# An Evaluation of the Psychometric Properties of an Advising Survey for Medical and Professional Program Students

Kenneth D. Royal<sup>1</sup> & Liara M. Gonzalez<sup>1</sup>

<sup>1</sup> Department of Clinical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, NC, USA

Correspondence: Kenneth D. Royal, Department of Clinical Sciences, College of Veterinary Medicine, North Carolina State University, 1060 William Moore Dr., Raleigh, North Carolina, 27607, USA. E-mail: kdroyal2@ncsu.edu

Received: March 8, 2016	Accepted: March 17, 2016	Online Published: April 7, 2016
doi:10.5539/jedp.v6n1p195	URL: http://dx.doi.org/10.5	539/jedp.v6n1p195

# Abstract

The purpose of this study was to evaluate the psychometric properties of a newly developed instrument intended to measure faculty competence as it pertains to their role as advisors, particularly in medical and professional programs. A total of 166 students completed the Faculty Advisor's Skills and Behaviors Inventory (FASBI). The psychometric properties of the FASBI were evaluated using the Rasch Rating Scale Model. Results indicate the FASBI is a psychometrically-sound instrument capable of producing valid and reproducible measures.

Keywords: advising, faculty, psychometrics, measurement, instrument, medical education, higher education

# 1. Introduction

# 1.1 Introduction of Problem

Student advising plays a critical role in student development (Light, 2001; Pizzolato, 2008; Reinarz & Ehrlich, 2002) and good advising has continually been linked to students' satisfaction with an institution (Baker & Griffin, 2010; Elliott & Healy, 2001; Freeman, 2008), and academic success (Campbell & Nutt, 2008; Museus & Ravello, 2010). Beggs, Bantham, and Taylor (2008) reported that college and university advising often influences students' career decisions and direction as well. Although colleges and universities often indicate a commitment to advising, extant research suggests advising is often "uneven in quality and ultimately ineffective" (Hossler, Ziskin, & Gross, 2009, p. 8). To this end, research by Hossler and colleagues (2009) and Kramer (2003) note that most colleges and universities in the United States do not assess advising. This is particularly unfortunate because the failure to evaluate the quality of advising at one's institution conveys a message to the faculty that advising is of low priority or is undervalued.

# 1.2 Purpose of the Present Study

Programmatic assessments are an important tool to actively evaluate successes, shortcoming, and outright failures of a program. This is essential for ongoing programmatic improvement. Given faculty advising is a topic of critical importance for students and their education, it is necessary that faculty advisors routinely be evaluated. Objective assessments of faculty strengths and weaknesses can inform the development of training programs or other interventions designed to help faculty in this vitally important role. Thus, the purpose of this study was to evaluate the psychometric properties of a newly developed instrument intended to measure faculty competence as it pertains to their role as advisors, particularly in medical and professional programs (e.g., veterinary medicine, pharmacy, dentistry, physical therapy, etc.).

# 2. Methods

# 2.1 Participants

Survey participants involved students at the North Carolina State University College of Veterinary Medicine. A census sampling of all students enrolled across each of the four years in a Doctor of Veterinary Medicine (DVM) program were invited to participate in the study. Of the 394 students invited, 167 provided valid responses for the 25 items evaluating the skills and behaviors of faculty advisors. Students' ages ranged from 21 to 49 years

old, with the average student being 26 (SD = 3.9). The median age for the sample was 25. A complete breakdown of student demographic characteristics are presented in Table 1.

	Ν	%
Sex		
Male	22	13.2
Female	145	86.8
Race		
White	124	74.3
Black	11	6.6
Other	32	19.2
Ethnicity		
Hispanic/Latino(a)	22	13.2
DVM Year		
Class of 2016	41	24.6
Class of 2017	39	23.4
Class of 2018	39	23.4
Class of 2019	48	28.7

	Τ	a	bl	e	1.	D	emograp	hic	c c	harac	teris	tics	of	sam	ple	е
--	---	---	----	---	----	---	---------	-----	-----	-------	-------	------	----	-----	-----	---

