetics assessment research has focussed on the *does* level of Miller's Pyramid. Much of this work has occurred in the work-integrated learning setting and results indicate that authentic assessment is desired by

trainees. Authentic, real-world assessment provides deep learning and develops clinical decision-making skills.¹² Although the setting for assessment is important, so too is trainee involvement and engagement. Trainee involvement creates a sense of empowerment and enhances assessment relevance whilst developing lifelong learning skills necessary for practice.^{12,15} A small number of studies in this review provided a limited description of trainee involvement in assessment with further research warranted because trainees have an important role and should hold responsibility.¹⁵ Credible external feedback and promotion of internal self-reflection fosters performance improvement and develops graduate attributes for lifelong learning.^{12,76} The findings of our systematic review support the understanding that feedback is highly valued by trainees and is needed to develop professional and personal competence.⁵

A limitation of this systematic review is the application of the critical appraisal tools. The tools are limited by the type and scope of questions and rely on the information reported in the studies.²³ Although the MERSQI is widely used to appraise health education literature, it does not consider the influence of the initiative on practice²³ or evaluate against best practice. For this reason, we used Miller's Pyramid to appraise the assessment literature and illustrate where research has focused, enabling comparison with other health professions. A novel criteria was developed to evaluate studies using the principles of programmatic assessment, allowing for an analysis against best practice assessment,



PRACTICE IMPLICATIONS

What Is the Current Knowledge on this Topic?

Assessment is key to health professional training. Programmatic assessment is best practice yet single quantitative means are common. Approaches to assess dietetics trainees have not been evaluated.

How Does this Research Add to Knowledge on this Topic?

Assessment focusses on single quantitative tools with low level outcomes. Few studies consider the trainee or practitioner role in assessment. Current practices do not reflect best practice.

How Might this Knowledge Influence Current Dietetics Practice?

There is a gap between current and best practice assessment. Initiatives must consider holistic programmatic assessment, quantitative and qualitative approaches, and outcomes for care recipients.

thus identifying gaps in knowledge and exposing future research opportunities. The inclusion of qualitative data enhanced and complemented the interpretation of quantitative data. This was especially important because a meta-analysis could not be conducted with the quantitative data due to heterogeneity. Unpublished assessment initiatives may not have been included in this review due to the search strategy which may have limited the breadth of identified assessment methods and instruments.

CONCLUSIONS

The findings of this systematic review offer valuable insight into current assessment practices for dietetics trainees and illustrate directions for future research. Results are similar to that observed by other professions and suggest a commonality in assessment practices for health professionals. The large body of evidence identified in this review illustrates significant research activity by the profession into trainee assessment methods and practices. To date, the focus has been on the development of single instruments applied within one institution and outcomes that focus on trainee satisfaction and education. There is a gap between current assessment and best practice that provides an opportunity for innovation. The dietetics profession needs to move beyond describing the psychometric properties of single instruments and instead design programs of assessment that purposefully combine quantitative and qualitative data to provide maximal trainee feedback whilst enabling credible and defensible assessment decisions. The subjective nature of assessment needs to be embraced through investment in assessor training, group discussion, and consensus building. Robust evaluation methods must be adopted to comprehensively understand the influence of assessment across all levels of Kirkpatrick's Hierarchy, including all recipients of dietetics practice. Such transformation will provide the rigor and evidence-based needed to produce competent dietitians who work across multiple practice settings to meet the current and future health needs of the population.

References

1. Rhea M, Bettles C. Future changes driving dietetics workforce supply and demand: Future scan 2012-2022. *J Acad Nutr Diet.* 2012;112(3 suppl):S10-S24.
2. Palermo C. Growing professional competence in nutrition and dietetics. *Nutr Diet.* 2015;72(2):96-100.
3. Kicklighter JR, Cluskey MM, Hunter AM, Nyland NK, Spear BA. Council on future practice visioning report and consensus agreement for moving forward the continuum of dietetics education, credentialing, and practice. *J Acad Nutr Diet.* 2013;113(12):1710-1732.
4. Schuwirth LWT, Van der Vleuten CPM. Programmatic assessment: From assessment of learning to assessment for learning. *Med Teach.* 2011;33(6):478.
5. Boud D. *Assessment 2020: Seven Propositions for Assessment Reform in Higher Education.* Sydney, Australia: Australian Learning and Teaching Council; 2010.
6. Swanwick T. *Understanding Medical Education: Evidence, Theory and Practice.* 1st ed. New Jersey, NJ: Wiley-Blackwell; 2010.
7. Wass V, Shatzer J, Jones R. Assessment of clinical competence. *Lancet.* 2001;357(9260):945-949.
8. Holmboe ES. Realizing the promise of competency-based medical education. *Acad Med.* 2015;90(4):411-413.

