

DEVELOPMENT OF A SKILLS-BASED INSTRUMENT TO MEASURE SCIENTIFIC REASONING IN MEDICINE ACROSS DIFFERENT LEVELS OF EXPERTISE

**Daniela Luminița Barz,
Andrei Achimaș-Cadariu**

Introduction

Scientific reasoning has been studied across several broad areas – cognitive sciences, education, developmental psychology, even artificial intelligence – that have tried to identify its underlying mechanisms. A common approach to defining scientific reasoning refers to the mental processes used when reasoning about scientific facts or when engaged in scientific inquiry. Philosophers of science delimit scientific inquiry within the bounds of a common belief system which scientists hold about the nature of phenomena (Giacomini, 2009). On this common belief system rests scientific research with its main constituting elements – variables, hypotheses and models of explanation. Scientific reasoning skills have come into focus along with the evidence-based practice framework and the competency-based medical education (Barz & Achimaș, 2016). Both of these paradigms rely vastly on skill development and equipping the learner with the relevant competencies to adapt to the fast-changing health needs of individuals. Along worldwide efforts adapted to the new requirements of healthcare education, reasoning skills have been operationalized into learning objectives such as applying quantitative reasoning to decision making, understanding the scientific method and integrating basic scientific knowledge into clinical reasoning (Amara & Smyth, 2015; Carraccio, Englander, Van Melle, Ten, Lockyer, Chan, Frank & Snell, 2016).

Early research on scientific reasoning has brought forth Klahr and Dunbar's Model of Scientific Discovery (1988, SDDS – Scientific Discovery as Dual Search), which has been the most preeminent attempt to integrate both knowledge acquisition, as well as cognitive mechanisms in order to provide a framework for the development of scientific reasoning. Experimental research conducted in the developmental psychology field found specific skills involved in scientific thinking, which included the isolation and control of variables, producing factorial combinations in multivariable tasks, selecting an experimental design, collecting data, generating a theory through



JOURNAL
OF BALTIC
SCIENCE
EDUCATION

ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

Abstract. *Scientific reasoning skills have been defined as mental processes used when engaged in scientific inquiry. Research aimed to develop and validate a Scientific Reasoning in Medicine (SRM) instrument through a psychometric approach which included a preliminary phase with 60 medical students and physicians, followed by a revision phase and subsequent re-research with 209 medical students and physicians. Research focused on determining the extent to which item content contributed significantly to explaining the variance in SRM, if the level of scientific reasoning differed in relation to medical expertise and if individuals who were inclined to a more rational thinking style showed higher scientific reasoning. Results indicated that item content explained 47% of the variance in SRM, there were significant differences in scientific reasoning depending on expertise and participants who scored higher on the Cognitive Reflection Test and the Need for Cognition scale, also scored higher on the SRM instrument.*

Keywords: *Cognitive Reflection Test, educational assessment, scientific reasoning, quantitative reasoning.*

**Daniela Luminița Barz,
Andrei Achimaș-Cadariu**
*„Iuliu Hațieganu” University of Medicine
and Pharmacy, Romania*



inductive reasoning and reconciling evidence that either confirmed or contradicted hypotheses (Kuhn, 2010; Zimmerman, 2007; Han, 2013). Over time, research on scientific reasoning has widened its scope more and more to integrate work on scientific reasoning in higher education, in biomedical sciences and STEM education in general.

A literature review of previously developed instruments aimed at measuring scientific reasoning indicated that there were several scales created to assess the level of scientific literacy within the general public (Lovelove & Brickman, 2013; Lawson, 1978; Lawson, 2000; Gromally, Brickman & Lutz, 2012; Drummond & Fischhoff, 2015). Motivation for developing a skills-based instrument for medical students' scientific reasoning skills comes from studies which show that quantitative reasoning skills are essential for novice understanding of complex concepts and their interrelation (Marsan, D'Arcy & Olimpo, 2016). Moreover, studies on assessment of clinical and diagnostic reasoning show that students who are taught to integrate basic sciences into their reasoning processes outperform students who are not exposed to reflection tasks on causal mechanisms (Kulasegaram, Manzone, Ku, Skye, Wadey & Woods, 2015; Lisk, Agur & Woods, 2016). For the purposes of the present study, we used, as guiding framework, research on *cognitive models of clinical performance* (Patel, Arocha & Zhang, 2005; Patel, Yoskowitz, Arocha & Shortliffe, 2009) and the *dual-processing theory* (Crosskerry, 2009, Crosskerry, Singhal & Mamede, 2013). As stated by Patel, Arocha and Zhang (2005), medical knowledge can be divided into two main categories: basic science knowledge and clinical knowledge. The traditional perspective on this division is that basic science knowledge provides the bases for further acquisition of clinical knowledge and reasoning development. So, ideally, the path from novice to expert would follow a straight line, with expertise gradually developed based on knowledge and skill. However, studies have identified a non-linear path from novice to expert, indicating an "intermediate effect" – performance seems to fall before it reaches a higher level of expertise. Novices rely more on biomedical knowledge than clinical experts who form more accurate representations of clinical cases using cognitive shortcuts (Patel & Groen, 1986). Explanations with regard to expertise development through the use of "mental shortcuts" have also been the focus of the dual-processing theory, with System 1 being responsible for intuitive reasoning and the use of heuristics, and System 2 for analytical reasoning and a resource intensive type of response (Crosskerry, 2009). One significant insight into the dual-processing system is that repetitive processing performed within System 2, with time, can lead to a System 1 type of response. This finding may explain why experienced doctors, although using a System 1 approach, elicit more accurate diagnoses, while intuitive thinking without enough experience often leads to errors in diagnostic reasoning (Crosskerry, Singhal & Mamede, 2013).

