

Global Health and Graduate Medical Education: A Systematic Review of the Literature

Corey B. Bills, MD, MPH
James Ahn, MD

ABSTRACT

Background Global health (GH) interest is increasing in graduate medical education (GME). The popularity of the GH topic has created growth in the GME literature.

Objective The authors aim to provide a systematic review of published approaches to GH in GME.

Methods We searched PubMed using variable keywords to identify articles with abstracts published between January 1975 and January 2015 focusing on GME approaches to GH. Articles meeting inclusion criteria were evaluated for content by authors to ensure relevance. Methodological quality was assessed using the Medical Education Research Study Quality Instrument (MERSQI), which has demonstrated reliability and validity evidence.

Results Overall, 69 articles met initial inclusion criteria. Articles represented research and curricula from a number of specialties and a range of institutions. Many studies reported data from a single institution, lacked randomization and/or evidence of clinical benefit, and had poor reliability and validity evidence. The mean MERSQI score among 42 quantitative articles was 8.87 (2.79).

Conclusions There is significant heterogeneity in GH curricula in GME, with no single strategy for teaching GH to graduate medical learners. The quality of literature is marginal, and the body of work overall does not facilitate assessment of educational or clinical benefit of GH experiences. Improved methods of curriculum evaluation and enhanced publication guidelines would have a positive impact on the quality of research in this area.

Introduction

Interest in global health (GH) education is a relatively recent phenomenon in academic medicine. Over the last 2 decades interest in GH in graduate medical education (GME) has steadily grown, with many residencies currently offering curricula and experiences in GH.^{1,2} The popularity of GH, especially in disciplines such as emergency medicine, family medicine, pediatrics, and internal medicine, has led to similar growth in the literature on GH education in GME settings.³ While there is no single, agreed-on definition of global health, our understanding rests on a definition provided by Koplan et al,⁴ that GH is “an area for study, research, and practice that places a priority on improving health and achieving equity in health for all people worldwide.”

While there is mounting agreement about the merits of GH, there is little consensus about how to educate the next generation of residents and fellows interested in GH.⁵ Preparing an effective workforce to meet worldwide clinical and educational priorities

requires research into GH-related educational needs in GME.⁶

Academic medical centers have aimed to improve educational experiences by standardizing GH education through establishing curricula and formal competency requirements, including the development of GH electives, tracks, and didactic study.^{7,8} Fellowship training in GH is also available, especially in pediatrics, internal medicine, and emergency medicine.⁹

Additionally, authors have proposed distinct educational strategies¹⁰ and standardized medical school-based curriculum guides.¹¹ At present, however, course offerings in GH tend to vary by academic centers. There are few previously published reviews synthesizing the literature on the overall quality and effectiveness of specific educational approaches in GH education in GME.

Given the lack of existing summaries, we aim to provide a systematic review of published approaches to GH in the GME literature. Our aim is to provide insight on the current state of the literature on GH education, synthesize current trends, and to formulate recommendations on the future development of GH education through analysis of the variety of developed curricula, teaching approaches, and evaluation methods.

DOI: <http://dx.doi.org/10.4300/JGME-D-15-00774.1>

Editor's Note: The online version of this article contains a table of interclass correlation coefficient between 2 reviewers for each of 10 domains in MERSQI scores, and the list of articles meeting inclusion criteria for the review.

Methods

We planned, executed, and reported this systematic review in adherence to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses).¹²

Literature Search

We conducted a systematic review of the literature via PubMed using variable terms to identify articles with abstracts published between January 1975 and January 2015 focused on GH and GME. Key search terms included variable iterations of the following keywords: medical, education, graduate, international, global, health, training, curriculum, residency, course, standards, electives, and competencies.

Eligibility Criteria

Studies were limited to those focused on GH education in GME, specifically focused on clinicians in training including residents and fellows, and excluded studies focused on medical student learners and learners in other allied health fields (such as pharmacy, nursing, or public health).

Studies were additionally required to meet the following inclusion criteria: English language; inclusion of an abstract; representative of original research or programming (nonabstract opinion pieces or letters to the editor were excluded); program description had to be part of a GME (residency or fellowship) program centered in the United States or Canada; and majority of study participants had to be enrolled in a GME program (residency or fellowship). We considered all articles regardless of methodological approach, including both qualitative and/or quantitative methodologies.