9. Bacon R, Holmes K, Palermo C. Exploring subjectivity in competency-based assessment judgements of assessors. *Nutr Diet.* 2017;74(4):357-364.
10. Palermo C, Beck EJ, Chung A, et al. Work-based assessment: Qualitative perspectives of novice nutrition and dietetics educators. *J Hum Nutr Diet.* 2014;27(5):513-521.
11. van der Vleuten CPM, Schuwirth LWT, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Med Teach.* 2012;34(3):205-214.
12. Lockyer J, Carraccio C, Chan MK, et al. Core principles of assessment in competency-based medical education. *Med Teach.* 2017;39(6):609-616.
13. van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: From methods to programmes. *Med Educ.* 2005;39(3):309-317.
14. Holmboe ES, Batalden P. Achieving the desired transformation: Thoughts on next steps for outcomes-based medical education. *Acad Med.* 2015;90(9):1215-1223.
15. Palermo C, Gibson SJ, Dart J, Whelan K, Hay M. Programmatic assessment of competence in dietetics: A new frontier. *J Acad Nutr Diet.* 2017;117(2):175.
16. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ.* 2015;350:g7647.
17. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ.* 2011;343:d5928.
18. *Endnote X8.* Philadelphia, PA: Clarivate Analytics; 2017.
19. *Covidence Systematic Review Software.* Melbourne, Australia: Veritas Health Innovation; 2018.
20. *Excel.* Redmond, WA: Microsoft Corp; 2016.
21. Harden RM, Grant J, Buckley G, Hart IR. BEME guide no 1: Best evidence medical education. *Med Teach.* 1999;21(6):553-562.
22. Butler A, Hall H, Copnell B. A guide to writing a qualitative systematic review protocol to enhance evidence-based practice in nursing and health care. *Worldviews Evid Based Nurs.* 2016;13(3):241-249.
23. Reed DA, Cook DA, Beckman TJ, Levine RB. Association between funding and quality of published medical education research. *JAMA.* 2007;298(9):1002-1009.
24. Reed DA, Beckman TJ, Wright SM, Levine RB, Kern DE, Cook DA. Predictive validity evidence for medical education research study quality instrument scores: Quality of submissions to JGIM's Medical Education special issue. *J Gen Intern Med.* 2008;23(7):903-907.
25. de Jong J, Visser M, Van Dijk N, van der Vleuten C, Wieringa-de Waard M. A systematic review of the relationship between patient mix and learning in work-based clinical settings. A BEME systematic review: BEME guide No. 24. *Med Teach.* 2013;35(6):E1181.
26. Ahmed R, Farooq A, Storie D, Hartling L, Oswald A. Building capacity for education research among clinical educators in the health professions. A BEME systematic review of the outcomes of interventions: BEME Guide No. 34. *Med Teach.* 2016;38(2):123.
27. Cook DA, Reed DA. Appraising the quality of medical education research methods: The Medical Education Research Study Quality Instrument and the Newcastle-Ottawa Scale-Education. *Acad Med.* 2015;90(8):1067-1076.
28. Better Value Healthcare Ltd. Critical Appraisal Skills Programme (CASP) checklists. <https://casp-uk.net/wp-content/uploads/2018/03/CASP-Qualitative-Checklist-Download.pdf>. Accessed April 20, 2017.
29. Cruess RL, Cruess SR, Steinert Y. Amending Miller's pyramid to include professional identity formation. *Acad Med.* 2016;91(2):180-185.
30. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.* 1990;(suppl):S63-S67.
31. Littlewood S, Dornan T, Ypinazar V, Margolis SA, Scherpbier A, Spencer J. Early practical experience and the social responsiveness of clinical education: Systematic review. *BMJ.* 2005;331(7513):387-391.
32. Sullivan GM. Deconstructing quality in education research. *JGME.* 2011;3(2):121-124.
33. Chan A, Purcell A, Power E. A systematic review of assessment and intervention strategies for effective clinical communication in culturally and linguistically diverse students. *Med Educ.* 2016;50(9):898-911.
34. Terry R, Hing W, Orr R, Milne N. Do coursework summative assessments predict clinical performance? A systematic review. *BMC Med Educ.* 2017;17(40).
35. The Joanna Briggs Institute. *Joanna Briggs Institute Reviewers' Manual: 2014 edition.* Adelaide, Australia: The Joanna Briggs Institute of the University of Adelaide; 2014.
36. Thomas H, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Med Res Methodol.* 2008;8(45):1-10.
37. Windish DM, Reed DA, Boonyasai RT, Chakraborti C, Bass EB. Methodological rigor of quality improvement curricula for physician trainees: A systematic review and recommendations for change. *Acad Med.* 2009;84(12):1677-1692.
38. Vosburgh PM, Tower JB, Peckos PS, Mason M. Development of a rating scale to measure learning in clinical dietetics. II. Pilot test. *J Am Diet Assoc.* 1976;68(5):446-449.
39. Wenberg BG, Ingersoll RW. Medical dietetics. II. The development of evaluative techniques. *J Am Diet Assoc.* 1965;47(4):298-300.
40. Wenberg BG, Ingersoll RW, Dohner CW. Evaluation of dietetic interns. *J Am Diet Assoc.* 1969;54(4):297-301.
41. Olive MS, Tseng RY, McProud LM. Comparison of two methods of educating dietitians. *J Am Diet Assoc.* 1985;85(12):1607-1611.
42. Pope J, Gines DJ. Correlates of success of graduates of a coordinated undergraduate program. *J Am Diet Assoc.* 1986;86(8):1022-1027.
43. Tower JB, Vosburgh PM. Development of a rating scale to measure learning in clinical dietetics. I. Theoretical considerations and method of construction. *J Am Diet Assoc.* 1976;68(5):440-445.
44. Gibson SJ, Davidson ZE. An observational study investigating the impact of simulated patients in teaching communication skills in preclinical dietetic students. *J Hum Nutr Diet.* 2016;29(4):529-536.
45. Palermo C, Davidson ZE, Hay M. A cross-sectional study exploring the different roles of individual and group assessment methods in assessing public health nutrition competence. *J Hum Nutr Diet.* 2016;29(4):523-528.
46. Schwartz VS, Rothpletz-Puglia P, Denmark R, Byham-Gray L. Comparison of standardized patients and real patients as an experiential teaching strategy in a nutrition counseling course for dietetic students. *Patient Educ Couns.* 2015;98(2):168-173.
47. Lennie SC, Juwah C. Exploring assessment for learning during dietetic practice placements. *J Hum Nutr Diet.* 2010;23(3):217-223.
48. Chambers MJ, Hubbard RM. Assessing achievement for minimal academic competency. II. Validity and reliability. *J Am Diet Assoc.* 1978;73(1):31-35.
49. Cochran SB, Spears MC. Critical incident evaluations in the clinical component of a CUP: Student self-assessment and instructors' ratings: A comparison. *J Am Diet Assoc.* 1980;76(3):253-257.
50. Farahat E, Rice G, Daher N, Heine N, Schneider L, Connell B. Objective Structured Clinical Examination (OSCE) improves perceived readiness for clinical placement in nutrition and dietetic students. *J Allied Health.* 2015;44(4):208-214.
51. Fiedler KM, Beach BL, Hayman J. Dietetic performance evaluation: Establishment of validity and reliability. *J Am Diet Assoc.* 1981;78(2):149-153.
52. Henry BW, Smith TJ. Evaluation of the FOCUS (Feedback on Counseling Using Simulation) instrument for assessment of client-centered nutrition counseling behaviors. *J Nutr Educ Behav.* 2010;42(1):57-62.
53. Hipskind P, DeChicco R, Gallagher R, Hamilton C. Successful use of objective structured clinical examinations as part of a comprehensive education program to train registered dietitians and dietetic interns in a large health system to diagnose malnutrition. *J Acad Nutr Diet.* 2013;113(9):A-39.
54. Horacek TM, Salomon JE, Nelsen EK. Evaluation of dietetic students' and interns' application of a lifestyle-oriented nutrition-counseling model. *Patient Educ Couns.* 2007;68(2):113-120.
55. Ingalsbe N, Spears MC. Development of an instrument to evaluate critical incident performance. *J Am Diet Assoc.* 1979;74(2):134-140.
56. Johnson CA, Hurley RS. Design and use of an instrument to evaluate students' clinical performance. *J Am Diet Assoc.* 1976;68(5):450-453.

57. Litchfield RE, Oakland MJ, Anderson J. Promoting and evaluating competence in on-line dietetics education. *J Am Diet Assoc.* 2002;102(10):1455.
58. Novascone MA. Evaluating student performance in clinical dietetics. *J Am Diet Assoc.* 1985;85(6):706-710.
59. Schumacher JR. Critical-thinking dispositions among dietetic interns at the completion of their internship. *J Fam Consum Sci.* 2014;106(2):55-57.
60. Shanklin CW, Beach BL. Implementation and evaluation of a competency-based dietetic program. *J Am Diet Assoc.* 1980;77(4):450-455.
61. Turner RE, Evers WD, Wood OB, et al. Computer-based simulations enhance clinical experience of dietetics interns. *J Am Diet Assoc.* 2000;100(2):183-190.
62. Brennan KM, Lennie SC. Students' experiences and perceptions of the use of portfolios in UK preregistration dietetic placements: A questionnaire-based study. *J Hum Nutr Diet.* 2010;23(2):133-143.
63. Earl SE. Self-appraisal: A formalised tool for improving dietitians' communication. *Hum Nutr Appl Nutr.* 1984;38(2):119-125.
64. Lambert L, Pattison DJ, De Looy AE. Dietetic students' performance of activities in an objective structured clinical examination. *J Hum Nutr Diet.* 2010;23(3):224-229.
65. Pender FT, Looy AE. Monitoring the development of clinical skills during training in a clinical placement. *J Hum Nutr Diet.* 2004;17(1):25-34.
66. Pender FT, Looy AE. The testing of clinical skills in dietetic students prior to entering clinical placement. *J Hum Nutr Diet.* 2004;17(1):17-24.
67. Whitehead KA, Langley-Evans SC, Tischler VA, Swift JA. Assessing communication skills in dietetic consultations: The development of the reliable and valid DIET-COMMS tool. *J Hum Nutr Diet.* 2014;27(Suppl. 2):321-332.
68. Daniels LA, Magarey A. The educational and vocational role of peer assessment in the training and professional practice of dietitians. *Nutr Diet.* 2000;57(1):18-22.
69. Volders E, Tweedie J, Anderson A. Advancements in nutrition and dietetics teaching and learning: evaluation of the student portfolio. *Nutr Diet.* 2010;67(2):112-116.
70. Hawker JA, Walker KZ, Barrington V, Andrianopoulos N. Measuring the success of an objective structured clinical examination for dietetic students. *J Hum Nutr Diet.* 2010;23(3):212-216.
71. Isenring E. Facilitating deep student learning with online quizzes despite an increased class size in a Master of Dietetics studies program. *Top Clin Nutr.* 2014;29(2):150-157.
72. Lake RE. Trying out the competency profile for entry-level nutrition/dietetics practice. *J Can Diet Assoc.* 1980;41(2):98-101.
73. Lordly D. Dietetic prior learning assessment: Student and faculty experiences. *Can J Diet Pract Res.* 2007;68(4):207-212.
74. Karupaiah T, Reinhard T, Krishnasamy S, Tan S-P, Se C-H. Incorporating the Nutrition Care Process Model into dietetics internship evaluation: A Malaysian university experience. *Nutr Diet.* 2016;73(3):283.
75. Bacon R, Kellett J, Dart J, et al. A Consensus Model: shifting assessment practices in dietetics tertiary education. <https://onlinelibrary.wiley.com/doi/abs/10.1111/1747-0080.12415>. Accessed April 18, 2018.
76. 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.
77. Jamieson J, Jenkins G, Beatty S, Palermo C. Designing programmes of assessment: A participatory approach. *Med Teach.* 2017;39(11):1182-1188.
78. Driessen EW, van Tartwijk J, Govaerts M, Teunissen P, van der Vleuten CPM. The use of programmatic assessment in the clinical workplace: A Maastricht case report. *Med Teach.* 2012;34(3):226.
79. Haji F, Morin MP, Parker K. Rethinking programme evaluation in health professions education: Beyond 'did it work?'. *Med Educ.* 2013;47(4):342-351.
80. Van Melle E, Gruppen L, Holmboe ES, Leslie F, Oandasan I, Frank JR. Using contribution analysis to evaluate competency-based medical education programs: It's all about rigor in thinking. *Acad Med.* 2017;92(6):752-758.
81. Paz-Ybarnegaray R, Douthwaite B. Outcome Evidencing: a method for enabling and evaluating program intervention in complex systems. *Am J Eval.* 2017;38(2):275-293.