Following these two major directions for the study of medical cognition, the present research aims to identify how scientific reasoning is developed throughout medical training by constructing a valid assessment method for evaluating medical students' thinking and reasoning competencies and to approach the challenge of operationalizing these specific skills across multiple levels of expertise. Although prior research findings have outlined the importance of critical thinking and quantitative reasoning skills in medical training, educational assessment methods have focused mostly on measuring diagnostic reasoning through clinical scenarios, while skills such as scientific reasoning, which may play a significant role in the advancement of medical expertise, have been overlooked. Furthermore, the purpose of the present research is to explore the development path of scientific reasoning, from novice to expert, by identifying its trajectory and its correlates. Hypotheses were directed at confirming a non-linear trajectory, prone to the "intermediate effect" and at expectations that individuals who rely more on the rational-analytical type of information processing will also demonstrate higher scientific reasoning, with a distinction found with clinicians – who should exhibit tendencies toward intuitive reasoning.

Methodology of Research

The present research included two distinct phases of questionnaire development and refinement, with data being collected electronically, as follows: first phase from 15/02 to 21/02/2015, second phase from 24/05 to 08/06/2016. All participating students in the present research were enrolled in the General Medicine programme at the "Iuliu Hațieganu" University of Medicine and Pharmacy Cluj-Napoca, while residents and clinicians were selected through the available contact information for staff at the Clinical Municipal Hospital Cluj-Napoca. The initial instrument was sent to a randomly selected sample of students, from a list of personal contact information with available email addresses, and the revised instrument was sent to all undergraduate students with available email addresses. Undergraduate students from the Medicine programme were also contacted with the help of the student representatives for each academic year and the appointed secretarial staff. Residents and clinicians were approached through email invitations, but also through on-site visits at the Clinical Municipal Hospital Cluj-Napoca.



For expert review of the instrument, 9 professors and primary physicians were asked for input and revision of item content, regardless of their clinical specialty. During both phases of instrument development, participants were asked to fill-out a feedback form meant to quantify perceived quality of SRM. The feedback form asked participants to rate perceived difficulty of the items, ambiguity of item content, relevance to the medical field and overall quality of the questionnaire. Administration of the revised SRM and the additional cognitive measures were accompanied by an optional open-ended feedback designated section for additional input.

Sample

The sample size for the pretesting phase of the Scientific Reasoning in Medicine instrument (SRM) was 60, with 47 female participants (78%) and 13 male participants (22%). The sample included 30 undergraduate students (33%), from the General Medicine programme of the "Iuliu Hațieganu" University of Medicine and Pharmacy Cluj-Napoca, 28 postgraduate students (47%), and 12 clinicians with PhD degrees from the Clinical Municipal Hospital Cluj-Napoca (20%).

The sample size for the revision phase of the Scientific Reasoning in Medicine instrument (SRM) was 209, selected on a voluntary basis, by email invitations sent out to all undergraduate students enrolled in the Medicine programme at the "Iuliu Hațieganu" University of Medicine and Pharmacy Cluj-Napoca, with 98 female participants (47%) and 111 male participants (53%). Participants included 152 undergraduate students (73%), 40 postgraduate students (19%), who also hold a clinical position (e.g. residents) and 16 clinicians, without a PhD degree (8%). The undergraduate students were further categorized into three cohorts: 1st and 2nd year students (15%), 3rd year students (25%), 4th, 5th and 6th year students (33%). Age groups included 123 participants (59%) between the ages of 18 to 24 years old, 55 participants (26%) between the ages of 25 and 30 years old and 31 participants (15%) between the ages of 31 and 45 years old.

Instrument Development and Procedure

Figure 1 illustrates the research design followed for the development of The Scientific Reasoning in Medicine (SRM) instrument. Initially, a list of core skills was generated, which included basic scientific reasoning skills and scientific reasoning applied in biomedical research and clinical settings. The design of the instrument included 2 major dimensions; the first one incorporated 5 sub-dimensions, while the second one 8 sub-dimensions. Basic scientific reasoning was considered to include an understanding of the scientific inquiry process, argumentation, measurement and quantification, constructing syllogisms and analogies. The operational definition of scientific reasoning skills applied in biomedical research consisted of an understanding of research designs, experimental control, sample variability, data representation, causal effects, prediction, hypotheses construction and evidence evaluation.

The first version of the instrument included 20 multiple-choice questions, with 5 possible responses, which were elaborated after a step-by-step review and were considered adequate for pretesting. Subsequently, individual item analysis and statistical tests led to the retention of 8 items which outperformed the rest and could be further improved. Two of the distractors were dropped for each of the 8 items and two multiple-choice questions were reworded. Following the initial hypothesized structure of the instrument 2 more items were included for the revision phase. The revision phase consisted of the administration of the 10-item revised SRM along with validated measures – Cognitive Reflection Test (Frederick, 2005), Need for Cognition and Faith in Intuition (Epstein, Pacini, Denes-Raj & Heier, 1996) and controlling for academic/professional level, age and an interest for biomedical research. Interest for biomedical research was assessed using a 5-point Likert scale (with 1 – very low interest to 5 – very high interest).