Title and Abstract Review

Both study authors reviewed the titles and abstracts of articles. If the abstract was insufficient to provide decision for inclusion, then the full text was reviewed independently by each of the 2 authors. In the event that the 2 authors disagreed on inclusion, the full text was re-reviewed, the article discussed, and consensus agreement was reached based on the predetermined inclusion or exclusion criteria.

Study Review/Data Extraction and Analysis

We developed a data extraction form that would allow for easy analysis based on the Best Evidence Medical Education Collaboration.¹³

The form was initially piloted in a review of 16 randomly selected studies (each reviewed by both authors) to confer ease of relevance and consistency in

filling out the form. Data abstracted from the studies included in this review consisted of 4 domains: (1) journal (name and impact factor); (2) study characteristics (publication year, study design, number, name, and location of participating institution[s]); (3) participants (number, level of training, profession, and medical specialty); and (4) outcome measures.

Methodological quality among articles employing a quantitative analysis was assessed using the Medical Education Research Study Quality Instrument (MERSQI), which has demonstrated relevance and validity.¹⁴ The MERSQI score was originally designed to evaluate methodological quality of medical education research. It consists of 10 items, reflecting 6 domains of study quality: study design, sampling, type of data, validity, data analysis, and outcomes.¹⁴ MERSQI scores were calculated for all quantitative studies included according to the methods described by the authors of the instrument. Each article meeting inclusion criteria was evaluated for content and scored by both authors to ensure relevance with possible summed scores ranging from 4.5 to 18.

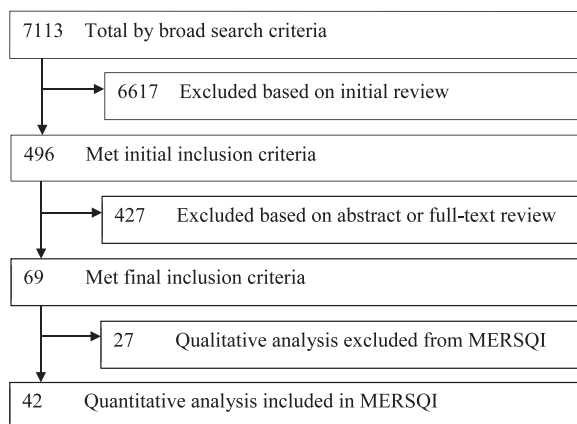
Educational outcomes were also classified using Kirkpatrick's hierarchy.¹⁵ Kirkpatrick level (KL) 0 provides no assessment of impact; KL-1 assesses learner reaction; KL-2 assesses attitude and knowledge or skills; KL-3 assesses changes in behavior; and KL-4 assesses changes in patient or systems-based outcomes.

Data Analysis

Interrater reliability was determined for the elements of the MERSQI scores by calculating an interrater correlation coefficient. Disagreements between raters in assigning quantitative scores were resolved by discussion (only after the interrater reliability was determined), re-review of the article, and achievement of a final consensus rating. A consensus was derived in all field types for each of the articles. We used the consensus mean and median MERSQI scores with standard deviations/interquartile range to describe the overall quality of included studies. All analyses were performed in Stata version 13 (StataCorp LP, College Station, TX).

Results

Initial broad search criteria led to 7113 unique articles (FIGURE). This number was then refined to 496 based on title and abstract analysis alone. Further full article review led to 69 articles meeting the inclusion criteria, and 42 of them meeting criteria for MERSQI analysis (FIGURE; article list provided as online supplemental material).

**FIGURE**

Selection Process Used in Systematic Review of Global Health Education in Graduate Medical Education Articles (1975–2014)

Abbreviation: MERSQI, Medical Education Research Study Quality Instrument.

Original work was published in a large variety of journals, including general GME and clinical specialty publications. The average impact factor of the journals in which the articles were published was 2.44 (SD = 1.3). The majority of scholarship was recent, with 97% (n = 67) published after 2000, and more than 70% (n = 50) published since 2010 (TABLE 1). Institutional representation was variable, with the majority of articles arising from university-affiliated academic medical centers (n = 65, 94%). Research meeting inclusion criteria represented 13 different medical specialties and subspecialties. Among studies that reported the number of participants (n = 45), the median (interquartile range) number was 26 (13–66).