AUTHOR INFORMATION

J. Jamieson is a doctoral degree candidate, Department of Nutrition and Dietetics, Monash University, Notting Hill, Victoria, Australia, and a lecturer, School of Medical and Health Sciences, Edith Cowan University, Joondalup, Western Australia, Australia. C. Palermo is an associate professor, Department of Nutrition and Dietetics, and an associate professor, Monash Centre for Scholarship in Health Education, Monash University, Notting Hill, Victoria, Australia. M. Hay is a professor, Monash Institute for Health and Clinical Education, Faculty of Medicine, Department of Nursing and Health Sciences, Monash University, Clayton, Victoria, Australia. S. Gibson is a senior lecturer, Department of Nutrition and Dietetics, Monash University, Notting Hill, Victoria, Australia.

Address correspondence to: Janica Jamieson, Grad Dip Diet, School of Medical and Health Sciences, Edith Cowan University, 270 Joondalup Dr, Joondalup, Western Australia, 6027 Australia. E-mail: jjamieson@ecu.edu.au

STATEMENT OF POTENTIAL CONFLICT OF INTEREST

No potential conflict of interest was reported by the authors.

FUNDING/SUPPORT

None to report.

AUTHOR CONTRIBUTIONS

J. Jamieson was responsible for all components of the research and manuscript. C. Palermo, S. Gibson, and M. Hay contributed to the research protocol and reviewed the manuscript. J. Jamieson, C. Palermo, and S. Gibson screened the studies. J. Jamieson and S. Gibson undertook the qualitative analysis. C. Palermo, S. Gibson, and M. Hay reviewed the extracted data for accuracy.

Participant		Intervention		
MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE and Versions(R)				
Dietetic/	AND	Students(+)/	AND	Assess*
OR		OR		OR
Nutritionist/		Student*		Apprais*
OR		OR		OR
Dietitian*		Learner*		Evaluat*
OR		OR		OR
Dietician*		Intern		Examination*
OR		OR		OR
Dietetic*		Interns		Exam
		OR		OR
	Internship*	Exams		
		OR		
		Competen*		
EMBASE (Ovid)				
Dietitian/	AND	Student (+)/	AND	Assess*
OR		OR		OR
Dietetic*		Student*		Apprais*
OR		OR		OR
Dietitian*		Learner*		Evaluat*
OR		OR		OR
Dietician*		Intern		Examination*
		OR		OR
		Interns		Exam
		OR		OR
	Internship*	Exams		
		OR		
		Competen*		
CINAHL Plus (EBSCO)				
Dietitians/	AND	Students (+)/	AND	Assess*
OR		OR		OR
Dietitian*		Student*		Apprais*
OR		OR		OR
Dietician*		Learner*		Evaluat*
OR	OR	OR		

(continued on next page)

Figure 1. Assessment practices and outcomes for dietetics trainees: Database search strategy. *Indicates truncation.

Participant		Intervention		
Dietetic*		Intern		Examination*
OR		OR		OR
Nutritionist*		Interns		Exam
		OR		OR
		Internship*		Exams
				OR
				Competen*
ERIC (ProQuest)				
Dietetics(+)/	AND	Students (+)/		Assess*
OR		OR		OR
Dietitian*		Student*		Apprais*
OR		OR		OR
Dietician*		Learner*		Evaluat*
OR		OR		OR
Dietetic*		Intern		Examination*
		OR		OR
		Interns		Exam
		OR		OR
	Internship*		Exams	
			OR	
			Competen*	

Figure 1. (continued) Assessment practices and outcomes for dietetics trainees: Database search strategy. *Indicates truncation.

PICO ^a	Inclusion	Exclusion
Participants	<ul style="list-style-type: none"> Dietetics trainees regardless of type of degree or stage in their studies 	<ul style="list-style-type: none"> Nondietetics trainees Dietetics assistants Dietetics trainees forming a mixed-discipline cohort where their data cannot be isolated
Intervention	<ul style="list-style-type: none"> Assessment intervention regardless of type, frequency, duration, or timing 	<ul style="list-style-type: none"> No assessment intervention Teaching intervention Learning intervention
Outcomes	<ul style="list-style-type: none"> Trainee participation Trainee learning (attitudes, perceptions, knowledge, and skills) Trainee behavior change Organization practice change Patient/client benefits Other relevant outcomes 	<ul style="list-style-type: none"> No outcomes reported
Research design	<ul style="list-style-type: none"> Primary data studies of quantitative, qualitative, and mixed-method design 	<ul style="list-style-type: none"> Secondary data studies
Setting	<ul style="list-style-type: none"> No restriction 	
Language	<ul style="list-style-type: none"> English language Language other than English where translation can be undertaken 	<ul style="list-style-type: none"> Unable to translate

Figure 2. Assessment practices and outcomes for dietetics trainees: Study eligibility criteria. ^aPICO=patient, problem, or population; intervention; comparison, control, or comparator; and outcome framework.

Criterion	Explanation	Questions
Multiple items	Multiple assessment moments are needed to construct a reliable profile of a trainee's performance. Single assessment items can be influenced by the context, can assess only one level of Miller's Pyramid, and do not provide information on progress	Is the assessment part of a program of assessment? Are there multiple assessments over time to capture performance and development?
Instrument validity	Assessment at the first 3 levels of Miller's Pyramid can be standardized through instrument design. Reliability is considered as agreement between assessors or a correlation coefficient ≥ 0.5 . Validity includes demonstrated face, content, concurrent, predictive, content, curricular, construct, or criterion.	Does the assessment demonstrate validity and reliability?
User validity	Nonstandardized assessment at the top levels of Miller's Pyramid occurs in the real world and, therefore, standardization cannot be achieved. Instead, the user of the assessment becomes more important than the instrument itself. Users require training, facilitation, feedback, and skill development in the assessment (more than providing assessment orientation information) and need to develop expertise, have sufficient time to complete the assessment, and be committed to the task. Assessments that target the <i>does</i> and <i>is</i> levels are considered for this criterion.	Does the assessment consider the user? Are the users trained? Do they receive feedback on their assessments? Is the training and feedback sufficient to develop expert users? How seriously do assessors take the assessment? Do assessors have sufficient time for the assessment?
Stakes-based	Assessment is viewed on a continuum from low to high stakes. Low-stake assessment has minimal consequences for the trainee and can be based on a single assessment. High-stake assessment has significant consequences for the trainee such as certification or graduation and should be based on multiple data points. Low-stake assessments can be combined to inform high-stake decisions.	Are significant decisions, for the trainee, based on multiple assessment data points? Are single assessments low stakes?
Driving learning	Assessment drives learning and as such should align with curriculum objectives and facilitate deep learning practices. Assessment should provide meaningful information for the trainee, including quantitatively and qualitatively. The exception is mastery skills that require certification (eg, resuscitation) although feedback should be provided.	Does the assessment task align with curriculum objectives? Are the curriculum objectives described? Does the assessment task provide meaningful feedback to the trainee? Does it encourage feedback seeking and giving attributes for the trainee?

