Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____	_____
Christopher DeStephano	4/18/2011

Providing 360 Degree Multisource Feedback to Nurse Educators in the Country of Georgia: A Pilot Study

By

Christopher Carl DeStephano
Degree to be awarded: Master of Public Health

Hubert Department of Global Health

_____

Roger Rochat, M.D.
Committee Chair

_____

Judy Wold, PhD, RN
Committee Member

Providing 360 Degree Multisource Feedback to Nurse Educators in the Country of Georgia: A Pilot Study

By

Christopher Carl DeStephano

B.S., University of Notre Dame, 2006

Thesis Committee Chair: Roger Rochat, M.D.

An abstract of
a thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
In Hubert Department of Global health
2011

Abstract

Providing 360 Degree Multisource Feedback to Nurse Educators in the Country of Georgia: A Pilot Study

By Christopher Carl DeStephano

Due to insufficient nursing education standards in the country of Georgia, fifteen nurse educators participated in a train-the-trainer program. These educators are now offering vocational education courses with plans to teach 2500 Georgian nurses over 3 years. Using a 360 degree multisource feedback model, self, video, student, peer, and program coordinator evaluations of teaching effectiveness were completed. After nurse educators reviewed their results and identified areas for improvement, a questionnaire on the perceived acceptability of teacher evaluation was completed. Of the 15 nurse educators, 93.3% indicated that nurse educators should receive feedback from self, student, peer, and video evaluations, while 100% indicated that nurse educators should receive feedback from the program coordinator. The accuracy and usefulness of the program coordinator evaluation were rated the highest while the peer evaluation was rated the lowest on these domains. This pilot study revealed that multisource feedback was acceptable to Georgian nurse educators.

Providing 360 Degree Multisource Feedback to Nurse Educators in the Country of Georgia: A Pilot Study

By

Christopher Carl DeStephano

B.S., University of Notre Dame, 2006

Thesis Committee Chair: Roger Rochat, M.D.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Hubert Department of Global Health
2011

Table of Contents

Chapter 1: Introduction

Introduction and Rationale

Insufficient nursing education standards in the country of Georgia have contributed to low professional status for nurses and sublevel nursing/medical care. Although "nursing colleges" exist in the country, nursing education does not take place at institutions of higher education which precludes program graduates from teaching in a university level baccalaureate nursing education program (Wold, Walker, & Partskhladze, 2007). Therefore, nurses are primarily governed and trained by physicians. This has resulted in poor morale and lack of vision for the future of nursing due to nurses not having a substantial voice in their profession (Walker & Wold, 2009). The Ministry of Education and Science of Georgia is responsible for the oversight and accreditation of "nursing colleges" but only conducts institutional accreditation (Wold et al., 2007). Since nursing is not regarded as a profession in Georgia, licensure and nursing standards are non-existent and regulations for nursing care are limited. This contributes to significant variation in nursing curricula among schools (Walker & Wold, 2009). Students are deemed qualified to take a nursing position at a hospital or clinic after passing a school-constructed, non-standardized examination (Wold et al., 2007). Once in a hospital or clinic, nurses are allowed to make few independent decisions regarding patient care and are trained at a level that falls between that of United States nursing assistants and licensed practical nurses (Ivanov & Paganpegara, 2003; Walker & Wold, 2009; T. Zimmerman, Putland, & Hughes, 1997).

If nurses are not educated and prepared to meet the challenges of a modern system of health care, patient morbidity, mortality, and patient satisfaction are adversely affected

(Scott, Sochalski, & Aiken, 1999). Despite the importance of adequate nursing levels, Georgia has a low nurse to doctor ratio compared to Western European countries (Gamkrelidze, Atun, & Gotsadze, 2002). In 1999, there were 1.18 nurses for every doctor in Georgia compared to 3.1 in the United Kingdom, 7 in Finland, 1.67 in France, 2.3 in Switzerland, and 2.7 in Germany (Gamkrelidze et al., 2002; WHO, 1999). The emphasis on specialty care and lack of primary care under the former Soviet Semashko model of care cannot be supported by the current health budget (Sanders, 2007). International donor organizations and the Georgian government recognize that nurse education needs to be strengthened so that nurses will be better prepared to take on an expanded role at hospitals and clinics across the country (Walker & Wold, 2009; Wold et al., 2007). Improved nursing practice has significant implications including higher quality and lower cost health services to underserved populations through health promotion, disease prevention, and management of chronic conditions (Wold et al., 2007).

To expand the role of nurses, improve their vision for the profession, and develop nurse leaders, Partners for International Development (PfID) consisting of Emory University School of Medicine, Emory University School of Nursing, and Partners for Health developed an intensive faculty development program for nurse educators from November 2009 to February 2010. PfID faculty taught general adult and pediatric nursing to fifteen Georgian healthcare professionals (trained previously as nurses or physicians). In addition, the curriculum included specialty care areas (perioperative, labor and delivery, critical care, and emergency nursing), teaching strategies, and nursing management and leadership. In February 2010, nurse educators trained by the program

began offering month long vocational education courses to nurses from partner hospitals. The program plans to teach 2500 nurses over three years with an emphasis on recruiting nurses from rural areas and regions affected by the August 2008 conflict when feasible (Walker & Wold, 2009). Half of the vocational education coursework is didactic with the other half being simulation and clinical experience. Simulation with mannequins is used to teach and practice basic life support, sterile technique, and placement of urinary foley catheters, nasogastric tubes, and intravenous catheters. Clinical experience is provided on a multi-profile unit where students learn and practice administration of medications, documentation, and the physical exam. Concurrently with this vocational education program, select nurse educators are involved in preparation of relevant courses for a four-year baccalaureate degree nursing school.

Problem Statement

Strengthening nurse education in the country of Georgia will be essential for improving nursing practice in the country. Higher education institutions in North America and Western Europe utilize teacher evaluation methods that provide feedback to educators and administrators with the goal of improving the quality of education (R. Berk, 2006). Although some studies question the validity and reliability of teacher evaluation methods, teacher ratings in the clinical setting are associated with student performance ratings (Blue, Griffith, Wilson, Sloan, & Schwartz, 1999; Griffith, Georgesen, & Wilson, 2000; Roop & Pangaro, 2001). The acceptability, usefulness, and accuracy of these methods in the country of Georgia and other post-Soviet countries are unknown. This limits the information available for improving educational standards and may prevent the benefits of improved nursing practice from being realized. This pilot

study of teacher evaluation in the country of Georgia lays the foundation for longitudinal evaluation of teaching effectiveness. Strengths and areas for improvement of nurse educators can be continuously identified, so that the teaching effectiveness and nursing educational standards in the country of Georgia and other post-Soviet countries are improved.

Purpose Statement and Research Questions

To provide feedback to nurse educators and program coordinators of a continuing education program in the country of Georgia, I used a 360 degree multisource feedback (MSF) model of evaluation to provide a complete picture of teacher performance. Five different sources of teacher evaluations were obtained for 15 nurse educators trained by PfID. The sources of evaluation consisted of the educators evaluating themselves (self evaluation), the educators evaluating their colleagues (peer evaluation), student evaluations of the educators (student evaluations), evaluations by the program coordinators (program coordinator evaluation), and videotape evaluations of the educators performed by the principal investigator of the study (video evaluation). After I provided the results to the nurse educators and asked them to identify areas for improvement, they completed a self-administered survey on the acceptability (accuracy, usefulness, and interest in future use) of receiving teacher evaluations from multiple sources. The purpose of this project was to answer the following questions:

1. What did ratings from the evaluation methods identify as teaching strengths and areas for improvement of nurse educators trained by PfID?
2. What areas for improvement were identified by nurse educators after they received results from the evaluations?

3. Was MSF acceptable for use in the country of Georgia?
4. Was the acceptability (usefulness, accuracy, and interest in future use) of specific sources of evaluation (peer, student, video, self, and program coordinator) more highly rated by nurse educators than other sources of evaluation?

Significance Statement

Answers to these questions will provide information that is relevant for numerous audiences. Most importantly, the ratings and teacher evaluations collected from MSF were meant to build the confidence of nurse educators trained by PfID and identify areas for improvement. Although the evidence is mixed, some studies in the medical literature show that providing MSF to physicians and medical trainees promotes positive behavior change resulting in improvements of medical practice (Miller & Archer, 2010). Similarly, by providing information to nurse educators, we hoped they could improve their educational methods which would improve the practice of nursing by their students.

The following review of the literature shows that information about faculty evaluation is non-existent in the country of Georgia and other post-Soviet countries. This project therefore has the potential to stimulate further study and implementation of faculty evaluation which could influence the standards of higher education institutions in the region. Furthermore, this project provides a model of evaluation that has not yet been used extensively for teacher evaluation or nursing education. There are numerous potential applications of MSF that can be explored further by future investigators. These applications include student evaluation, baccalaureate nursing program accreditation, and evaluation of practicing nurses

Chapter 2: Comprehensive Review of the Literature

The following review of the literature will begin by describing the historical context influencing the healthcare system and the role of nursing in the country of Georgia. I will then discuss the efforts undertaken in Georgia to strengthen nurse education followed by a summary of studies on faculty evaluation in nursing and medical education. The review will conclude by describing the MSF model and its use in medical practice.

Country Profile- Country of Georgia

The country of Georgia is a small mountainous country of 5.4 million people bordered by Turkey and Armenia to the South, Russia to the North, the Black Sea in the West, and by Azerbaijan in the East (Figure 1) (Sanders, 2007). The country also includes the breakaway regions of the autonomous Republic of Abkhazia and the former autonomous regions of Ajara and South Ossetia (Sanders, 2007).

Figure 1: Map of the country of Georgia (CIA, 2011)

Until the dissolution of the Soviet Union, Georgia has been occupied for much of its history. The country was initially settled by the ancient kingdoms of Colchis and Kartli-Iberia and then came under Roman and Christian influences in the first centuries A.D. (CIA, 2011). Following the collapse of the Roman Empire, Persian, Arabic, and Turkish groups occupied the country until a Georgian golden age (11^{th}-13^{th} centuries) (CIA, 2011). The Mongol invasion of 1236 interrupted this era (CIA, 2011). The Ottoman and Persian empires heavily influenced the country until Georgia was absorbed into the Russian Empire in the 19^{th} century (CIA, 2011). Georgia gained independence from 1918-1921, but was incorporated into the USSR by force in 1921 until the Soviet Union dissolved in 1991 (CIA, 2011).

Following independence on April 9, 1991, Georgia was one of the strongest post-Soviet Republics in regards to per capita income and living standards due to bustling tourism, strong industry, and a prosperous agricultural sector (Atun, 1999; Skarbinski et al., 2002). Universal education led to Georgia having the highest adult literacy rate (approximately 100%) in the world (UN, 2009). However, high inflation, civil war from 1991-1994 with the separatist regions of Abkhazia and South Ossetia, and an influx of 300,000 internally displaced people into the major cities resulted in economic collapse with gross domestic product dropping by 75% (Atun, 1999; Skarbinski et al., 2002). Despite macroeconomic and political stabilization in 1995, infrastructure deteriorated, corruption was commonplace, and the per capita GDP was $983 in 1999 (Skarbinski et al., 2002). After the peaceful overthrow of the Shevardnadze government in 2003 (the Rose Revolution), Mikhail Saakashvili was elected as president (Walker & Wold, 2009). As president, Saakashvili has reduced corruption in Georgia, strengthened the

infrastructure, and implemented market-based reforms. Improvements in living standards and infrastructure have been realized, however, Georgia remains a middle income country with a per capita GDP of $4400 (149th in world), an unemployment rate of 16.4% (156th in world), and 31% of the population below the poverty line (CIA, 2011).

Healthcare System in the Country of Georgia

Under the Soviet model ("Semashko"), the government controlled the healthcare system through normative planning financed by a global budget raised from centralized resources (Skarbinski et al., 2002). With hospitals and specialists dominating healthcare delivery and an emphasis on curative over preventive care, the system was redundant, expensive, and required large numbers of healthcare providers (Sanders, 2007). Data from the 1980s and 1990s demonstrate the excess capacity that resulted from centralized planning. Under the Soviet model in the early 1980s, Georgia was required to have 60,000 hospital beds to serve its population (Atun, 1999). In 1999, Georgia had 287 hospitals with 25,000 hospital beds resulting in a ratio of 4.5 beds per 1000 population compared to 2.5 per 1000 population on average in Organisation for Economic Co-operation and Development (OECD) countries (Atun, 1999). Average occupancy rates were low at 28% with around 100 hospitals having occupancy rates less than 10% (Atun, 1999). There was also a surplus of physicians with one physician per 245 people in the population (compared to 1:400 in OECD) (Atun, 1999).

