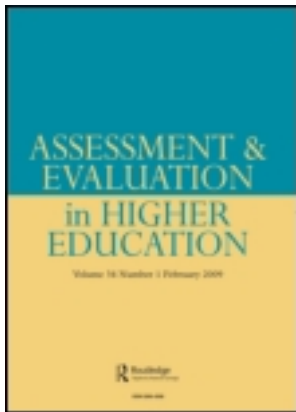


This article was downloaded by: [University Library Utrecht]

On: 24 January 2012, At: 11:32

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Assessment & Evaluation in Higher Education

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/caeh20>

A Mastery Rubric: guiding curriculum design, admissions and development of course objectives

Rochelle E. Tractenberg^a, Jason G. Umans^{b c} & Robert J. McCarter^{d e}

^a Departments of Neurology, Biostatistics, Biomathematics & Bioinformatics, and Psychiatry, Georgetown University School of Medicine, Washington, DC, USA

^b Division of Medicine, Georgetown University School of Medicine, Washington, DC, USA

^c MedStar Research Institute, Hyattsville, MD, USA

^d Biostatistics and Informatics, Children's Research Institute, Children's National Medical Center, Washington, DC, USA

^e Departments of Pediatrics and Epidemiology & Biostatistics, The George Washington University Schools of Medicine and Public Health, Washington, DC, USA

Available online: 14 Jan 2009

To cite this article: Rochelle E. Tractenberg, Jason G. Umans & Robert J. McCarter (2010): A Mastery Rubric: guiding curriculum design, admissions and development of course objectives, *Assessment & Evaluation in Higher Education*, 35:1, 15-32

To link to this article: <http://dx.doi.org/10.1080/02602930802474169>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

A Mastery Rubric: guiding curriculum design, admissions and development of course objectives

Rochelle E. Tractenberg^{a*}, Jason G. Umans^{b,c} and Robert J. McCarter^{d,e}

^aDepartments of Neurology, Biostatistics, Biomathematics & Bioinformatics, and Psychiatry, Georgetown University School of Medicine, Washington, DC, USA; ^bDivision of Medicine, Georgetown University School of Medicine, Washington, DC, USA; ^cMedStar Research Institute, Hyattsville, MD, USA; ^dBiostatistics and Informatics, Children's Research Institute, Children's National Medical Center, Washington, DC, USA; ^eDepartments of Pediatrics and Epidemiology & Biostatistics, The George Washington University Schools of Medicine and Public Health, Washington, DC, USA

This article describes a 'Mastery Rubric' (MR) used to design both the curriculum and the assessments in a new two-year certificate programme intended to train physicians in clinical research skills. The MR for clinical research skills is built around a set of core research skills: critical review of literature; articulation of research objective; development of research design; development of analysis plan; implementation of the study; implementation of the analysis plan and presentation of results. Four distinct levels of performance are described for each skill: beginning, novice, competent and proficient. This rubric outlines and provides a path to mastery of the clinical research skills the certificate programme was designed and funded to target. Using the rubric to design the curriculum ensures that courses will provide instruction in key domains, promotes assessment that demonstrates development in the target skills and knowledge, and encourages reflection and cognitive self-monitoring in the students. It is a flexible, criterion-referenced definition of 'success' for students as well as the programme itself. The criteria are characterised in terms of the skills, habits of mind and organisational principles that can foster excellence in clinical research, but the approach can be generalised.

Keywords: curriculum development, rubrics; higher education; proficiency; research skills

Introduction

There is an urgent need to increase the recruitment of physicians and non-physician clinicians to productive careers in clinical research and to enhance the rigor of their training programmes so that they will succeed as independent investigators (Nathan and Wilson 2003). The accelerating pace of basic science discoveries potentially benefiting human health makes it critical that we expand the cadre of clinical researchers, who are central to bringing the fruits of this progress 'from the bench to the bedside'. This need has been exacerbated by the limited formal research training afforded in medical training, during residency and fellowship as well as the potential paucity of role models and mentors (Kahn et al. 2001).

*Corresponding author. Email: ret7@georgetown.edu

This situation was recognised by the National Institutes of Health (NIH), who concluded that development and inclusion of high-quality multidisciplinary didactic training would complement a mentored research experience, which is key to early research career development. The NIH National Center for Research Resources (NCRR) issued a Program Announcement in 2003 specifically to promote this type of development at academic institutions nationwide (for more information about the award, see <http://grants.nih.gov/grants/guide/rfa-files/RFA-HL-04-004.html>; for information about the programme, see <http://grants2.nih.gov/training/K30.htm>). One author was recently funded by the NCRR to develop a clinical research training programme for physicians, situated within the School of Medicine. The challenge was to design a postgraduate curriculum – for what was expected to be a wide range of experience levels – that would promote development of the requisite skills and that would also enable trainees to demonstrate their progress in authentic, interpretable ways. For this curriculum, we chose what could be characterised as a systems-based (Brown and Knight 1994, chap. 10; see also Toohey 1999, 51–2) approach. Brown and Knight provide four compelling reasons to choose the overall (system) level as the focal point for curriculum (and assessment) development: it supports breadth and balance in the curriculum; it promotes student progression by integrating multiple opportunities to use and exhibit the target skills of the programme; by pervading the programme, it can minimise resistance to the ‘novelty’ of the programme and its assessment; a systems-based approach can be more efficient when one approach is identified and adapted within multiple courses; similarly, the focus on the entire curriculum reflects the intended purpose of the support (grant funds in this case) and finally, designing a curriculum as a system promotes transparency and accountability for the programme as a whole (Brown and Knight 1994, 122–4). In describing the challenges of characterising the meaning of undergraduate assessment or evaluation, Heywood (2000) noted, ‘(i)t might ... be supposed that there exist simple definitions of what a graduate with first-class honours or a grade point average of 4.5 should be able to do’ (249). In a postgraduate context such as our clinical research training programme, however, we felt that we might create not only a list of what our programme’s graduates should be able to do, but also general descriptions of the manner in which they did it. Thus, we created the Mastery Rubric (MR) for Clinical Research.