(continued on next page)

Figure 3. Assessment practices and outcomes for dietetics trainees: Programmatic assessment framework.¹¹

Criterion	Explanation	Questions
Expert judgment	Competence is complex and interpreting performance involves human judgment. Although subjectivity can be minimized through assessment design it cannot be completely removed. Expert judgment is necessary at each single assessment point, when combining assessment data and when making high-stake decisions based on multiple assessment data points. Subjectivity can be ameliorated by sufficient sampling and procedural measures (eg, critique of assessment data, obtaining additional information, and deliberation processes).	Does the assessment contain expert judgment? Is it acknowledged and discussed?

Figure 3. (continued) Assessment practices and outcomes for dietetics trainees: Programmatic assessment framework.¹¹

Author(s) and year	MERSQI score	Validity and reliability ^a	Key quantitative outcomes	Programmatic Assessment Criteria ^b					
				Multiple items	Instrument validity	User validity	Stakes-based	Driving learning	Expert judgment
Brennan and Lennie ⁶² 2010	8	—	Evaluation: Neutral trainee satisfaction. Portfolio burdensome and ineffective at measuring performance and competence. Trainees wanted more information. Trainee assessment scores: B placement 97.4% pass, 0.9% fail, and 1.7% in progress C placement 55.2% pass, 5.3% fail, and 36.8% in progress	Y	N	N	Y	N	N
Chambers and Hubbard ⁴⁸ 1978	13.5	Content validity: demonstrated Curricular validity: not demonstrated Concurrent validity: poor/ moderate (Test A vs UG ^c GPA ^d $r=0.519$ and vs course GPA $r=0.654$. Test B vs UG GPA $r=0.386$ and vs course GPA 0.613)	—	N	Y	—	N	N	N

(continued on next page)

Figure 6. Assessment practices and outcomes for dietetics trainees: Medical Education Research Study Quality Instrument (MERSQI)²³ score; validity and reliability; quantitative outcomes, including change (Δ); and programmatic assessment framework for included studies (1965-2016). ^aCoefficient interpretation: <0.5 poor, 0.5-0.75 moderate; 0.75-0.90 good; >0.90 excellent. ^bRefer to Figure 3. ^cUG=undergraduate. ^dGPA=grade point average. ^eND=no difference. ^fOSCE=Objective Structure Clinical Examination. ^gCVD=cardiovascular disease. ^hDM=diabetes mellitus. ⁱADIME=Assessment Diagnosis Intervention Monitoring and Evaluation. ^jPG=postgraduate. ^kACT=American College Test. ^lCCOG=Shortened Calgary Cambridge Observation Guide. ^mBECCI=Behaviour Change Counselling Index. ⁿCET=Cooperative English Tests. ^oWGCTA=Watson-Glasser Critical Thinking Appraisal. ^pOAIS=Opinion, Attitudes, and Interest Survey.

Author(s) and year	MERSQI score	Validity and reliability ^a	Key quantitative outcomes	Programmatic Assessment Criteria ^b					
				Multiple items	Instrument validity	User validity	Stakes-based	Driving learning	Expert judgment
		Interrater reliability: moderate (Test A $r=0.749$ and Test B $r=0.688$)							
Cochran and Spears ⁴⁹ 1980	10	—	Trainee vs assessor score: Agreement at 3 points during placement (first point 75.6%, second point 88.4%, third point 90.9%). ND ^e in agreement based on trainee GPA Trainee performance: Increase over placement	N	N	N	N	Y	N
Daniels and Magarey ⁶⁸ 2000	6.5	—	Evaluation: Trainees satisfied (87% overall agreement). Peer grades were appropriate (77% agreement) Trainees want to contribute to assessment criteria (71% agreement) Trainee assessment scores: Self-reported improvement (learning, confidence, self-evaluation, giving, and receiving feedback)	N	N	—	N	Y	N

(continued on next page)

Figure 6. (continued) Assessment practices and outcomes for dietetics trainees: Medical Education Research Study Quality Instrument (MERSQI)²³ score; validity and reliability; quantitative outcomes, including change (Δ); and programmatic assessment framework for included studies (1965-2016). ^aCoefficient interpretation: <0.5 poor, 0.5-0.75 moderate; 0.75-0.90 good; >0.90 excellent. ^bRefer to Figure 3. ^cUG=undergraduate. ^dGPA=grade point average. ^eND=no difference. ^fOSCE=Objective Structure Clinical Examination. ^gCVD=cardiovascular disease. ^hDM=diabetes mellitus. ⁱADIME=Assessment Diagnosis Intervention Monitoring and Evaluation. ^jPG=postgraduate. ^kACT=American College Test. ^lCCOG=Shortened Calgary Cambridge Observation Guide. ^mBECCI=Behaviour Change Counselling Index. ⁿCET=Cooperative English Tests. ^oWGCTA=Watson-Glasser Critical Thinking Appraisal. ^pOAIS=Opinion, Attitudes, and Interest Survey.

Author(s) and year	MERSQI score	Validity and reliability ^a	Key quantitative outcomes	Programmatic Assessment Criteria ^b					
				Multiple items	Instrument validity	User validity	Stakes-based	Driving learning	Expert judgment
Earl ⁶³ 1984	5.5	—	Trainee assessment scores: Self-reported improvement (group presentation ability and self-evaluation)	N	N	N	N	N	N
Farahat and colleagues ⁵⁰ 2015	9.5	Content validity: Demonstrated	Evaluation: Trainees satisfied. Superior to hospital setting (76% agreement), want OSCE ^f to continue (81%), realistic (92% agreement). Collaboration with other health professionals prepared trainees for practice (78% agreement). Trainee assessment scores: Self-reported improvement in readiness for practice, scale 1-10 (pre 4.9±2.5 and post 5.8±1.9; $P \leq 0.05$) Self-reported improvement: prioritizing work, leadership to achieve	Y	Y	—	N	Y	N

(continued on next page)

Figure 6. (continued) Assessment practices and outcomes for dietetics trainees: Medical Education Research Study Quality Instrument (MERSQI)²³ score; validity and reliability; quantitative outcomes, including change (Δ); and programmatic assessment framework for included studies (1965-2016). ^aCoefficient interpretation: <0.5 poor, 0.5-0.75 moderate; 0.75-0.90 good; >0.90 excellent. ^bRefer to Figure 3. ^cUG=undergraduate. ^dGPA=grade point average. ^eND=no difference. ^fOSCE=Objective Structure Clinical Examination. ^gCVD=cardiovascular disease. ^hDM=diabetes mellitus. ⁱADIME=Assessment Diagnosis Intervention Monitoring and Evaluation. ^jPG=postgraduate. ^kACT=American College Test. ^lCCOG=Shortened Calgary Cambridge Observation Guide. ^mBECCL=Behaviour Change Counselling Index. ⁿCET=Cooperative English Tests. ^oWGCTA=Watson-Glasser Critical Thinking Appraisal. ^pOAIS=Opinion, Attitudes, and Interest Survey.

