

RESEARCH ARTICLE

Open Access

# Towards a conceptual framework demonstrating the effectiveness of audiovisual patient descriptions (patient video cases): a review of the current literature

Damian Roland<sup>1\*</sup>, Tim Coats<sup>1</sup> and David Matheson<sup>2</sup>

## Abstract

**Background:** Technological advances have enabled the widespread use of video cases via web-streaming and online download as an educational medium. The use of real subjects to demonstrate acute pathology should aid the education of health care professionals. However, the methodology by which this effect may be tested is not clear.

**Methods:** We undertook a literature review of major databases, found relevant articles relevant to using patient video cases as educational interventions, extracted the methodologies used and assessed these methods for internal and construct validity.

**Results:** A review of 2532 abstracts revealed 23 studies meeting the inclusion criteria and a final review of 18 of relevance. Medical students were the most commonly studied group (10 articles) with a spread of learner satisfaction, knowledge and behaviour tested. Only two of the studies fulfilled defined criteria on achieving internal and construct validity. The heterogeneity of articles meant it was not possible to perform any meta-analysis.

**Conclusions:** Previous studies have not well classified which facet of training or educational outcome the study is aiming to explore and had poor internal and construct validity. Future research should aim to validate a particular outcome measure, preferably by reproducing previous work rather than adopting new methods. In particular cognitive processing enhancement, demonstrated in a number of the medical student studies, should be tested at a postgraduate level.

**Keywords:** Patient video clips, Methodology, Evaluation, Educational intervention

## Background

There are a plethora of educational programmes and implementation strategies aimed at improving the quality of care delivered by health care professionals. A number of these are delivered via information technology systems with the use of video as an educational medium well established [1-3]. A new educational tool, that has become possible through multimedia advances in the last decade, is the audio-visual demonstration of signs

and symptoms in patients, referred to as Patient Video Cases or PVCs [4]. They are easily displayed via online platforms, are widely used, and have been endorsed by the National Patient Safety Agency [5] as an example of good practice. However there is little academic study of their effectiveness. Given the financial pressures affecting all health care agencies, it is important to know if these resource intensive e-learning strategies give demonstrable benefit to patients or health care professionals.

Theoretical constructs exist to evaluate interventions designed to improve clinical performance, but no single approach is followed, due to the wide range of individual and organisational factors that affect the outcomes before, during and after the intervention [6]. Kirkpatrick's

\* Correspondence: dr98@le.ac.uk

<sup>1</sup>Emergency Medicine Academic Group, Emergency Department secretaries c/o Elizabeth Cadman-Moore, Leicester Royal Infirmary, Leicester LE1 5WW, UK

Full list of author information is available at the end of the article

training evaluation is defined by four distinct levels of outcome to be approached in a stepwise fashion [7]. The four key domains of the Kirkpatrick model are learner satisfaction, learner knowledge, learner behaviour change and organisational change. Although others have argued contextual factors not classified under these domains may be significant [6], the Kirkpatrick model still remains a valid methodology with systematic reviews using the process to examine training effectiveness [8]. A healthcare relevant modification of the Kirkpatrick model has been used in a study of inter-professional education in health and social care [9]. When using the Kirkpatrick model, or other relevant frameworks for assessing an educational or training intervention, the outcome measures and the methodology by which they are obtained must be valid. The concepts of internal and construct validity are classifications with direct relevance to outcome measures and are components of methodological quality used by the Campbell Collaboration [10,11].

- i. Internal Validity is the extent to which the intervention can reliably be ascribed to have affected the change
- ii. Construct Validity relates to the association between the concept being investigated and the measures used to test it i.e. does the data collected accurately reflect the outcome measure chosen?

Other forms of validity exist but are not directly relevant to the quality of the outcome measures chosen, for example good external validity would imply that using PVCs could be beneficial in different populations, but would not give any information if the initial outcome measure was fit for purpose.

The aim of this work is to answer the question "What is the validity and quality of outcome measures that have been used to evaluate interventions based on PVCs?". This literature review will be used to identify which outcome measures are most valid in the assessment of the clinical effectiveness of an intervention based on PVCs. It will also help identify areas where more methodological research is needed to enable future studies to demonstrate high internal and construct validity.

## Methods

This review was performed over three stages, the first stage collating relevant literature followed by individual study quality appraisal in stage two with a summation of the overall validity of the studies.

### Stage one

Stage one identified literature relevant to the use of PVCs in health care settings. The definition of Health Care Settings used was; 'any location or environment where

students or graduates are practising or learning medicine.' The definition of a PVC was; 'any pre-recorded or live video footage of a patient used for the purposes of demonstrating a sign or symptom'. It did not include footage recorded for the purposes of educating other patients or families. Inclusion criteria were:

- i. Humans
- ii. The study described the use of PVCs in a training, educational (undergraduate or postgraduate), implementation capacity or environment.

