Accepted Manuscript

The use of video vignettes to measure health worker knowledge. Evidence from Burkina Faso

Sheheryar Banuri, Damien de Walque, Philip Keefer, Haidara Ousmane Diadie, Paul Jacob Robyn, Maurice Ye

PII: S0277-9536(18)30413-1

DOI: 10.1016/j.socscimed.2018.07.046

Reference: SSM 11875

To appear in: *Social Science & Medicine*

Received Date: 14 November 2017

Revised Date: 22 May 2018

Accepted Date: 25 July 2018

Please cite this article as: Banuri, S., de Walque, D., Keefer, P., Diadie, H.O., Robyn, P.J., Ye, M., The use of video vignettes to measure health worker knowledge. Evidence from Burkina Faso, *Social Science & Medicine* (2018), doi: 10.1016/j.socscimed.2018.07.046.

The use of video vignettes to measure health worker knowledge.

Evidence from Burkina Faso

Sheheryar Banuri, PhD, Lecturer, University of East Anglia

Damien de Walque, PhD, Senior Economist, Development Research Group, The World Bank

Philip Keefer, PhD, Principal Economic Advisor, Institutions for Development Department, Inter-American Development Bank

Haidara Ousmane Diadie, MD, Senior Health Specialist, Health Nutrition and Population, The World Bank

Paul Jacob Robyn, PhD, Senior Health Specialist, Health Nutrition and Population, The World Bank

Maurice Ye, MD, MPH, PhD, Centre de Recherche en Santé de Nouna, Burkina Faso

Corresponding author: Damien de Walque, 1818 H Street NW, Washington DC, 20433, USA. + 12024732517, ddewalque@worldbank.org

Keywords: health care quality, health provider knowledge, vignette, video

Declarations of interest: none

Role of the funding source: This research is funded by the Health Results Innovation Trust Funds (HRITF) at the World Bank which funded the production of the video vignettes, the experiment set-up and the data collection.

Acknowledgments: We are extremely grateful for a fruitful collaboration with the Ministry of Health in Burkina Faso and in particular with Philippe Compaore, Baly Ouattara, Abdoul Karim Ouedraogo, Boukary Tapsoba and Paul Toe. We are also very grateful to all participants in the experiment and those who facilitated it, in particular the direction, staff and students from the Faculté de Médecine de l'Université de Ouagadougou, the Ecole Nationale de Santé Publique and the Ecole de Santé privée Sainte Edwige, as well as the direction and staff of the health facilities in the regions of Boucle du Mouhoun, Centre-Nord, Centre-Ouest and Nord. The data were collected by Centre Muraz with special thanks to Hervé Hien, Nicolas Meda and all staff implementing the lab in the field experiment. We are also grateful to Nouna Health Research Centre, through Maurice Ye who developed the clinical-based scenarios for the videos. Special thanks to Bambio Yiriyibin for coordinating and supervising the experiment and to Boubakar Diallo and the team from Les Films du Dromadaire for producing the videos. This study was approved by the Comité d'Ethique pour la Recherche en Santé, Ministère de la Santé and Ministère de la Recherche Scientifique et de l'Innovation, Burkina Faso. Decision# 2013-7-066, July 11, 2013. This research is funded by the Health Results Innovation Trust Funds (HRITF) at the World Bank. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank, its Executive Directors, or the countries they represent.

The use of video vignettes to measure health worker knowledge.

Evidence from Burkina Faso

Abstract

The quality of care is a crucial determinant of good health outcomes, but is difficult to measure. Survey vignettes are a standard approach to measuring medical knowledge among health care providers. Given that written vignettes or knowledge tests may be too removed from clinical practice, particularly where "learning by doing" may be an important form of training, we developed a new type of provider vignette. It uses videos presenting a patient visiting the clinic with maternal/early childhood symptoms. We tested these video vignettes with current and future (students) health professionals in Burkina Faso. Participants indicated that the cases used were interesting, understandable and common. Their performance was consistent with expectations. Participants with greater training (medical doctors vs. nurses and midwives) and experience (health professionals vs. students) performed better. The video vignettes can easily be embedded in computers, tablets and smart phones; they are a convenient tool to measure provider knowledge; and they are cost-effective instruction and testing tools.

Keywords: health care quality; health provider knowledge; vignette; video

Introduction

The quality of care is a crucial determinant of good health outcomes, but is difficult to measure (Hrisos et al. 2009). In particular, clinical quality is more difficult to measure compared to structural quality of the facilities. Facility surveys, knowledge tests and direct observations of medical consultations all exhibit limitations, ranging from bias to cost, that are particularly pronounced in poor country settings. To address these limitations, we introduce a new measurement device based on clinical video vignettes. We developed to assess the quality of care among medical practitioners in Burkina Faso. Our measure is low cost, more immersive and potentially less prone to bias when incentivized.

Standard facility surveys generally include indicators of structural quality focusing on the availability of equipment, drugs, and commodities, as well as on staffing (training, experience and absenteeism). Structural quality, however, while a prerequisite for the quality of health services, does not accurately measure the care actually delivered by providers.