Following independence, the economic plight of the country required the government to scale down its involvement in the health sector and redirect its activities from direct provision of care toward regulatory, licensing, and financing functions (Skarbinski et al., 2002; Wold et al., 2007). Health care spending decreased from $95.50

per capita in 1985 to $0.90 in 1994 (Skarbinski et al., 2002). A 1999 probability-to-size cluster survey of 248 households showed that this shift from collective to individual responsibility and out-of-pocket payments placed a severe burden on the population (Skarbinski et al., 2002). Of 306 household members who had been ill in the last 6 months, 93% reported that costs were the major deterrent to obtaining health care, 10% reported being unable to obtain health care because of high costs, 16% reported not being able to afford all the medications necessary to treat their illness, and 53% reported that total out-of-pocket expenditures were paid for by borrowing money or selling personal items (Skarbinski et al., 2002). The increased out-of-pocket expenditures may have increased mortality from deaths secondary to cardiovascular disease. The overall age-adjusted mortality rose by 18% from 1988 to 1994 and deaths secondary to cardiovascular disease increased by 35% from 1990 to 1999 (Atun, 1999).

The Georgian government began discussing health reform in 1994 with a strategy published in 1996 that changed the legal status of hospitals and polyclinics, resulted in a new case-based payment method for hospitals, created competition for providers, and formed a new intermediary agency to pay for health services (Atun, 1999). Legislation from this strategy allowed the government to remove 130,000 health workers from the budget, establish licensure of medical facilities, and provide for certification and licensure of physicians (Atun, 1999; Wold et al., 2007). The reform process has also begun to strengthen primary health care and prevention activities (Sanders, 2007). Although promising, achieving progress in providing lower cost basic health services to all who need them has been difficult due to economic circumstances and the lack of a “roadmap” for achieving health reform objectives (Gamkrelidze et al., 2002; Wold et al.,

2007). Health staff levels continue to be high, staff face difficulties achieving adequate income, patient access to health services continues to be limited by official and unofficial charges, and training and accreditation needs further strengthening to ensure public safety while the system is decentralized (Gamkrelidze et al., 2002).

Strengthening Nurse Education in the Country of Georgia

Although the limited economic resources for health have delayed the implementation of improvements in nursing care and establishment of a baccalaureate nursing program (Nishiyama, Wold, & Partskhladze, 2008), progress has been attained in a number of areas (Gamkrelidze et al., 2002). Since 1992, non-governmental organizations and international donors have provided support for improving nursing practice in the country of Georgia through international alliances with schools of nursing in the United States (Gamkrelidze et al., 2002). In coordination with the American International Health Alliance and Partners for Health, Partners for International Development (PfID)—formerly Georgia to Georgia Partnership—have attained a number of important accomplishments including the appointment of a nursing director to serve as the Chief of Nursing of the Ministry of Labour, Health and Social Affairs in 1996 and the opening of the Nursing Resource Center within the National Information Learning Center in 1998 (Nishiyama et al., 2008). Since 1999, PfID has worked with its Georgian partner, Partners for Health, to offer continuing education classes to nurses in Tbilisi designed to increase the level of clinical and administrative nursing practice (Nishiyama et al., 2008). In 2005, PfID, Open Society Georgia, and Emory University School of Nursing sponsored four Georgian physicians who had worked as nurses to complete a teaching program at Emory University School of Nursing in the United States (Nishiyama et al.,

2008). These physicians developed a culturally appropriate baccalaureate nursing curriculum and 40 modules for continuing nursing education in Georgia (Nishiyama et al., 2008). Upon return to Tbilisi, they coordinated continuing education classes provided to nurses at the Gudashauri national medical center, the Iashvili Children's Hospital, and other hospitals in Georgia (Nishiyama et al., 2008). Most recently, the United States Agency of International Development (USAID) provided a grant to PfID. This led to the intensive faculty development program for nurse educators that is the topic of this thesis.

The Role of Faculty Evaluation in Medical and Nursing Education

For education programs like the one developed by PfID to be successful at improving the knowledge, practice, and professional status of nurses, high quality teaching is essential. To determine teaching effectiveness and improve education quality, faculty assessment is a standard component of graduate medical and nurse education programs (Leach & Philibert, 2006). According to Peter Seldin's Changing practices in evaluating teaching, the literature contains over 15,000 studies on teaching effectiveness (Seldin, 1999). In the medical and nursing literature, studies have examined the teaching behaviors of effective and ineffective clinical teachers. Education scholars have used this information to develop instruments to evaluate and provide feedback to teachers. A recent systematic review of questionnaires for assessing different domains of clinical teaching in medical education states, "Excellent clinical teachers are described as physician role models, effective supervisors, dynamic teachers, and supportive individuals, possibly complemented by their role as assessors, planners, and resource developers"(Fluit, Bolhuis, Grol, Laan, & Wensing, 2010, p. 1337). In the nursing literature, a questionnaire of clinical teaching in nursing has been used in multiple

settings (Canada, Greece, United States, Hong Kong, Israel, Australia) to assess the teaching behaviors that nursing students value (Benor & Leviyof, 1997; Kotzabassaki et al., 1997; Lee, Cholowski, & Williams, 2002; Li, 1997; J. Mogan & Knox, 1987; Nehring, 1990). The 8 items rated highest by both the faculty and students in the original Mogan and Knox study (1987) of the questionnaire included “is a good role model”, “enjoys nursing”, “demonstrates clinical skills and judgment”, “enjoys teaching”, “is well prepared for teaching”, “takes responsibility for own actions”, “is approachable”, and “is self-confident” (p. 334).

The literature is limited with regards to whether an association between teacher evaluations and student outcomes exists. A PubMed search of “(NCLEX-RN)” revealed 195 studies with none of them examining whether teacher ratings are associated with student performance on the National Council Licensure Examination-Registered Nurses. However, there are multiple studies demonstrating an association between faculty ratings and medical student knowledge and practice. One study using linear regression and controlling for the score on the first medical board exam (United States Medical Licensing Exam Step 1) showed that medical students who trained under the highest rated attending physicians during the internal medicine rotation achieved significantly higher scores on the second step of the medical board exam than those who trained under the lowest rated attending physicians (score of 207 versus 199) (Griffith et al., 2000). Another study showed that although pre-clerkship grade point average accounted for 28% of the variance in final student performance on the internal medicine rotation, teaching behaviors accounted for 6% of the variance in final, student performance ($p<0.0001$) (Roop & Pangaro, 2001). Student performance was based on the clinical performance

score, the National Board of Medical Examiners Medicine subject examination, a 3-hour open book essay examination of analytic ability, and a multiple-choice test in the interpretation of laboratory values. Regarding surgical training, students working under the lowest rated attending physicians received significantly lower scores on the National Board of Medical Examiners surgery subject examination and the Objective Structured Clinical Examination (Blue et al., 1999).

It should be noted that student interest and enthusiasm may be confounding variables limiting the conclusions that can be drawn from the preceding studies. The students' enthusiasm and interest in internal medicine or surgery may have influenced student performance and ratings of faculty members. Those who are interested in the field likely had better experiences on the rotation and studied harder for the subject exam leading to better exam performance. This in turn could have resulted in higher faculty ratings unrelated to teaching effectiveness. Use of other faculty evaluation sources (e.g. external expert, resident evaluations, peer evaluations) may have strengthened the design of these studies by improving the identification of the most effective teachers.

Although researchers in the United States have extensively studied effective clinical teaching behaviors and teacher evaluation instruments, the literature is limited in Russia and non-existent in post-Soviet countries. One article from 2005 on medical education reform in Central Asian Republics states that "standardized evaluation of faculty by administration, peers, and students could be used to enhance faculty performance and to support the development of a performance-based reward system" (Conaboy, Nugmanova, Yeguebaeva, Jaeger, & Daugherty, 2005, p. 62). However, there is no evidence in the literature that countries implemented this recommendation.

To further define the extent of faculty evaluation in the United States compared to post-Soviet Republics, a PubMed search of "(United States) AND ((medical education) or (nursing education)) AND (faculty OR teacher) AND (ratings OR feedback OR evaluation OR assessment)" was performed and revealed 2933 articles. Use of "USSR"—the MeSH term that comprises Russia and each former Soviet republic--instead of "United States" in the search revealed only 15 articles (9 articles in English, 6 in other languages). Of the 9 articles in English, one of the articles was a national survey of electroconvulsive therapy in Russia (Nelson, 2005), one was a continuing education needs assessment of nursing leaders in the Latvian Republic (Kalnins & Kalnins, 1991), and seven discussed international medical and nursing education partnerships (twinning) between U.S. or Western European institutions and Russian or post-Soviet Republic institutions (Aghababian et al., 1995; Driever, Perfiljeva, Callister, & McGivern, 2005; Immonen, Anderssen, & Lvova, 2008; Jenkins, Brush, McGonagle, Vartanian, & Levy, 2000; Kalnins, Barkauskas, & Seskevicius, 2001; Levine & Perpetua, 2006; Wong & Agisheva, 2007). However, only one of these articles obtained and presented faculty evaluation results. The evaluations were acquired during a "developing teaching skills for medical educators in Russia" project (Wong & Agisheva, 2007). Teaching performance was assessed using a pretest/post-test questionnaire of faculty participants' self-reported ratings of teaching ability before the faculty development intervention and 1 and 12 months after the intervention (Wong & Agisheva, 2007). Global teaching performance and specific teaching behaviors improved significantly according to the self-reported ratings (Wong & Agisheva, 2007). Commitment to change statements were also written by participants following the intervention and successfully implemented by 71%

of the faculty (Wong & Agisheva, 2007). This study was limited by participant self-evaluation, which is not as objective as direct observation of teaching behaviors (Wong & Agisheva, 2007). The 6 studies in languages other than English did not discuss faculty evaluation in the English abstract.

Formative Versus Summative Decisions

Due to the lack of literature on faculty evaluation in post-Soviet Republics and the interest in developing a nursing program in the country of Georgia comparable to what is found in the United States, PfID included teacher evaluation as a monitoring indicator in its USAID grant proposal (Walker & Wold, 2009). In order to develop a system for faculty evaluation in the country of Georgia, it's important to first understand the difference between formative and summative decisions (R. Berk, 2006). Faculty make formative decisions to improve and shape their pedagogy (R. Berk, 2006). These decisions are made based on feedback from students and administrators throughout the year to plan and revise the curriculum. Summative decisions are made by administrators to "sum up" overall performance, which influences tenure, promotion, dismissal, and merit pay (R. Berk, 2006). Summative decisions are higher-stakes, personnel decisions, which if made appropriately and fairly should use valid and reliable instruments (R. Berk, 2006). PfID decided that summative decisions based on faculty evaluation would not be appropriate for the program due to the lack of previous reports about the acceptability or validity of teacher evaluation in the country.

360 Degree Multisource Feedback

Historically, the main (or only) source of teacher ratings for both formative and summative decisions in higher education were student evaluations (R. A. Berk, 2009). Since use of only one source can be unreliable and biased, Berk proposes use of additional evaluation sources: “Each source can supply unique information, but also is fallible, usually in a different way from the other sources…By drawing on three or more different sources of evidence, the strengths of each source can compensate for weaknesses of other sources” (R. Berk, 2006, p. 13). The bias and limited information that result from single sources of evaluation led to the use of 360-degree multiple source feedback (MSF) in medicine approximately 15 years ago (the first article in PubMed is from 1997) (R. A. Berk, 2009; Lockyer, 2003). MSF is a form of workplace based assessment method, since day-to-day practices are evaluated in the working environment (Miller & Archer, 2010). It was initially developed for use in management and industry to assess performance of employees from a variety of perspectives (R. A. Berk, 2009; Lockyer, 2003). In industry, competence for specific behaviors (e.g. aptitude, finishing tasks, interpersonal skills, leadership skills) is assessed using questionnaires that are completed by people at the same, lower, and higher levels of the organizational chart (e.g. coworkers, supervisor, subordinates, and clients) (R. Berk, 2006; Lockyer, 2003). The results are compared to self-ratings so that precise feedback is obtained and formative decisions can be made to improve job performance (R. Berk, 2006).