Author(s) and year	MERSQI score	Validity and reliability ^a	Key quantitative outcomes	Programmatic Assessment Criteria ^b					
				Multiple items	Instrument validity	User validity	Stakes-based	Driving learning	Expert judgment
			client outcomes, education and counseling, Nutrition Care Process, and referring to and being familiar with health professionals. No self-reported improvement: using evidence-based practice; communicating with health professionals, dietitians, assessors, and clients from diverse populations; answering client questions; monitoring and evaluation; and self-evaluation.						
Fiedler and colleagues ⁵¹ 1981	10.5	Interrater reliability: moderate ($r=0.72$)	—	N	Y	N	N	N	N
Gibson and Davidson ⁴⁴ 2015	11	<i>Abbreviated and modified Calgary-Cambridge</i> Face validity: Demonstrated	Trainee assessment scores: First OSCE 74% pass, 17% borderline and 9% failed	Y	Y	—	N	Y	Y

(continued on next page)

Figure 6. (continued) Assessment practices and outcomes for dietetics trainees: Medical Education Research Study Quality Instrument (MERSQI)²³ score; validity and reliability; quantitative outcomes, including change (Δ); and programmatic assessment framework for included studies (1965-2016). ^aCoefficient interpretation: <0.5 poor, 0.5-0.75 moderate; 0.75-0.90 good; >0.90 excellent. ^bRefer to Figure 3. ^cUG=undergraduate. ^dGPA=grade point average. ^eND=no difference. ^fOSCE=Objective Structure Clinical Examination. ^gCVD=cardiovascular disease. ^hDM=diabetes mellitus. ⁱADIME=Assessment Diagnosis Intervention Monitoring and Evaluation. ^jPG=postgraduate. ^kACT=American College Test. ^lCCOG=Shortened Calgary Cambridge Observation Guide. ^mBECCI=Behaviour Change Counselling Index. ⁿCET=Cooperative English Tests. ^oWGCTA=Watson-Glasser Critical Thinking Appraisal. ^pOAIS=Opinion, Attitudes, and Interest Survey.

Author(s) and year	MERSQI score	Validity and reliability ^a	Key quantitative outcomes	Programmatic Assessment Criteria ^b					
				Multiple items	Instrument validity	User validity	Stakes-based	Driving learning	Expert judgment
		OSCE Content validity: Demonstrated	Improvement in assessment score (Δ 2.5 \pm 4.2 marks; $P=0.000$). Trainees who failed the first OSCE had largest improvement (Δ 8.0 \pm 3.7 marks; $P=0.000$), followed by borderline (Δ 5.5 \pm 3.8 marks; $P=0.000$) and pass (Δ 1.3 \pm 3.6; $P=0.000$). ND between trainees with and without video recording of first OSCE						
Hawker and colleagues ⁷⁰ 2010	8.5	—	Evaluation: Trainees satisfied. Fair and objective assessment (93%), briefing useful (95.6% agreement), and understand station requirements (90.1%-100% agreement). Adequacy of station time varied (33.9%-98.7% agreement)	Y	Y	—	Y	N	Y

(continued on next page)

Figure 6. (continued) Assessment practices and outcomes for dietetics trainees: Medical Education Research Study Quality Instrument (MERSQI)²³ score; validity and reliability; quantitative outcomes, including change (Δ); and programmatic assessment framework for included studies (1965-2016). ^aCoefficient interpretation: <0.5 poor, 0.5-0.75 moderate; 0.75-0.90 good; >0.90 excellent. ^bRefer to Figure 3. ^cUG=undergraduate. ^dGPA=grade point average. ^eND=no difference. ^fOSCE=Objective Structure Clinical Examination. ^gCVD=cardiovascular disease. ^hDM=diabetes mellitus. ⁱADIME=Assessment Diagnosis Intervention Monitoring and Evaluation. ^jPG=postgraduate. ^kACT=American College Test. ^lCCOG=Shortened Calgary Cambridge Observation Guide. ^mBECCI=Behaviour Change Counselling Index. ⁿCET=Cooperative English Tests. ^oWGCTA=Watson-Glasser Critical Thinking Appraisal. ^pOAIS=Opinion, Attitudes, and Interest Survey.

Author(s) and year	MERSQI score	Validity and reliability ^a	Key quantitative outcomes	Programmatic Assessment Criteria ^b					
				Multiple items	Instrument validity	User validity	Stakes-based	Driving learning	Expert judgment
			Trainee assessment scores: OSCE score 45%-88% and clinical placement score 39%-96% OSCE lower than clinical placement scores (OSCE=71.7±7.7 and clinical placement 73.1±12.0; $P=0.014$)						
Henry and Smith ⁵² 2010	11.5	Criterion validity: moderate (CVD ^g $r=0.66$ and DM ^h $r=0.61$) Interrater reliability: moderate CVD case ($r=0.72$) and poor DM case ($r=0.18$)	Trainee assessment scores: DM simulation 74±11 and CVD 65±14 (out of 100)	N	N	—	N	Y	N
Hipskind and colleagues ⁵³ 2013	5	—	Evaluation: Trainees satisfied (94% agreement) Trainee performance: Self-reported improvement in malnutrition diagnosis (16% pre and 92% post)	N	Y	—	N	Y	N
Horacek and colleagues ⁵⁴ 2007	13	Face and content validity: Demonstrated	Trainee assessment scores: Self-reported improvement: nutrition knowledge, listening	Y	Y	Y	N	Y	Y

(continued on next page)

Figure 6. (continued) Assessment practices and outcomes for dietetics trainees: Medical Education Research Study Quality Instrument (MERSQI)²³ score; validity and reliability; quantitative outcomes, including change (Δ); and programmatic assessment framework for included studies (1965-2016). ^aCoefficient interpretation: <0.5 poor, 0.5-0.75 moderate; 0.75-0.90 good; >0.90 excellent. ^bRefer to Figure 3. ^cUG=undergraduate. ^dGPA=grade point average. ^eND=no difference. ^fOSCE=Objective Structure Clinical Examination. ^gCVD=cardiovascular disease. ^hDM=diabetes mellitus. ⁱADIME=Assessment Diagnosis Intervention Monitoring and Evaluation. ^jPG=postgraduate. ^kACT=American College Test. ^lCCOG=Shortened Calgary Cambridge Observation Guide. ^mBECCI=Behaviour Change Counselling Index. ⁿCET=Cooperative English Tests. ^oWGCTA=Watson-Glasser Critical Thinking Appraisal. ^pOAIS=Opinion, Attitudes, and Interest Survey.

Author(s) and year	MERSQI score	Validity and reliability ^a	Key quantitative outcomes	Programmatic Assessment Criteria ^b					
				Multiple items	Instrument validity	User validity	Stakes-based	Driving learning	Expert judgment
			skills, questioning skills, rapport building, client-oriented goal setting, philosophy for nutrition counseling, organization of session, empowerment skills, confidence, flexibility, and professionals (all <i>P</i> values <0.001) Trainee vs assessor score: Trainee self-assessment higher than assessors (24.85±3.05 compared with 24.09±1.99; <i>P</i> <0.05)						
Ingalsbe and Spears ⁵⁵ 1979	5	Face and content validity: Not demonstrated	Evaluation: Trainees satisfied	N	N	N	N	N	N
Isenring ⁷¹ 2014	7	—	Evaluation: Trainees satisfied (encouraged completion of pre-reading, enhanced topic understanding, facilitated active participation in learning)	N	N	—	Y	Y	N

(continued on next page)

Figure 6. (continued) Assessment practices and outcomes for dietetics trainees: Medical Education Research Study Quality Instrument (MERSQI)²³ score; validity and reliability; quantitative outcomes, including change (Δ); and programmatic assessment framework for included studies (1965-2016). ^aCoefficient interpretation: <0.5 poor, 0.5-0.75 moderate; 0.75-0.90 good; >0.90 excellent. ^bRefer to Figure 3. ^cUG=undergraduate. ^dGPA=grade point average. ^eND=no difference. ^fOSCE=Objective Structure Clinical Examination. ^gCVD=cardiovascular disease. ^hDM=diabetes mellitus. ⁱADIME=Assessment Diagnosis Intervention Monitoring and Evaluation. ^jPG=postgraduate. ^kACT=American College Test. ^lCCOG=Shortened Calgary Cambridge Observation Guide. ^mBECCI=Behaviour Change Counselling Index. ⁿCET=Cooperative English Tests. ^oWGCTA=Watson-Glasser Critical Thinking Appraisal. ^pOAIS=Opinion, Attitudes, and Interest Survey.