A total of 61 articles reported on curriculum development and evaluation with the majority focused on clinical electives (n = 36). Further educational approaches included health tracks (n = 9), fellowships (n = 6), didactic lectures (n = 4), and web- and simulation-based training (each n = 1).

Methodological Quality

Of the 69 articles, 27 articles were qualitative studies, and 42 used quantitative methods. Of the 42 studies that used quantitative methodology, the most common study designs were single group (n = 30, 71%), with comparison groups used in a small number of studies (n = 9, 21%; TABLE 2). No randomized controlled trials were represented in the dataset.

Many studies lacked multiple institution sampling (n = 31, 74%), and measures beyond self-assessment were used in a minority of publications (n = 13, 31%). Additionally, there was little reported validity evidence for instruments used in studies, with 64% of

included studies (n = 27) having a MERSQI summed validity score of 1 or less.

Analysis for the majority of articles was appropriate, given the study type (n = 38, 91%), but the sophistication of analysis was often limited to descriptive reporting only. Evidence supporting clinical benefit was also lacking, as no studies reported on patient- or health-related outcomes. The majority of publications reported on the following subjective outcomes: satisfaction, attitudes, perception, opinions, and general facts (n = 29, 69%). A minority assessed knowledge/skills (n = 7, 17%) and behaviors (n = 6, 14%).

The mean consensus MERSQI score was 8.87 ± 2.79 ($\pm$ SD) out of a possible score of 18. Although overall study quality was found to be poor, there were 4 articles with MERSQI scores above 13 that provided evidence of good study quality.^{16–19} The 4 articles were from 4 separate institutions, representing 4 distinct medical specialties, with the participant number ranging from 19 to 298. The articles were similar in that all reported objective data measures and provided an analysis that was beyond just descriptive and was appropriate for the data type.

For the scoring on the MERSQI, interclass correlation coefficient was excellent, with perfect agreement (interclass correlation coefficient = 1.0 [SD = 0.00]) for 5 items in the 10-item MERSQI scale. Items with limited disagreement included institutions (SD = 0.005), response rate (SD = 0.012), validity content (SD = 0.012), appropriateness of analysis (SD = 0.006), and sophistication of analysis (SD = 0.012). There were no items with some disagreement.

When evaluating Kirkpatrick's hierarchy of outcomes in studies that reported on an educational intervention, 81.2% (56 of 69) reported Level 1 outcomes ("learner satisfaction"), and only 10.1% (7 of 69) and 8.7% (6 of 69) of studies reported Level 2 ("change in attitude, knowledge, or skills") and Level 3 ("change in behaviors") outcomes, respectively. No studies reported Level 4 outcomes ("change in results").

Qualitative studies offered similar findings to the quantitative studies, with individuals perceiving learning as generally positive, and analysis limited to perceptions and attitudes of the benefits of GH curricula and experiences.

Discussion

The sizable increase of GH education publications after 2010 correlates with growing interest from residents and medical students in the United States and elsewhere.¹ This rapid growth may have

TABLE 1

Baseline Characteristics of Articles in a Systematic Review of Global Health Education in Graduate Medical Education (January 1975–January 2015)

Characteristic	No. (Average), N = 69	% (SD)
Year		
1997	1	1.4
1999	1	1.4
2002	1	1.4
2005	1	1.4
2006	2	2.9
2007	3	4.3
2008	8	11.6
2009	2	2.9
2010	5	7.2
2011	12	17.4
2012	9	13.0
2013	12	17.4
2014	10	14.5
2015	2	2.9
Journal		
<i>Academic Medicine</i>	6	8.7
<i>Academic Pediatrics</i>	2	2.9
<i>Academic Psychiatry</i>	2	2.9
<i>American Journal of Tropical Medicine and Hygiene</i>	7	10.1
<i>Archives of Surgery</i>	2	2.9
<i>BMC Medical Education</i>	2	2.9
<i>Family Medicine</i>	2	2.9
<i>Journal of the American College of Radiology</i>	2	2.9
<i>Journal of Graduate Medical Education</i>	3	4.3
<i>Journal of Surgical Education</i>	5	7.2
<i>Medical Education</i>	3	4.3
<i>Rhode Island Medical Journal</i>	2	2.9
<i>Teaching and Learning in Medicine</i>	3	4.3
Other (1 each only)	28	40.6
Institution		
University of California, San Francisco	5	7.2
University of Minnesota	4	5.8
Mt. Sinai	5	7.2
University of Washington	4	5.8
University of Wisconsin	3	4.3
Vanderbilt University	4	5.8
Yale University	3	4.3
Other (single institution)	41	59.4