As PVCs relate to demonstration of signs and symptoms in patients, studies using video to demonstrate verbal communication, non-lexical utterances or solely history taking between a patient and doctor or patient and patient were excluded as were non-English language papers which could not be translated. The full literature search was developed in conjunction with a senior NHS Librarian and is available on request. The following general search terms were used (Video\* OR Video record\* OR video clip OR digital\* record\* OR analogue recording OR *patient video clip*) and (Educat\* OR Train\* OR learn\* OR teach\* OR inservice training). The following databases were searched: Medline, British Nursing Index (BNI), EMBASE, Health Management Information Consortium (HMIC), CINAHL, NIHR Health Technology Assessment Programme (HTA), Database of Abstracts of Reviews of Effects (DARE), Scopus, The Cochrane Library and the Education Resources Information Centre (ERIC). Internet search engines and NHS evidence were used to identify publications or articles related to the search terms. The search strategy was not limited to any particular research methodology used in the articles. The last search performed was 27<sup>th</sup> July 2012 by the principal author. In all phases of the study any uncertainty as to classification or indexing of information was discussed with the collaborating authors.

Articles with a relevant abstract (any detail relating to the recording and utilisation of video clips of patients) had a complete paper review (as did any abstracts in which there was uncertainty about inclusion potential). Information on aim, health care user, educational purpose, modified Kirkpatrick training level domain, type of study, outcome measure and conclusions was extracted from each paper as shown in Table 1. The Educational purpose was subdivided into three categories:

### Stage two

To enable objective review of articles to determine the aspects of validity under study the following domains were used which represent features reducing the internal validity of studies. They have been amended from the list described by Farrington [12]. This work was chosen as it

**Table 1 Studies by Healthcare professional grouping**

Health care professional	Number of studies pertaining to that group
Undergraduates	11
Basic Postgraduate Training	3
Specialist Postgraduate Training	1
Undergraduate professionals allied to Medicine	2
Professional allied to Medicine	1
Trained Doctor Continuing Professional Development	2
Veterinary Students	1
<b>Total</b>	<b>21</b>

is based on Cook and Cambell's original work on methodological quality. Although other methodologies of analysis are available this is a widely used and accepted process which allows for an objective process to be applied.

1. Selection: Does the outcome measure allow for control between groups?
2. History: Does the outcome measure allow for the effects caused by some event occurring at the same time as the intervention?
3. Maturation: Does the outcome measure allow for natural progression in learning and knowledge?
4. Instrumentation: Is the outcome measure reproducible?
5. Testing: Does the outcome measure itself affect the results?
6. Differential attrition: Can the outcome measure control for differing numbers of participants in control or experimental groups (if present) or large drop out rates.

The extraction of information was undertaken by the principal author.

### Stage three

Once this process had occurred a number of more global questions were asked of each paper to determine whether the article's author had evaluated the outcome methods they had chosen and allow an assessment of the construct validity of the study.