The medical literature has increasingly described the MSF model for evaluating practicing physicians and medical residents. A PubMed search of “(multisource OR multi-source) AND (feedback)” resulted in 82 papers with 73 of them being from 2005-

2010. Limiting the search to "(multisource OR multi-source) AND (feedback) AND (education, medical)" produced 52 papers. A search of "(multisource OR multi-source) AND (feedback) AND (education, nursing)" only resulted in 1 paper on "learning mechanisms to limit medication administration" that did not discuss MSF (Drach-Zahavy & Pud, 2010).

The medical education studies have predominantly examined whether MSF is effective in promoting changed behaviors of medical residents and practicing physicians. A 2010 systematic review of workplace based assessment methods concluded that the evidence is mixed (Miller & Archer, 2010). To provide a framework for how to use MSF, the subsequent segment of the literature review is a sample of the studies that have been completed for general practitioners, surgeons, and medical trainees.

Of 113 family medicine physicians in a pilot study of MSF, 69 (61%) reported that they planned to make or had already made changes in response to the feedback (J. M. Sargeant et al., 2003). Communication with patients, medical colleagues, and co-workers made up three fourths of the changes identified by the physicians (J. M. Sargeant et al., 2003). A follow-up study revealed that participants who responded negatively to MSF did not agree with their feedback and were not planning on using it for practice improvement (J. Sargeant, Mann, & Ferrier, 2005). These reactions were based on perceptions of accuracy, credibility, and usefulness of feedback which were influenced by the recruitment of credible reviewers, ability of reviewers to make objective assessments, and the specificity of the feedback (J. Sargeant et al., 2005). In another study of family physicians who received feedback from 6 peers, 6 referring physicians, 6 co-workers, and 25 patients about 55 aspects of their medical practices, 217 (83%) of 255

physicians contemplated a change and 168 (66%) reported initiating a change for at least one aspect of practice (Fidler, Lockyer, Toews, & Violato, 1999). The two aspects of practice that physicians most frequently initiated a change for were "communication with patients" and "support of patients" (Fidler et al., 1999). Changes were more likely to be contemplated or initiated if physicians had lower mean ratings (Fidler et al., 1999). Regarding whether MSF resulted in improved ratings over time, a study of 250 physicians receiving MSF 5 years apart showed a significant increase in ratings between the two assessments from medical colleagues and co-workers, but not between the two patient surveys (Violato, Lockyer, & Fidler, 2008).

In a study of surgeons, 144 (71.6%) of 201 surgeons in Canada who received MSF from medical colleagues, coworkers, and patients reported that they had contemplated or initiated change on the basis of the feedback about collaboration, office systems, stress management, and communication style with patients and colleagues (Violato, Lockyer, & Fidler, 2003). However, another study of surgeons using a Likert scale for the likelihood of initiating change based on MSF (1=Not considering implementing, 5=Very likely to implement) revealed that the mean scores for each practice characteristic were low—ranging from 1.63 for selecting the appropriate treatment to 2.12 for maintaining appropriate medical records (Lockyer, Violato, & Fidler, 2003).

Regarding MSF for medical trainees, two studies again show mixed results. A randomized controlled trial in Canada of 18 pediatric residents who received MSF compared to 18 who received standard feedback showed that the MSF group's ratings increased significantly more than the control group's ratings (Brinkman et al., 2007). A

study of 679 United Kingdom foundation house officers (equivalent to medical residents in the United States), supervisors (those who facilitate providing feedback to trainees), and MSF raters reported an overall mean (standard deviation) of 3.98 (.90) on a a 5-point Likert scale (1=Strongly disagree, 5=Strongly agree) for the item stating "MSF is a good idea in principle" (Burford, Illing, Kergon, Morrow, & Livingston, 2010). However, only 31% of trainees agreed or strongly agreed with the statement that feedback from MSF had been "useful and valuable to their learning" (Burford et al., 2010). In responding to whether "I [trainees] have changed/will change" different behaviors (e.g. relationships with patients, working with colleagues, clinical care, medical knowledge, teaching and training skills, attitude and approach to job, professional skills), mean ratings were between 2.31-2.91 suggesting that they are unlikely to change based on feedback from MSF (Burford et al., 2010). It should be noted that the overall response rate for the questionnaire on MSF was only 45% which could have resulted in selection bias.

Although there is mixed evidence regarding whether MSF promotes changed behaviors, PfID decided that use of MSF would be most appropriate for evaluating and providing feedback to the program's recently trained nurse educators. The program made this decision because MSF is designed to frame a more complete picture of performance than could be provided from a single-source, top-down, supervisor-only approach (R. Berk, 2006). The literature is limited with regards to using MSF for faculty evaluation, but a 2009 article by Berk on using the 360 degree multisource feedback model to evaluate teaching provides a framework. He proposes characteristics of MSF for faculty evaluation:

> "…teaching behaviors defined for each source may be different and complementary…different Likert-type scales for the different type of raters can be developed…the quality of many homegrown scales varies from very good to poor, but commercial student rating scales are better…the professor can track changes and progress in teaching improvement from the different sources of evidence and document improvement" (R. A. Berk, 2009, pgs. 1076-1077).

He also proposes that the sources of evaluation (figure 2) could come from self evaluations, other colleagues, peer observers, department chairs, external experts, other personnel, secretaries, and students (R. A. Berk, 2009).

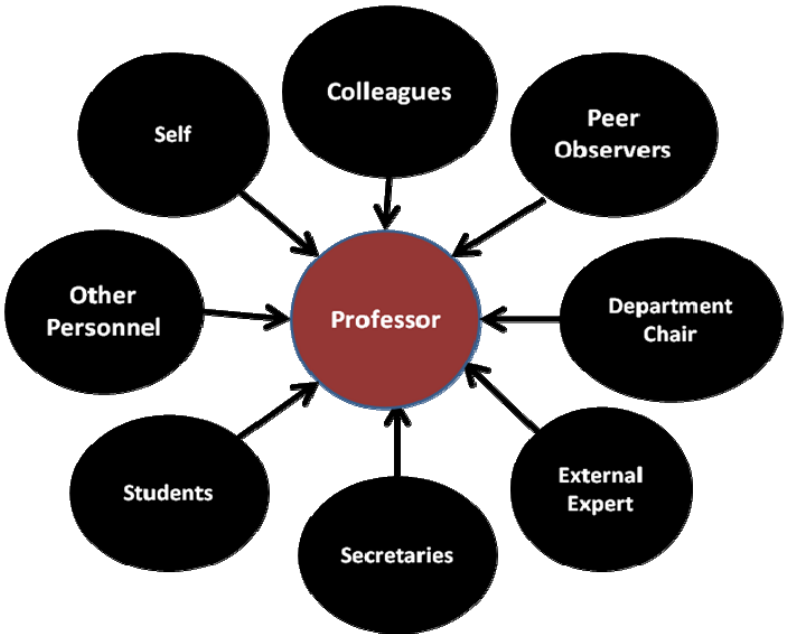

Figure 2: MSF of a professor. Reprinted with permission. (R. A. Berk, 2009, p. 1077)

From this framework, PfID decided that peer, self, student, video, and program coordinator evaluations of lectures could be obtained and used to provide feedback to the nurse educators.

Instruments for Nursing Faculty Evaluation

After PfID program coordinators reviewed 5 instruments available in the literature (R. A. Berk, Naumann, & Appling, 2004; J. Mogan & Knox, 1987; Judith Mogan & Warbinek, 1994; Thompson & Sheckley, 1997; L. Zimmerman & Westfall, 1988), the Nursing Clinical Teaching Effectiveness Inventory (NCTEI) and Johns Hopkins School of Nursing Peer Evaluation Scale were selected as the most appropriate instruments for the program (R. A. Berk et al., 2004; Knox & Mogan, 1985). The designers granted permission to use the instruments for the PfID program.

Mogan and Knox (1987) described the NCTEI following their article about effective and ineffective behaviors of clinical teachers as perceived by nursing students and faculty members (Knox & Mogan, 1985). The instrument is a 48-item checklist of teaching behaviors on a 7 point Likert scale (1=Least Descriptive of Teacher, 7=Most Descriptive of Teacher) (J. Mogan & Knox, 1987). Although the instrument was designed to evaluate "clinical teaching," the characteristics rated by students are relevant to didactic teaching as well. The instructions for use of the instrument state that alpha reliability coefficients range from .79-.89 for five categories of teaching characteristics. These characteristics include teaching ability, nursing competence, evaluation, interpersonal relationships, and personality. The questionnaire also had high test-retest reliability when completed twice by baccalaureate nursing students in Canada four weeks apart (no significant differences between the teacher evaluations from the first and second testing). Content validity (the adequacy and representativeness of content) was established by obtaining descriptions of effective and ineffective teaching behaviors from the literature and faculty and students at the institution that the instrument was developed

(Knox & Mogan, 1985; J. Mogan & Knox, 1983). Importantly, the instrument has been studied in multiple settings around the world (Canada, Greece, United States, Hong Kong, Israel, Australia), which increases its external validity (Beitz & Wieland, 2005; Benor & Leviyof, 1997; Kotzabassaki et al., 1997; Lee et al., 2002; Li, 1997; J. Mogan & Knox, 1987; Nehring, 1990).

The Johns Hopkins University School of Nursing Peer Observation Scale uses a 4 point Likert scale (1=Needs improvement, 2=Good, 3=Very good, 4=Excellent) to evaluate 7 domains of teaching behaviors: “Content and Organization”, “Communication Skills”, “Questioning Skills”, “Critical Thinking Skills”, “Rapport with Students”, “Learning Environment”, and “Teaching Methods.” Only one article has been written about use and development of the scale, and reliability and validity were not established (R. A. Berk et al., 2004). Although the research backing of the scale is not as strong as the NCTEI, it was appealing to PfID as the questionnaire’s items were relevant to the videotaped lectures of the PfID nurse educators (before and after the “train-the-trainer” program). The scale also covers aspects of teaching that peers and external experts are better qualified to evaluate than students (e.g. accurate knowledge of content and use of innovative teaching methods) (R. A. Berk et al., 2004).

Summary

In summary, the medical and nursing literature describes faculty evaluation in North America and Western Europe, but not in the country of Georgia or other post-Soviet countries. Use of MSF for evaluating physicians and medical trainees has increased significantly in the last decade, but the literature is limited in terms of articles that describe MSF for nurse education programs or teacher evaluation. Therefore, the

following pilot study will add to the literature by providing information on the acceptability of teacher evaluation in the country of Georgia and describe an example of MSF used for teacher evaluation in a nurse education program.

Chapter 3: Manuscript

For submission to: **International Journal of Nursing Education Scholarship**

[Short Title: Acceptability of Multisource Feedback]

[Category: Original Research]

Providing 360 Degree Multisource Feedback to Nurse Educators in the Country of Georgia: A Pilot Study

Christopher C. DeStephano, BS [a],

Kim A. Crawford, RN, MSN, MPH, FNP-C [b]

Maia Jashi, M.D. [b]

Judith L. Wold, PhD, RN[c]

[a] MD/MPH Candidate, Rollins School of Public Health, Emory University and Mayo Medical School, Mayo Clinic College of Medicine

[b] Program coordinators, Partners for International Development

[c] Clinical Professor, Lillian Carter Center for International Nursing, Emory University

Address correspondence and reprint requests to:

Christopher DeStephano
200 First St. SW
Rochester, MN, 55905
Phone: (904)703-0152
Email: destephano.christopher@mayo.edu

FUNDING

Support for field activities was received through the Rollins School of Public Health Global Field Experience from the NIH Global Field Frameworks grant.