TABLE 1 (continued)

Characteristic	No. (Average), N = 69	% (SD)
Country		
Canada	6	8.7
United States	61	88.4
Both	2	2.9
Program site (country)		
Africa ^a	16	23.2
Central and South America	7	10.1
Middle East	1	1.4
Multiple sites	18	26.1
Not applicable ^b	27	39.1
Specialty		
Emergency medicine	4	5.8
Family medicine	8	11.6
General and subspecialty surgery	17	24.6
Graduate medical education (cross discipline)	17	24.6
Internal medicine	7	10.1
Neurology	1	1.4
Pediatrics	10	14.5
Psychiatry	3	4.3
Radiology	2	2.9
Type of study		
Curriculum development	24	34.8
Curriculum development and evaluation	3	4.4
Curriculum evaluation	34	49.3
Global health center development	2	2.9
Institutional collaboration	6	8.7
Curriculum subtype ^c		
Clinical and health elective	36	59.0
Didactic	4	6.6
Fellowship	6	9.8
Other curriculum	6	9.8
Track	9	14.8

^a Single individual country with the most programmatic work was Uganda (n = 6).

^b No country established as training site (ie, global health curriculum delivered at home institution), or the country was not included in the study report.

^c Out of 61 studies reporting of curriculum evaluation, development, or both.

contributed to heterogeneity in curricula, with no single “accepted” strategy for teaching GH in GME. The quality of analysis of the existing educational platforms is less than desirable, with particular weakness in reporting on the validity evidence of evaluation instruments. These findings corroborate other medical education studies that have found

TABLE 2

MERSQI Scores for Each of 10 Domains in Systematic Review of Global Health Education in Graduate Medical Education Articles (January 1975–January 2015)

Domain	Item	Item Score	Max Score	Average	Manuscripts (N = 42)	
					SD	n (%)
Total			18	8.87	2.79	
Study design						
1. Study design			3	1.25	0.42	
	Single group cross-sectional	1				30 (71.4)
	Single group pretest and posttest	1.5				3 (7.1)
	Nonrandomized, 2 group	2				9 (21.4)
	Randomized controlled experiment	3				0 (0)
Sampling			3	1.61	0.75	
2. Institutions			1.5	0.71	0.38	
	Single institution	0.5				31 (73.8)
	2 institutions	1				4 (9.5)
	More than 2 institutions	1.5				7 (16.7)
3. Response rate			1.5	0.89	0.62	
	Not applicable	0				11 (26.2)
	Response rate < 50% or not reported	0.5				4 (9.5)
	Response rate 50%–74%	1				10 (23.8)
	Response rate ≥ 75%	1.5				17 (40.5)
Type of data						
4. Type of data			3	1.67	0.95	
	Assessment by study subject	1				28 (66.7)
	Objective measurement	3				14 (33.3)
Validity of evaluation instruments' scores			3	1.00	1.10	
5. Internal structure			1	0.33	0.48	
	Not reported/NA	0				28 (66.7)
	Reported	1				14 (33.3)
6. Content			1	0.48	0.51	
	Not reported/NA	0				22 (52.4)
	Reported	1				20 (47.6)
7. Relationships to other variables			1	0.19	0.40	
	Not reported/NA	0				34 (81.0)
	Reported	1				8 (19.0)
Data analysis			3	2.12	0.55	
8. Appropriateness of analysis				0.90	0.30	
	Data analysis inappropriate	0				4 (9.5)
	Data analysis appropriate	1				38 (90.5)
9. Sophistication of analysis				1.21	0.42	
	Descriptive analysis only	1				33 (78.6)
	Beyond descriptive analysis	2				9 (21.4)
Outcome						
10. Outcome			3	1.23	0.37	
	Satisfaction, attitudes, perceptions	1				29 (69.5)
	Knowledge, skills	1.5				7 (16.7)
	Behaviors	2				6 (14.3)
	Patient/health care outcome	3				0 (0)