The Cognitive Reflection Test (CRT) is a 3-item measure of reflectiveness (Frederick, 2005) or a type of cognitive ability, which has been defined as a disposition to inhibit the first response that comes to mind, an intuitive response, in favour of a rational, thought-out, response. Since Frederick described this measure, there has been an extensive debate regarding whether CRT measures a basic numerical ability, a rational thinking style or a tendency toward open-mindedness (Campitelli & Gerrans, 2014). Studies exploring the processes behind CRT performance found that the task implies, to a certain extent, both a monitoring of System 1, responsible for intuitive responses, as well as numeracy skills, which are linked to System 2 (Campitelli & Gerrans, 2014; Welsh, Burns & Delfabbro, 2013; Pennycook, Cheyne & Koehler, 2015).



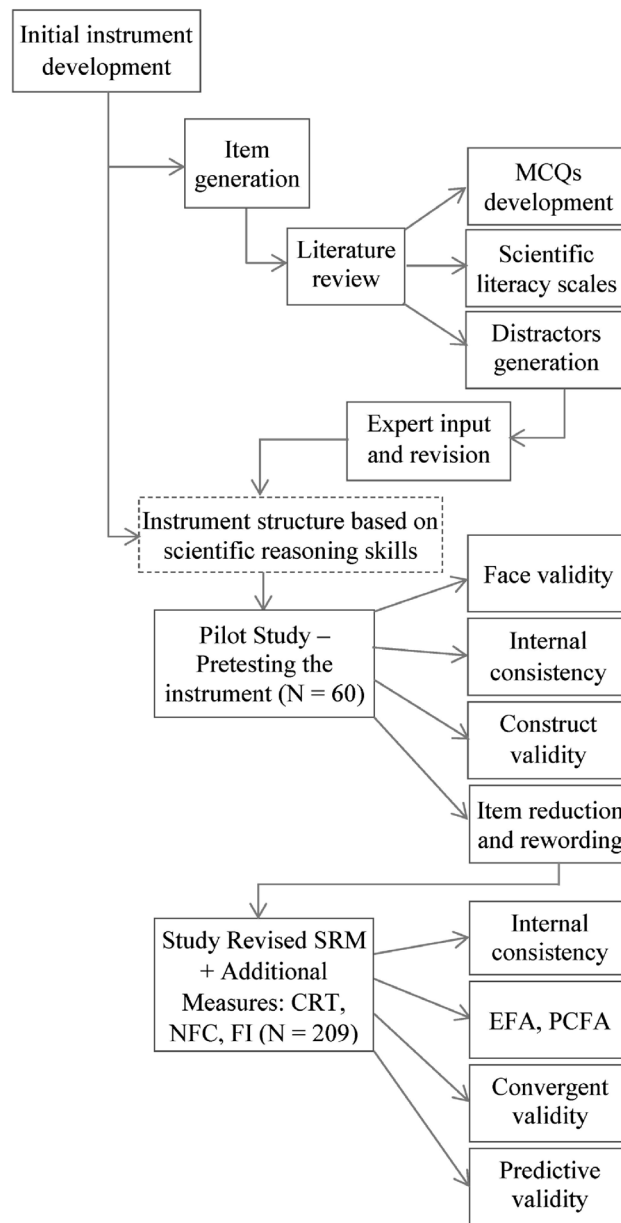


Figure 1: Flow chart for the development process of the SRM instrument.

Need for Cognition (NFC) and Faith in Intuition (FI) have both been broadly used in research to assess rational-intuitive thinking styles and their psychometric properties have been demonstrated as satisfactory in multiple studies (Epstein, Pacini, Denes-Raj & Heier, 1996). The 5-item version of the scales were used in the present study, with reported internal consistency Cronbach coefficients of .73 (NFC) and .72 (FI). Our hypotheses posited that a need for cognition and cognitive reflectiveness or reflection will positively correlate with scientific reasoning. Faith in intuition was thought to reveal distinct patterns of association with scientific reasoning particularly for clinicians.

Statistical Analyses

In the pretesting development phase of the scientific reasoning instrument (Study 1), analyses included frequencies and four chi-square tests to test for differences in face validity (perceived difficulty, ambiguity, relevance



to the medical field, overall quality) in relation to the academic/professional level of the participant. After inspecting the distributions, several dimension reduction procedures were performed to test for factorability. Kaiser-Meyer-Olkin measure of sampling adequacy along with a Bartlett's test of sphericity were performed to establish suitability for an exploratory factor analysis. Cronbach's α was computed for the remaining items (Costello & Osborne, 2005). Levene's test was used to check for equal variances and t-test for mean difference between undergraduate students and clinicians. Also, a Cohen's d was preferred for estimating the effect size for the difference in means (Fritz, Morris & Richler, 2012).