AUTHOR CONTRIBUTIONS

CCD provided substantial contributions to the conception, design, and execution of the study, acquisition of data, and interpretation and analysis of the data. He drafted the manuscript, revised it critically for intellectual content, and provided final approval of the version being submitted. As the Partners for International Development program coordinators, KAC and MJ reviewed, provided feedback, and approved the design and execution of the study. They reviewed the manuscript, revised it critically for intellectual content, and provided final approval of the version being submitted. As the first author's thesis adviser and research mentor, JLW provided substantial contributions to the conception and design of the study and interpretation of the data. She revised the manuscript critically for intellectual content, and provided final approval of the version being submitted.

ABSTRACT

BACKGROUND: Due to insufficient nursing education standards in the country of Georgia, fifteen nurse educators participated in a train-the-trainer program. These educators are now offering vocational education courses with plans to teach 2500 Georgian nurses over 3 years.

METHODS: Using a 360 degree multisource feedback model, self, video, student, peer, and program coordinator evaluations of teaching effectiveness were completed. After nurse educators reviewed their results and identified areas for improvement, a questionnaire on the perceived acceptability of teacher evaluation was completed.

RESULTS: Of the 15 nurse educators, 93.3% indicated that nurse educators should receive feedback from self, student, peer, and video evaluations, while 100% indicated that nurse educators should receive feedback from the program coordinator. The accuracy and usefulness of the program coordinator evaluation were rated the highest while the peer evaluation was rated the lowest on these domains.

CONCLUSION: This pilot study revealed that multisource feedback was acceptable to Georgian nurse educators.

INTRODUCTION

Insufficient nursing education standards in the country of Georgia have contributed to low professional status for nurses and sublevel nursing/medical care. Although "nursing colleges" exist in the country, nursing education does not take place at institutions of higher education. This precludes program graduates from teaching in a university level baccalaureate nursing education program (Wold et al., 2007). Therefore, nurses are primarily governed and trained by physicians. This has resulted in poor morale and lack of vision for the future of nursing due to nurses not having a substantial voice in their profession (Walker & Wold, 2009). The Ministry of Education and Science of Georgia is responsible for the oversight and accreditation of "nursing colleges" but only conducts institutional accreditation (Wold et al., 2007). Since nursing is not regarded as a profession in Georgia, licensure and nursing standards are non-existent and regulations for nursing care are minimal. This has contributed to significant variation between nursing curricula (Walker & Wold, 2009). Students are deemed qualified to take a nursing position at a hospital or clinic after passing a school-constructed, non-standardized examination (Wold et al., 2007). Once in a hospital or clinic, nurses are allowed to make few independent decisions regarding patient care and are trained at a level that falls between that of nursing assistants and licensed practical nurses in the United States (Ivanov & Paganpegara, 2003; Walker & Wold, 2009; T. Zimmerman et al., 1997).

If nurses are not educated and prepared to meet the challenges of a modern system of health care, patient morbidity, mortality, and patient satisfaction are adversely affected

(Scott et al., 1999). To expand the role of nurses, improve their vision for the profession, and develop nurse leaders, a partnership of academic institutions in the United States and the country of Georgia, funded by USAID, developed an intensive faculty development program for nurse educators from November 2009 to February 2010. Nurses from the United States taught courses in general and pediatric nursing to fifteen Georgian healthcare professionals (trained previously as nurses or physicians). In addition, the curriculum included specialty care areas (perioperative, labor and delivery, critical care, and emergency nursing), teaching strategies, and nursing management and leadership. In February 2010, the trained nurse educators began offering month long vocational education courses with plans to teach 2500 nurses over three years (Walker & Wold, 2009). Concurrently with this vocational education program, select nurse educators are involved in preparation of relevant courses for a four-year baccalaureate degree nursing school.

Although higher education institutions in North America and Western Europe use teacher evaluation methods that provide feedback to educators and administrators to improve the quality of education (R. Berk, 2006), researchers have not evaluated these methods in Georgia or other post-Soviet countries. This limits the information that is available for improving educational standards and nursing practice. Therefore, the purpose of this pilot study was to provide feedback to nurse educators and program coordinators of the vocational education program and evaluate the perceived acceptability (as rated by nurse educators) of teacher evaluation in the country of Georgia. This project lays the foundation for longitudinal evaluation of teaching effectiveness, so that strengths and areas for improvement of nurse educators can be continuously identified.

Literature Review

The literature contains over 15,000 studies on teaching effectiveness (Seldin, 1999). In the nursing and medical literature, many of these studies have examined what makes an effective "clinical teacher" with the intent of developing instruments to evaluate and provide feedback to teachers (Fluit et al., 2010; Knox & Mogan, 1985; Li, 1997; J. Mogan & Knox, 1983; J. Mogan & Knox, 1987; Nehring, 1990; Raingruber & Bowles, 2000; L. Zimmerman & Westfall, 1988). A recent systematic review of questionnaires for assessing teachers summarized the different domains of clinical teaching in medical education: "Excellent clinical teachers are described as physician role models, effective supervisors, dynamic teachers, and supportive individuals, possibly complemented by their role as assessors, planners, and resource developers" (Fluit et al., 2010, p. 1337). Ratings of clinical teaching effectiveness are associated with medical student performance on the National Board of Medical Examiners subject examinations and the United States Medical Licensing examination (Blue et al., 1999; Griffith et al., 2000; Roop & Pangaro, 2001). The students who trained under the highest rated attending physicians scored significantly higher than those who trained under the lowest rated physicians.

In the nursing literature, investigators have used the Nursing Clinical Teacher Effectiveness Inventory (NCTEI) in multiple settings (Canada, Greece, United States, Hong Kong, Israel, Australia) to assess the teaching behaviors that nursing faculty and students value (Benor & Leviyof, 1997; Kotzabassaki et al., 1997; Lee et al., 2002; Li, 1997; J. Mogan & Knox, 1987; Nehring, 1990). In the original Mogan and Knox study

(1987) using the instrument, faculty and students rated the following 8 items highest: “is a good role model”, “enjoys nursing”, “demonstrates clinical skills and judgment”, “enjoys teaching”, “is well prepared for teaching”, “takes responsibility for own actions”, “is approachable”, and “is self-confident” (p. 334). Studies of whether these ratings are associated with student performance on the National Council Licensure Examination-Registered Nurses (NCLEX-RN) have not been performed.

Although investigators have extensively studied effective clinical teaching behaviors and teacher evaluation instruments in the United States, the literature is limited in Russia and non-existent in other post-Soviet countries. While a PubMed search of “(United States) AND ((education, medical) OR (education, nursing)) AND (faculty OR teacher) AND (ratings OR feedback OR evaluation OR assessment))” revealed 2933 articles, use of “USSR”—the MeSH term that comprises Russia and other former Soviet Republics—instead of “United States” only revealed 15 articles. Of the 9 articles in English, only 1 reported faculty evaluation results (Wong & Agisheva, 2007). This study assessed teaching performance using a pretest/post-test questionnaire of faculty participants’ self-reported ratings of teaching ability before and after a “developing teaching skills for medical educators in Russia” faculty development intervention (Wong & Agisheva, 2007). The 6 studies in languages other than English did not discuss faculty evaluation in the English abstract.

The literature describes two types of decisions that result from faculty evaluation: formative and summative decisions. Faculty make formative decisions to improve and shape their pedagogy (R. Berk, 2006). These decisions are made based on feedback from students and administrators throughout the year to plan and revise the curriculum (R.

Berk, 2006; R. A. Berk, 2009). Administrators make summative decisions to "sum up" overall performance, which influences tenure, promotion, dismissal, and merit pay (R. Berk, 2006; R. A. Berk, 2009).

Higher education has historically used student evaluations as the main (or only) source of teacher ratings for both formative and summative decisions in higher education (R. Berk, 2006; R. A. Berk, 2009). However, investigators have recently proposed use of additional evaluation sources, since use of only one source is often unreliable and biased (R. Berk, 2006; R. A. Berk, 2009; R. A. Berk et al., 2004). The bias and limited information that result from single sources of evaluation led to the use of 360-degree multiple source feedback (MSF) in medicine approximately 15 years ago (the first article in PubMed is from 1997) (R. A. Berk, 2009; Lockyer, 2003). MSF was initially developed for use in management and industry to assess performance of employees from a variety of perspectives (R. A. Berk, 2009; Lockyer, 2003). In industry, competence for specific behaviors (e.g. aptitude, finishing tasks, interpersonal skills, leadership skills) is assessed using questionnaires that are completed by people at the same, lower, and higher levels of the organizational chart (R. Berk, 2006; Lockyer, 2003). The results are compared to self-ratings so that precise feedback is obtained and formative decisions can be made to improve job performance (R. Berk, 2006). The medical literature has increasingly described the MSF model for evaluating practicing physicians and medical residents. The evidence is mixed regarding whether MSF promotes changed behaviors (Miller & Archer, 2010). A PubMed search of "(multisource OR multi-source) AND (feedback)" resulted in 82 papers with 73 of them being from 2005-2010. However, only

one of these described use of MSF for teacher evaluation and none of them for nursing education (R. A. Berk, 2009).

Conceptual Framework

Since the literature on faculty evaluation in post-Soviet countries is non-existent and summative decisions should be limited to instruments for which validity and reliability are established, we decided that evaluation of the nurse educators should be used for formative decisions. We also determined that MSF would be most appropriate for evaluating and providing feedback to the program's recently trained nurse educators, since the model was designed to frame a more complete picture of performance than could be provided from a single-source, top-down, supervisor-only approach (R. Berk, 2006). Although the literature is limited regarding use of MSF for either faculty evaluation or nursing education, Berk (2009) provides a framework for using the 360 degree multisource feedback model to evaluate teaching. Sources of evaluation that were proposed for faculty evaluation included colleagues, peer observers, department chairs, external experts, other personnel, secretaries, and students (R. A. Berk, 2009). As shown in figure 3, peer, video, student, program coordinator, and self evaluations of the educators were chosen for the PfID vocational education program.

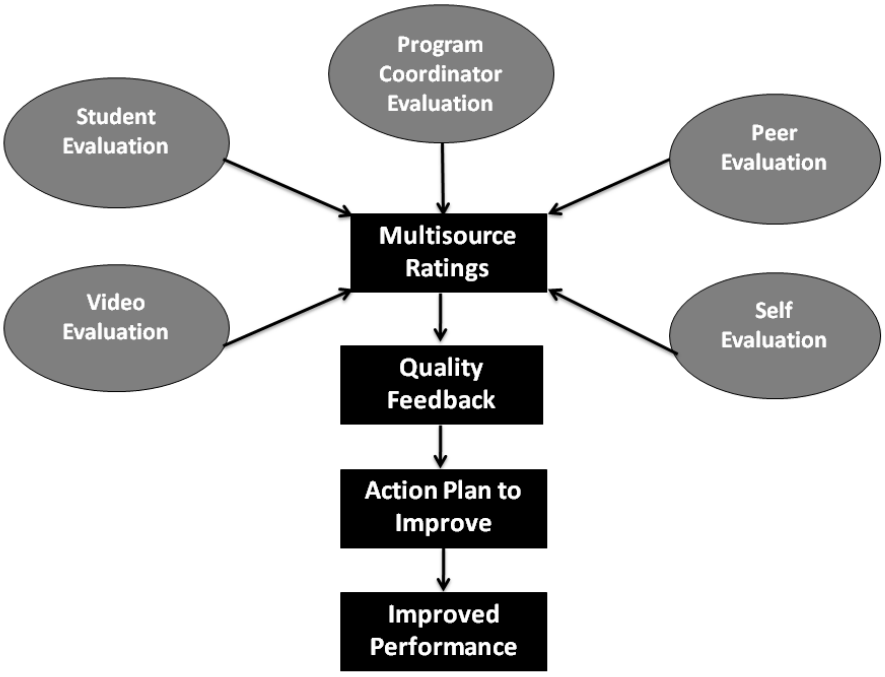

Figure 3: Conceptual framework of teacher evaluation using multisource feedback in a country of Georgia vocational education program

METHODS

Intervention

In June 2010, 15 Georgian nurse educators who previously participated in a faculty development program between November 2009 and February 2010 taught vocational education general nursing courses in the capital city (Tbilisi) of the country of Georgia. The educators were divided into 4 teams of 3-4 teachers with each team responsible for 19-24 students (practicing nurses from around the country). The nurse educators developed lectures and examinations in English to be reviewed and edited for content and accuracy by program coordinators. The lectures and examinations were translated to Georgian and presented to students. The program coordinators included a nurse practitioner with a master of public health trained in the United States and a

Georgian trained physician who worked as a nurse educator in the country of Georgia and completed coursework at a nursing teaching institute in the United States. Half of the vocational education coursework was didactic with the other half being simulation and clinical experience.