#### 2.2 Instrumentation

A number of instruments are available for assessing faculty advising. However, most instruments focus on elements that are typically less relevant to faculty in medical and health-related professional programs. For example, most instruments assessing advising ask questions that focus on the extent to which faculty were able to help students identify a major and select appropriate/relevant courses to ensure degree requirements are fulfilled and students graduate in a timely matter. For medical and health-related programs, students have already selected their major, and typically navigate the curriculum in a cohort manner. Thus, there is little need to assess these facets. Similarly, instruments that focus on advising as it pertains to graduate and doctoral students (e.g., the Graduate Advising Survey for Doctoral Students) tends to focus on many of the same issues, but also perceptions of the department's climate, peer influence, and the advisor's role as someone training researchers. Given medical and health-related programs have somewhat different needs, we developed a new survey based on elements we believed to be particularly relevant for these types of programs (see Zimmerman & Mokma, 2004; Barnes et al., 2011; Belcheir, 2000). This resulted in the development of the Faculty Advisor's Skills and Behaviors Inventory (FASBI).

#### 2.3 Validation Framework

Much has been written in the psychometric literature about the limitations of traditional statistical approaches for validation studies. Royal (2010) lists six major weaknesses of traditional approaches, which include: 1) problems associated with erroneously treating ordinal ratings as interval level measures, 2) erroneously perceiving all items of equal importance, 3) erroneously assuming error is equal across all measures, 4) sample-dependency problems, 5) parametric statistical approaches require normally distributed data, and 6) how missing data can seriously threaten score validity. Many measurement scholars have declared Rasch models the "gold standard" for psychometric validation studies because they overcome the limitations of traditional statistical models and are the only measurement models that have the property of invariance (Royal, 2010; Salzberger, 2002; Wright, 1997).

Rasch models are particularly attractive for analyzing survey research data as they are able to separate person measures (e.g., knowledge, ability, skills, etc.) and item data (e.g., difficulty of item, difficulty of task, etc.) and explore how these two facets interact with one another. In the present study, the latent trait being measured is students' tendency to endorse (agree with) a variety of survey items. Rasch models produce linear measures (called "logits") and create a common linear continuum onto which both person and item measures are mapped. Because Rasch models are probabilistic models, the likelihood of a student endorsing an item can be modeled as a logistic function of the distance between a student and a survey item. Readers interested in learning more about Rasch measurement models are encouraged to see Engelhard (2014) and Bond and Fox (2007).

For the present study we opted to use the Rasch Rating Scale Model (RRSM) (Andrich, 1978) for the analysis of survey data as it is particularly well-suited for polytomous data. According to the RRSM model, the probability of a person n responding in category x to item i, is given by:

$$P_{xni} = \frac{\exp \sum_{j=0}^{x} [\beta_n - (\delta_i + \tau_j)]}{\sum_{k=0}^{m} \exp \sum_{j=0}^{k} [\beta_n - (\delta_i + \tau_j)]} \qquad x = 0, l, ..., m$$

where  $\tau_0 = 0$  so that  $\exp \sum_{j=0}^{0} [\beta_n - (\delta_i + \tau_j)] = 1$   $\beta_n$  is the person's position on the variable,  $\delta_i$  is the scale

value (difficulty to endorse) estimated for each item i and  $\tau_1, \tau_2, ..., \tau_m$  are the *m* response thresholds estimated for the *m* + 1 rating categories. Winsteps measurement software (Linacre, 2016) was used to perform the data analysis. Parameters were estimated using joint maximum likelihood estimation procedures (Wright & Masters, 1982).

#### 3. Results

Before the results of the psychometric validation are presented, it is necessary to first explore summative descriptive statistics. Table 2 presents a summary of mean and Standard Deviation (SD) values for each of the 25 items.