Different methods have been used to assess the quality of interaction between provider and patient (Leonard and Maestad 2016). Patient exit interviews seek to obtain the patients' feedback about their experience. They can reveal useful information about respect, courtesy, or waiting times, but because patients are rarely well informed about standard medical protocols, exit interviews are usually less reliable about the content of care. They might also be affected by recall bias, if they are conducted long after the consultation, or by response bias, for example if conducted in or close to the health facility and where patients might feel embarrassed to express their true opinions.

Written survey vignettes are another standard approach to measuring the quality of interaction between provider and patient. They present surveyed staff with a typical case with symptoms and ask them to provide a diagnosis and recommend a treatment course (Glassman et al. 2000; Peabody, Luck et al. 2000 and 2004; Peabody, Tozija et al. 2004; Das and Hammer 2005; Veloski et al. 2005). However, survey vignettes measure knowledge and not actual practice. They are not able to capture effort and therefore miss the "know-do" gap. In addition, written vignettes or knowledge tests may be too removed from clinical practice, particularly in poor countries. First, "learning by doing" may be an important form of training, so that, second, health workers might be uncomfortable with the written form.

Direct observations of medical consultations allow measurement of actual health care practice, but they are likely to be influenced by Hawthorne effects as providers increase their level of effort when they realize that they are being observed (Leonard and Masatu, 2005 and 2006). The standardized patient method can address these concerns: actor patients, trained to feign a specific illness and record the provider's action, present themselves unannounced at a health facility (Das et al. 2012). However, this method is by its nature limited to health conditions which can easily be feigned and do not require an invasive procedure to be diagnosed. As such, it is quite difficult to apply to maternal and child health. Direct clinical observations and standardized patients are also relatively costly, logistically (hiring and training standardized patients can be a long process that is also difficult to scale given inconsistencies in performance, fatigue, etc.) and in monetary terms (standardized patients need specialized training which tend to increase the costs in terms of ability and time).

Acted-out case study vignettes are lower cost than standardized patient method and are similarly advantageous when health workers come from a variety of educational backgrounds. Acted-out

vignettes measure the competence of health workers in a real clinical setting better than written vignettes (Leonard and Maestad 2016). However, they still present similar, if lower, logistical and monetary burdens, compared to standardized patients.

We therefore developed a new type of provider vignette using videos presenting a patient with maternal/early childhood symptoms visiting the clinic. Compared to vignettes acted-out live, the video vignettes do not allow an assessment of the doctor-patient interaction. However, video vignettes confront substantially lower logistical obstacles. They are particularly easy to implement in contexts where tablet or phone-based surveys are now the norm. Compared to acted-out vignettes, they offer an exceptional ability to standardize presentation for measurement and research. Compared to written vignettes, they offer an opportunity to present complex signs and symptoms that might be hard to describe in a written vignette, but easier to demonstrate with an actor.

The video vignettes were applied within the context of a "lab-in-the-field experiment" in Burkina Faso that tests effects of various incentive measures on provider performance within the context of a "laboratory" setting. The lab experiment is one study component of a larger impact evaluation on a Performance Based Financing (PBF) pilot in Burkina Faso.

This paper describes our instrument and presents evidence of its validity. By validity, we refer to how well the video vignettes measure health worker knowledge (Sullivan 2011). Since our interest was precisely in investigating the quality of care related to medical conditions that are difficult for actors to simulate, we could not easily validate the measure by comparing results with the standardized patient method. Instead, we focus on two necessary conditions for

validity. First, participants with more experience and training should perform better than those with less. Second, and as a corollary, the measure should reflect the "real world" of significant heterogeneity of practitioner knowledge. Hence, practitioners should exhibit considerable dispersion in their performance as evidenced by variance in their responses. These validity conditions, linking the results of the vignettes to established measures of knowledge and capacity (medical qualifications and experience) are recommended in the literature (Downing 2003; Cook and Beckman 2006).

Methods

Recruitment: We recruited two types of participants in the study. First, in February-March 2014, we recruited 1,029 medical professionals who were attending training sessions for a new program of PBF in the health sector. The sessions were organized at the regional level in the cities of Gourcy, Kaya, Koudougou, Nouna and Ouahigouya in Burkina Faso. At the sessions, participants were invited to take part in a lab-in-the-field activity which included the video vignettes. Each health facility (including primary care facilities and district/regional hospitals) in the particular region were expected to send at least one representative from their facility to attend the training sessions. Most of the facilities provide primary health care services.

Second, in April-May 2014, we recruited 1,113 future health professionals (nurses, midwives, and doctors). Nursing and midwifery students were in their last year, and came from a public (Ecole Nationale de Santé Publique) and a private (Ecole de Santé Privée Sainte Edwige) nursing and midwifery school. Medical students were in their 5^{th} or 6^{th} year, and came from a public university (Faculté de Médecine, Université de Ouagadougou) in Ouagadougou, the capital city.

Video vignettes:

We designed video vignettes representing cases related to maternal and early childhood care. The cases followed symptom description and treatment protocols as defined by Burkina Faso's national therapeutic guidelines as well as WHO IMCI guidelines. The cases were developed and tested with nurses at the two nursing schools in Ouagadougou in January 2014. Cases with adequate variation in responses were kept while others were replaced with new cases.