The first part of the course provided a survey of nursing practice, which included lectures on health promotion, patient education, nursing ethics, physical assessment, hygiene, and basic procedures. The next part of the course provided an introduction to the pathophysiology of common diseases. Simulation with mannequins was used to teach and practice basic life support, sterile technique, and placement of urinary foley catheters, nasogastric tubes, and intravenous catheters. Clinical experience was provided on a multi-profile unit where students learn and practice administration of medications, documentation, and the physical examination. Students completed assignments and weekly examinations to assess their understanding of the material. Students were also required to complete a final examination and receive a score of 80% or above to pass the course.

Teacher Evaluation Process

To evaluate the teaching effectiveness of the nurse educators, we used five different sources (self, video, student, peer, and program coordinator evaluations) between November 2009 and June 2010. At the end of the course, an external evaluator summarized these results and provided a table with the mean (95% confidence interval) ratings from each to the educators. The table provided the lowest and highest ratings from each source and the educators' overall mean (95% CI) ratings. After reviewing the summary report, we asked nurse educators to identify three areas of teaching they wanted

to improve. We then administered a questionnaire on the acceptability of the teacher evaluation methods to the educators. The study protocol was submitted to our institution's institutional review board, which determined it to be exempt.

Video evaluation Procedure and Instrumentation

The principal investigator completed video evaluations in May 2010. Educators were videotaped giving lectures in English at the beginning (November 2009) and the end (February 2010) of the faculty development course. The educator chose the topic for the first lecture, and the program coordinators selected the topic ("What it means to be a nurse") for the final lecture. The principal investigator adapted and used the Johns Hopkins University School of Nursing Peer Observation Scale with permission to evaluate these videotaped presentations (R. A. Berk et al., 2004). The instrument uses a 4-point Likert scale (1=Needs improvement, 2=Good, 3=Very good, 4=Excellent) to evaluate 7 domains of teaching behaviors: "Content and Organization", "Communication Skills", "Questioning Skills", "Critical Thinking Skills", "Rapport with Students", "Learning Environment", and "Teaching Methods." As the videotaped lectures were 5 minutes, did not involve interaction with students, and used one teaching method (PowerPoint lectures), only the "content and organization" and "communication skills" domains were used for the 17-item, modified instrument.

Self, peer, student, and program coordinator evaluation procedure and instrumentation

The self evaluations consisted of the nurse educators rating themselves at the beginning of the vocational education course. At the halfway point of the course, the

educators from each team completed peer evaluations by evaluating their fellow team members. Each educator completed 2-3 evaluations depending on the size of the team. To ensure that students had adequate exposure to each educator, students completed evaluations for each educator after the educator had taught at least 12 hours of lecture. To ensure anonymity, the principal investigator required the nurse educator to leave the classroom prior to having the students complete the evaluation. At the end of the June 2010 general nursing vocational education course, the nurse practitioner trained in the United States completed the program coordinator evaluation.

We adapted and used with permission the Nursing Clinical Teaching Effectiveness Inventory (NCTEI) for the self, peer, student, and program coordinator evaluations. Mogan and Knox (1987) described the NCTEI following their article about effective and ineffective behaviors of clinical teachers as perceived by nursing students and faculty members (Knox & Mogan, 1985). The instrument is a 48-item checklist of teaching behaviors on a 7-point Likert scale (1=Least Descriptive of Teacher, 7=Most Descriptive of Teacher). Investigators have used the instrument in multiple settings (Canada, Greece, United States, Hong Kong, Israel, and Australia) with results presented as mean $\pm$ standard deviation (Beitz & Wieland, 2005; Benor & Leviyof, 1997; Kotzabassaki et al., 1997; Lee et al., 2002; Li, 1997; J. Mogan & Knox, 1987; Nehring, 1990). The instructions for use of the instrument state that alpha reliability coefficients range from .79-.89 for five categories of teaching characteristics. These characteristics include teaching ability, nursing competence, evaluation, interpersonal relationships, and personality. To establish content validity, Mogan and Knox (1987) used expert review,

use of the literature, and focus groups with students and faculty about effective and ineffective teaching behaviors.

After a discussion with nurse educators and program coordinators about which items were most relevant for evaluation of educators in the country of Georgia, we shortened the instrument to 21 items with the anchors: 1=Not at all descriptive of teacher and 7=Very descriptive of teacher. The instrument was translated to Georgian and back translated into English to assure accuracy of translation. As all of the nurse educators were fluent in Georgian and English, they completed the self and peer evaluations in English. The student evaluation was completed in Georgian.

Acceptability Instrument

We developed and administered a 23-item questionnaire to nurse educators that assesses the acceptability of teacher evaluation in the country of Georgia. The first section asked for background information: age of the nurse educators, whether professors (university, nursing school, medical school) in Georgia receive evaluations of teaching performance, types of evaluations received by professors, previous teaching experience, and whether the educator received evaluations of teaching performance in the past. The next sections evaluated the accuracy, usefulness, and interest in future use of the different sources of teacher evaluation. For each source of feedback, a 5-point Likert scale (1=Strongly disagree, 2=Disagree, 3=Undecided, 4=Agree, 5=Strongly Agree) was used to assess agreement with the following two statements: "The evaluation provided accurate feedback about my teaching effectiveness" and "The evaluation provided feedback that I will use to improve teaching effectiveness." Educators then marked "yes" or "no" to the following question for each source of feedback: "Should nurse educators

receive feedback from the evaluation in the future?" At the end of the questionnaire, educators were asked to "rank from 1 to 5 which evaluations are the most helpful in providing information to improve teaching effectiveness (1=least helpful, 5=most helpful)." The directions stated that each number should only be used once. Space was provided at the end of the questionnaire for the educators to provide any other comments or recommendations that they have about the evaluation instruments.

Statistical Analysis

Although the survey data are ordinal and negatively skewed, mean (95% confidence interval) ratings are presented for Likert-scale results. The mean is more discriminating than the median and more understandable to the nurse educators who have limited experience with statistical methods. Because of the extremely high ratings received by the nurse educators, we were less concerned about the weakness of using the mean to describe negatively skewed data--portraying lower class ratings than occurred (R. Berk, 2006). Number (%) was used for nominal data. Statistical analysis was performed using a paired t-test for comparing the highest and lowest rated sources of feedback and the mean ratings from the video evaluation before and after the faculty development course. Although a systematic analysis of qualitative data was not planned as part of this evaluation, quotes are included as representative of educators' and students' comments.

RESULTS

Background of Nurse Educators

Peer, student, video, self, and program coordinator evaluations were completed for fifteen nurse educators. The mean (standard deviation) age of the educators was 37.3 (6.5) years. Of the 15 educators, 6 (40%) reported being trained first as a nurse while 9 (60%) reported being trained in a medical specialty. Having previous teaching experience was reported by 5 (33.3%) of the nurses, which consisted of 2 educators previously teaching at a family medicine center, 1 teaching at a private medical school with classes conducted in English, 1 teaching in an emergency room, and 1 teaching in a nursing school and a high school. Six (40%) of the educators reported that professors (of universities, medical schools, and nursing schools) receive evaluations of their teaching performance; four (26.7%) reported that teachers receive student evaluations, and two (13.3%) reported that teachers receive evaluations from other teachers. Only one nurse educator had previously received an evaluation of her teaching performance; the evaluation was from students.

Teacher Evaluation Results

The nurse educators were rated highly by the student, self, peer, and program coordinator evaluations. Due to the number of evaluation sources and the 38 total items evaluated, mean (95% confidence interval) ratings are provided for the highest (table 1) and lowest (table 2) rated items to reveal the range of responses from the 5 sources of feedback.

Table 1: Highest mean (95% confidence interval) ratings for items from self, student, peer, program coordinator, and video evaluations

Highest Rated Item	Evaluation Source	Mean (95% CI) Rating
"Enjoys Teaching"	Self	7.00 (all responses)†
	Program Coordinator	6.73 (6.40-7)†
"Is well-prepared for teaching"	Peer	6.68 (6.45-6.90)†
"Explains clearly"	Student	6.94 (6.91-6.97)†
"Corrects students' mistakes without belittling them"	Student	6.94 (6.91-6.97)†
"Spoke with adequate volume"	Video Before Faculty Development Course	3.40 (3.05-3.75)‡
	Video After Faculty Development Course	3.75 (3.51-3.99)‡
† Self, peer, student, and program coordinator ratings were from a 7-point Likert scale (1=Not at all descriptive of teacher, 7=Very descriptive of teacher) ‡ Video ratings were from a 4-point Likert Scale (1=Needs improvement, 2=Good, 3=Very good, 4=Excellent)		

Table 2: Lowest mean (95% confidence interval) ratings for items from self, student, peer, program coordinator, and video evaluations

Lowest Rated Items	Evaluation Source	Mean (95% CI) Rating
"Demonstrates extensive knowledge in nursing"	Self	5.93 (5.54-6.32)†
	Peer	6.05 (5.74-6.36)†
"Encourages active participation in discussion"	Student	6.72 (6.62-6.82)†
"Identifies students' strengths and limitations fairly"	Program Coordinator	5.07 (4.46-5.68)†
"Presented overview of class content/objectives"	Video Before Faculty Development Course	1.07 (0.92-1.21)‡
	Video After Faculty Development Course	1.75 (1.09-2.41)‡
† Self, peer, student, and program coordinator ratings were from a 7-point Likert scale (1=Not at all descriptive of teacher, 7=Very descriptive of teacher) ‡ Video ratings were from a 4-point Likert Scale (1=Needs improvement, 2=Good, 3=Very good, 4=Excellent)		

The overall mean (95% CI) ratings for the 21 item 7-point Likert scale (1=Not at all descriptive of teacher, 7=Very descriptive of teacher) was 6.61 (6.41-6.81) from the self evaluation, 6.54 (6.29-6.78) from the peer evaluation, 6.88 (6.84-6.93) from the student evaluation, and 6.08 (5.62-6.54) from the program coordinator evaluation. The video evaluations used a 4-point Likert scale (1=Needs improvement, 2=Good, 3=Very Good, 4=Excellent), and were lower than the student, self, peer, and program coordinator evaluations. The overall mean (95% CI) rating for the lecture videotaped after the faculty development course was 3.18 (2.91-3.45), which was significantly higher than the 2.34 (1.99-2.69) rating of the lecture prior to the faculty development course (paired t-test p-value<0.0001). Table 3 provides the ratings for the highest and lowest rated educators from each source of evaluation.

Table 3: Mean (95% CI) ratings for highest and lowest rated educators from each source of evaluation

	Rating for Lowest Rated Educator	**Rating for Highest Rated Educator**
Self Evaluation	6.00 (5.71-6.29) †	7.0 (all responses) †
Peer Evaluation	4.71 (4.47-4.94) †	6.95 (6.85-7) †
Student Evaluation	6.77 (6.68-6.86) †	7.0 (all responses) †
Program Coordinator Evaluation	4.19 (3.48-4.91) †	6.95 (6.85-7.00) †
Video Before Faculty Development Course	1.41 (1.15-1.67) ‡	3.41 (2.93-3.89) ‡
Video After Faculty Development Course	2.53 (2.01-3.05) ‡	3.94 (3.82-4.00) ‡
† Self, peer, student, and program coordinator ratings used a 7-point Likert scale (1=Not at all descriptive of teacher, 7=Very descriptive of teacher) ‡ Video ratings used a 4-point Likert Scale (1=Needs improvement, 2=Good, 3=Very good, 4=Excellent)		

Self and student evaluations provided the highest ratings and the ratings with the least difference between the highest and lowest rated educators. The peer, program

coordinator, and video evaluations provided lower ratings and more discrimination between the educators.

Areas for improvement identified by nurse educators

After the nurse educators received their evaluations, they were asked to identify areas for improvement. The areas identified were mainly specific to the ratings received by the individual educators. For example, two of the educators received the lowest rating from their peers on the item "Is approachable." Both of the educators identified this as an area they would like to improve. There were also two areas that were identified by a majority of the educators. Of the 15 educators, 8 made statements about "encouraging more active participation" from students and 7 hoped to "improve their knowledge of nursing."