The curriculum design for this programme has proceeded on the basis of the rubric that is the topic of this manuscript. Course topics were identified that would provide didactic opportunities to develop the target skill set as articulated in the rubric. When the project was funded, the curriculum had to be formalised *de novo*; a rubric (Stevens and Levi 2005) describing the dimensions and performance of ‘the ideal Fellow’ at the start and finish of our programme was developed, and is described here. This approach was chosen without reference to any theoretical structure or method for curriculum development, where curriculum is defined as ‘a planned educational experience’ (Kern et al. 1998, 1). We were generally guided by the three key features of assessment identified by Messick (1994): What is/are the knowledge, skills and abilities (KSAs) that students should possess (at the end of the curriculum)? What actions/behaviours by the students will reveal these KSAs? What tasks will elicit these specific actions or behaviours? These features reflect, and are informed by, the ‘outcomes-based’ approach to education outlined by Ralph Tyler in the mid-twentieth century (Tyler 1949). As Marzano (2001) noted, in this approach, ‘... a program or an instructional intervention was evaluated on the extent to which it

had accomplished its explicit goals ...' (2). Toohey (1999) refers to the outcomes-based approach as a systems- or performance-based approach. Although unintentional, our postgraduate curriculum was developed in sync with these widely accepted tenets of curriculum and course development.

A rubric is '... a set of ordered categories to which a given piece of work can be compared. Scoring rubrics specify the qualities or processes that must be exhibited in order for a performance to be assigned a particular evaluative rating' (McDaniel 1993). Generally, rubrics are developed and used for specific courses or assignments (e.g. Stevens and Levi 2005); this manuscript describes a programme-level rubric. The rubric reflects the specific outcomes of instruction that we intended the programme to provide, but by adding the performance characterisations (reminiscent of, although developed wholly without reference to, the structure of the observed learning outcome (SOLO) taxonomy (Biggs and Collis 1982; discussed and cited by Toohey 1999, 171–2)), our approach goes beyond an articulated set of learning objectives for the curriculum as a whole. As is described below, performance of the Clinical Research Training curricular objectives at a proficient level is the ultimate learning objective for any student entering our programme.

Methods

We developed a Mastery Rubric to guide the design of both the curriculum and the assessments in a two-year certificate programme intended to train physicians in clinical research skills. The component skills in the MR were identified based on nine years' worth of one-on-one consultation with, and mentoring of, researchers, together with 'research manuals' such as are published by the American Psychological Association and American Medical Association (APA 2001). The content areas ('rows') of our MR constitute the needs assessment of our target learners (Diamond 1998; Kern et al. 1998). We examined the elements of publishable research reports and successful grant applications, and formulated the content areas on this basis together with our own individual experience as consultants, researchers and reviewers.

For flexibility in our curriculum, we created performance-level descriptions for each skill corresponding to exemplars from our collective experiences mentoring researchers. The intention was to provide enough structure for all course developers to create content and assessments that would enable the students to demonstrate 'where they were' in the rubric. For example, courses in experimental design, biostatistics, epidemiology or ethical human subjects research could all have different assignments reflecting the first skill in the rubric (Table 1): 'Critically review the literature and evaluate the quality of evidence relating to an important research question'.

We identified the key skills that correspond to the best science in our experience with graduate students, medical students, clinical researchers and in our roles as scientists and as reviewers of research manuscripts and grants. These skills represent the general framework for research and ignore content- or domain-specific skills (such as laboratory techniques or diagnostic tests). These skills represent a standard methodological repertoire, for example consistent with the sections of an NIH grant application or manuscript for publication (see also APA 2001). The skills and performance levels are presented in the rubric shown in Table 1.

The key skills in, or dimensions of, clinical research on which our programme is based are:

Table 1. The clinical research Mastery Rubric.

	<i>Beginning</i>	<i>Novice</i>	<i>Competent</i>	<i>Proficient</i>
Level <i>Description of performance</i>	Complete inconsistency in aims/goals/approach; poor articulation/communication; general unfamiliarity with scientific method and research methods	Inconsistency present in several, but not all study elements; developing skills of articulation of argument and flow; communication skills developing; developing familiarity with scientific method and research methods	Solid consistency stated aims/articulation of aims which are concrete and achievable. Good Strong communication; skilful description of and compliance with scientific methods	Complete consistency in terms of aims/goals/approach. Excellent articulation of aims which are concrete and achievable. Strong communication; skilful description of and compliance with scientific method and research methods
<i>General description of work</i>	Unreadable, unratable, very difficult reading/evaluation	Readable and ratable; novice standing obvious	Readable and providing a solid framework for editorial commentary and improvement	Excellent work, interesting read, editorial input specific and targeted
Dimension				
Critically review the literature and evaluate the quality of evidence relating to an important research question	Rambling review; irrelevant literature; inability to articulate an argument as to how the literature supports question; no consideration of quality of evidence	Review that emphasises 'completeness' over concision; main points of reviewed work established together with many less relevant points. Beginnings of evaluation of quality of evidence	Good review of most of the relevant literature – a starting point from which reviewers can augment (with pertinent references). Quality of evidence considered. Generally, well structured and engaging enough that reader is interested in question posed	Concise review of the relevant literature – with assessment of greater (cited/described) and lesser (cited as 'but see ...') evidentiary weight which engages reader and leads naturally to posed questions. Careful argument followed directly to question of interest
Articulate a research objective that follows from a critical review of the literature and is achievable specific aims and perhaps testable hypotheses that address the objective	A series of incompatible or unanswerable questions; failure to connect literature to questions of interest. No clear research objective, no testable hypotheses. No statement of scientific impact of objectives	Developing ability to articulate an argument as to how the literature supports the objective(s); generally poor articulation of research question; beginnings of hypothesis statements. No clear scientific impact of objectives (possibly hunted at)	Research objectives generally stated, and generally supported by literature review. Specific aims vague but reader engaged enough to continue reading with general idea of hypotheses to be tested – and the potential impact	Research objectives clear, clearly stated and well supported by literature review. Aims concrete, specific, achievable and scientifically sound. Impact and relevance clearly stated

Table 1. (Continued).