For ability measurement, we used 4 cases: a pregnant woman experiencing pre-term labor, a 6-month-old child with pneumonia, a lactating woman with mastitis and a 3-year-old child with measles. In order to capture the variety of ways in which patients present themselves to practitioners, each case consisted of a video lasting either 60 or 100 seconds. In all cases, the same actress was used to play the role of the patient, and always in the same setting (an average clinic). Two of the videos lasted 60 seconds each, and represented "simple" cases in which the actress used normal language and had a standard appearance. These cases were the pre-term labor and the mastitis cases. The 100 second videos represented cases with more symptoms and, in addition, the actress used rambling language and appeared to be poorer and less educated. This more disadvantaged socio-economic status was conveyed through the patient's dress and the vocabulary and language she used to describe her illness history and symptoms. The two 100 second cases were the child with pneumonia and the child with measles

Table 1 reports the average time taken and the average score obtained for each of the four cases. The order of the 4 cases was set and not randomized. Since learning about the tasks was likely, we therefore expect the first case to have taken longer and to have yielded a lower score; this was indeed the case. In addition, subjects averaged less time on case 3, with a 60 second video, than

for cases 2 and 4, with 100 second videos. The table indicates that the score increased across the four cases. While this steady increase was potentially driven by learning effects, it is important to emphasize that the four cases differed with respect to diagnosis, length and complexity, and are therefore not easy to compare.

The videos were shot and edited by locally-renowned director Boubakar Diallo. Care was taken to select an actress of neutral appearance in terms of ethnicity. Subjects viewed the video and were asked to select the correct (1) diagnosis, (2) treatment, (3) follow-up schedule, and (4) alternate treatment in case the patient did not respond well to the initial treatment course. For each question, they selected their answer from a multiple-choice list (see example in Figure 1). Each answer set had one correct response, two nearly correct responses, and two wholly incorrect responses, yielding additional variation. “Wholly incorrect” responses would be inconsistent with the symptoms and data provided in the case while “nearly correct” responses would be consistent with most (but not all) of the symptoms provided to the health professional. Correct responses were those that followed the national guidelines. .

Subjects were given 4 cases, in the same sequence, to diagnose and treat. In addition to the videos, they received information on the computer screen about patient vital signs and other relevant facts (see the cases in the appendix for details). They could take as long as they liked to respond to the questions and were allowed to pause, rewind or fast forward the video. They were asked to provide responses to the following questions for each case:

i) What is the most likely diagnosis?

ii) What is the most appropriate treatment?

iii) When would you plan to see the patient again for follow-up visit after the initial treatment?

iv) What is the best alternative treatment for the patient (if, for example, the patient's condition does not improve)?

The text of the cases is included in an online appendix and the video vignettes are further described and available for download at:

http://www.rbfhealth.org/resource/video-vignettes-lab-field-experiment-burkina-faso

To mimic "real life" interactions with patients, and to incentivize subject effort and attention during the study, subjects earned 100 CFA (US$0.20) for each correct response. The average earnings from participation in this component amounted to 772.59 CFA (US$1.55), which is roughly the equivalent of two lunches in the type of restaurant that subjects might frequent. The use of incentives is important in countering potential Hawthorne effects (Falk and Heckman, 2009), as they make the decision environment "rich" and interactive, in line with induced value theory (Smith, 1976). That is, the use of incentives adds costs to the misrepresentation of preferences.

Statistical analysis

We calculate a raw score for each participant by giving a 1 to each correct answer and 0 otherwise, with a maximum of 16 (4 cases with 4 sub-questions each).

We look at the distribution of raw and time-adjusted scores by respondent characteristics (student or health professional, medical doctor or other medical qualification and other socio-demographic characteristics) using descriptive statistics and linear regression models estimated with Stata 13.0.

Results

Table 2 describes our sample of study participants. Out of the 1,113 students, 400 were students in their 5th or 6th year of medicine, and 454 and 259 were in their last year of study to become a nurse or a midwife, respectively. Out of the 1,029 recruited health professionals, 15 medical doctors participated along with 552 nurses (state nurses (infirmiers d'état) or certified nurses (infirmiers brevetés), 124 midwives (maïeuticiens d'état) and 338 professionals with other qualifications. The other qualifications included: health assistants (attachés de santé), nurse aides (agents itinérants de santé) and assistant-midwives (accoucheuses auxiliaires). 60.7% of students and 59.2% of health professionals were female. Table 1 further reports the distribution of estimated monthly family income for the students and an average of 139,332 CFA Francs (corresponding to approximately 232 USD) as monthly income for the health professionals.

We also investigate whether the participants found the instructions to be clear and the medical cases and related questions to be interesting, fair and common. Cases were considered by the participants as very interesting or interesting (93%), as very fair or fair (83%) and as common (strongly agree/agree: 77%). Sixty-five percent of participants found the instructions clear most of the time (or more).

Our first and main test of validity is whether participants with more experience and training perform better than those with less. Figures 2 to 5 display the distribution of raw scores among participants. In all cases, variation is significant. Better-trained and more experienced subjects scored better. Figure 2 separates the scores among students and health professionals. The mode is 6 among students and 7 among health professionals; scores tend to be higher among the health professionals, reflecting their greater experience and training. Figures 3 and 4 separately display

the distributions by qualification for students and health professionals. Overall, both distributions suggest higher scores for medical doctors and medical students, consistent with the greater intensity of their training compared to that of the other professionals. Figure 5 presents s the distribution for all participants, indicating a modal score of 7 (out of 16). It displays substantial variance in responses, our second test of validity, following logically from the heterogeneity of performance by qualification and experience evidenced in figures 2 to 4.