Item	Mean	SD
1. Is proactive in reaching out to meet with me	2.48	1.13
2. Is easy to get in touch with	3.25	0.84
3. Responds to my emails/calls in a timely manner	3.36	0.78
4. Gives me as much time as I need when we meet	3.44	0.71
5. Encourages me to seek him/her for help	3.14	0.95
6. Takes a personal interest in me	3.13	0.94
7. Is a good listener	3.45	0.69
8. Respects my opinions and feelings	3.54	0.56
9. Is knowledgeable about academic policies and procedures in the college	3.44	0.63
10. Provides me with accurate information and answers to my questions	3.48	0.59
11. Refers me to other resources for assistance, when necessary	3.51	0.62
12. Encourages me to achieve my educational goals	3.51	0.63
13. Helps me identify the obstacles I need to overcome to reach my academic and/or	3.21	0.79
professional goals		
14. Makes his or her self available to meet with me	3.27	0.87
15. Helps me to examine my needs, interests, and values	3.24	0.81
16. Is familiar with my academic background	3.17	0.85
17. Helps me explore careers in my area of interest	3.14	0.93
18. Is approachable and easy to talk to	3.39	0.75
19. Encourages me to ask questions and discuss any concerns	3.35	0.74
20. Shows concern for my personal growth and development	3.26	0.81

Table 2. Results of traditional statistical analysis

21. Encourages me to discuss personal issues	2.82	1.01
22. Has a lot in common with me	2.88	0.78
23. Is helpful	3.30	0.75
24. Is someone I consider a mentor	3.02	1.00
25. Is someone I would recommend as an advisor to other students	3.12	0.97

#### 3.1 Dimensionality

Virtually all survey data sets consist of multiple dimensions. The question, however, is to what degree are various dimensions present, and is there evidence of a single, primary underlying construct being measured. To answer this question, a Rasch-based Principal Components Analysis (PCA) of standardized residual correlations was performed to assess dimensionality. In total, 64.4% of the Rasch dimension was explained, with 20.8% being attributed to the items. The largest secondary dimension explained 4.4% of the variance and had an Eigenvalue of 3.0, indicating a strength of about 3 items. The ratio of the variance explained by the items (20.8%) and the largest secondary dimension (4.4%) is about 5:1. Thus, for all practical purposes the data were sufficiently unidimensional for a Rasch measurement analysis.

#### 3.2 Reliability

Reliability was assessed using multiple methods. First, the traditional measure of reliability (Cronbach's alpha) was calculated, then reliability measures were calculated from the Rasch measurement framework using "real" and "modeled" measures. Cronbach's  $\alpha$  reliability was .972 for the 25 items. Rasch-based reliability estimates were .94 for "real", and ".95" for modeled, suggesting true reliability is likely somewhere in between. All three measures of reliability indicate highly reproducible measures (Royal & Hecker, 2015). Separation, which refers to the number of statistically distinguishable levels within the data, was also assessed. The separation statistic was 4.27, indicating approximately four statistically distinguishable levels were present in the data.

#### 3.3 Rating Scale Effectiveness

Rating scale diagnostics were assessed to determine the extent to which students were able to appropriately interpret and make use of each rating scale category (see Table 3). Results indicate students made use of the full rating scale, although fewer students tended to provide ratings of disagreement. Infit and outfit mean square fit statistics were in appropriate range, the category measures and structure calibration measures each advanced in a stepwise manner as anticipated (Linacre, 2002). Collective results provide evidence the rating scale functioned very well for this instrument.

Dating Catagory	10	0/	INFIT	OUTFIT	Structure	Category
Rating Category	11	70	MnSq	MnSq	Calibration	Measure
(1) Strongly Disagree	186	5	1.00	1.06	None	-3.50
(2) Disagree	407	12	1.13	1.22	-2.27	-1.49
(3) Agree	1322	38	.95	.87	67	1.17
(4) Strongly Agree	1603	46	.97	.94	2.95	4.07

#### Table 3. Rating scale diagnostics

#### 3.4 Item and Person Measure Quality

A variety of statistical indicators were used to assess item and person measure quality (see Table 4). First, the variability of item difficulty measures were investigated. Results indicate logit measures ranged between -1.21 to 2.38 with an average standard error of .17 (SD = .01) in magnitude. These values indicate adequate variability with sufficiently small standard errors to ensure statistically stable measures. Infit and outfit mean square fit statistics were evaluated using the recommendations provided by Wright and Linacre (1994), noting ideal values should range between .60-1.40, and values exceeding 2.00 may be indicative of a "noisy" (potentially problematic) item. Of the 25 items, 23 fell within ideal range for fit statistic values. Item #23-[My advisor] is helpful yielded fit statistics slightly below .60 indicating these responses were a bit predictable. Item #1-[My

advisor] is proactive in reaching out to meet with me yielded inflated fit statistics (1.83 and 2.26, respectively), thus indicating a potentially problematic item. Point-measure correlations were all (high) positive values indicating excellent discriminatory abilities (Linacre, 2015a).