	<i>Beginning</i>	<i>Novice</i>	<i>Competent</i>	<i>Proficient</i>
Develop a research design and protocol that provides an efficient and effective framework and data to meet each of the study aims and hypotheses	Research design does not match aims. Hypotheses not tested with stated framework. Aims redundant or unrelated to one another; a coherent scientific argument is not made. Proposed data to be collected will not address questions (as far as can be determined)	Research design partially matches aims. Hypotheses may be testable within stated framework but this may be unclear. Aims redundant or unrelated to one another but they are more clearly stated so corrections are more straightforward. The beginnings of a coherent scientific argument are made. Proposed data to be collected will generally address questions, but additional data points may be necessary	Research design generally addresses stated aims and hypotheses; data collection mechanisms are not as efficient as they could be (in terms of interpretability, alternative specifications). Coherent scientific argument is made; this may not be as competently connected to aims and design as would be desired	Research design carefully constructed to specifically and concretely address each stated aim and hypothesis. Scientific argument is clear and specific, and is fully consistent with aims and designs. Alternatives and contingencies specified and interpretations of outcomes well-considered
Develop an analysis plan and estimate the sample size that will enable the study to evaluate the evidence from the data to address each of the study aims and hypotheses	No specific analyses are planned. Any analyses that are planned are not appropriate for stated aims, as far as can be determined. If included, power calculations are inappropriate given design and stated aims; effect sizes are wildly optimistic	Planned analyses do not fit study design or specific aims. Specific analyses that are planned are not appropriate for some stated aims. Power calculations are inappropriate given design and stated aims; effect sizes are optimistic and not well justified	Planned analyses generally fit study design and specific aims. Power calculations are appropriate given design and stated aims; effect sizes are optimistic and not well justified	Planned analyses are optimal for study design and specific aims. Power calculations are appropriate given design and stated aims; effect sizes are well justified and alternative formulations are conceptualised (i.e. effect sizes and power calculations are based on least powerful version of design)
Direct the implementation of the study design and protocol assuring the quality and completeness of the data	Study carried out by others who make decisions without input from author. Decisions not understood or known of; these are therefore not included in the write up. Study does not conform to IRB	Study carried out by others who make decisions without input from author, or whose influence in decision-making is obvious. Decisions generally not included in the write up. Study generally conforms	Decision points in study implementation were generally identified <i>a priori</i> , limiting the number of decisions that must be made along the way. Staff are generally trained (e.g. based on their respective	Decision points in study implementation were thought out <i>a priori</i> , limiting the number of decisions that must be made along the way to truly unforeseeable. All foreseeable decision points are included in the protocol and staff are

Table 1. (Continued).

	<i>Beginning</i>	<i>Novice</i>	<i>Competent</i>	<i>Proficient</i>
Oversee the implementation of the analysis plan, assemble the evidence and draw inferences from the evidence regarding each study aim and hypothesis	<p>application(s) and/or consent forms. This individual is unlikely to take lead role in write up; resulting reports tend to be inaccurate in terms of how the data were collected and the extent to which critical elements such as randomisation, blinding and objectivity in ratings are maintained. Weight of evidence generated in this study is decremented in unknown ways by the implementation</p>	<p>to IRB application(s) and/or consent forms. This individual is unlikely to take lead role in write up but will have guidance from an attentive mentor; resulting reports are accurate in terms of how the data were collected and the extent to which critical elements such as randomisation, blinding and objectivity in ratings are maintained. Weight of evidence generated in this study is decremented in generally known ways by the implementation</p>	<p>backgrounds) to follow the protocol. Decisions made during study are described and quantified for inclusion in the write up. Study conforms to IRB application(s) and/or consent forms. This individual will take the lead role in write up and will have guidance from an attentive mentor; resulting reports are accurate in terms of how the data were collected and the extent to which critical elements such as randomisation, blinding and objectivity in ratings are maintained. Weight of evidence generated in this study is generally as originally intended by design</p>	<p>appropriately and formally trained specifically to follow the protocol. All decisions are documented with a system established ahead of time so that these can be described and quantified for inclusion in the write up. Study fully conforms to IRB application(s) and/or consent forms as well as to study aims. This individual will take the lead role in write up; resulting reports are accurate in terms of how the data were collected and the extent to which critical elements such as randomisation, blinding and objectivity in ratings are maintained. Weight of evidence generated in this study is as originally intended by design and concepts like this are included in write up</p>
	<p>Analysis plan incomplete and/or unachievable. Unarticulated key elements are not acknowledged. Collection of evidence is not in accordance with stated aims, methods generally do not conform to best practices and/or demonstrate serious threats to ethical standards, validity</p>	<p>Analysis plan complete and yet may be unachievable. Collection of evidence is generally in accordance with stated aims, although methods may not conform to best practices. Potential threats to ethical standards, validity and interpretability of results can be identified</p>	<p>Analysis plan is complete, and possibly 'over-sophisticated' analytic approach is proposed. Failure to appreciate the contribution that qualitative analyses can make to the interpretability of the results may be apparent. Assumptions not tested; nonparametric versions of methods may be warranted.</p>	<p>Complete, and completely appropriate, analysis plan. Analyses include both qualitative and quantitative elements, as appropriate. Quest for <i>p</i>-value is not purpose of analyses; only inferences that are supported by literature, theory, as well as the collected evidence, are drawn. Where warranted, both a parametric and a robust</p>

Table 1. (Continued).