Figures 2 – 4 provide only illustrative evidence of our second test of validity. Table 3 reports results that verify that, in fact, better trained and more experienced subjects obtained higher scores. Table 3 investigates the determinants of the raw score and the time-adjusted score in multivariate linear regressions. The first 4 columns present the results for the pooled sample of all participants, while columns 5 and 6 separate the results for the students and the health professionals, respectively. Columns 2-4 indicate that health professionals obtain significantly better scores than less-experienced students, increasing the score by close to 1 point in column 4 (a 0.23 standard deviation increase in the multivariate specification, 0.22 standard deviation increase in a bivariate specification). Medical doctors and midwives (especially student midwives) perform better than the nurses, the omitted category. Staff with auxiliary qualifications (grouped under “other”) perform worse. Among students, those studying medicine or to become a midwife score better than nurses. The contrasts by qualification among health professionals are less significant (except for auxiliary staff performing worse), but the sample of medical doctors is limited to 15 individuals.

In terms of socio-demographic characteristics, men tend to score better than women by about half a point and scores increase with age. Family income (for students), personal income (for professionals) and the self-described state of personal finances seem to have little impact on the

scores. Individuals who perceived the cases and the questions as fair and who found the instructions to be clear obtained higher scores.

Discussion

We develop a new type of provider vignette using videos that present patients with maternal/early childhood symptoms who visit a clinic. We tested the vignettes with health practitioners and students in Burkina Faso. Participants indicated that the cases used were interesting, understandable and common. Results indicated that participants who are expected to perform better, based on ex ante criteria such as their training (medical doctors vs. other qualifications) or their experience (health professionals vs. students), actually did perform better. As a consequence, overall results displayed a substantial variance in performance, consistent with expectations, given the heterogeneity in qualifications and experience among study participants.

While we hope that video vignettes improve on written vignettes or knowledge tests, which may be too abstract from actual clinical practice, we acknowledge limitations to the methodology. First, the video vignettes remain essentially tests of provider knowledge and as such cannot help in measuring the effort exerted by providers. As with direct observations, they suffer from the Hawthorne effect since participants know that their performance is recorded and assessed and this instrument might therefore overstate the ability of health workers with more training (Leonard and Masatu 2005), although the Hawthorne effect might be mitigated if the vignettes are incentivized (Falk and Heckman, 2009). Finally, in contrast with standardized patients or with case-study vignettes that would be acted-out live, the video vignettes that we developed are also missing the back-and-forth of a real live consultation where there are questions and an

interaction between the provider and the patient. But this could potentially be improved by using more advanced interactive technologies.

Conclusion

Despite these limitations, the video vignettes, which can easily be embedded on computers, tablets and smart phones, offer interesting possibilities. In particular, they offer the possibility to standardize presentation for research and measurement, and to present complex signs and symptoms that might be hard to describe in a written vignette but easier to demonstrate with an actor. They could thus be used as a convenient tool to measure provider knowledge, as well as cost-effective instruction and testing tools. They could also be linked to incentives with the objective of improving quality of care in primary health care settings, such as in the case of routine supervision and quality assessment visits conducted by district or provincial/regional coordination teams or other regulatory agents in the health system.

References

Cook, David A. and Thomas J. Beckman. 2006. Current concepts in validity and reliability for psychometric instruments. Am J Med.;119(2):166e7–166e16.

Das Jishnu , Alaka Holla, Veena Das, Manoj Mohanan, Diana Tabak and Brian Chan. 2012. In Urban And Rural India, A Standardized Patient Study Showed Low Levels Of Provider Training And Huge Quality Gaps. Health Affairs, 12: 2774-2784

Das, Jishnu Das and Jeffrey Hammer. 2005. Which doctor? Combining vignettes and item response to measure clinical competence. Journal of Development Economics 78: 348–383

Downing, Steven. 2003. Validity: on the meaningful interpretation of assessment data. Med Educ.;37(9):830–837.

Falk, A. and Heckman, J.J., 2009. Lab experiments are a major source of knowledge in the social sciences. Science, 326(5952), pp.535-538.

Glassman PA, Luck J, O'Gara EM, Peabody JW. 2000. Using standardized patients to measure quality: evidence from the literature and a prospective study. Jt Comm J Qual Improv. 26(11):644-53.

Hrisos, Susan, Martin Eccles, Jill Francis, Heather Dickinson, Eileen Kaner, Fiona Beyer and Marie Johnston. 2009. Are there valid proxy measures of clinical behaviour? A systematic review. *Implementation Science*. 4:37

Leonard, Kenneth an Ottar Maestad. 2016. "Chapter 7. Measuring the Performance of Health Workers" in Scheffler, Richard M., Christopher H. Herbst, Christophe Lemiere, and Jim Campbell, editors. *Health Labor Market Analyses in Low- and Middle-Income Countries: An Evidence-Based Approach.* Directions in Development. Washington, DC: World Bank.

Leonard, K.L. and M.C. Masatu, 2005. The Use of Direct Clinician Observation and Vignettes for Health Services Quality Evaluation in Developing Countries. *Social Science & Medicine* 61 (9).