Abbreviations: MERSQI, Medical Education Research Study Quality Instrument; Max, maximum; NA, not available.

substandard methodology and validity reporting.^{14,20,21}

We sought to add to the conceptualization of the quality of the GH literature by reporting the impact factor of the journals in which articles were published. For the articles in our review, the average impact factor of the journals in which they were published was higher than the median impact factor for education journals as whole (0.902).²²

Deficiencies in medical education research are widely acknowledged, and the GH literature does not differ. Global health-related articles have a lower mean MERSQI score than previously studied medical education manuscripts (8.9 versus 10.7).²³ Our study found that the majority of studies were single institution, reducing generalizability, and the lack of comparison group in many of these studies weakens the data.

When assessed via Kirkpatrick's hierarchy of outcomes, the majority of studies reported Level 1 outcomes (learner satisfaction only),¹⁵ and none of the representative articles reported on patient- or health-related outcomes. Moreover, the conclusions that study authors offered regarding the success or weakness of educational opportunities rested solely on learners' perceptions of their experience, rather than other, more robust outcomes.

We found a large and diverse number of educational approaches to GME global health. Though models of didactic learning and experiential learning appear to predominate, similar to prior research, there appears to be no agreed on consensus regarding what constitutes a successful technique for teaching GH to residents and fellows. The highest-quality articles reinforce the relationship between learning, often using core competencies and practice. The current crop of GME GH learners hail from the millennial generation, and the current model of didactic learning may need to change. The educational approach moving forward should share a common framework, including core competencies that incorporate generational preferences for e-learning technologies, and techniques such as "flipping the classroom," in which the typical homework and lecture elements are reversed.^{24,25}

Recent proposals have reinforced the responsibility of sponsoring organizations in GH training experiences, with a need to focus on education and local needs, and the priorities of host institutions.²⁶ We have argued a similar position elsewhere, and promote efforts to focus on local needs, as well as evaluate local outcomes in the analysis of approaches to GH educational research.²⁷ The linking of patient outcomes to trainee performance has been proposed previously.²⁸ Examples of local outcomes include data on changes in number and type of patients seen,

effectiveness of resident participation in clinical care, and changes in mortality and morbidity related to local demographics.

Our analysis of research on the current status of GH and medical education has several limitations, beginning with only including English-language articles from North American institutions indexed in PubMed. Second, while the authors were initially blinded to each other's scoring of the articles, they were not blinded to author(s), title, or journal name of the articles. Third, the authors used the MERSQI to assess study quality, and this metric does not encompass all aspects of study quality or allow for comment on an article's conceptual framework or the importance of the research question. Fourth, we reported on aspects of the journals in which articles were published, including the impact factor. While a general metric of the relative importance of a journal, impact factor does not necessarily serve as a direct indication of a given article's importance in its field.

Conclusion

In exploring the GH medical education literature, we identified heterogeneity in published studies, and 3 areas lacking in rigor and methodological quality. First, few of the studies included a comparison group. Second, studies need to be explicit about report validity evidence of evaluation instruments. Third, there is a need for more robust methods of GH curriculum evaluation, with a focus on levels 3 and 4 of Kirkpatrick's hierarchy of educational evaluation outcomes.

References

1. AAMC. *Medical School Graduation Questionnaire*. Washington, DC: Association of American Medical Colleges; 2011.
2. AAMC. *Medical School Graduation Questionnaire*. Washington, DC: Association of American Medical Colleges; 2001.
3. Kerry VB, Walensky RP, Tsai AC, et al. US medical specialty global health training and the global burden of disease. *J Glob Health*. 2013;3(2):020406.
4. Koplan JP, Bond TC, Merson MH, et al. Towards a common definition of global health. *Lancet*. 2009;373(9679):1993–1995.
5. Panosian C, Coates T. The new medical "missionaries": grooming the next generation of global health workers. *N Engl J Med*. 2006;354(17):1771–1773.
6. Beaglehole R, Bonita R, Horton R, et al. Public health in the new era: improving health through collective action. *Lancet*. 2004;363(9426):2084–2086.