In the second development phase of the scientific reasoning instrument (Study 2), a Principal Components Analysis with an oblique method of extraction (Oblimin with Kaiser Normalization) was chosen, due to the correlation between components. In addition to the set of statistical analyses performed in the first phase, a partial confirmatory factor analysis was performed (Gignac, 2009) using a Maximum Likelihood rotation. One-way analyses of variance between variables in Study 2 and demographic characteristics were conducted, in particular with academic/professional levels. Levene's test showed significance for SRM and the two extracted components ($p < .05$), so analyses of variance included robust tests of equality of means (Welch and Brown-Forsythe); Welch's F was reported in those cases, followed by Games-Howell post-hoc tests, which were recommended for unequal variances (Shingala & Rajyaguru, 2015) and eta-square measure for effect size. Pearson correlations and multiple hierarchical regression analyses were preceded by scatter plots for prior identified significant relationships; auto-correlation analysis with Durbin-Watson test which showed that the data met the assumption of independent errors (Durbin-Watson value = 1.5, recommended values between $1.5 < \text{value} < 2.5$); multicollinearity was tested by inspecting VIF values ($VIF < 5$), which indicated that multicollinearity was not a concern (Ernst & Albers, 2016). Alpha levels of .01 and .05 were considered for all statistical tests and were highlighted separately. Statistical analyses were carried out using SPSS package, version 23.

Results of Research

Results for the Pretesting Phase of SRM

Preliminary individual item analysis was performed by inspecting variance, internal consistency and factorability of SRM items. After preliminary analysis, 8 items were retained and were considered adequate for further development in the second phase of SRM development, based on factor loadings, inter-item correlation, corrected item-total correlation and internal consistency. SRM scores were computed by summing up the correct responses per item, ranging from 0 to 8 ($N = 60$, $M = 3.62$, $SD = 2.20$). Feedback results on a 5-point Likert scale (from 1 - very low to 5 - very high) showed perceived difficulty to rate as high by 42% of participants, ambiguity was assessed as low or very low by 49% of participants, while relevance was considered high by 57% participants. Overall quality ranged from neither high, nor low (20%) to high (15%) and very high (65%), with 0 participants perceiving it at a very low or low level. Item difficulty ranged from 10% to 80%, with undergraduate students being able to correctly answer 36% of the questions, postgraduate students 41% of the questions and clinicians with PhD degree 51% of the questions. The Kaiser-Meyer-Olkin measure of sampling adequacy was .61, above the recommended value of .6, and Bartlett's test of sphericity was significant ($c^2(28) = 55.25$, $p < .05$). PCA with orthogonal rotation (Varimax) indicated a 1 factor solution which explained 29% of the variance, with factor loadings ranging from .46 to .65 and item-total correlations ranging from .25 to .44. t-test for independent groups was conducted to determine differences in mean scores between undergraduate students and clinical specialists, which showed that undergraduate students had significantly lower scores ($M = 2.85$, $SD = 1.78$) than clinicians ($M = 6.5$, $SD = 1.08$), $t(30) = -6.38$, $p < .001$, $d = -2.47$. Internal consistency was assessed by Cronbach's α at .64.

Results for the Revision Phase of SRM

Cronbach's α for the revised 10-item structure of SRM was $\alpha = .74$, suggesting that the items have satisfactory internal consistency. A principal components analysis (PCA) was conducted on the 10 revised items with oblique rotation (Oblimin with Kaiser Normalization). The Kaiser-Meyer-Olkin measure of sampling adequacy was .75, above the recommended value of .6, and Bartlett's test of sphericity was significant ($c^2(28) = 488.65$, $p < .001$), indicating that correlations between items were sufficiently large to explore dimensionality. Two components presented with eigenvalues over the recommended criterion and explained 47% of the variance. Item loadings above .40



were looked for and then the items were examined for interpretability based on the scientific reasoning construct.

Items loading on the first component, which explained 33% of variance, seemed to all relate to one's reasoning in complex biomedical research, issues concerning bias and the development of scientific studies. Factor 1 loadings ranged from .64 to .76, with inter-item correlations from .31 to .54 and item-total correlations from .46 to .60. Factor 1 was labelled Scientific Reasoning in Biomedical Research (SRR) for further reference. Items loading on the second component, explaining 13% of the variance, seemed to relate to a basic quantitative reasoning found with scientific data interpretation, such as numeracy skills. Factor 2 loadings showed one problematic item, with a factor loading of .38, which was also impacting internal consistency of SRM. Item 7 was eliminated from further analysis, but was kept for future development of SRM. The remaining items loading on the second component ranged from .43 to .83, with inter-item correlations from .18 to .50 and item-total correlations from .28 to .50. Factor 2 was labelled Basic Quantitative Reasoning (BQR). Table 1 represents the main findings of the exploratory analysis, which were then tested within a partial confirmatory factor analysis (PCFA). PCFA consisted of extracting the two factors retained after PCA using the Maximum Likelihood method of extraction and then calculating model fit parameters, $\chi^2(26) = 86.68$ (ratio chi-square to degrees of freedom should be less than or equal to 2 or 3) and a normed fit index of NFI = .82 (value should be greater than .95 for a good model fit), suggesting the model would benefit from further exploration.