Leonard, K.L. and M.C. Masatu, 2006. "Outpatient Process Quality Evaluation and the Hawthorne Effect. *Social Science & Medicine* 63 (9).

Peabody, John, Jeff Luck, Peter Glassman, Timothy Dresselhaus, and Martin Lee. 2000. Comparison of Vignettes, Standardized Patients and Chart Abstraction. A Prospective Validation Study of 3 Methods for Measuring Quality. *JAMA* 283(13): 1715-1722

Peabody, John, Jeff Luck, Peter Glassman, Sharad Jain, Joyce Hansen; Maureen Spell, and Martin Lee. 2004. Measuring the Quality of Physician Practice by Using Clinical Vignettes: A Prospective Validation Study. *Ann Intern Med.*141:771-780.

Peabody, John, Fimka Tozija, Jorge Muñoz, Robert Nordyke, and Jeff Luck. 2004. Using Vignettes to Compare the Quality of Clinical Care Variation in Economically Divergent Countries, HSR: Health Services. 39:6.

Smith, V.L., 1976. Experimental economics: Induced value theory. The American Economic Review, 66(2), pp.274-279.

Sullivan, Gail 2011. A Primer on the Validity of Assessment Instruments. *Journal of Graduate Medical Education.*3(2): 119-20.

Veloski Jon, Stephen Tai, Adam Evans and David Nash. 2005. Clinical Vignette-Based Surveys: A Tool for Assessing Physician Practice Variation. *American Journal of Medical Quality:* 20: 151

Table 1: Average time and score for each case				
	Case 1 (60 seconds video, simple)	Case 2 (100 seconds video, more symptoms, rambling, appearance of poverty)	Case 3 (60 seconds video, simple)	Case 4 (100 seconds video, more symptoms, rambling, appearance of poverty)
Medical issue	Preterm labour	Pneumonia (child)	Mastitis	Measles (child)
Average time (seconds)	360.9	346.8	257.9	305.3
Average time (std. dev.)	197.04	194.46	142.69	163.51
p-value for difference with preceding case	n.a.	<0.01	<0.01	<0.01
Average score (out of 4)	1.45	1.88	1.94	1.99
Average score (std. dev.)	1.01	0.99	0.92	0.95
p-value for difference with preceding case	n.a.	<0.01	<0.04	<0.06

Table 2: Summary Statistics

		Students	Professionals
Observations		1,113	1,029
Age (mean)		27.75 (5.36)	35.81 (6.33)
Income (CFA) - Professionals only		-- (--)	139,332.00 (176.207.50)
Female (%)		61%	59%
Interest in task (% responding Very Interesting/Interesting)		94%	93%
Clarity of instructions (% responding Always/Most of the time)		70%	59%
Questions about cases were fair? (% responding Very Fair/Fair)		85%	82%
Are the cases common? (% responding Strongly Agree/Agree)		83%	72%
Current state of personal finances (% responding Excellent/Good)		12%	10%
Qualifications			
	Nurses (%)	41%	54%
	Midwife (%)	23%	12%
	Doctor (%)	36%	1%
	Other (%)	--	33%
Family income			
	Less than 10,000 CFA per month (%)	24%	--
	Between 10K and 50K per month (%)	28%	--
	Between 50K and 100K per month (%)	22%	--
	Between 100K and 200K per month (%)	14%	--
	More than 200K per month (%)	12%	--

Table 3: Dependent variable: Raw score

	All I	All II	All III	All IV	Students V	Professionals VI
Female (D)	-0.564***	-0.556***	-0.479***	-0.431***	-0.443***	-0.367***
	(0.09)	(0.08)	(0.10)	(0.10)	(0.14)	(0.13)
Age (years)	0.056***	0.031***	0.035***	0.035***	0.073***	0.006
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Professional (D)		0.627***	0.893***	0.920***		
		(0.10)	(0.11)	(0.11)		
Doctor (D)			0.404***	0.377***	0.618***	0.674
			(0.13)	(0.14)	(0.16)	(0.48)
Midwife (D)			0.448***	0.434***	0.739***	-0.294
			(0.12)	(0.12)	(0.16)	(0.19)
Other qualification (D)			-0.338**	-0.338**		-0.428***
			(0.14)	(0.14)		(0.14)
Interest in medical task				0.018	0.116	-0.040
				(0.07)	(0.10)	(0.09)
Clarity of instructions				0.067*	0.075	0.054
				(0.04)	(0.05)	(0.05)
Medical questions are fair				0.211***	0.220**	0.187**
				(0.07)	(0.10)	(0.09)
Cases are common				0.014	0.031	-0.009
				(0.03)	(0.04)	(0.04)
State of personal finances				0.060		
				(0.07)		
Family Income (categories)					0.065	
					(0.04)	
Own Income (CFA)						0.000
						(0.00)
Constant	5.849***	6.336***	5.918***	4.597***	2.769***	7.213***
	(0.20)	(0.21)	(0.25)	(0.47)	(0.69)	(0.61)
R-squared	0.059	0.076	0.088	0.095	0.091	0.042
Observations	2142	2142	2142	2142	1113	1029

Figure 1: Example of case questions

1) What is the most probable diagnosis?		
A	Ruptured uterus	Incorrect
B	Premature membranes ruptured	Almost
C	Eclampsia	Incorrect
D	Preterm labour	**Correct**
E	Vaginal candidiasis	Almost