7. Drain PK, Holmes KK, Skeff KM, et al. Global health training and international clinical rotations during residency: current status, needs, and opportunities. *Acad Med*. 2009;84(3):320–325.
8. Battat R, Seidman G, Chadi N, et al. Global health competencies and approaches in medical education: a literature review. *BMC Med Educ*. 2010;10:94.
9. Nelson B, Izadnegahdar R, Hall L, et al. Global health fellowships: a national, cross-disciplinary survey of us training opportunities. *J Grad Med Educ*. 2012;4(2):184–189.
10. Dowell J, Merrylees N. Electives: isn't it time for a change? *Med Educ*. 2009;43(2):121–126.
11. Brewer TF, Saba N, Clair V. From boutique to basic: a call for standardized medical education in global health. *Med Educ*. 2009;43(10):930–933.
12. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med*. 2009;151(4):264–269, W64.
13. Best evidence medical and health professional education. <http://www.bemecollaboration.org/>. Accessed June 3, 2016.
14. Reed DA, Cook DA, Beckman TJ, et al. Association between funding and quality of published medical education research. *JAMA*. 2007;298(9):1002–1009.
15. Kirkpatrick D. Great ideas revisited. Techniques for evaluating training programs: revisiting Kirkpatrick's four-level model. *Train Dev*. 1996;50(1):54–59.
16. Bazemore AW, Henein M, Goldenhar LM, et al. The effect of offering international health training opportunities on family medicine residency recruiting. *Fam Med*. 2007;39(4):255–260.
17. Bjorklund AB, Cook BA, Hendel-Paterson BR, et al. Impact of global health residency training on medical knowledge of immigrant health. *Am J Trop Med Hyg*. 2011;85(3):405–408.
18. Oettgen B, Harahsheh A, Suresh S, et al. Evaluation of a global health training program for pediatric residents. *Clin Pediatr (Phila)*. 2008;47(8):784–790.
19. Ton TG, Gladding SP, Zunt JR, et al. The development and implementation of a competency-based curriculum for training in global health research. *Am J Trop Med Hyg*. 2015;92(1):163–171.
20. Kerry VB, Ndung'u T, Walensky RP, et al. Managing the demand for global health education. *PLoS Med*. 2011;8(11):e1001118.
21. Reed DA, Beckman TJ, Wright SM. An assessment of the methodologic quality of medical education research studies published in The American Journal of Surgery. *Am J Surg*. 2009;198(3):442–444.
22. Azer SA, Holen A, Wilson I, et al. Impact factor of medical education journals and recently developed indices: can any of them support academic promotion criteria? *J Postgrad Med*. 2016;62(1):32–39.
23. Reed D, Beckman T, Wright SM, et al. Predictive validity evidence for medical education research study quality instrument scores: a quality of submission to JGIM's Medical Education Special Issue. *J Gen Intern Med*. 2008;23(7):903–907.
24. Roberts DH, Newman LR, Schwartzstein RM. Twelve tips for facilitating Millennial's learning. *Med Teach*. 2012;34(4):274–278.
25. Moffett J. Twelve tips for “flipping” the classroom. *Med Teach*. 2015;37(4):331–336.
26. Crump JA, Sugarman J; Working Group on Ethics Guidelines for Global Health Training (WEIGHT). Ethics and best practice guidelines for training experiences in global health. *Am J Trop Med Hyg*. 2010;83(6):1178–1182.
27. Bills CB, Ahn J. Global health education as a translational science in graduate medical education. *J Grad Med Educ*. 2015;7(2):166–168.
28. Haan CK, Edwards FH, Poole B, et al. A model to begin to use clinical outcomes in medical education. *Acad Med*. 2008;83(6):574–580.



Corey B. Bills, MD, MPH, is Assistant Clinical Professor, Department of Emergency Medicine, and Affiliate Faculty, Global Health Sciences, University of California, San Francisco; and **James Ahn, MD**, is Assistant Professor and Associate Program Director, Division of Emergency Medicine, University of Chicago School of Medicine.

Funding: The authors report no external funding source for this study.

Conflict of interest: The authors declare they have no competing interests.

Corresponding author: Corey B. Bills, MD, MPH, University of California San Francisco–San Francisco General Hospital, Department of Emergency Medicine, 1001 Potrero Avenue, San Francisco, CA 94110, 917.414.8899, cbills@gmail.com

Received December 9, 2015; revision received March 30, 2016; accepted May 4, 2016.