The identification of "encouraging more active participation" from students likely arose from this item being the lowest rated by the students. One educator stated, "I want to provide more encouragement to students so they feel comfortable figuring out questions without interruption from me…Continue to question their underlying knowledge and bring the material down a level so that they understand." Other educators reported that they would "ask more questions of students to test their understanding" and "will keep trying to engage audience and look out at them." The identification of "demonstrating extensive knowledge in nursing" corresponds with this being the lowest rated item on both the self and peer evaluations. Of the 7 statements, 5 of them were made by educators who were trained originally as physicians that began working as a nurse later. The educators with a predominantly medical background made statements such as, "I want to improve my 'nursing mentality' since I was a doctor before and do not

feel I have adequate knowledge in nursing." Another educator stated, "I would like to work on my theoretical knowledge of nursing because nursing is new. Even though I was trained in Britain as a physician, nursing is different." The two educators predominantly from a nursing background made these statements: "I would like to work on my knowledge of nursing so that I can be a role model to students. I'm still learning so I don't feel like I am a model yet" and "I need to improve knowledge of nursing for teaching."

Acceptability of Teacher Evaluation

After receiving feedback and identifying areas for improvement, educators rated the perceived accuracy and usefulness of the different sources of feedback. The mean (95% CI) ratings are provided in table 4.

Table 4: Mean (95% CI) ratings of the perceived accuracy and usefulness of the different sources of feedback on a 5-point Likert Scale

	Self (n=15)	**Student (n=15)**	**Peer (n=15)**	**Video (n=15)**	**Program Coordinator (n=15)**
"The evaluation provided feedback that I will use to improve my teaching effectiveness"	4.07 (3.62-4.51)	4.33 (4.06-4.60)	3.80* (3.49-4.11)	4.07 (3.74-4.40)	4.53* (4.25-4.82)
"The evaluation provided accurate feedback about my teaching effectiveness"	N/A	3.87 (3.46-4.28)	3.80 (3.49-4.11)	3.87 (3.67-4.06)	4.27 (4.01-4.52)
* Paired t-test p-value<0.05					

The accuracy and usefulness of the program coordinator evaluation were rated the highest while the peer evaluation was rated the lowest. The difference between the peer and

program coordinator ratings was statistically significant for accuracy (p=0.048) and usefulness (p=0.0032). When educators were asked to rank from 1 to 5 (1=Least helpful, 5=Most helpful) "which evaluations were the most helpful in providing information to improve teaching effectiveness," the evaluation from the program coordinator was again rated the highest while the evaluation from the peer evaluation was rated lowest. The rankings (mean, 95% CI) from highest to lowest are as follows: program coordinator evaluation (3.73, 3.02-4.44), student evaluation (3.60, 2.98-4.22), self evaluation (2.73, 1.99-3.47), video evaluation (2.60, 1.65-3.56), and peer evaluation (2.40, 1.74-3.06). Of the 15 nurse educators, 14 (93.3%) indicated that "nurse educators should receive feedback" from self, student, peer, and video evaluation while 15 (100%) indicated that "nurse educators should receive feedback" from the program coordinator.

Comments from the educators were generally positive: "Assessment is very important…Working hard and evaluation provides validation that we are doing well. It gives us an idea of what needs to be improved….This helps develop our professional skills." Of the 8 educators, 4 provided comments stating that assessments need to be completed "regularly" and "very often." One of the comments compared the evaluations and explained their reasoning for how she ranked the evaluations: "I appreciate feedback from Kim [program coordinator] because she is competent and knowledgeable about what makes good lecture and teacher. Self and student evaluations do not provide feedback on American model of nursing because we don't know what is good teaching." Another educator stated, "I think that my peers' and students' assessments are subjective. It is important to assess my job from program coordinator. She is best at assessing the quality of lesson plans."

DISCUSSION

This pilot study describes the use of MSF for teachers of a vocational education program for nurses in the country of Georgia. Despite minimal exposure to teacher evaluation methods, almost all of the educators indicated that "nurse educators should receive feedback" from each source of evaluation. It should be noted that educators were informed that the feedback was designed for formative decisions. Acceptability ratings may have differed if the evaluations were used for higher stakes, summative decisions.

Educators' interest in receiving feedback from MSF may have multiple explanations. The high ratings for the survey items, "the evaluation provided feedback that I will use to improve my teaching effectiveness" and "the evaluation provided accurate feedback about my teaching effectiveness" suggest that the educators perceived the evaluations to be fair and useful. This is important as a previous study of MSF for physicians revealed that participants who disagreed with the feedback were less likely to use it for practice improvement (J. Sargeant et al., 2005). When educators were provided feedback from the ratings, only one educator identified aspects of the feedback that were unfair. She reported that she was unhappy that her peer evaluations were so low, and stated that her ratings on the domains of "approachability" and "enthusiasm" were incorrect. Other than this case and another educator who was rated lower on the domain of "approachability" by her peers, the ratings received by educators from the different sources were generally high, especially the ratings from the self, peer, and student evaluations.

The reasons for the high student evaluation ratings warrant discussion, as it may represent the “halo effect” (Abruzzese, 1992). The lack of low ratings may have decreased the likelihood of educators contemplating or initiating changes to teaching practices. As shown in table 3, the lowest rated educator by the students received a mean of 6.77 out of 7 compared to the highest rated educator who received 7 on all responses. This is in contrast to the United States where low student ratings of teacher effectiveness have resulted in cynicism from some faculty toward student evaluations (R. Berk, 2006). One explanation for the high ratings was the novelty of continuing education for nurses in the country of Georgia. For a majority of the nurses in the vocational education courses, it was their first experience with continuing education and there appeared to be appreciation for the courses being offered which may have introduced social desirability bias. In addition, the lack of exposure to other teachers likely reduced the students’ ability to differentiate effective teaching from ineffective teaching.

Even though the educators received high ratings from students, 8 of 15 educators identified “encourages active participation in discussion” as an area for improvement, which was the lowest rated item from the student evaluations. This suggests that the numerical value given for a domain of teaching may be less important than the relative position of the item’s rating compared to the other items. Therefore, student evaluations in the country of Georgia may still be useful for providing formative feedback despite not providing information that can be used to differentiate the teaching effectiveness of educators. Future research will be undertaken to determine how additional exposure to vocational education will influence student ratings of teaching effectiveness. Students from the general nursing course described in this study now have the opportunity to take

courses in specialty areas of nursing including intensive care unit nursing, operating room nursing, emergency department nursing, and obstetric nursing.

Although we hypothesized prior to the study that sources providing higher ratings would be more acceptable than sources that provided lower ratings, this did not appear to be the case. Despite educators receiving their lowest ratings from the program coordinator and the video evaluations, the acceptability of the program coordinator evaluation was rated highest while the rating for the video evaluation was similar to the other sources of evaluation. The quotes from educators about their desire to "learn the American model of nursing" provide one explanation. The fact that evaluations from the program coordinator and video were completed by evaluators originally from the United States may have influenced the acceptability ratings. However, another explanation is that lower ratings provide more insight about changes that need to be made to teaching practice. Supporting this explanation is a previous study of 255 physicians showing that physicians who contemplated or initiated practice changes received lower mean MSF ratings (Fidler et al., 1999).

Since educators ranked the program coordinator evaluation as the most helpful, the question is raised whether a top-down, single source of feedback approach would have been sufficient. The possibility exists that MSF diluted the impact of lower rated domains. Although the educator may have received a low rating from one source, she could have discounted the accuracy of the rating due to receiving higher ratings on that item from other sources. Additional research will be required, but the program coordinators and investigators of this study found MSF to provide richer information than evaluations from a single source. By combining the sources, the strengths of one source

can compensate for weaknesses of the other sources (R. Berk, 2006). In the case of this study, the program coordinators were able to evaluate the accuracy of the content that the Georgian nurse educators used for their lectures. However, they had difficulty evaluating the interpersonal skills of the educators and the ability of the educators to work as a team. This is why having feedback from the peer evaluations was helpful in identifying two educators that were thought to be "less approachable." Similarly, as students gain more exposure to continuing education from different teachers, it's hoped that student evaluations will be able to assess teaching performance in the classroom better than a program coordinator can by observing the class over a limited period of time.

Conclusions

This pilot study revealed that MSF was acceptable to nurse educators in a vocational education program in the country of Georgia. Further research is necessary to assess the validity and reliability of different rating scales, determine how ratings change over time, and ultimately evaluate whether MSF improved teaching performance of nurse educators and the nursing practice of the program's participants.

REFERENCES

Abruzzese, R. S. (1992). *Nursing Staff Development: Strategies for Success*. New York: Mosby-Year Book.

Beitz, J. M., & Wieland, D. (2005). Analyzing the teaching effectiveness of clinical nursing faculty of full- and part-time generic BSN, LPN-BSN, and RN-BSN nursing students. *J Prof Nurs, 21*(1), 32-45.

Benor, D. E., & Leviyof, I. (1997). The development of students' perceptions of effective teaching: the ideal, best and poorest clinical teacher in nursing. *J Nurs Educ, 36*(5), 206-211.

Berk, R. (2006). *Thirteen Strategies to Measure College Teaching: A Consumer's Guide to Rating Scale Construction, Assessment, and Decision Making for Faculty, Administrators, and Clinicians*. Sterling, Virginia: Stylus Publishing.

Berk, R. A. (2009). Using the 360 degrees multisource feedback model to evaluate teaching and professionalism. *Med Teach, 31*(12), 1073-1080.

Berk, R. A., Naumann, P. L., & Appling, S. E. (2004). Beyond student ratings: peer observation of classroom and clinical teaching. *Int J Nurs Educ Scholarsh, 1*, Article10.

Blue, A. V., Griffith, C. H., 3rd, Wilson, J., Sloan, D. A., & Schwartz, R. W. (1999). Surgical teaching quality makes a difference. *Am J Surg, 177*(1), 86-89.

Fidler, H., Lockyer, J. M., Toews, J., & Violato, C. (1999). Changing physicians' practices: the effect of individual feedback. *Acad Med, 74*(6), 702-714.

Fluit, C. R., Bolhuis, S., Grol, R., Laan, R., & Wensing, M. (2010). Assessing the quality of clinical teachers: a systematic review of content and quality of questionnaires for assessing clinical teachers. *J Gen Intern Med, 25*(12), 1337-1345.

Griffith, C. H., 3rd, Georgesen, J. C., & Wilson, J. F. (2000). Six-year documentation of the association between excellent clinical teaching and improved students' examination performances. *Acad Med, 75*(10 Suppl), S62-64.

Ivanov, L. L., & Paganpegara, G. (2003). Public health nursing education in Russia. *J Nurs Educ, 42*(7), 292-295.

Knox, J. E., & Mogan, J. (1985). Important clinical teacher behaviours as perceived by university nursing faculty, students and graduates. *J Adv Nurs, 10*(1), 25-30.

Kotzabassaki, S., Panou, M., Dimou, F., Karabagli, A., Koutsopoulou, B., & Ikonomou, U. (1997). Nursing students' and faculty's perceptions of the characteristics of 'best' and 'worst' clinical teachers: a replication study. *J Adv Nurs, 26*(4), 817-824.

Lee, W. S., Cholowski, K., & Williams, A. K. (2002). Nursing students' and clinical educators' perceptions of characteristics of effective clinical educators in an Australian university school of nursing. *J Adv Nurs, 39*(5), 412-420.

Li, M. K. (1997). Perceptions of effective clinical teaching behaviours in a hospital-based nurse training programme. *J Adv Nurs, 26*(6), 1252-1261.

Lockyer, J. (2003). Multisource feedback in the assessment of physician competencies. *J Contin Educ Health Prof, 23*(1), 4-12.

Miller, A., & Archer, J. (2010). Impact of workplace based assessment on doctors' education and performance: a systematic review. *Bmj, 341*, c5064.

Mogan, J., & Knox, J. (1983). Students' perceptions of clinical teaching. *Nurs Pap, 15*(3), 4-20.

Mogan, J., & Knox, J. E. (1987). Characteristics of 'best' and 'worst' clinical teachers as perceived by university nursing faculty and students. *J Adv Nurs, 12*(3), 331-337.