	<i>Beginning</i>	<i>Novice</i>	<i>Competent</i>	<i>Proficient</i>
Assemble the evidence in the form of tables and graphs, and present the results together with the study methods orally and in writing	Difficult to follow, conflicting and inconsistent statements of purposes, results, interpretations and conclusions. Conclusions not supported by results; conclusions contradict stated aims; conclusions are not discussed in terms of potential alternative explanations; probability and mathematical foundations of inference relied on with no appeal to assumptions, requirements or robustness of methods. Resulting work is not useful at all for other scientists (i.e. evidentiary weight is non-existent due to unknowable, but probably large, threats to validity of design, data collection and analysis)	Disjointed, possibly complete, reporting of analyses. Clear subjectivity in discussion, and especially conclusions. No or limited attention paid to possible alternative interpretations and outcomes. Impact of stated interpretation is not discussed, alternatives are not mentioned or integrated into discussion. No real contextualisation of results for the reader. Resulting work is generally not useful for other scientists (i.e. evidentiary weight is low due to unknowable threats to validity of design, data collection and analysis). Claims in conclusions are generally not supported by results, and description of study is so poor/incoherent that even if claims are supported by results, it is too difficult to make this judgement with confidence	Straightforward reporting of analyses, with limited subjectivity in discussion. Subjectivity is present in interpretations, low level of attention paid to possible alternative interpretations and outcomes. Impact of stated interpretation is discussed, alternatives may be mentioned but are not fully integrated into discussion. Minimal contextualisation of results. Resulting work is generally useful for other scientists (i.e. evidentiary weight is generally high due to validity of design). Claims in conclusions are generally supported by results	(nonparametric or nonlinear) approach are used; determination of 'result' is based on the synthesis of all relevant results. No threats to ethical standards, validity and interpretability of results Straightforward reporting of exactly what analyses found. Clear interpretations, well-grounded in terms of possible alternative interpretations and outcomes. Impact of supported interpretation, as well as alternatives, clearly stated and discussed. Ideas and findings described in review are revisited so that present results can be contextualised by reader. Resulting work is clearly useful in future reviews (i.e. evidentiary weight is high due to validity of design and clarity of writing). Claims in conclusions are exactly supported by results. Relevance to science is cautiously stated, and next step(s) clearly articulated
and interpretability of results	from the plan and thereby, addressed	No threats to ethical standards or validity of results, although interpretability may be questioned		

- critically review the literature and evaluate the quality of evidence relating to an important research question;
- articulate a research objective that follows from a critical review of the literature and develop achievable specific aims and perhaps testable hypotheses that address the objective;
- develop a research design and protocol that provides an efficient and effective framework and data to meet each of the study aims and hypotheses;
- develop an analysis plan and estimate the sample size that will enable the study to evaluate the evidence from the data to address each of the study aims and hypotheses;
- direct the implementation of the study design and protocol assuring the quality and completeness of the data;
- oversee the implementation of the analysis plan, assemble the evidence and draw inferences from the evidence regarding each study aim and hypothesis; and
- assemble the evidence in the form of tables and graphs, and present the results together with the study methods orally and in writing.

Each of these skills can be performed at one of four levels:

- (1) *Beginning*: Unratable; very difficult reading/evaluation – no clear point or multiple inconsistencies that make assigning a rating difficult or impossible.
- (2) *Novice*: Ratable; points may be of insufficient depth or representing incomplete knowledge or awareness of the area of interest; novice standing obvious.
- (3) *Competent*: Providing a solid framework for editorial commentary and improvement.
- (4) *Proficient*: Excellent work, interesting read, editorial input specific and targeted.

When applying for a place in the programme, candidates' mentors or supervisors are asked to place the applicant on the rubric, citing specific examples of work that support the placement. Candidates are also asked to place themselves in the rubric, and are also asked for supporting examples for their argument. The rows of the rubric represent our instructional objectives; so potential applicants who are shown to have already attained mastery – whether this was by experience, training or other means – would be ineligible for our programme.

Once admitted to the programme, the placement of an individual within the rubric is based on instructors' assessment of the current level of the applicant's work within each of the skill dimensions. Thus, mastery, which we define as reaching the 'proficient' level, on each of the skills, is related to the observable outcomes (assignments or tests) from courses. In this way, progress through the programme of coursework becomes progress to mastery. The assignments were intended to be authentic, and the rubric influenced course development both in content and in requiring that course assignments and tests would provide evidence supporting a claim of proficiency with respect to the skill or objective (Mislevy 2003).

When new courses are considered for inclusion in the curriculum, their potential to provide evidence supporting one or more of the MR skills is evaluated. When course materials and tests/assignments are being developed, the potential for the work to support a claim about placement (or movement) within the rubric is used to design/

select items and assignments. We are able to evaluate courses within the programme, assignments within courses and ancillary (independent) research, in terms of their contributions to students' progress through the rubric.

The dimensions of the rubric and performance levels are explicit, and are made available to all stakeholders. The rubric is extremely flexible, so that evidence of achieving any given level of performance can take many forms (and would given the broad range of experiences and perspectives of our first cohort of Clinical Research Fellows). Learners – and their needs for new skills – should be considered when the curriculum is developed (Diamond 1998; Kern et al. 1998). However, our approach considers the curriculum itself and the goals of our programme, and then places the learners within the rubric. In this respect, the development of our clinical research skills curriculum via the MR is closer to the approach outlined by the Middle States Commission on Higher Education (2003) for the assessment of student learning. Within that framework, the MR is a list of 'intended learning and educational experiences' (55). That is, the MR presents the curriculum in its most elemental form. Since Fellows can only take a limited number of courses, this rubric can help in the design of the 'optimal' courses – in terms of providing opportunities for instruction, performance and assessment across relevant dimensions and levels of achievement. The rubric is not so specific as to limit evidence of performance level or achievement, however, and so is suitably flexible.

Results

The MR was used both to evaluate applicants to the programme and to develop the courses and their assessments. Experience with each of these is described below.

Evaluation

We have used the rubric to invite applicants to the programme in the first two cohorts (2006, 2007). The consortium programme (<http://dcrtc.org/>), sponsored by NIH, offers a Clinical Research Fellowship covering tuition, books and instruction for two years. The MR was given to all interested applicants and their nominators, with instructions to write an essay describing their current performance levels (with evidence, as needed) on each of the skills in the rubric. In this way, we were able to select a group of 13 individuals from the 35 or so applications in the first cohort, and 9 of 17 applications in the second cohort, who argued *and demonstrated* a need to participate because they were all generally performing the key skills at the novice level (applications dropped in the second year because recruiting emphasised and required explicit commitments of protected time by the sponsoring institution for any selected fellow).