2) What is the most appropriate treatment?		
A	Magnesium sulphate IV	Incorrect
B	Tocolysis with Salbutamol IV	**Correct**
C	Nystatine tablets	Almost
D	Amoxicillin tablets	Almost
E	Caesarean-section	Incorrect

3) When should you see the patient for a follow-up after the completion of the initial treatment?		
A	4 days	Incorrect
B	7 days	**Correct**
C	10 days	Incorrect
D	30 days	Incorrect
E	A follow-up visit is not necessary	Incorrect

4) What is likely to be the best alternative treatment for the patient (for example, if the patient's condition does not improve)?		
A	Refer to emergency unit	Incorrect
B	Deliver the woman	**Correct**
C	Amphotericin B tablets	Almost
D	Blood transfusion	Incorrect
E	Close surveillance	Almost

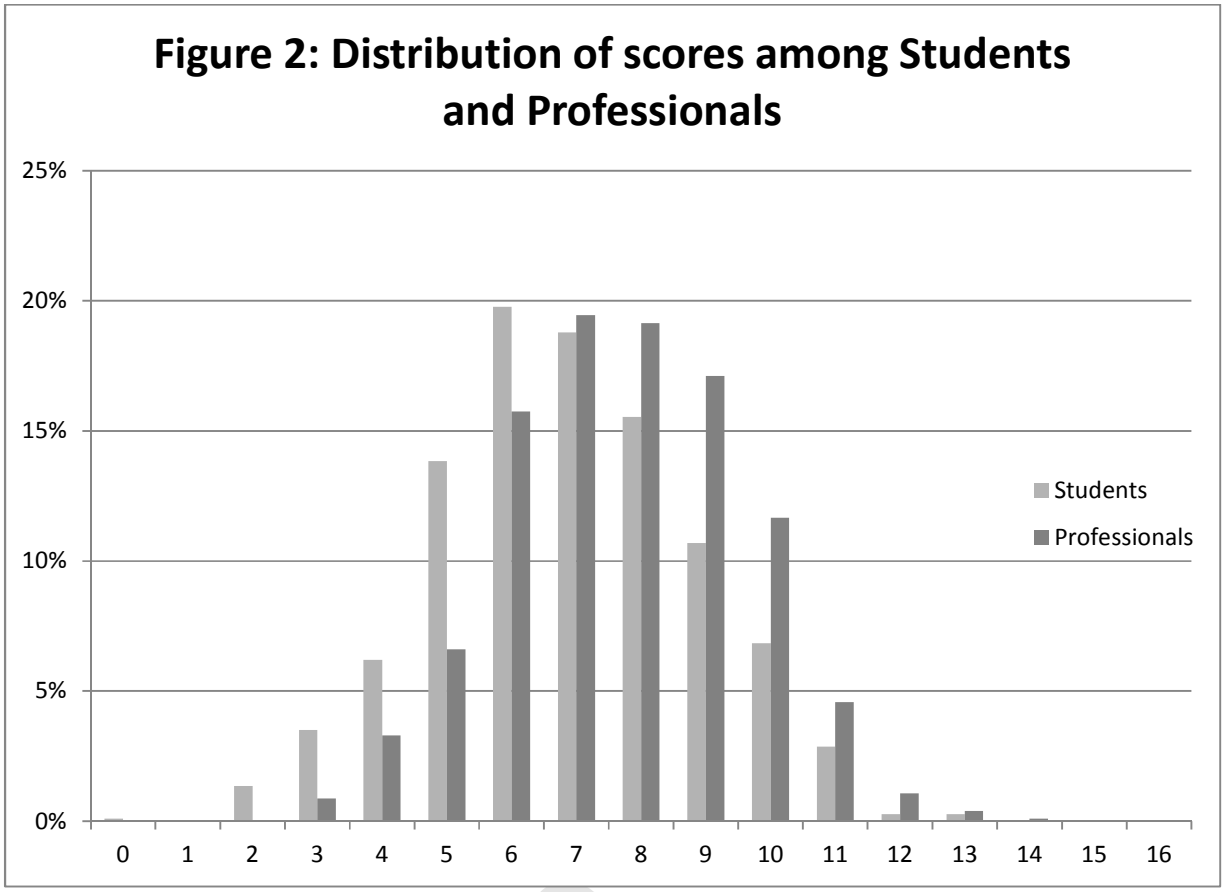
Figure 2: Distribution of scores among Students and Professionals
25%
20%
15%
10%
5%
0%
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Students
Professionals