Table 1. Principal components analysis of revised SRM items

Items	F1* loadings	F2** loadings
Item 1. <i>Understanding the mechanism by which drug resistance develops</i>	.76	
Item 5. <i>Understanding the use of random sampling</i>	.75	
Item 9. <i>Understanding bias as a source of study design</i>	.74	
Item 10. <i>Understanding causation versus association</i>	.66	
Item 8. <i>Understanding the effects of confounding variables</i>	.64	
Item 3. <i>Estimating correlation coefficients</i>		.83
Item 4. <i>Estimating data variance</i>		.70
Item 6. <i>Estimating probabilities</i>		.47
Item 2. <i>Using graphical representations for research data</i>		.43
Item 7. <i>Estimating research significance</i>		.38
% Variance explained	33%	13%
Eigenvalues	3.34	1.32
Cronbach's α	.77	.60***

Note: SRM = Scientific Reasoning in Medicine; N = 209; *F1 = Scientific Reasoning Applied in Biomedical Research (SRR); **F2 = Basic Quantitative Reasoning (BQR); *** Cronbach's α after eliminating Item 7.

KMO = .75; Bartlett's test: $\chi^2_{45} = 488.65, p < .001$; Total variance explained 47%; Cronbach's α SRM = .74

Correlations between the revised SRM and the convergent measures showed significant patterns of association, as presented in Table 2. Participants who scored higher on the SRM had higher scores on the Cognitive Reflection Test ($r = .56, p < .01$), which seemed to be the strongest association, comparing all convergent measures. Interest for biomedical research (IBR) and Need for Cognition (NFC) seemed to demonstrate the same degree of association to SRM ($r = .47, p < .01$), while Faith in Intuition (FI) did not produce significant patterns of association. Scores on the Basic Quantitative Reasoning (BQR) factor indicated a higher association with scores on NFC, but lower with scores on IBR. Results further showed that scores found with the Scientific Reasoning in Biomedical Research (SRR) factor were highly correlated with scores on the CRT and IBR. Neither factor showed significant correlations with FI.



Table 2. Means, standard deviations and correlation matrix of variables (Revised SRM).

Variable	M	SD	1	2	3	4	5	6
1. IBR	3.56	1.06						
2. NFC	18.08	3.39	.49**					
3. FI	18.02	3.65	.25**	.22				
4. CRT	1.86	0.98	.32**	.34**	.10			
5. SRM	5.83	2.61	.47**	.47**	.19	.56**		
6. BQR	3.36	1.42	.38**	.43**	.11	.42**	.81**	
7. SRR	2.47	1.68	.42**	.37**	.20	.51**	.86**	.41**

Note: IBR = Interest for biomedical research; NFC = Need for cognition; FI = Faith in Intuition; CRT = Cognitive Reflection Test; SRM = Scientific Reasoning in Medicine; BQR = Basic Quantitative Reasoning; SRR = Scientific Reasoning Applied in Biomedical Research; N = 209; ** $p < .01$

The first step in analysing predictive validity of revised SRM consisted of exploratory analyses of the relationships between all variables included in Study 2 and demographic characteristics. An analysis of variance showed that the effect of medical expertise (academic/professional level) on scientific reasoning (SRM) was significant, *Welch's F* (4, 70.17) = 29.80, $p < .01$, $\eta^2 = .39$ (large effect > .26). Post-hoc analyses using Games-Howell procedure showed that 3rd year students ($N = 53$, $M = 3.68$, $SD = 2.34$) scored significantly lower than all other cohorts, except for 1st and 2nd year students ($N = 31$, $M = 4.84$, $SD = 2.60$), who scored significantly lower only in comparison with the postgraduate students who hold a clinical position ($N = 40$, $M = 8.38$, $SD = 1.90$). Additional analyses of variance on the effect of medical expertise on scientific reasoning applied in biomedical research (SRR) showed that 1st and 2nd year students ($N = 31$, $M = 2.61$, $SD = 1.81$) demonstrated significantly lower SRR scores than 4th, 5th and 6th year students ($N = 68$, $M = 3.77$, $SD = 1.41$), postgraduate students ($N = 40$, $M = 4.68$, $SD = .62$) and clinicians ($N = 16$, $M = 3.93$, $SD = .85$), effect which was not found with basic quantitative reasoning (BQR). An analysis of variance was also performed to determine the effect of age on SRM scores and found that age had a significant effect on scientific reasoning, *Welch's F* (3, 26) = 26.00, $p < .01$, $\eta^2 = .27$ (large effect > .26). Gender did not produce any significant differences in relation to scientific reasoning.

Table 3. Summary of hierarchical regression analysis for variables predicting scientific reasoning in medicine (SRM).

Variable	Model 1			Model 2			Model 3			Model 4			Model 5		
	B	SE B	β	B	SE B	β	B	SE B	β	B	SE B	β	B	SE B	β
CRT	1.51	.15	.56	1.42	.13	.53**	1.23	.13	.46**	1.12	.14	.42**	1.04	.14	.39**
Age				1.36	.16	.43**	1.19	.16	.37**	0.72	.24	.23**	0.60	.25	.19*
IBR							0.57	.13	.23**	0.64	.13	.26**	0.53	.14	.21**
Academic level										0.43	.17	.19*	0.46	.17	.20**
NFC													0.10	.04	.13*
R ²		.32			.50			.54			.56			.57	
ΔR^2		.32			.18			.05			.01			.01	
F		96.31**			74.11**			20.23**			6.15*			5.55*	

Note: CRT = Cognitive Reflection Test; IBR = Interest for biomedical research; NFC = Need for Cognition; N = 209; * $p < .05$, ** $p < .01$.