A four-person review panel reviewed applications using the MR to make their selections independently. Agreement on the incoming classes has been 100% in each year. Furthermore, decisions about acceptance and rejection are concrete and clearly explainable to all applicants. Thus, all Fellows have been/can be characterised using the same criterion-referenced tool, even though they come from a variety of disciplines, have highly variable educational backgrounds and research interests and experiences. For our purposes, we seek applicants who classified themselves at the 'novice' level; as we track our Fellows we can show that, given their performance levels upon entry to the programme, their progress through the rubric (and the programme) will be defined by the performance levels they attain. This is contrasted with typical certificate

programmes, where the summative assessment is in the form of a letter grade or whether or not courses were taken with satisfactory grades. The MR provides greater detail as a summary of proficiency than course grades.

Courses and assessment

The programme was developed based on learning goals consistent with the skills within the rubric. That is, the rubric provided a unifying structure on which the overall learning goals could be based for the diverse set of courses designed for students in the programme. To date, four courses have been developed (and continue to be refined): ethics, biostatistics, issues and considerations in experimental design, and a survey in clinical research. The content, learning goals and assessments in each of these courses was also tailored to the research skills outlined in the MR, so that assignments within each course could be utilised as evidence of performance level for a given skill where applicable. For example, problems in the ethics course require answers contextualised in narrative form, so that an outside evaluator can assess the extent to which the work product from an ethics assignment represents performance in one of the target dimensions. After the assignments are turned in, the topic is discussed, and students have the opportunity to revise their submissions, keeping in mind the discussion as well as features of the rubric relating to performance.

Similarly, at the end of the semester long course in biostatistics, the students prepare a novel analysis plan (according to a template they are provided). They then rate their own plan using the component of the MR to which this assignment was developed to correspond. Appendix 1 contains the directions for students, together with the rubric selection, for self-review; one student agreed to share this review of her analysis plan. The plan, and the review, provides fairly direct evidence demonstrating forward momentum (in terms of performance) in the MR. Instructor-derived evidence would be integrated with the student's use of the rubric to create a reasonable argument for achieving any particular performance level for each of the skills in the Rubric.

Discussion

We identified a set of key skills that generally represent 'research skills' apart from content- or domain-specific knowledge. These skills, which represent a basic research repertoire and are not anchored to any particular discipline or methodology, were then described at each of four distinct performance levels, creating a 7×4 (skill \times level) rubric for curriculum development and both curriculum and student assessment (Figure 1). The courses for the programme have been developed based on learning goals consistent with the skills within the rubric. That is, the rubric provides a unifying structure on which the overall learning goals of the Fellowship programme were based for the diverse set of courses (ranging from human subjects protection/research ethics to principles of biostatistics). The content of the courses was also tailored to the research skills outlined in the MR, so that assignments within each course could be utilised as evidence of performance level for a given skill. These data are still being analysed, but will yield simple, yet concrete, in summaries for grant progress reports (starting May 2008). Two examples appear in Appendix 1; one self-evaluation by a fellow (biostatistics) and one instructor evaluation of a fellow (ethics).

Although the MR was developed without knowledge of, or appeal to, the literature on curriculum development, our 'instincts' were consistent with traditional (e.g. Tyler

1949) and more modern (e.g. Biggs and Collis 1982) educational considerations. An emphasis on learning goals and valid, authentic assessment is always a critical feature in effective instruction (Mislevy 2003), but for postgraduate education in particular, students will naturally differ depending on their time in school, motivation for the current topic and level of engagement with material (as noted by Kern et al. 1998). The programme we developed, and continue to refine, around the MR is designed specifically to meet our Fellowship programme learning goals – to provide students the opportunities to develop each of the skills identified in the rubric based on any experience or work product that can be placed (rated) within the MR. Additionally, because they were developed with the MR in mind, each course provides multiple opportunities for students to demonstrate concretely the level of performance (in the rubric) they have attained after completing the course.

Since the rubric is available to students from the point at which they apply, they are able to reflect on their own work at all times in the programme, and develop the meta-cognitive skills required to monitor their performance so that their work will reflect the highest possible performance level on the rubric. These are highlighted as features of learning experiences that tend to result in the most effective learning outcomes (Middle States Commission on Higher Education 2003, 77–80). Since we have just completed the first year with this curriculum, we have no ‘baseline’ against which to compare the learning of our first Fellowship cohort. The courses were universally well-regarded and in their evaluations Fellows expressed an appreciation for the emphasis that the programme placed on their ‘real learning’.

The way our programme has used the MR, potential applicants whose materials suggest that they have already attained mastery – whether this was by experience, training or other means – would be ineligible for our programme. However, the rubric approach could easily be adapted for use as a specific placement tool, or to determine whether preliminary courses for a programme are needed by a given student. Many of our Fellows have commented that, although ‘on paper’ they might appear not to need additional coursework such as our programme provides, the ability to place themselves in the rubric, and in particular the requirement for evidence of performance at any level, has allowed them to argue that they would, in fact, benefit from our training programme.

Using the rubric means that success in our programme can be characterised, not in terms of completing a series of courses but in terms of documenting that they have developed the habits of mind and organisational principles that can foster excellence in clinical research. That is, as it has been designed, individuals who enter our programme will be characterised at a certain level on skills that were identified as critical to successful research; when they complete the curriculum, the intention is that all students will be firmly within the ‘proficient’ column on all of the skills, and that claims of proficiency will be supported with concrete evidence (Mislevy 2003). When students have moved towards the proficient side of the rubric, we will be able to report the success of our programme in explicit terms, providing concrete characterisations, based on work products from the courses, rather than subjective ratings or other, more general (less descriptive) means. Our first progress report with data from students will be submitted in July 2008.