Similar procedures used to assess person measure quality. Person measures ranged from -4.01 to 6.52 with an average standard error of .52 (SD = .22) in magnitude. These values also indicate excellent variability with sufficiently small standard errors to ensure stable measurements. Only 9 students (5.4%) from the sample had at least one fit statistic exceeding 2.0 indicating possible misfit.

Item	Difficulty	Error	INFIT	OUTFIT Mean	Point Measure
	Measure		Mean Square	Square	Correlation
Q1	2.38	.14	1.83	2.26	.75
Q2	17	.17	1.48	1.48	.71
Q3	38	.18	1.67	1.72	.64
Q4	56	.19	1.31	1.52	.65
Q5	.06	.17	1.11	.98	.81
Q6	.23	.16	.84	.72	.85
Q7	71	.19	.87	.80	.74
Q8	-1.21	.20	.61	.44	.75
Q9	61	.19	1.29	1.34	.62
Q10	86	.19	.67	.61	.76
Q11	86	.20	.64	.52	.76
Q12	-1.13	.20	.63	.65	.77
Q13	.33	.18	1.07	1.07	.76
Q14	21	.18	.87	.73	.81
Q15	.28	.18	.78	.69	.81
Q16	.37	.18	1.05	1.11	.77
Q17	.57	.17	.99	.90	.80
Q18	73	.18	.83	.71	.78
Q19	37	.19	.80	.72	.79
Q20	11	.18	.67	.64	.84
Q21	1.52	.17	1.23	1.34	.79
Q22	1.52	.16	.99	1.23	.77
Q23	24	.18	.52	.46	.85
Q24	.57	.17	.77	.74	.86
Q25	.33	.17	.78	.69	.85

Table 4. Item quality indicators

Two items appeared to violate assumptions for Local Item Dependence (LID). Specifically, item #2—Is easy to get in touch with, and item #3—Responds to my emails/calls in a timely manner, shared a standardized residual correlation of .70 indicating high statistical dependence. This indicates a student's response to one of these items will likely correlate highly with their response to the other item.

Differential Item Function (DIF) analyses were performed to assess if the construct remained invariant across relevant subgroups, particularly class year and gender. The iterative-logit (Rasch-Welch) method presented in Linacre (2015b) was performed. Because multiple comparisons were made across 25 items, a Bonferroni correction was necessary to control for compounding error. This resulted in the p-value normally set at 0.05 being reduced to 0.002 in order to detect statistically significant differences. Results indicate no statistically significant differences across items per subgroup.

## 3.5 Construct Hierarchy

In psychometrics, the manner in which items are ordered along a linear continuum is considered a construct. In essence, most substantive results emanating from a Rasch measurement analysis can be found in the construct map (also known as a Wright Map). In short, a construct map presents a visual snapshot of the "psychometric ruler" onto which person and items measures have been placed (see Figure 1). In the present study, students

appear on the left side of the map, with those near the top representing individuals that had the least difficulty endorsing the items and those appearing at the bottom having the greatest difficulty endorsing the items. Likewise, items appear on the right side of the map, with those most difficult to endorse at the top and those easiest to endorse at the bottom. Here, students had the most difficulty endorsing item Q1—[My advisor] is proactive in reaching out to meet with me, and the least difficulty endorsing items Q12—[My advisor] encourages me to achieve my educational goals, and Q8—[My advisor] respects my opinions and feelings.



Figure 1. Construct hierarchy

# 4. Discussion

#### 4.1 Psychometric Properties of the FASBI

Samuel Messick's (1989) framework for construct validity provides a useful guide for interpreting validity evidence. According to Messick, validity is the integration of any evidence that impacts the interpretation or meaning of a score. Messick's framework consists of six unique "aspects" of validity: substantive, content, generalizability, structural, external and consequential. It is from this framework that we appraise validity evidence present in the aforementioned results.