The second step in establishing the instrument's predictive validity was to conduct multiple hierarchical regression analyses for SRM, BQR and SRR. Table 3 shows the results of the regression of cognitive reflectiveness



(CRT), age, interest in biomedical research (IBR), academic/professional level, need for cognition (NFC) on scientific reasoning (SRM). Based on prior tested relationships, variables were added to the regression model starting with cognitive reflectiveness (CRT), which explained 32% of the variance and the estimated regression coefficient was significant ($B = 1.51, \beta = .56, p < .01$). When entered into the model, age explained 18% of the variance ($B = 1.51, \beta = .56, p < .01$), while interest in biomedical research ($B = 0.57, \beta = .23, p < .01$), academic level ($B = 0.43, \beta = .19, p < .05$) and need for cognition ($B = 0.10, \beta = .13, p < .05$) explained an additional 7% of the total variance.

Table 4. Summary of hierarchical regression analysis for variables predicting basic quantitative reasoning (BQR).

Variable	Model 1			Model 2			Model 3		
	B	SE B	β	B	SE B	β	B	SE B	β
Age	0.77	.11	.44**	0.72	.10	.41**	0.61	.10	.35**
CRT				0.58	.08	.40**	0.48	.09	.33**
NFC							0.09	.03	.21**
R ²		.20			.35			.39	
ΔR^2		.20			.16			.03	
F		50.17**			49.91**			11.30**	

Note: CRT = Cognitive Reflection Test; NFC = Need for Cognition;

N = 209; * $p < .05$, ** $p < .01$.

Table 4 represents the results for the multiple regression analysis conducted on the Basic Quantitative Reasoning (BQR) component. Age significantly predicted BQR scores ($B = 0.77, \beta = .44, p < .01$), accounting for 20% of the variance. Cognitive reflectiveness was a significant estimated regression coefficient ($B = 0.58, \beta = .40, p < .01$) and explained 16% of the variance, while need for cognition (NFC) added 3% ($B = 0.09, \beta = .21, p < .01$). Finally, Table 5 shows the multiple regression modelling results for Scientific Reasoning applied in Biomedical Research (SRR). As opposed to the regression model for BQR, significant predictors for SRR were found to be cognitive reflectiveness ($B = 0.88, \beta = .51, p < .01$), academic/professional level ($B = 0.54, \beta = .37, p < .01$) and interest for biomedical research ($B = 0.42, \beta = .27, p < .01$), which together explained 46% of the total variance.

Table 5. Summary of hierarchical regression analysis for variables predicting scientific reasoning applied in biomedical research (SRR).

Variable	Model 1			Model 2			Model 3		
	B	SE B	β	B	SE B	β	B	SE B	β
CRT	.88	.10	.51**	.75	.10	.44**	.60	.10	.35**
Academic level				.54	.08	.37**	.53	.08	.36**
IBR							.42	.09	.27**
R ²		.26			.40			.46	
ΔR^2		.26			.13			.06	
F		74.06**			44.55**			24.08**	

Note: CRT = Cognitive Reflection Test; IBR = Interest in Biomedical Research;

N = 209; * $p < .05$, ** $p < .01$.



Discussion

The purpose of the current research was to propose a valid assessment method for scientific reasoning skills across multiple levels of medical expertise by developing a skills-based instrument and providing evidence of its validity through multiple revisions. The validation process was conducted on the basis of the following main hypotheses: content should contribute significantly to explaining the variance in SRM, the level of scientific reasoning should differ in relation to medical expertise and individuals who were more inclined to rational, resource-intensive responses were expected to show higher scientific reasoning. Research hypotheses were built on findings regarding clinical performance and differences in information processing between novices and experts. So, significant differences in SRM in relation to academic/professional, significant associations with other cognitive measures and different prediction patterns for the two interrelated components – basic quantitative reasoning and scientific reasoning in biomedical research – were considered evidence of validity.

Results on the pretesting phase of SRM showed satisfactory psychometric qualities, with multiple stages of revision which allowed for a retention of 10 corrected items with 3 options (2 distractors). Correct response rate was surprisingly low in the first phase, even in the case of participants who had both clinical and research expertise, but improved after eliminating 2 distractors. Research on internal consistency showed Cronbach's alpha values tend to be lower with heterogeneous domains and in the case of instruments with a small number of items (Drummond & Fischhoff, 2015), so we expected lower values, which increased with content revision. Results on the revision phase of SRM demonstrated significant differences between levels of medical expertise and scores obtained on the SRM measure. With an increase in the number of participants in the revision phase, results allowed for a more detailed representation of differences – postgraduate students who hold a clinical position obtained the highest overall score, followed by clinicians, undergraduate students enrolled in their 4th through 6th academic year, undergraduate students enrolled in their 1st and 2nd academic year and lastly, the 3rd year students. Cohorts were analyzed separately on the bases of comparable sample sizes and descriptive preliminary analysis, which showed differences in performance for 3rd year undergraduate students. The decline in performance may be due to the particularly demanding 3rd academic year, defined by the introduction of clinical disciplines and practical stages into the curriculum. Explanations may also be found with literature on medical expertise development, which demonstrated that intermediates may exhibit a drop in performance, caused by the continuous acquisition of knowledge during a phase in which adequate knowledge-based associations have not yet been achieved (Patel & Groen, 1986).