Author(s) and year	MERSQI score	Validity and reliability ^a	Key quantitative outcomes	Programmatic Assessment Criteria ^b					
				Multiple items	Instrument validity	User validity	Stakes-based	Driving learning	Expert judgment
			Trainee assessment scores: 90% passed						
Johnson and Hurley ⁵⁶ 1976	10	Content validity: Demonstrated Interrater reliability: Moderate ($r=0.75$)	—	N	Y	N	N	N	N
Karupaiah and colleagues ⁷⁴ 2016	11.5	Face and content validity: Demonstrated Reliability (Cronbach's alpha): good/ excellent (r values between 0.819 and 0.963) Interrater reliability (intraclass coefficient): good/ excellent (r values between 0.706 and 0.927)	Trainee assessment scores: Improvement in Nutrition Care Plan and documentation (ADIME) ⁱ ($P<0.001$)	N	Y	N	N	N	N
Lake ⁷² 1980	5	—	—	N	N	N	N	Y	N
Lambert and colleagues ⁶⁴ 2010	8	Face and content validity: Demonstrated	Evaluation: Trainees satisfied. Adequate preparation; appropriate design, time (stations 1, 2, and 6), environment, facilities resources, and	N	Y	—	N	N	Y

(continued on next page)

Figure 6. (continued) Assessment practices and outcomes for dietetics trainees: Medical Education Research Study Quality Instrument (MERSQI)²³ score; validity and reliability; quantitative outcomes, including change (Δ); and programmatic assessment framework for included studies (1965-2016). ^aCoefficient interpretation: <0.5 poor, 0.5-0.75 moderate; 0.75-0.90 good; >0.90 excellent. ^bRefer to Figure 3. ^cUG=undergraduate. ^dGPA=grade point average. ^eND=no difference. ^fOSCE=Objective Structure Clinical Examination. ^gCVD=cardiovascular disease. ^hDM=diabetes mellitus. ⁱADIME=Assessment Diagnosis Intervention Monitoring and Evaluation. ^jPG=postgraduate. ^kACT=American College Test. ^lCCOG=Shortened Calgary Cambridge Observation Guide. ^mBECCI=Behaviour Change Counselling Index. ⁿCET=Cooperative English Tests. ^oWGCTA=Watson-Glasser Critical Thinking Appraisal. ^pOAIS=Opinion, Attitudes, and Interest Survey.

Author(s) and year	MERSQI score	Validity and reliability ^a	Key quantitative outcomes	Programmatic Assessment Criteria ^b					
				Multiple items	Instrument validity	User validity	Stakes-based	Driving learning	Expert judgment
			instructions (100% agreement). Insufficient time at some stations Trainee assessment scores: Station 1, 70.9%±12.6%; station 2, 72.4%±10.1%; station 3, 63.0%±10.4%; station 4, 62.5%±11.9%; station 5, 54.6%±17.5%; station 6, 41.9%±16.6%						
Lennie and Juwah ⁴⁷ 2010	11	—	Evaluation: Assessor evaluation. Positive weak correlation between length of training and knowledge of applying assessment instruments ($P=0.000$; $r=0.393$); department training and depth of knowledge of applying instruments ($P=0.000$; $r=0.598$); department and personal experience of instrument and personal knowledge of how to apply ($P=0.031$; $r=0.218$ and $P=0.000$; $r=0.335$, respectively).	—	—	—	—	—	—

(continued on next page)

Figure 6. (continued) Assessment practices and outcomes for dietetics trainees: Medical Education Research Study Quality Instrument (MERSQI)²³ score; validity and reliability; quantitative outcomes, including change (Δ); and programmatic assessment framework for included studies (1965-2016). ^aCoefficient interpretation: <0.5 poor, 0.5-0.75 moderate; 0.75-0.90 good; >0.90 excellent. ^bRefer to Figure 3. ^cUG=undergraduate. ^dGPA=grade point average. ^eND=no difference. ^fOSCE=Objective Structure Clinical Examination. ^gCVD=cardiovascular disease. ^hDM=diabetes mellitus. ⁱADIME=Assessment Diagnosis Intervention Monitoring and Evaluation. ^jPG=postgraduate. ^kACT=American College Test. ^lCCOG=Shortened Calgary Cambridge Observation Guide. ^mBECCI=Behaviour Change Counselling Index. ⁿCET=Cooperative English Tests. ^oWGCTA=Watson-Glasser Critical Thinking Appraisal. ^pOAIS=Opinion, Attitudes, and Interest Survey.

Author(s) and year	MERSQI score	Validity and reliability ^a	Key quantitative outcomes	Programmatic Assessment Criteria ^b					
				Multiple items	Instrument validity	User validity	Stakes-based	Driving learning	Expert judgment
			The longer instruments use the greater belief that assessors were consistent ($P=0.02$; $r=0.302$) and instruments ability to differentiate trainees ($P=0.001$; $r=0.340$). Agreement that assessors are consistent in application of instruments ($P=0.05$). Agreement for a national assessment instrument (87% agreement)						
Litchfield and colleagues ⁵⁷ 2002	12	—	Trainee assessment scores: Online instruction higher than without, only nutrition support and pediatric nutrition significant ($P=0.01$ and $P=0.03$, respectively). Variables: ND between with and without online instruction scores and USA National Registration	N	N	—	N	N	N

(continued on next page)

Figure 6. (continued) Assessment practices and outcomes for dietetics trainees: Medical Education Research Study Quality Instrument (MERSQI)²³ score; validity and reliability; quantitative outcomes, including change (Δ); and programmatic assessment framework for included studies (1965-2016). ^aCoefficient interpretation: <0.5 poor, 0.5-0.75 moderate; 0.75-0.90 good; >0.90 excellent. ^bRefer to Figure 3. ^cUG=undergraduate. ^dGPA=grade point average. ^eND=no difference. ^fOSCE=Objective Structure Clinical Examination. ^gCVD=cardiovascular disease. ^hDM=diabetes mellitus. ⁱADIME=Assessment Diagnosis Intervention Monitoring and Evaluation. ^jPG=postgraduate. ^kACT=American College Test. ^lCCOG=Shortened Calgary Cambridge Observation Guide. ^mBECCI=Behaviour Change Counselling Index. ⁿCET=Cooperative English Tests. ^oWGCTA=Watson-Glasser Critical Thinking Appraisal. ^pOAIS=Opinion, Attitudes, and Interest Survey.