Nehring, V. (1990). Nursing Clinical Teacher Effectiveness Inventory: a replication study of the characteristics of 'best' and 'worst' clinical teachers as perceived by nursing faculty and students. *J Adv Nurs, 15*(8), 934-940.

Raingruber, B., & Bowles, K. (2000). Developing student evaluation instruments to measure instructor effectiveness. *Nurse Educ, 25*(2), 65-69.

Roop, S. A., & Pangaro, L. (2001). Effect of clinical teaching on student performance during a medicine clerkship. *Am J Med, 110*(3), 205-209.

Sargeant, J., Mann, K., & Ferrier, S. (2005). Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness. *Med Educ, 39*(5), 497-504.

Scott, J. G., Sochalski, J., & Aiken, L. (1999). Review of magnet hospital research: findings and implications for professional nursing practice. *J Nurs Adm, 29*(1), 9-19.

Walker, K., & Wold, J. (2009). Emory University School of Medicine: Proposal for Georgia Nursing Education in Tbilisi, Georgia: United States Agency for International Development/Caucasus.

Wold, J. L., Walker, K., & Partskhladze, N. (2007). Establishing a medical and nursing education program in the Republic of Georgia. In C. Holtz (Ed.), *Global Health*. Sudbury, Massachusetts: Jones and Bartlett.

Wong, J. G., & Agisheva, K. (2007). Developing teaching skills for medical educators in Russia: a cross-cultural faculty development project. *Med Educ, 41*(3), 318-324.

Zimmerman, L., & Westfall, J. (1988). The development and validation of a scale measuring effective clinical teaching behaviors. *J Nurs Educ, 27*(6), 274-277.

Zimmerman, T., Putland, L. C., & Hughes, J. R. (1997). Russian and local nurse educators exchange ideas. *Nurs Spectr (Washington D.C.), 7*(14), 9.

Chapter 4: Discussion, Conclusions, and Recommendations

This pilot study was designed to provide formative feedback to nurse educators in the country of Georgia, determine whether teacher evaluation is acceptable, establish a foundation for future teacher evaluation, and inform studies on nurse education program monitoring and evaluation. The implications of the project and MSF on public health depend on whether the results of the evaluation are used to guide future action in planning and carrying out educational interventions (Bastable, Gramet, Jacobs, & Sopczyk, 2011). MSF has the potential for expanded use in the field of monitoring and evaluation of public health and health education programs. The following discussion will describe potential applications of MSF within the Roberta Straessle Abruzzese (RSA) evaluation model (figure 4) developed for nursing staff development (Abruzzese, 1992). The model conceptualizes total program evaluation as encompassing a hierarchical pyramid of evaluations that increase in complexity from bottom to top. In order of increasing complexity, the evaluations include process, content, outcome, and impact evaluations (Abruzzese, 1992).

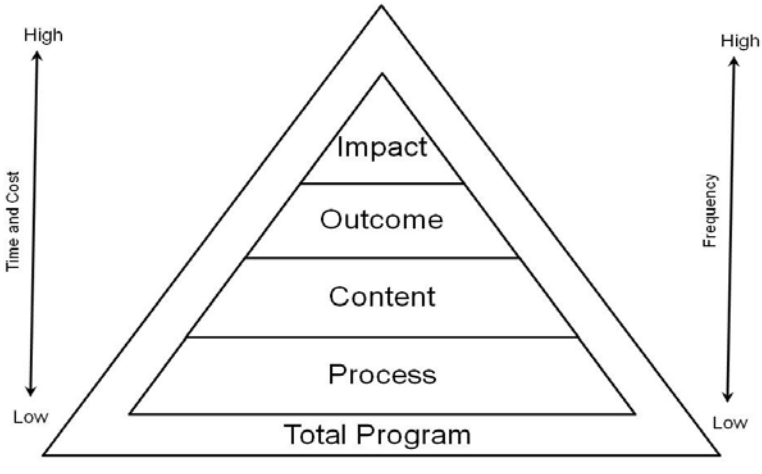

Figure 4: RSA Evaluation Model. Reprinted with permission (Abruzzese, 1992, p. 238)

The first level of the evaluation hierarchy is process evaluation which is used for making adjustments in personnel, materials, facilities, learning objectives, or teaching style (Abruzzese, 1992; Bastable et al., 2011). Process evaluation is ongoing throughout an educational activity and identifies areas where teaching can be improved to facilitate learning (Abruzzese, 1992; Bastable et al., 2011). The scope is usually limited to a specific learning activity (e.g. class or workshop) and includes aspects of the learning experience such as teacher behavior, learner-teacher interaction, and learner response to teaching methods (Abruzzese, 1992; Bastable et al., 2011).

The use of student, self, peer, video, and program coordinator evaluations by PfID for the program's month-long vocational education courses is an example of using MSF for a process evaluation. Educators received feedback on their teaching effectiveness from 5 sources and developed an action plan for improvement. Further research is required to determine whether use of MSF translates into improved teaching and whether these improvements translate into better nursing practice. The study by Fidler et al. (1999) on whether physicians made changes to their medical practices following MSF provides a framework for assessing whether MSF resulted in teaching changes by the PfID nurse educators. A survey with the items from the PfID teacher evaluations can be administered; next to each item (e.g. "encourages active participation"), nurse educators can be asked whether they "contemplated change," "initiated change," or "did not need to make change." The results can then be compared to the teacher evaluation results to determine whether there is an association between the ratings from MSF and the teaching behaviors that were reportedly changed over the course of the year. As the nurse educators gain more teaching experience and students take additional vocational

education courses, MSF using peer, student, self, and program coordinator evaluation should be completed after each vocational education course to determine whether the ratings increase, decrease, or remain the same. Over time, the process of assessment and making changes based on feedback will potentially improve nurse education standards and the accountability of educators. Improved teaching standards and accountability have the potential to strengthen the knowledge base and practice of nursing in the country of Georgia.

The next level of the hierarchy consists of content evaluations which determine the extent to which learners have gained knowledge or skills during the learning experience (Bastable et al., 2011). This form of evaluation focuses on immediate, short-term outcomes, is limited to a specific learning experience, determines whether specifically stated objectives for the experience were met, and occurs immediately following the learning experience (Bastable et al., 2011). An example of a content evaluation was the post-test administered to students following the PfID vocational education program. The multiple choice test was designed to determine whether the learning objectives of the course were met, provide information to PfID about which students "passed" the course, and identify content that needs to be taught better by nurse educators in future courses. Although multiple choice tests are commonly used, many teacher-made multiple-choice examinations only test the immediate recall of knowledge and are neither valid nor reliable (Abruzzese, 1992). Previous research has shown that asking learners what they learned from an experience and how they intend to use the information is more predictive of whether the content will be used in clinical practice than administering a multiple-choice examination (Abruzzese, 1992). Requiring learners

to apply knowledge to cases through objective structured clinical exams and/or oral case studies has been shown to improve clinical competency scores, course satisfaction, and preceptor evaluations of advanced nursing students (Kurz, Mahoney, Martin-Plank, & Lidicker, 2009). Therefore, combining student evaluation approaches using a MSF model may provide a more accurate content evaluation of the knowledge and skills that are acquired by learners than a post-test alone. Teacher evaluation ratings used for process evaluation can be compared to MSF of students for identifying whether effective teaching identified by MSF is associated with better student outcomes.

Outcome evaluations are used to determine the effects of a learning experience by measuring whether long-term change persists (Abruzzese, 1992; Bastable et al., 2011). The scope and indicators used for these evaluations depend on the objectives for the educational activity. Therefore, a pre/post-test can be used. However, if the objective states that the knowledge obtained from the course will be incorporated into clinical practice, then the health professional's knowledge or practice should be evaluated after he or she has returned to his or her original practice setting (Bastable et al., 2011). Outcome evaluations have not yet been planned for the nurses who completed the PfID vocational education program. Although MSF was used in the PfID program for teacher evaluation, it could also be used to complete outcome evaluations of the nurses in practice. This would be similar to the use of MSF for evaluation of physicians, medical trainees, and administrators in Western Europe, Canada, and the United States (Archer, McGraw, & Davies, 2010; Cohen, Farrant, & Taibjee, 2009; Fidler et al., 1999; Garman, Tyler, & Darnall, 2004). Checklists completed by other nurses, doctors, and patients could be used to determine whether learning objectives are being met 1 month before, 1

month after, and 6 months after the vocational education course. Potential indicators could include procedural skills, inpatient hygiene, patient satisfaction, and the roles taken by nurses once they return to their former hospitals or clinics.

The final level of evaluation determines the effect of the educational intervention on broader, longer-term, institutional or public health indicators (Bastable et al., 2011). Thus the purpose of impact evaluation is to answer the question: was the intervention worth the cost (Bastable et al., 2011)? A randomized controlled trial would be ideal for an impact evaluation, however, the expense and time required often preclude the use of this design. Using the PfID vocational education program, an impact evaluation would compare hospitals and clinics with PfID trained nurses to hospitals with nurses who did not complete the vocational education program. Other smaller scale randomization approaches could be employed such as randomizing nurses from specific hospital units to receive vocational education. Relevant health care quality indicators could then be compared between the two groups of nurses (with versus without vocational education) from hospitals or nursing units. Potential indicators include catheter infections, in-hospital mortality, hospital length of stay, or adverse drug events.

At this point, the possible benefits of MSF for downstream public health effects are speculative due to studies on MSF focusing on its use for process, content, and outcome evaluations. However, it's possible that the more complete picture that MSF provides of teacher and learner performance may confirm that a project was well-executed. This makes it more likely that the intervention would lead to positive public health outcomes. If the intervention did not lead to positive public health outcomes, investigators could be confident that the intervention was ineffective and not a result of

poor execution. Ultimately, impact evaluations will be required to determine whether projects that incorporate MSF are more effective than projects that use single sources of evaluation.

Limitations

This study is limited by being a pilot study evaluating a small number of educators (n=15). This limits the statistical power to compare the acceptability of the different sources of evaluation. Due to the specific study population, generalizability (external validity) to other contexts is limited to Georgian nurse educators receiving evaluations for formative decisions. Although the survey for the student, peer, and self evaluations was previously validated in multiple settings, the high ratings suggest bias from the "halo effect" due to the lack of experience with Likert scales (Abruzzese, 1992). Adding survey items requiring negative responses or open-ended questions may have decreased this source of bias.

Conclusions and Recommendations

This study revealed that MSF from self, student, peer, program coordinator, and video evaluations was acceptable to nurse educators in the country of Georgia. Program coordinator evaluations were ranked as most helpful while peer evaluations were perceived as the least helpful. Additional research is required to determine whether educators contemplated or initiated changes in their teaching practice following MSF. Studies on usefulness and accuracy of MSF for evaluations of students on clinical rotations and nurses in clinical practice are also needed. Depending on the results of

future studies, use of MSF could be considered in a variety of health care and global health settings to improve public health and the quality of care delivered.

References

Abruzzese, R. S. (1992). *Nursing Staff Development: Strategies for Success*. New York: Mosby-Year Book.

Aghababian, R. V., Levy, K., Moyer, P., Mottley, L., Ciottone, G., Freitas, R., et al. (1995). Integration of United States emergency medicine concepts into emergency services in the New Independent States. *Ann Emerg Med, 26*(3), 368-375.

Archer, J., McGraw, M., & Davies, H. (2010). Assuring validity of multisource feedback in a national programme. *Arch Dis Child*, *95*(5), 330-335.

Atun, R. (1999). *Georgia: Health Briefing Paper*. London: United Kingdom Institute for Health Sector Development.

Bastable, S., Gramet, P., Jacobs, K., & Sopczyk, D. (2011). *Health Professional as Educator: Principles of Teaching and Learning*. Sudbury, MA: Jones and Bartlett Learning.

Beitz, J. M., & Wieland, D. (2005). Analyzing the teaching effectiveness of clinical nursing faculty of full- and part-time generic BSN, LPN-BSN, and RN-BSN nursing students. *J Prof Nurs, 21*(1), 32-45.

Benor, D. E., & Leviyof, I. (1997). The development of students' perceptions of effective teaching: the ideal, best and poorest clinical teacher in nursing. *J Nurs Educ, 36*(5), 206-211.