Results on SRM's relation to other measures confirmed strong positive associations with interest in biomedical research, need for cognition and cognitive reflectiveness or reflection. The latter seemed to have a greater influence on scientific reasoning than a rational-analytical cognitive style. Research on CRT performance found that this measure may involve both processing systems, which might explain its predictive value for scientific reasoning in medicine and biomedical research. Complex reasoning processes have been thought to also involve an intuitive thinking style to a certain extent, particularly with experienced doctors. An intuitive thinking style has been negatively associated with performance in decision-making, yet positively associated with experience keeping in mind that these associations have proven to be highly dependent on the type of task (Phillips, Fletcher, Marks & Hine, 2016). Results indicated that, although, experienced clinicians showed higher scores on the FI measure, faith in intuition did not reveal significant patterns of association with scientific reasoning.

Regarding differences in prediction patterns, basic quantitative reasoning and scientific reasoning in biomedical research seem to be explained by a different combination of factors, which we have taken into consideration as evidence of validity of SRM instrument. Age, cognitive reflection and need for cognition seemed to be the variables which accounted for basic quantitative reasoning, while scientific reasoning applied in biomedical research proved to be influenced by cognitive reflection, medical expertise and an interest in biomedical research. Taken together, these findings align with research hypotheses on the development of scientific reasoning, operationalized through skills like applying quantitative reasoning and scientific discovery principles in clinical practice.

The present research has several limitations related to sample sizes, reliability and generalizability. Establishing validity represents a continuous process of improvement, with several steps taken so far, but with more evidence needed still. Further validation is required before it is possible to confidently recognize these particular skills as evidence of scientific reasoning. The SRM model with two components could not be confirmed, further item refinement and exploratory analysis is needed to ensure adequate modelling. Another limitation comes from the initial development of the instrument, which was revised by a group of experts selected using convenience sampling. Furthermore, valuable input could have come from a larger number of clinicians or experienced practitioners.



Conclusions

Present research findings can be analyzed in relation to two important research directions which address key aspects regarding the development of scientific reasoning in medical education. The first line of research refers to cognitive factors which contribute to medical expertise, while the second line of research deals with challenges in respect to educational assessments throughout medical training in the context of global efforts to implement competency-based curricula.

Findings on the validity of the skills-based instrument demonstrate that the constructed items are able to discriminate between levels of expertise and its structure corresponds to hypothesized relevant reasoning skills. Scientific reasoning in medical education incorporates both basic quantitative reasoning, which is associated with skills such as estimating probabilities and data analysis, as well as a more complex research-driven scientific reasoning, which translates to an understanding of causal mechanisms and essential research components. Although the development trajectory of scientific reasoning can be traced along medical expertise, findings show that, specifically, the more complex type of scientific thinking is significantly associated with expertise. This comes in support of research into diagnostic reasoning and suggests that even though basic science knowledge doesn't seem to be explicitly applied in clinical case workups, complex research-driven reasoning may play a critical part in the diagnostic process of certain cases which don't fit to standard criteria and should be considered when exploring medical cognition. Cognitive reflection has also proven to be a particularly strong factor in the development of scientific reasoning and has been found to influence both numerical skills as well as complex scientific thinking. As cognitive reflection has been linked to both types of information processing systems, inferences can be drawn concerning the nature of clinical reasoning as an intertwined mechanism of analytical processing and the use of heuristics. Being able to reflect upon clinical data critically and deliberately should be a core skill to be developed throughout medical training, thus shaping future reflective practitioners.

Skill assessment within a competency-based medical education is pivotal to the professional development of medical trainees. Although critiques of evidence-based medicine (EBM) have paved the way for Person Centered Medicine (PCM) as the new paradigm in clinical practice, medical trainees need to develop specific scientific reasoning skills in order to be able to appraise new evidence, critically evaluate its significance and determine its suitability in certain clinical scenarios. Future research should direct efforts at understanding how scientific reasoning becomes integrated and how it is employed in clinical contexts, what is its role in clinical training, how it influences performance in residency and if there are scientific reasoning strategies which would reduce medical errors.