Author(s) and year	MERSQI score	Validity and reliability ^a	Key quantitative outcomes	Programmatic Assessment Criteria ^b					
				Multiple items	Instrument validity	User validity	Stakes-based	Driving learning	Expert judgment
			Examination ($P=0.90$). Positive relationship between GPA and Δ in nutrition support calculation score ($P<0.03$); and USA National Registration Examination and posttest score on nutrition support calculation ($P<0.01$)						
Lordly ⁷³ 2007	8	—	—	Y	Y	Y	Y	Y	Y
Novascone ⁵⁸ 1985	7	Face and content validity: Demonstrated	—	Y	Y	Y	N	N	Y
Olive and colleagues ⁴¹ 1985	13	<i>Performance evaluation instrument:</i> Face and content validity: Demonstrated Reliability: Excellent (clinical questions, $r=0.910$) and moderate (community questions, $r=0.835$)	Trainee assessment scores: Assessor scores higher for UG than PG ^j . Trainee vs assessor score: ND between trainee and assessor scores. Variables: Trainee self-rating correlated with UG internship satisfaction, PG placement satisfaction, PG work	N	Y	N	N	N	N

(continued on next page)

Figure 6. (continued) Assessment practices and outcomes for dietetics trainees: Medical Education Research Study Quality Instrument (MERSQI)²³ score; validity and reliability; quantitative outcomes, including change (Δ); and programmatic assessment framework for included studies (1965-2016). ^aCoefficient interpretation: <0.5 poor, 0.5-0.75 moderate; 0.75-0.90 good; >0.90 excellent. ^bRefer to Figure 3. ^cUG=undergraduate. ^dGPA=grade point average. ^eND=no difference. ^fOSCE=Objective Structure Clinical Examination. ^gCVD=cardiovascular disease. ^hDM=diabetes mellitus. ⁱADIME=Assessment Diagnosis Intervention Monitoring and Evaluation. ^jPG=postgraduate. ^kACT=American College Test. ^lCOG=Shortened Calgary Cambridge Observation Guide. ^mBECCI=Behaviour Change Counselling Index. ⁿCET=Cooperative English Tests. ^oWGCTA=Watson-Glasser Critical Thinking Appraisal. ^pOAIS=Opinion, Attitudes, and Interest Survey.

Author(s) and year	MERSQI score	Validity and reliability ^a	Key quantitative outcomes	Programmatic Assessment Criteria ^b					
				Multiple items	Instrument validity	User validity	Stakes-based	Driving learning	Expert judgment
			experience, and UG trainee age. Correlation between examination scores and trainee-self rating and preceptor scores.						
Palermo and colleagues ⁴⁵ 2016	11.5	—	Trainee assessment scores: Difference between group and individual assessment score (79.7% [52.5%-93.1%] and 75.0% [20.0%-100.0%]; $P=0.003$). Trainees with lower individual score achieved higher group score and trainees with higher individual score achieved lower group score	Y	Y	Y	Y	Y	Y
	5			N	N	—	N	N	Y

(continued on next page)

Figure 6. (continued) Assessment practices and outcomes for dietetics trainees: Medical Education Research Study Quality Instrument (MERSQI)²³ score; validity and reliability; quantitative outcomes, including change (Δ); and programmatic assessment framework for included studies (1965-2016). ^aCoefficient interpretation: <0.5 poor, 0.5-0.75 moderate; 0.75-0.90 good; >0.90 excellent. ^bRefer to Figure 3. ^cUG=undergraduate. ^dGPA=grade point average. ^eND=no difference. ^fOSCE=Objective Structure Clinical Examination. ^gCVD=cardiovascular disease. ^hDM=diabetes mellitus. ⁱADIME=Assessment Diagnosis Intervention Monitoring and Evaluation. ^jPG=postgraduate. ^kACT=American College Test. ^lCCOG=Shortened Calgary Cambridge Observation Guide. ^mBECCI=Behaviour Change Counselling Index. ⁿCET=Cooperative English Tests. ^oWGCTA=Watson-Glasser Critical Thinking Appraisal. ^pOAIS=Opinion, Attitudes, and Interest Survey.

Author(s) and year	MERSQI score	Validity and reliability ^a	Key quantitative outcomes	Programmatic Assessment Criteria ^b					
				Multiple items	Instrument validity	User validity	Stakes-based	Driving learning	Expert judgment
Pender and de Looy ⁶⁶ 2004		Content validity: Not demonstrated	Evaluation: Trainees satisfied (95% agreement) Trainee assessment scores: 11% failed in at least 1 skill area						
Pender and de Looy ⁶⁵ 2004	12.5	Face and content validity: Demonstrated	Trainee assessment scores: Scores improved over placement period for written, interviewing and dietary assessment (all P values ≤ 0.001) but not oral/presentation skills ($P \leq 0.119$)	N	Y	Y	N	N	Y
Pope and Gines ⁴² 1986	12.5	—	Variables: Moderate relationship between Registration Examination scores and ACT ^k English score ($r=0.54$; $P<0.005$); ACT Natural Science ($r=0.58$; $P<0.001$); overall GPA ($r=0.67$; $P<0.001$); GPA science ($r=0.60$; $P<0.001$); GPA nutrition ($r=0.53$; $P<0.001$); final GPA ($r=0.57$; $P<0.001$)	N	N	—	N	N	N

(continued on next page)

Figure 6. (continued) Assessment practices and outcomes for dietetics trainees: Medical Education Research Study Quality Instrument (MERSQI)²³ score; validity and reliability; quantitative outcomes, including change (Δ); and programmatic assessment framework for included studies (1965-2016). ^aCoefficient interpretation: <0.5 poor, 0.5-0.75 moderate; 0.75-0.90 good; >0.90 excellent. ^bRefer to Figure 3. ^cUG=undergraduate. ^dGPA=grade point average. ^eND=no difference. ^fOSCE=Objective Structure Clinical Examination. ^gCVD=cardiovascular disease. ^hDM=diabetes mellitus. ⁱADIME=Assessment Diagnosis Intervention Monitoring and Evaluation. ^jPG=postgraduate. ^kACT=American College Test. ^lCOG=Shortened Calgary Cambridge Observation Guide. ^mBECCI=Behaviour Change Counselling Index. ⁿCET=Cooperative English Tests. ^oWGCTA=Watson-Glasser Critical Thinking Appraisal. ^pOAIS=Opinion, Attitudes, and Interest Survey.

Author(s) and year	MERSQI score	Validity and reliability ^a	Key quantitative outcomes	Programmatic Assessment Criteria ^b					
				Multiple items	Instrument validity	User validity	Stakes-based	Driving learning	Expert judgment
Schumacher ⁵⁹ 2014	7	—	Trainee assessment scores: Strong skills in inquisitiveness (58% trainees). Weak truth seeking (58% trainees). Neither strong nor weak skills for open-mindedness (83% trainees), analytical (83%), systematicity (67%), confidence (75%), and maturity (67%)	N	N	N	N	N	N
Schwartz and colleagues ⁴⁶ 2015	13.5	—	Trainee assessment scores: ND between simulated and real patient in the first (CCOG ^l 1.76±0.22 and 1.63±0.29; <i>P</i> =0.053; BECCI ^m 3.45±0.57 and 3.16±0.73; <i>P</i> =0.096) and second encounter (CCOG 1.72±0.27 and 1.64±0.28; <i>P</i> =0.241; BECCI 3.47±0.54 and 3.24±0.64; <i>P</i> =0.079). Trainees performed well on both the CCOG and BECCI in the first and second encounters ND in score Δ between	N	Y	Y	N	Y	N

(continued on next page)

Figure 6. *(continued)* Assessment practices and outcomes for dietetics trainees: Medical Education Research Study Quality Instrument (MERSQI)²³ score; validity and reliability; quantitative outcomes, including change (Δ); and programmatic assessment framework for included studies (1965-2016). ^aCoefficient interpretation: <0.5 poor, 0.5-0.75 moderate; 0.75-0.90 good; >0.90 excellent. ^bRefer to Figure 3. ^cUG=undergraduate. ^dGPA=grade point average. ^eND=no difference. ^fOSCE=Objective Structure Clinical Examination. ^gCVD=cardiovascular disease. ^hDM=diabetes mellitus. ⁱADIME=Assessment Diagnosis Intervention Monitoring and Evaluation. ^jPG=postgraduate. ^kACT=American College Test. ^lCCOG=Shortened Calgary Cambridge Observation Guide. ^mBECCI=Behaviour Change Counselling Index. ⁿCET=Cooperative English Tests. ^oWGCTA=Watson-Glasser Critical Thinking Appraisal. ^pOAIS=Opinion, Attitudes, and Interest Survey.