Berk, R. (2006). *Thirteen Strategies to Measure College Teaching: A Consumer's Guide to Rating Scale Construction, Assessment, and Decision Making for Faculty, Administrators, and Clinicians*. Sterling, Virginia: Stylus Publishing.

Berk, R. A. (2009). Using the 360 degrees multisource feedback model to evaluate teaching and professionalism. *Med Teach, 31*(12), 1073-1080.

Berk, R. A., Naumann, P. L., & Appling, S. E. (2004). Beyond student ratings: peer observation of classroom and clinical teaching. *Int J Nurs Educ Scholarsh, 1*, Article10.

Blue, A. V., Griffith, C. H., 3rd, Wilson, J., Sloan, D. A., & Schwartz, R. W. (1999). Surgical teaching quality makes a difference. *Am J Surg, 177*(1), 86-89.

Brinkman, W. B., Geraghty, S. R., Lanphear, B. P., Khoury, J. C., Gonzalez del Rey, J. A., Dewitt, T. G., et al. (2007). Effect of multisource feedback on resident communication skills and professionalism: a randomized controlled trial. *Arch Pediatr Adolesc Med, 161*(1), 44-49.

Burford, B., Illing, J., Kergon, C., Morrow, G., & Livingston, M. (2010). User perceptions of multi-source feedback tools for junior doctors. *Med Educ, 44*(2), 165-176.

CIA. (2011). The World Fact Book: Georgia. Retrieved April 10, 2011, from https://www.cia.gov/library/publications/the-world-factbook/geos/gg.html

Cohen, S. N., Farrant, P. B., & Taibjee, S. M. (2009). Assessing the assessments: U.K. dermatology trainees' views of the workplace assessment tools. *Br J Dermatol, 161*(1), 34-39.

Conaboy, K. A., Nugmanova, Z., Yeguebaeva, S., Jaeger, F., & Daugherty, R. M. (2005). Central Asian republics: a case study for medical education reform. *J Contin Educ Health Prof, 25*(1), 52-64.

Drach-Zahavy, A., & Pud, D. (2010). Learning mechanisms to limit medication administration errors. *J Adv Nurs, 66*(4), 794-805.

Driever, M. J., Perfiljeva, G., Callister, L. C., & McGivern, S. (2005). Creating a context for professional dialogue between United States and Russian nurses: design of an international conference. *J Contin Educ Nurs, 36*(4), 168-174.

Fidler, H., Lockyer, J. M., Toews, J., & Violato, C. (1999). Changing physicians' practices: the effect of individual feedback. *Acad Med, 74*(6), 702-714.

Fluit, C. R., Bolhuis, S., Grol, R., Laan, R., & Wensing, M. (2010). Assessing the quality of clinical teachers: a systematic review of content and quality of questionnaires for assessing clinical teachers. *J Gen Intern Med, 25*(12), 1337-1345.

Gamkrelidze, A., Atun, R., & Gotsadze, G. (2002). Health Care System in Transition: Georgia. from http://www.euro.who.int/__data/assets/pdf_file/0017/80702/e75489.pdf

Garman, A. N., Tyler, J. L., & Darnall, J. S. (2004). Development and validation of a 360-degree-feedback instrument for healthcare administrators. *J Healthc Manag, 49*(5), 307-321; discussion 321-302.

Griffith, C. H., 3rd, Georgesen, J. C., & Wilson, J. F. (2000). Six-year documentation of the association between excellent clinical teaching and improved students' examination performances. *Acad Med, 75*(10 Suppl), S62-64.

Immonen, I., Anderssen, N., & Lvova, M. (2008). Project work across borders in the arctic Barents region: practical challenges for project members. *Nurse Educ Today, 28*(7), 841-848.

Ivanov, L. L., & Paganpegara, G. (2003). Public health nursing education in Russia. *J Nurs Educ, 42*(7), 292-295.

Jenkins, D., Brush, K., McGonagle, E., Vartanian, M., & Levy, K. (2000). Expanding the role of nurses in Armenia. *Journal for Nurses in Staff Development, 16*(3), 138-143.

Kalnins, I., Barkauskas, V. H., & Seskevicius, A. (2001). Baccalaureate nursing education development in 2 Baltic countries: Outcomes 10 years after initiation. *Nurs Outlook, 49*(3), 142-147.

Kalnins, I., & Kalnins, Z. P. (1991). Continuing education needs assessment of nursing leaders in the Latvian Republic, 1990. *J Contin Educ Nurs, 22*(4), 167-169.

Knox, J. E., & Mogan, J. (1985). Important clinical teacher behaviours as perceived by university nursing faculty, students and graduates. *J Adv Nurs, 10*(1), 25-30.

Kotzabassaki, S., Panou, M., Dimou, F., Karabagli, A., Koutsopoulou, B., & Ikonomou, U. (1997). Nursing students' and faculty's perceptions of the characteristics of 'best' and 'worst' clinical teachers: a replication study. *J Adv Nurs, 26*(4), 817-824.

Kurz, J. M., Mahoney, K., Martin-Plank, L., & Lidicker, J. (2009). Objective structured clinical examination and advanced practice nursing students. *J Prof Nurs, 25*(3), 186-191.

Leach, D. C., & Philibert, I. (2006). High-quality learning for high-quality health care: getting it right. *Jama, 296*(9), 1132-1134.

Lee, W. S., Cholowski, K., & Williams, A. K. (2002). Nursing students' and clinical educators' perceptions of characteristics of effective clinical educators in an Australian university school of nursing. *J Adv Nurs, 39*(5), 412-420.

Levine, M. A., & Perpetua, E. M. (2006). International immersion programs in baccalaureate nursing education: professor and student perspectives. *J Cult Divers, 13*(1), 20-26.

Li, M. K. (1997). Perceptions of effective clinical teaching behaviours in a hospital-based nurse training programme. *J Adv Nurs, 26*(6), 1252-1261.

Lockyer, J. (2003). Multisource feedback in the assessment of physician competencies. *J Contin Educ Health Prof, 23*(1), 4-12.

Lockyer, J., Violato, C., & Fidler, H. (2003). Likelihood of change: a study assessing surgeon use of multisource feedback data. *Teach Learn Med, 15*(3), 168-174.

Miller, A., & Archer, J. (2010). Impact of workplace based assessment on doctors' education and performance: a systematic review. *Bmj, 341*, c5064.

Mogan, J., & Knox, J. (1983). Students' perceptions of clinical teaching. *Nurs Pap, 15*(3), 4-20.

Mogan, J., & Knox, J. E. (1987). Characteristics of 'best' and 'worst' clinical teachers as perceived by university nursing faculty and students. *J Adv Nurs, 12*(3), 331-337.

Mogan, J., & Warbinek, E. (1994). Teaching behaviours of clinical instructors: an audit instrument. *Journal of Advanced Nursing, 20*(1), 160-166.

Nehring, V. (1990). Nursing Clinical Teacher Effectiveness Inventory: a replication study of the characteristics of 'best' and 'worst' clinical teachers as perceived by nursing faculty and students. *J Adv Nurs, 15*(8), 934-940.

Nelson, A. I. (2005). A national survey of electroconvulsive therapy use in the Russian Federation. *J Ect, 21*(3), 151-157.

Nishiyama, M., Wold, J. L., & Partskhladze, N. (2008). Building competencies for nurse administrators in the Republic of Georgia. *Int Nurs Rev, 55*(2), 179-186.

Raingruber, B., & Bowles, K. (2000). Developing student evaluation instruments to measure instructor effectiveness. *Nurse Educ, 25*(2), 65-69.

Roop, S. A., & Pangaro, L. (2001). Effect of clinical teaching on student performance during a medicine clerkship. *Am J Med, 110*(3), 205-209.

Sanders, J. (2007). A family medicine training program in the Republic of Georgia: incorporating a model of chronic disease management. *J Am Board Fam Med, 20*(6), 557-564.

Sargeant, J., Mann, K., & Ferrier, S. (2005). Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness. *Med Educ, 39*(5), 497-504.

Sargeant, J. M., Mann, K. V., Ferrier, S. N., Langille, D. B., Muirhead, P. D., Hayes, V. M., et al. (2003). Responses of rural family physicians and their colleague and coworker raters to a multi-source feedback process: a pilot study. *Acad Med, 78*(10 Suppl), S42-44.

Scott, J. G., Sochalski, J., & Aiken, L. (1999). Review of magnet hospital research: findings and implications for professional nursing practice. *J Nurs Adm, 29*(1), 9-19.

Seldin, P. (Ed.). (1999). *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions*. Bolton, MA: Anker Publishing.

Skarbinski, J., Walker, H. K., Baker, L. C., Kobaladze, A., Kirtava, Z., & Raffin, T. A. (2002). The burden of out-of-pocket payments for health care in Tbilisi, Republic of Georgia. *Jama, 287*(8), 1043-1049.

Thompson, C., & Sheckley, B. G. (1997). Differences in classroom teaching preferences between traditional and adult BSN students. *J Nurs Educ, 36*(4), 163-170.

UN. (2009). *Human Development Report: Georgia*. Tbilisi, Georgia: United Nations (UN) Development Programme.

Violato, C., Lockyer, J., & Fidler, H. (2003). Multisource feedback: a method of assessing surgical practice. *Bmj, 326*(7388), 546-548.

Violato, C., Lockyer, J. M., & Fidler, H. (2008). Changes in performance: a 5-year longitudinal study of participants in a multi-source feedback programme. *Med Educ, 42*(10), 1007-1013.

Walker, K., & Wold, J. (2009). Emory University School of Medicine: Proposal for Georgia Nursing Education in Tbilisi, Georgia: United States Agency for International Development/Caucasus.

WHO. (1999). *World Health Organization (WHO) Regional Office for Europe health for all database*. Retrieved April 10, 2011. from http://data.euro.who.int/hfadb/.

Wold, J. L., Walker, K., & Partskhladze, N. (2007). Establishing a medical and nursing education program in the Republic of Georgia. In C. Holtz (Ed.), *Global Health*. Sudbury, Massachusetts: Jones and Bartlett.

Wong, J. G., & Agisheva, K. (2007). Developing teaching skills for medical educators in Russia: a cross-cultural faculty development project. *Med Educ, 41*(3), 318-324.

Zimmerman, L., & Westfall, J. (1988). The development and validation of a scale measuring effective clinical teaching behaviors. *J Nurs Educ, 27*(6), 274-277.

Zimmerman, T., Putland, L. C., & Hughes, J. R. (1997). Russian and local nurse educators exchange ideas. *Nurs Spectr (Washington D.C.), 7*(14), 9.

EMORY UNIVERSITY

Institutional Review Board

TO: Christopher DeStephano
Principal Investigator

DATE: April 9, 2010

RE: **Notification of Submission Determination: No IRB Review Required**
IRB00040945
A Formative Evaluation of Nurse Educators Teaching Effectiveness in the Republic of Georgia

The above-referenced study has been vetted by the Institutional Review Board (IRB), and it was determined that it does not require IRB review because it does not meet the definition of "Research" or the definition of "Clinical Investigation" under applicable federal regulations. Accordingly, IRB review is not required.

45 CFR Section 46.102(d) defines "Research" as follows:

Research means a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge. Activities which meet this definition constitute research for purposes of this policy, whether or not they are conducted or supported under a program which is considered research for other purposes. Based on the information included in this submission, the purpose of this formative evaluation is to improve the quality of future training of nurse educators provided by Partners for International Development (PfID) and inform decisions about use of formative feedback instruments. The results of this evaluation will be reported directly to PfID and will not be generalizable outside of their nurse educator training program.

The IRB has determined that this study does not constitute "Research" under the foregoing definition.

In addition, the IRB has determined that the study is not a "Clinical Investigation" under applicable Food & Drug Administration regulations because it does not involve a test article and does not otherwise meet the requirements of the definition of "Clinical Investigation" as set forth in 21 CFR Section 50.3(c).

Please note that any changes to the protocol could conceivably alter the status of this research under the federal regulations cited above. Accordingly, any substantive changes in the protocol should be presented to the IRB for consideration prior to their implementation in the research.

Sincerely,

Carol Corkran, MPH
Senior Research Protocol Analyst
This letter has been digitally signed