References

- Altman, N., & Krzywinski, M. (2016). Regression diagnostics. *Nature Methods*, 13 (5), 385-386.
- Amara, F., & Smyth, D. D. (2015). Developing a scientific basis of medicine framework in curriculum renewal: The ideal model. *Medical Science Educator*, 25 (4), 365-554.
- Barz, D. L., & Achimaş-Cadariu, A. (2016). The development of scientific reasoning in medical education: A psychological perspective. *Clujul Medical*, 89 (1), 32-7.
- Campitelli, G., & Gerrans, P. (2013). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory and Cognition*, 42 (3), 434-447.
- Carraccio, C., Englander, R., Van Melle, E., Ten, C. O., Lockyer, J., Chan, M., Frank, J. R., Snell, L. S., (2016). Advancing competency-based medical education: A charter for clinician educators. *Academic Medicine*, 91 (5), 645-9.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*, 10 (7), 1-9.
- Croskerry, P. (2009). A universal model of diagnostic reasoning. *Academic Medicine*, 84 (8), 1022-1028.
- Croskerry, P., Singhal, G., & Mamede, S. (2013). Cognitive debiasing 1: origins of bias and theory of debiasing. *BMJ Quality & Safety*, 2 (22), 58-64.
- Drummond, C., & Fischhoff, B. (2015). Development and validation of the scientific reasoning scale. *Journal of Behavioral Decision Making*, 30 (1), 26-38.
- Dunbar, K., Fugelsang J. (2005) Scientific thinking and reasoning. In K. J. Holyoak & R. G. Morrison (Eds.) *The Cambridge handbook of thinking and reasoning* (pp. 705,720.). Cambridge University Press.
- Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology*, 71 (2), 390-405.
- Ernst, A., & Albers, C. (2016). Regression assumptions in clinical psychology research practice - A systematic review of common misconceptions. *PeerJ Preprints*, 4 (1), 1-23.
- Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, 19 (4), 25-42.



- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141 (1), 2-18.
- Giacomini, M. (2009). *Theory based medicine and the role of evidence: Why the emperor needs new clothes, again*. McMaster University, Department of Clinical Epidemiology and Biostatistics. Ontario: Centre for Health Economics and Policy Analysis.
- Gignac, G. E. (2009). Partial confirmatory factor analysis: Described and illustrated on the NEO-PI-R. *Journal of Personality Assessment*, 91 (1), 40-47.
- Gormally, C., Brickman, P., & Lutz, M. (n.d.). Developing a test of Scientific Literacy Skills (TOSLS): Measuring undergraduates' evaluation of scientific information and arguments. *CBE - Life Sciences Education*, 11 (4), 364-377.
- Han, J. (2013) Scientific reasoning: Research, development and assessment. *PhD Thesis*. Ohio: The Ohio State University.
- Kulasegaram, K., Manzone, J. C., Ku, C., Skye, A., Wadey, V., & Woods, N. N. (2015). Cause and effect: Testing a mechanism and method for the cognitive integration of basic science. *Academic Medicine*, 90 (11), 63-69.
- Kuhn, D. (2010). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *The Wiley Blackwell handbook of childhood cognitive development* (pp. 2-27). 2nd ed. Oxford.
- Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching*, 15 (1), 11-24.
- Lawson, A. E. (2000). *Classroom test of scientific reasoning*. Test Revision, Arizona State University, Arizona.
- Lisk, K., Agur, A. M. R., & Woods, N. N. (2016) Exploring cognitive integration of basic science and its effect on diagnostic reasoning in novices. *Perspectives in Medical Education*, 5, 147-153.
- Lovelace, M., & Brickman, P. (2013). Best practices for measuring students' attitudes toward learning science. *CBE - Life Sciences Education*, 12, 606-617.
- Marsan, L. A., D'Arcy, C. E., & Olimpo, J. T. (2016). The impact of an interactive statistics module on novices' development of scientific process skills and attitudes in a first-semester research foundations course. *Journal of Microbiology & Biology Education*, 17 (3), 436-443.
- Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science*, 10, 91-116.
- Patel, V. L., Arocha, J. F., & Zhang, J. (2005). Thinking and reasoning in medicine. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 727-750). New York: Cambridge University Press.
- Patel, V. L., Yoskowitz, N. A., Arocha, J. F., & Shortliffe, E. H. (2009). Cognitive and learning sciences in biomedical and health instructional design: A review with lessons for biomedical informatics education. *Journal of Biomedical Informatics*, 42 (1), 176-197.
- Pennycook, G., Cheyne, J., & Koehler, D. (2015). Is the cognitive reflection test a measure of both reflection and intuition? *Behavior Research Methods*, 48 (1), 341-348.
- Phillips, W. J., Fletcher, J. M., Marks, A. D., & Hine, D. W. (2016). Thinking styles and decision making: A meta-analysis. *Psychological Bulletin*, 142 (3), 260-290.
- Shingala, M. C., & Rajyaguru, A. (2015). Comparison of post hoc tests for unequal variance. *International Journal of New Technologies in Science and Engineering*, 2 (5), 22-33.
- Trimble, M., & Hamilton, P. (2016). The thinking doctor: Clinical decision making in contemporary medicine. *Clinical Medicine*, 16 (4), 343-346.
- Welsh, M. B., Burns, N. R., & Delfabbro, P. H. (2013). The Cognitive Reflection Test: How much more than Numerical Ability? *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 1587-1592). Cognitive Science Society.
- Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, 20, 99-149.

Received: April 04, 2017

Accepted: May 12, 2017

Daniela Luminița Barz

PhD Candidate, Department of Medical Education, „Iuliu Hațieganu” University of Medicine and Pharmacy Cluj-Napoca, 8 Victor Babes Street, Cluj-Napoca, Romania.
E-mail: Luminita.Barz@umfcluj.ro

Andrei Achimaș-Cadariu

MD, PhD, Professor, „Iuliu Hațieganu” University of Medicine and Pharmacy, Cluj-Napoca, Romania.
E-mail: aachimas@umfcluj.ro