Author(s) and year	MERSQI score	Validity and reliability ^a	Key quantitative outcomes	Programmatic Assessment Criteria ^b					
				Multiple items	Instrument validity	User validity	Stakes-based	Driving learning	Expert judgment
			simulated and real patients for CCOG and BECCI						
Shanklin and Beach ⁶⁰ 1980	8.5	Content and curricular validity: Demonstrated	Trainee assessment scores: Scores improved during program with all achieving the minimum level of acceptable performance at the end of program examination. Variables: No relationship among scores and GPA, ACT, and course grade	N	Y	—	N	Y	N
Tower and Vosburgh ⁴³ 1976	6	Face and content validity: Demonstrated	—	N	Y	N	N	N	Y
Turner and colleagues ⁶¹ 2000	14.5	Face and content validity: Demonstrated	Trainee assessment scores: Scores improved over 8-wk placement Variables: Trainees with a delayed internship received higher ratings	N	Y	N	N	Y	N

(continued on next page)

Figure 6. (continued) Assessment practices and outcomes for dietetics trainees: Medical Education Research Study Quality Instrument (MERSQI)²³ score; validity and reliability; quantitative outcomes, including change (Δ); and programmatic assessment framework for included studies (1965-2016). ^aCoefficient interpretation: <0.5 poor, 0.5-0.75 moderate; 0.75-0.90 good; >0.90 excellent. ^bRefer to Figure 3. ^cUG=undergraduate. ^dGPA=grade point average. ^eND=no difference. ^fOSCE=Objective Structure Clinical Examination. ^gCVD=cardiovascular disease. ^hDM=diabetes mellitus. ⁱADIME=Assessment Diagnosis Intervention Monitoring and Evaluation. ^jPG=postgraduate. ^kACT=American College Test. ^lCCOG=Shortened Calgary Cambridge Observation Guide. ^mBECCI=Behaviour Change Counselling Index. ⁿCET=Cooperative English Tests. ^oWGCTA=Watson-Glasser Critical Thinking Appraisal. ^pOAIS=Opinion, Attitudes, and Interest Survey.

Author(s) and year	MERSQI score	Validity and reliability ^a	Key quantitative outcomes	Programmatic Assessment Criteria ^b					
				Multiple items	Instrument validity	User validity	Stakes-based	Driving learning	Expert judgment
Volders and colleagues ⁶⁹ 2010	7	—	<p>Evaluation:</p> <p>Trainees satisfied. Useful aspects: Development of self-evaluation (87%-95% agreement), counseling skill guides (97%-97.5% agreement), regular assessor feedback (95%-97.5% agreement), regular clinical assessor feedback (92%-97.5% agreement), facilitating goal setting (92%-95% agreement), and logging workload (90%-92% agreement). Third-year trainees found checklists helpful to understand portfolio requirements (100% agreement) and fourth-year trainees considered portfolio valuable (92% agreement).</p> <p>Assessors satisfied.</p> <p>Supported assessment, guided feedback and goal setting, clear instruction (95%- 100%</p>	Y	N	N	Y	Y	Y

(continued on next page)

Figure 6. (continued) Assessment practices and outcomes for dietetics trainees: Medical Education Research Study Quality Instrument (MERSQI)²³ score; validity and reliability; quantitative outcomes, including change (Δ); and programmatic assessment framework for included studies (1965-2016). ^aCoefficient interpretation: <0.5 poor, 0.5-0.75 moderate; 0.75-0.90 good; >0.90 excellent. ^bRefer to Figure 3. ^cUG=undergraduate. ^dGPA=grade point average. ^eND=no difference. ^fOSCE=Objective Structure Clinical Examination. ^gCVD=cardiovascular disease. ^hDM=diabetes mellitus. ⁱADIME=Assessment Diagnosis Intervention Monitoring and Evaluation. ^jPG=postgraduate. ^kACT=American College Test. ^lCCOG=Shortened Calgary Cambridge Observation Guide. ^mBECCI=Behaviour Change Counselling Index. ⁿCET=Cooperative English Tests. ^oWGCTA=Watson-Glasser Critical Thinking Appraisal. ^pOAIS=Opinion, Attitudes, and Interest Survey.

Author(s) and year	MERSQI score	Validity and reliability ^a	Key quantitative outcomes	Programmatic Assessment Criteria ^b					
				Multiple items	Instrument validity	User validity	Stakes-based	Driving learning	Expert judgment
			agreement). Supported regular assessment and improved trainee knowledge and ability (72%-85% agreement). Foodservice and patient log less useful (56%-59% agreement)						
Vosburgh and colleagues ³⁸ 1976	6	Face and content validity: Demonstrated	—	Y	Y	Y	Y	Y	Y
Wenberg and Ingersoll ³⁹ 1965	9	—	Trainee assessment scores: Trainees scored higher than national average for the CET ⁿ (>67th percentile) and WGCTA ^o (57th percentile). OAIS ^p scores suggest trainees have strong motivation to achieve	N	N	N	Y	N	N
Wenberg and colleagues ⁴⁰ 1969	11.5	—	Trainee assessment scores: No change in WGCTA, CET, or OAIS scores Variables: None or weak correlations between External Criterion and GPA for WGCTA, CET, or OAIS	N	N	N	N	N	N

(continued on next page)

Figure 6. (continued) Assessment practices and outcomes for dietetics trainees: Medical Education Research Study Quality Instrument (MERSQI)²³ score; validity and reliability; quantitative outcomes, including change (Δ); and programmatic assessment framework for included studies (1965-2016). ^aCoefficient interpretation: <0.5 poor, 0.5-0.75 moderate; 0.75-0.90 good; >0.90 excellent. ^bRefer to Figure 3. ^cUG=undergraduate. ^dGPA=grade point average. ^eND=no difference. ^fOSCE=Objective Structure Clinical Examination. ^gCVD=cardiovascular disease. ^hDM=diabetes mellitus. ⁱADIME=Assessment Diagnosis Intervention Monitoring and Evaluation. ^jPG=postgraduate. ^kACT=American College Test. ^lCCOG=Shortened Calgary Cambridge Observation Guide. ^mBECCI=Behaviour Change Counselling Index. ⁿCET=Cooperative English Tests. ^oWGCTA=Watson-Glasser Critical Thinking Appraisal. ^pOAIS=Opinion, Attitudes, and Interest Survey.

Author(s) and year	MERSQI score	Validity and reliability ^a	Key quantitative outcomes	Programmatic Assessment Criteria ^b					
				Multiple items	Instrument validity	User validity	Stakes-based	Driving learning	Expert judgment
Whitehead and colleagues ⁶⁷ 2013	14.5	Content, face, predictive and construct validity: Demonstrated Interrater reliability: Moderate ($r=0.59$) Intrarater reliability (Spearman's): Good ($r=0.898$)	—	N	Y	Y	N	Y	Y

Figure 6. (continued) Assessment practices and outcomes for dietetics trainees: Medical Education Research Study Quality Instrument (MERSQI)²³ score; validity and reliability; quantitative outcomes, including change (Δ); and programmatic assessment framework for included studies (1965-2016). ^aCoefficient interpretation: <0.5 poor, 0.5-0.75 moderate; 0.75-0.90 good; >0.90 excellent. ^bRefer to Figure 3. ^cUG=undergraduate. ^dGPA=grade point average. ^eND=no difference. ^fOSCE=Objective Structure Clinical Examination. ^gCVD=cardiovascular disease. ^hDM=diabetes mellitus. ⁱADIME=Assessment Diagnosis Intervention Monitoring and Evaluation. ^jPG=postgraduate. ^kACT=American College Test. ^lCCOG=Shortened Calgary Cambridge Observation Guide. ^mBECCI=Behaviour Change Counselling Index. ⁿCET=Cooperative English Tests. ^oWGCTA=Watson-Glasser Critical Thinking Appraisal. ^pOAIS=Opinion, Attitudes, and Interest Survey.