First, a Rasch-based PCA of standardized residual correlations indicated the data were primarily unidimensional, indicating a single, primary latent trait was being measured. This evidence speaks to the substantive aspect of validity. Measures of reliability consistently exceeded .90 indicating highly reproductive measures. This evidence speaks to the generalizability aspect of validity. An evaluation of rating scale diagnostics indicated the rating scale functioned appropriately. This speaks to the structural aspect of validity. An evaluation of item and person measures confirmed the measures were psychometrically-sound. This speaks to the content aspect of validity. Additional validity evidence was discernible by way of DIF analyses that confirmed measures were invariant across class year and sex subpopulations. This speaks to the systematic aspect of validity and provides additional support for the generalizability aspect of validity. Results of this study have not been correlated with other studies, thus we present no evidence that speaks to the consequential aspect of validity. Finally, because the FASBI has not be used previously we cannot speak to the consequential aspect of validity which involves consequences (positive or negative) resulting from the use of the instrument (Royal & Puffer, 2014). In sum, there is an abundance of validity evidence to support the psychometric quality of the FASBI and its ability to produce high-quality measures.

## 4.2 Implications

The purpose of this study was to evaluate the psychometric properties of a newly developed instrument intended to measure faculty skills and behaviors as it pertains to their role as advisors, particularly in medical and professional programs (e.g., veterinary medicine, pharmacy, dentistry, physical therapy, etc.). Results of a thorough investigation of the FASBI's psychometric properties indicates the instrument is psychometrically-sound. Therefore, persons interested in evaluating faculty advising in medical, health, and various professional programs are especially encouraged to use this instrument. Of course, the FASBI may be relevant and appropriate for persons interested in evaluating in other disciplines as well.

The FASBI may be particularly helpful for evaluators as it addresses a wide-variety of topics of concern for most academic programs. Having insights about how students feel with respect to each of the items may be particularly informative with regard to identify attributes of more and less successful faculty advisors. Furthermore, identifying these strengths and weaknesses would be particularly helpful for preparing faculty training to become more effective advisors.

Finally, another implication of this study is methodological in nature. The techniques presented in this study involve what many measurement experts consider to be "gold standard" methods for conducting survey validation studies. Further, the use Messick's framework for evaluating and organizing validity evidence may help other researchers better identify how to present validity evidence to others.

#### 4.3 Limitations

Of course, this study is not without its limitations. While this study intentionally made no effort to speak to substantive findings as that was beyond the scope of this paper, it remains a potential limitation that the survey data were of a self-reported nature. The extent to which students may have provided socially desirable responses, or that non-response bias may be an issue remains unknown. With respect to the sample frame, students were by and large female, which is typically the norm for veterinary medical programs. Despite the underrepresentation of male respondents, comparatively speaking, it is important to note that tests for DIF were conducted to determine if students appear to respond differently to the FASBI items based on sex. Results of the DIF analysis concluded that participant's sex does not affect students' responses to any FASBI items in any systematic way, t(833) = .00, p = 1.000. Finally, veterinary medicine, much like all medical professional programs, has significant issues with regard to diversity. Future research should evaluate the functioning of FASBI items based on race and ethnicity as well.

# 5. Conclusion

Effective faculty advising remains a problem in many colleges and universities, including medical and health professions programs. To date, relatively few institutions having actively evaluated faculty advising. We suspect this may be in part due to a lack of appropriate instrumentation. Thus, this study sought to investigate the psychometric properties of an instrument intended to assess faculty competence as it pertains to advising. Results of the psychometric validation study provide considerable amount of validity evidence to support the quality of the FASBI. We encourage others to adopt the FABSI for similar evaluations of faculty advising.