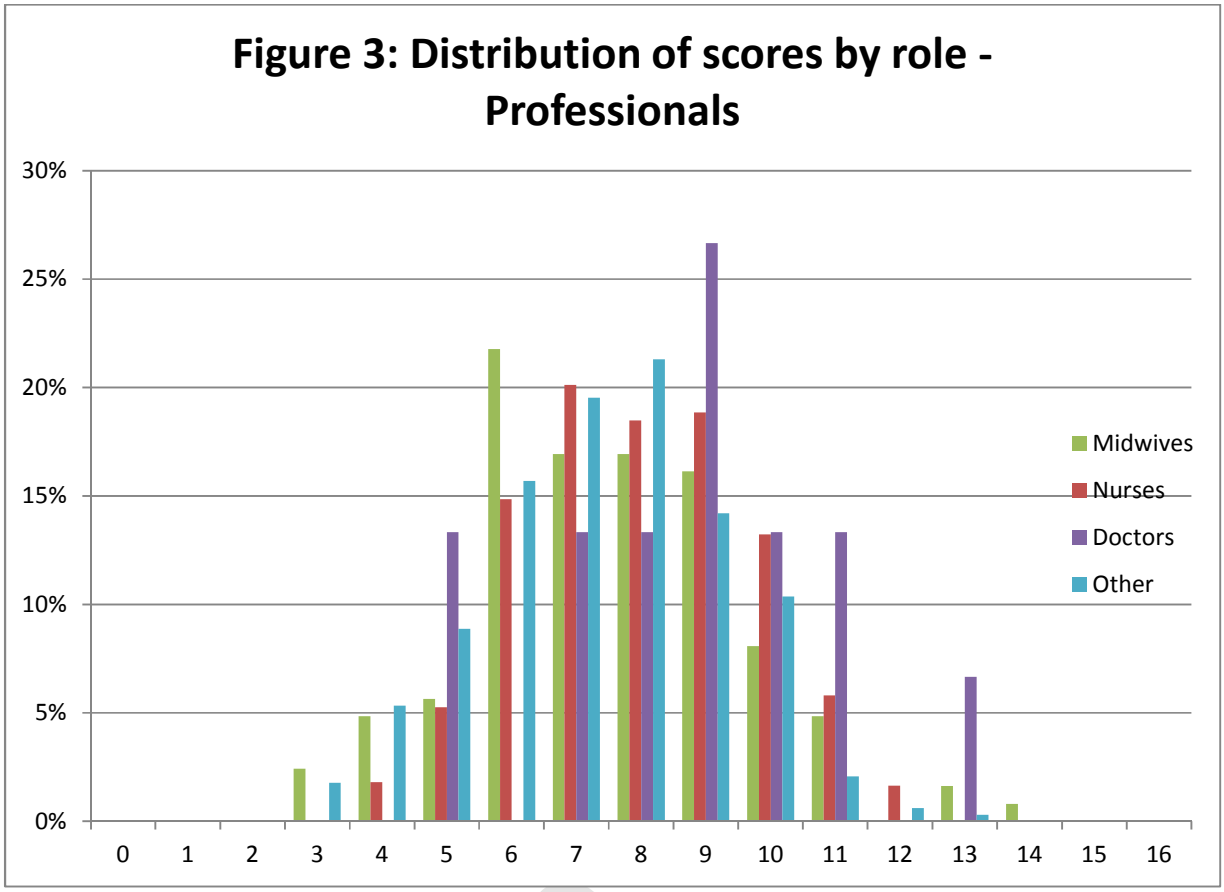
Figure 3: Distribution of scores by role - Professionals
30%
25%
20%
15%
10%
5%
0%
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Midwives
Nurses
Doctors
Other

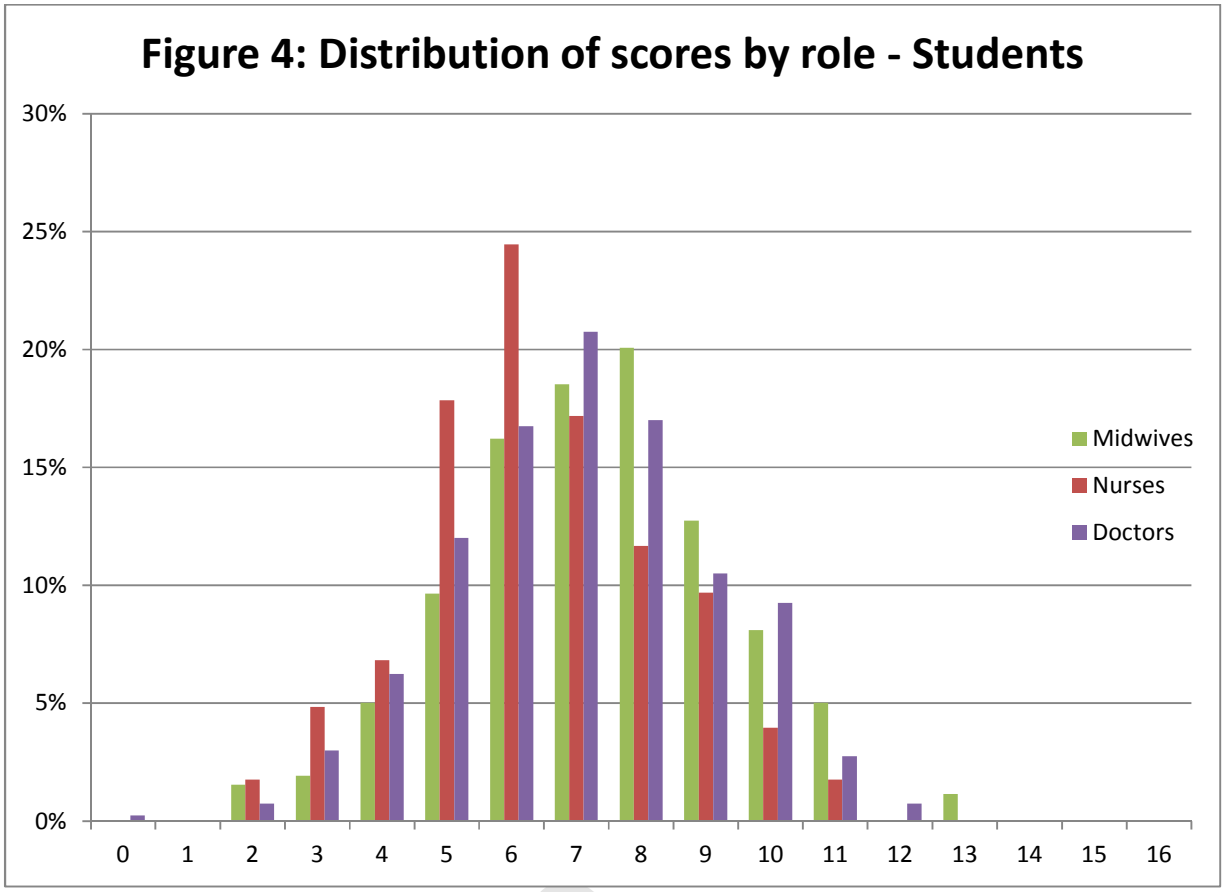
Figure 4: Distribution of scores by role - Students
30%
25%
20%
15%
10%
5%
0%
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Midwives
Nurses
Doctors