It is important to note that, as clinical researchers, the main (authentic) work product for fellows is written material: manuscripts, reviews and grants. Thus, any individual who develops a novel method would not be evaluated on the basis of the novelty or utility of the method but rather, on their ability to convey the importance

and use of the method, in writing or orally, to others. The number of studies completed, level of technical skill, number of cases treated or other outcomes that represent work or skills, but cannot be (or typically are not) *described*, are difficult to incorporate into this rubric. This reflects our intention: performing that work or exhibiting those skills must be translated into what we consider the main work product(s) for clinical research, taking the form of written materials. This approach could be easily adapted to reflect other types of work or to accommodate more quantitative evidence. Schunn and Anderson (2001) summarised their research into the nature of the development of ‘expertise’ in science, reporting that skills representing ‘expert’ level in science were not specifically taught in undergraduate research methods courses (within a psychology department with a research orientation). With our curriculum, and specifically our approach to course and curriculum development, we not only sought to represent the specific skills for careful science but also to encourage movement from novice to competent use of skills required for good clinical research. Thus, the MR combines the powerful guidance of an outcomes (Tyler 1949) or construct (Messick 1994) based approach with the informational and evidentiary support that a rubric- or SOLO-type taxonomy (Biggs and Collis 1982) can provide.

Schunn and Anderson (2001) noted that, across five (undergraduate) courses in research methods, there was little agreement in the ‘outcomes’ – those skills that the students should possess at the end of the course in research methods (109–10). Our rubric has the advantage of being set with a more professional level of student than in the undergraduate setting, possibly simplifying decisions about critical skills to include. Nevertheless, we believe that our approach to the development (or evaluation) of curricula, articulating both key skills and the desired performance levels, can be adapted across higher educational contexts and domains.

Placement in the MR describes the level of functioning on key research skills more completely and concretely than a letter grade can – because placement is more specific. Tagg (2003, 28) noted that ‘assessment of student work in courses tends to be powerfully biased toward the simply quantifiable’ and that, when a single summary such as the GPA is used to characterise the totality of student achievement within a curriculum, ‘(t)he trajectory of the student through the curriculum, as well as the consistency or lack thereof in the student’s work, is invisible’ (27). The MR makes the trajectory explicit, such that the summary of a student’s progress through the curriculum is a characterisation, not a number.

Any programme could adapt this rubric to accommodate the specific intentions/learning goals and work products that best reflect it. In fact, the concept of an MR, and its use as described here, is completely applicable in other domains and skill sets (see e.g. Heywood 2000, 275–315). For example, a new MR for Physicianship for medical students is being developed currently. As with all rubrics, the purpose and goal of the rubric’s use dictates its form and the evidence (and performance-level descriptions) comprising the body of the rubric.

In summary, the MR for clinical research skills that we describe here not only guided the development (and evaluation) of our new research skills curriculum, but also streamlined and facilitated the admissions process. Furthermore, we have a consistent, criterion-referenced, yet flexible method for reporting individual Fellow’s progress for the entire two-year programme. As an instructional design tool, the MR supports, and focuses, the design of specific courses and the types of assignments within those courses to support claims that material has been ‘learned’, rather than emphasising what material has been covered (McKeachie and Svinicki 2006, 11).

Practice points

Instructors sometimes use rubrics to facilitate/grade assignment in a course; rubrics are excellent tools for clarifying assessment goals and learning objectives. This article describes a rubric that comprises a clinical research training programme.

The MR for Clinical Research entails a set of seven core clinical research skills: critical review of literature; articulation of research objective; development of research design; development of analysis plan; implementation of the study; implementation of the analysis plan and presentation of results.

The MR describes four distinct levels of performance on each of the skills: beginning, novice, competent and proficient. Our Fellowship programme seeks to move fellows from the novice end toward the proficient end; performance-level descriptors permit far greater characterisation of achievement than letter grades or other quantitative summaries.

This rubric outlines and provides a path to mastery of the clinical research skills the programme was designed and funded to target, and can be adapted to other contexts or programmes.

The rubric facilitates a flexible, criterion-referenced definition of 'success' for students as well as the programme itself. The criteria are characterised in terms of the skills, habits of mind and organisational principles that can foster excellence in clinical research.

Acknowledgements

This research was supported by K30 RR022272 (to J.G.U.) from the National Center for Research Resources.

Notes on contributors

Rochelle E. Tractenberg directs the curriculum described in this manuscript, an NIH funded certificate-granting programme at Georgetown University Medical Center with the objective of training physicians in clinical research. She is a research methodologist (focusing on Alzheimer's disease, cognitive aging, and hypertension) and has been working in pedagogy in higher education since 1994.

Jason G. Umans is an associate program director of the General Clinical Research Center at Georgetown and the principal investigator of the NIH grant funding the curriculum described in this manuscript. He served as the director of education in the Georgetown Division of Nephrology and Hypertension, developing the curriculum in an NIH-funded training program in Nephrology and Hypertension.

Robert J. McCarter directs the Biostatistics and Informatics Core of the General Clinical Research Center at the Children's National Medical Center and is a professor in the Departments of Pediatrics and Epidemiology & Biostatistics, The George Washington University Schools of Medicine and Public Health.

References

- American Psychological Association (APA). 2001. *Publication manual*. 5th ed. Washington, DC: APA.
- Biggs, J.B., and K.F. Collis. 1982. *Evaluating the quality of learning: The SOLO taxonomy – Structure of the observed learning outcome*. New York: Academic Press.
- Brown, S., and P. Knight. 1994. *Assessing learners in higher education*. London: Routledge-Falmer.