#### References

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573. http://dx.doi.org/10.1007/BF02293814
- Baker, V. L., & Griffin, K. A. (2010). Beyond mentoring and advising: Toward understanding the role of faculty "developers" in student success. *About Campus*, 14(6), 2-8. http://dx.doi.org/10.1002/abc.20002
- Barnes, B. J., Chard, L. A., Wolfe, E. W., Stassen, M. L. A., & Williams, E. A. (2011). An evaluation of the psychometric properties of the Graduate Advising Survey for Doctoral Students. *International Journal of Doctoral Studies*, 6, 1-17.
- Beggs, J. M., Bantham, J. H., & Taylor, S. (2008). Distinguishing the factors influencing college students' choice of major. *College Student Journal*, 42, 381-394.
- Belcheir, M. J. (2000). *An evaluation of advising programs* (Research Report). Boise State University. Retrieved from http://eric.ed.gov/?id=ED443338
- Bond, T. G., & Fox, C. M. (2007). Applying the Rasch Model. In *Fundamental measurement in the human sciences* (2nd ed.). Lawrence Erlbaum Associate. http://dx.doi.org/10.4324/9781410614575
- Campbell, S. M., & Nutt, C. L. (2008). Academic advising in the new global century: Supporting student engagement and learning outcomes achievement. *Peer Review*, 10(1), 4-7.
- Elliott, K. M., & Healy, M. A. (2001). Key factors influencing student satisfaction related to recruitment and retention. *Journal of Marketing for Higher Education*, 10(4), 1-11. http://dx.doi.org/10.1300/J050v10n04\_01
- Engelhard, G. (2013). *Invariant measurement: Rasch measurement in the social, behavioral and health sciences*. Routledge.
- Freeman, L. C. (2008). Establishing effective advising practices to influence student learning and success. *Peer Review*, *10*(1), 12-14.
- Hossler, D., Ziskin, M., & Gross, J. P. K. (2009). Getting serious about institutional performance in student retention: Research-based lesson on effective policies and practices. *About Campus*, 13(6), 2-11. http://dx.doi.org/10.1002/abc.271
- Kramer, G. L. (2003). *Faculty advising examined: Enhancing the potential of college faculty as advisors*. Bolton, MA: Anker.
- Light, R. J. (2001). *The power of good advice for students*. Chronicle of Higher Education. Retrieved from http://chronicle.com/article/The-Power-of-Good-Advice-for/9193/
- Linacre, J. M. (2015a). *Correlations: Point-biserial, point-measure, residual*. Retrieved from http://www.winsteps.com/winman/correlations.htm
- Linacre, J. M. (2015b). *Differential item functioning DIF pairwise*. Retrieved from http://www.winsteps.com/winman/table30\_1.htm, 2015.
- Linacre, J. M. (2016). WINSTEPS® (Version 3.81). Computer Software. Beaverton, OR: Winsteps.com.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85-106.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). New York: Macmillan.
- Museus, S. D., & Ravello, J. N. (2010). Characteristics of academic advising that contribute to racial and ethnic minority student success at predominantly white institutions. NACADA Journal, 30, 47-58. http://dx.doi.org/10.12930/0271-9517-30.1.47
- Pizzolato, J. E. (2008). Advisor, teacher, partner: Using the learning partnerships model to reshape academic advising. *About Campus*, 13(1), 25. http://dx.doi.org/10.1002/abc.243
- Reinarz, A. G., & Ehrlich, N. J. (2002). Assessment of academic advising: A cross-sectional study. NACADA Journal, 22, 50-65. http://dx.doi.org/10.12930/0271-9517-22.2.50
- Royal, K. D. (2010). Making meaningful measurement in survey research: A demonstration of the utility of the Rasch model. *IR Applications*, 28, 1-16.

- Royal, K. D., & Hecker, K. G. (2015). Understanding reliability: A review for veterinary educators. *Journal of Veterinary Medical Education*. http://dx.doi.org/10.3138/jvme.0315-030R
- Royal, K. D., & Puffer, J. C. (2014). The consequential validity of ABFM examinations. *Journal of the American Board of Family Medicine*, 27(3), 430. http://dx.doi.org/10.3122/jabfm.2014.03.140089
- Salzberger, T. (2012). The illusion of measurement: Rasch versus 2-PL. *Rasch Measurement Transactions*, *16*(2), 882.
- Wright, B. D. (1997). Fundamental measurement. Rasch Measurement Transactions, 11(2), 558.
- Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*, 370.
- Wright, B. D., & Masters, G. N. (1982). Rating scale analysis: Rasch Measurement. Chicago: MESA Press.
- Zimmerman, A., & Mokma, A. (2004). Developing a student evaluation of advising survey instrument. *NACTA Journal*, 48(3), 2-6.

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).