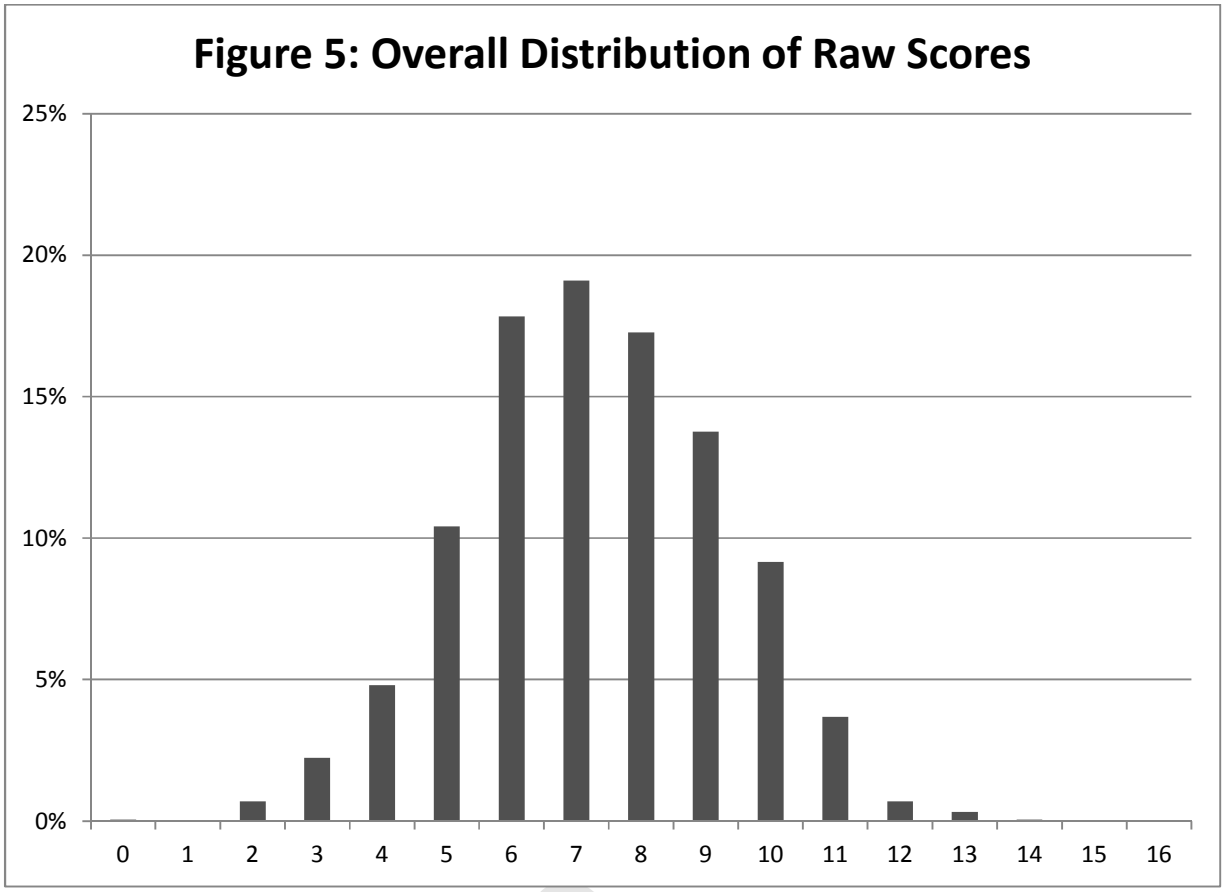
Figure 5: Overall Distribution of Raw Scores
25%
20%
15%
10%
5%
0%
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

The use of video vignettes to measure health worker knowledge.

Evidence from Burkina Faso

Research Highlights

- To measure health worker knowledge, we tested a new type of provider vignette using videos
- The videos presented a patient with maternal/childhood symptoms visiting the clinic.
- Results displayed significant variance in performance, consistent with expectations.
- Participants expected to perform better (doctors, experienced professionals) did perform better.