- Diamond, R.M. 1998. *Designing and assessing courses and curricula: A practical guide*. Rev. ed. San Francisco, CA: Jossey-Bass.
- Heywood, J. 2000. *Assessment in higher education*. London: Jessica Kingsley.
- Kahn, J.A., S.K. Parsons, P.A. Pizzo, J.W. Newburger, and C.J. Homer. 2001. Work-family issues and perceptions of stress among pediatric faculty and house staff. *Ambulatory Pediatrics* 1, no. 3: 141–9.
- Kern, D.E., P.A. Thomas, D.M. Howard, and E.B. Bass. 1998. *Curriculum development for medical education: A six-step approach*. Baltimore, MD: Johns Hopkins University Press.
- Marzano, R.J. 2001. *Designing a new taxonomy of educational objectives*. Thousand Oaks, CA: Corwin Press.
- McDaniel, E. 1993. *Understanding educational measurement*. Dubuque, IA: William C. Brown.
- McKeachie, W.J., and M. Svinicki. 2006. *McKeachie's teaching tips*. 12th ed. Boston, MA: Houghton Mifflin.
- Messick, S. 1994. The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher* 23, no. 2: 13–23.
- Middle States Commission on Higher Education. 2003. *Student learning assessment: Options and resources*. Philadelphia, PA: Author.
- Mislevy, R.J. 2003. Substance and structure in assessment arguments. *Law, Probability, and Risk* 2, no. 4: 237–58.
- Nathan, D.G., and J.D. Wilson. 2003. Clinical research and the NIH – A report card. *New England Journal of Medicine* 349, no. 19: 1860–5.
- Schunn, C.D., and J.R. Anderson. 2001. Acquiring expertise in science: Explorations of what, when, and how. In *Designing for science: Implications from everyday, classroom, and professional settings*, ed. K. Crowley, C.D. Schunn, and T. Okada, 83–114. Mahwah, NJ: Lawrence Erlbaum.
- Stevens, D.D., and A.J. Levi. 2005. *Introduction to rubrics: An assessment tool to save grading time, convey effective feedback and promote student learning*. Portland, OR: Stylus Publishing.
- Tagg, J. 2003. *The learning paradigm college*. Bolton, MA: Anker.
- Toohey, S. 1999. *Designing courses for higher education*. Philadelphia, PA: Society for Research into Higher Education and Open University Press.
- Tyler, R.W. 1949. *Basic principles of curriculum and instruction*. Chicago, IL: University of Chicago Press.

Appendix 1

A. Selection from Mastery Rubric for self-evaluation of final project, and one example of a self-evaluation by Clinical Research Training Fellow

Fellows were given the assignment to write an analysis plan according to a detailed template. They were further instructed to use the following excerpt from the Mastery Rubric to argue what rating their work merits. The analysis plans are long, detailed outlines including background, references and literature reviews; they typically represent manuscripts or grant proposals in preparation, and thus are not available for publication (however, interested readers can email the corresponding author to discuss the assignment and/or template).

Following is a matrix (rubric) for you to use in evaluating your own work before you turn it in. In addition to your analysis plan, you must 'review your plan using the rubric' – turn in the plan WITH a one page (not less than one paragraph) review, including the LEVEL at which you think the plan was executed and specific examples or evidence from your plan that led you to this judgement.

Analysis plan rubric evaluation <student name withheld>

In evaluating <name of project omitted>, I find the analysis plan to be at the competent level. To begin with, the introduction appropriately frames the issues and challenges involved with the care of critically ill paediatric cardiac patients. It discusses what the mechanism of action of ketamine is with respect to reducing drug-induced hyperalgesia and outlines several studies that have supported its use for this purpose. Furthermore, it describes the lack of studies in the proposed patient population that this study intends to address.

In the methods discussion, the study's experimental design is clearly outlined. This includes clear definitions of the patient population with inclusion and exclusion criteria as well as the type of study to be done. The interventions and time course of the study are also well defined. The outcome measures are clearly stated and realistic endpoints to obtain.

One weakness of the analysis plan is the lack of pilot data available to help determine sample size. Based on the previously mentioned studies, a decrease in morphine consumption of 20% is reasonable. However, this will need to be confirmed in the proposed patient population. With this data, we can then determine sample size for the study with 80% power and an alpha level of .05.

The description of analysis again clearly states the outcomes to be measured and the analysis to be performed. It outlines the need to verify that variances are equal and the different contingencies for determining equivalence of means based on this. It also addresses other statistical methods that could be used if all assumptions are not met for this analytic method. Variable definitions and descriptive statistics are also appropriate.

Finally, the null hypotheses and hypothesis testing fit the study design and address the specific aims proposed. The stated power of 80% and the alpha level of .05 are desirable, however, the sample size to achieve this is currently unknown. In light of the fact that we are testing for a decrease in morphine consumption in the ketamine group relative to the control group and this is an initial study, a one-sided test is appropriate.

In general, this analysis plan is consistent in its stated goals and scientific approach to achieve them. While the issue of sample size needs to be addressed with pilot data, the outcomes proposed are realistic and the statistical analysis appears to be appropriate. Overall, this plan is at the competent level in the rubric.

B. Instructor evaluation, using Mastery Rubric, of the progress of one Clinical Research Training Fellow, in the course on Ethics/Responsible Conduct of Research

Fellows were regularly given assignments to write a structured analysis of case studies. The instructor used the Mastery Rubric to justify an argument of change in the work of Fellows from the start to the end of the course. The example shown represents the instructor's assessment of the student based on the evaluation, and comparison, of the earliest and last assignment.

Level <i>Description of performance</i>	<i>Beginning</i>	<i>Novice</i>	<i>Competent</i>	<i>Proficient</i>
<i>General description of work</i>	Complete inconsistency in aims/goals/approach; poor articulation/communication; general unfamiliarity with scientific method and research methods	Inconsistency present in several, but not all study elements; developing skills of articulation of argument and flow; communication skills developing; developing familiarity with scientific method and research methods	Solid consistency stated aims/goals/approach. Good articulation of aims which are concrete and achievable. Strong communication; skilful description of and compliance with scientific methods and research methods	Complete consistency in terms of aims/goals/approach. Excellent articulation of aims which are concrete and achievable. Strong communication; skilful description of and compliance with scientific method and research methods
Develop an analysis plan that will enable the study to evaluate the evidence from the data to address each of the study aims and hypotheses	Unreadable, unratable, very difficult reading/evaluation	Readable and ratable; novice standing obvious	Readable and providing a solid framework for editorial commentary and improvement	Excellent work, interesting read, editorial input specific and targeted
	No specific analyses are planned. Any analyses that are planned are not appropriate for stated aims, as far as can be determined. If included, power calculations are inappropriate given design and stated aims; effect sizes are wildly optimistic	Planned analyses do not fit study design or specific aims. Specific analyses that are planned are not appropriate for some stated aims. Power calculations are inappropriate given design and stated aims; effect sizes are optimistic and not well justified	Planned analyses generally fit study design and specific aims. Power calculations are appropriate given design and stated aims; effect sizes are optimistic and not well justified	Planned analyses are optimal for study design and specific aims. Power calculations are appropriate given design and stated aims; effect sizes are well justified and alternative formulations are conceptualised (i.e. effect sizes and power calculations are based on least powerful version of design)

Level	<i>Beginning</i>	<i>Novice</i>	<i>Competent</i>	<i>Proficient</i>
<i>Description of performance</i>	Complete inconsistency in aims/goals/approach; poor articulation/communication; general unfamiliarity with scientific method and research methods	Inconsistency present in several, but not all study elements; developing skills of articulation of argument and flow; communication skills developing; developing familiarity with scientific method and research methods	Solid consistency stated aims/goals/approach. Good articulation of aims which are concrete and achievable. Strong communication; skilful description of and compliance with scientific methods	Complete consistency in terms of aims/goals/approach. Excellent articulation of aims which are concrete and achievable. Strong communication; skilful description of and compliance with scientific method and research methods
<i>General description of work</i>	Unreadable, unratable, very difficult reading/evaluation	Readable and ratable; novice standing obvious	Readable and providing a solid framework for editorial commentary and improvement	Excellent work, interesting read, editorial input specific and targeted
Direct the implementation of the study design and protocol assuring the quality and completeness of the data	Study carried out by others who make decisions without input from author. Decisions not understood or known of; these are therefore not included in the write up. Study does not conform to IRB application(s) and/or consent forms. This individual is unlikely to take lead role in write up; resulting reports tend to be inaccurate in terms of how the data were collected and the extent to which critical	Study carried out by others who make decisions without input from author, or whose influence in decision-making is obvious. Decisions generally not included in the write up. Study generally conforms to IRB application(s) and/or consent forms. This individual is unlikely to take lead role in write up but will have guidance from an attentive mentor;	Decision points in study implementation were generally identified <i>a priori</i> , limiting the number of decisions that must be made along the way. Staff are generally trained (e.g. based on their respective backgrounds) to follow the protocol. Decisions made during study are described and quantified for inclusion in the write up. Study conforms to IRB application(s) and/or consent forms. This individual will take the	Decision points in study implementation were thought out <i>a priori</i> , limiting the number of decisions that must be made along the way to truly unforeseeable. All foreseeable decision points are included in the protocol and staff are appropriately and formally trained specifically to follow the protocol. All decisions are documented with a system established ahead of time so that

(Continued).

<p>elements such as randomisation, blinding and objectivity in ratings are maintained. Weight of evidence generated in this study is decremented in unknown ways by the implementation</p>	<p>resulting reports are more accurate in terms of how the data were collected and the extent to which critical elements such as randomisation, blinding and objectivity in ratings are maintained. Weight of evidence generated in this study is decremented in generally known ways by the implementation</p>	<p>lead role in write up and will have guidance from an attentive mentor; resulting reports are accurate in terms of how the data were collected and the extent to which critical elements such as randomisation, blinding and objectivity in ratings are maintained. Weight of evidence generated in this study is generally as originally intended by design</p>	<p>these can be described and quantified for inclusion in the write up. Study fully conforms to IRB application(s) and/or consent forms as well as to study aims. This individual will take the lead role in write up; resulting reports are accurate in terms of how the data were collected and the extent to which critical elements such as randomisation, blinding and objectivity in ratings are maintained. Weight of evidence generated in this study is as originally intended by design and concepts like this are included in write up</p>
--	---	--	---

Because the assignments represent detailed analyses by the student of cases that are also detailed, these are not shown below (however, interested readers can email the corresponding author to discuss the assignments).

The relevant component of the rubric is shown on the next two pages; the instructor's evaluation follows.

Ethical treatment of human subjects

Course evaluation for <student's name withheld>

In order to develop a research design and protocol that provides an efficient and effective framework and data to meet each of the study aims and hypotheses, the design and protocol must include mechanisms that protect the rights and welfare of human participants. Only then can a study protocol be implemented that assures the quality and completeness of data. Early in the course, <student's name withheld> failed to suggest that provisions for rescue medications, careful supervision of vulnerable subjects and special procedures for consent be included in a potential comparison of medication and placebo in <the first case study> patients. Later submissions from the same student showed careful considerations of these aspects of design when they were pertinent to an assigned protocol.

In the analysis of the <first case> study, <the student> said, 'The patients in the induced group and the control groups are going to have some adverse symptoms as a result of this study protocol. The probability that these patients will have long-term effects is not clear. What is also not clear is the severity of the symptoms. These need to be addressed'. This analysis reflects a 'Novice' level of awareness; the response represents a failure to appreciate the role of the investigator in the decision-making process – including the knowledge of exactly what decisions must be made. No mention in the response (from which the above is a quote) was made about the consent forms, nor was any mention of the IRB and its role present in the student's analysis. The first analysis is usually more cursory, and tends not to recognise the safety issues and vulnerabilities of particular patient populations, and as might be expected at the start of the ethics course, this student would be unlikely or unwilling to take lead role in ensuring responsible conduct – or oversight – of this research project, deferring to others. However, in the analysis of <the last> study, <the student> summarised the analysis with, 'The research institution and the IRB needs to protect the study subjects and ensure adequate surveillance of the researchers and the subjects. Careful follow up and safeguards to ensure that subjects are not left to monitor the long-term effects without access to medical care'. This excerpt contrasts with the first, and the analysis itself reflects a level of awareness that is consistent with a rating of 'Competent'. Importantly, <the student> did not simply evolve over the course to write longer analyses but became able to convey a clear and consistent consideration of both the goals of the research and a responsible, ethically sound approach to both completing the research and protecting the patients. The last analysis reflects an awareness of the roles of individual investigators and institutional bodies in the scientific enterprise, without abdicating responsibility or suggesting that others involved in the research were more liable than the investigator.

<Student's name withheld> has not only completed the work and participated actively in the discussions of this course, but overall I can rate <student's name> at the 'Competent' level – possibly halfway between 'Competent' and 'Proficient'.