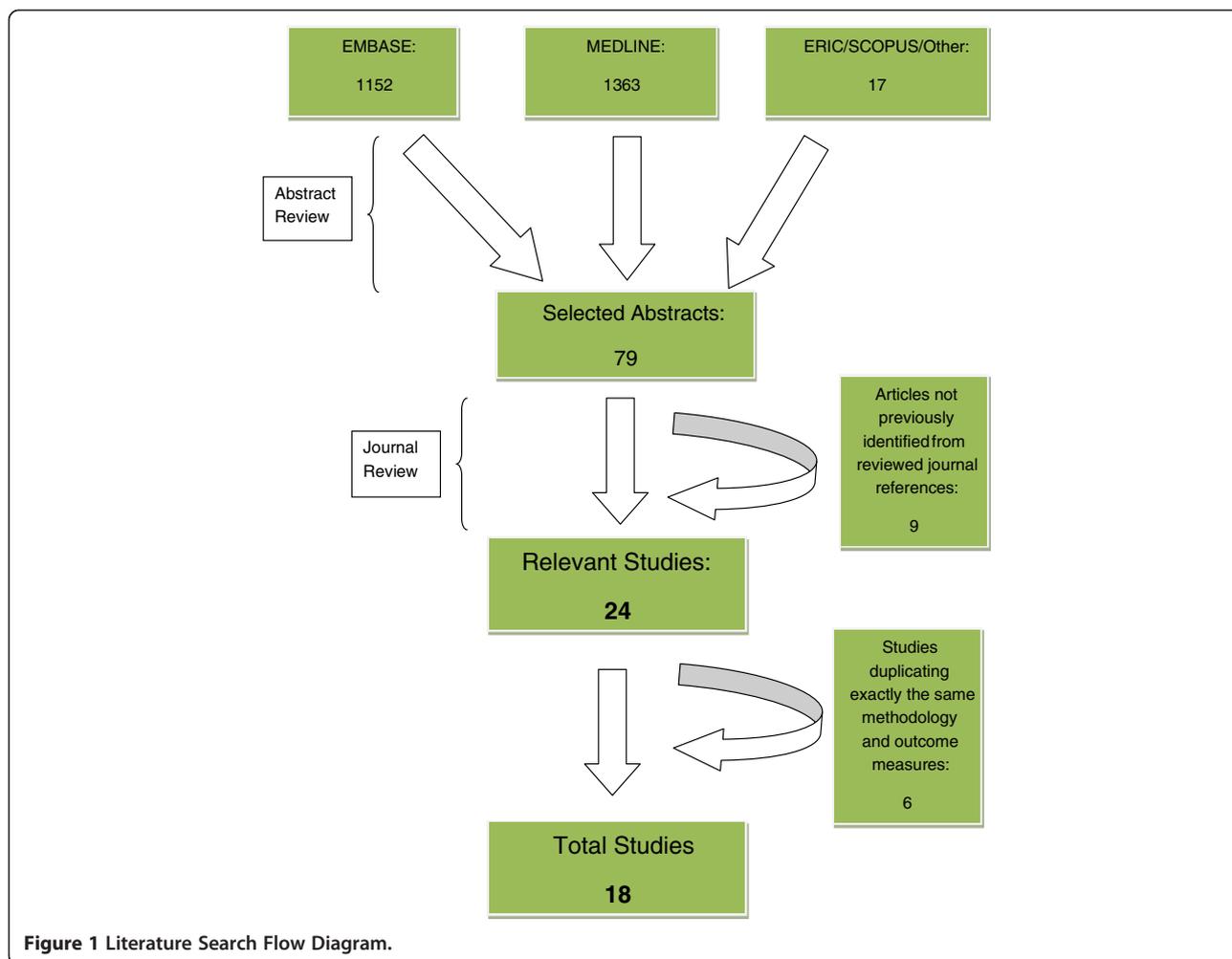
- a) How was the choice of outcome measure justified?
- b) Did the choice determine the results the study aims to investigate?
- c) To what extent were the writers aware of the disadvantages as well as the advantages of the outcome measures chosen?
- d) How did they overcome the disadvantages?

### Results

Figure 1 shows the flow of journals from the initial search to the final selection of articles. The types of healthcare professionals studied is demonstrated in Table 1 and the number of studies classified by educational purpose and Kirkpatrick level shown in Table 2. Two studies evaluated both undergraduate and basic postgraduate trainees leading to a total of 21 studies of health care professional groups and two studies evaluated both learner knowledge and learner behaviour leading to a total of 20 studies of the relevant Kirkpatrick level.

The purpose of this work was to be as inclusive as possible so as to capture all outcome measures used. Although twenty-two articles (twenty-three studies) underwent a thorough analysis in stage two, half of these require further clarification as to the reasons for their inclusion. These articles were all reviewed by all three authors and a collaborative decision reached on their inclusion. Under the inclusion criteria it had not been the intention to include animal studies in the protocol. However one, in the field of veterinary medicine [13], studied PVCs in precisely the context human patients clips would be used with an accompanying relevant and feasible methodology. It has been included in the final review as it was decided methodology rather than context was being investigated. The search was repeated removing the 'human only' limitation but no other veterinary journals of relevance were found.

One study examining an intervention to improve the physical examination component of a medical student exam via a web-based video did not specifically use abnormal or normal clinical signs [14]. The study looked at outcomes across a whole year group in a *before and after cohort* design. This study has been included as the methodology could have been easily used in a PVC-related intervention. A study using video to demonstrate a specific clinical examination was also included although it could be argued that the precise aim of the



tool was not demonstrating specific clinical signs but a methodology of elucidating them. The methodology used, a Solomon four-group design [15], was considered relevant to defining robust outcome measures in future PVC studies.

Finally six studies [16-21], although in different patient groups (ankylosing spondylitis, rheumatoid arthritis, fibromyalgia) used exactly the same methodology as two initial studies into osteoarthritis by the same investigators. These were studies in the

validation of an examination methodology in both medical students and consultants. Although the actual data was different, the papers used exactly the same introduction, methods and discussion. In terms of the narrative review, these eight journal articles represent only one methodological approach in two different cohorts of participants. It was felt due to the lack of difference in the wording of the arthritic publications these should be considered as two studies, one representing undergraduates and the other

**Table 2 Classification of studies**

Educational purpose	Number of studies	Kirkpatrick level	Number of studies
Knowledge Gain	8	Learner Reaction (level 1)	5
Testing Methods	3	Learner Knowledge (2a)	0
Patient Examination Skills	5	Learner Knowledge (2b)	8
Overall Clinical Care	2	Learner Behaviour	7
<b>Total</b>	<b>18</b>	<b>Total</b>	<b>20</b>

**Table 3 Identification of health care settings in which educational patient video clips have been utilised**

Paper	Aim or hypothesis	Health care user	Educational purpose	Kirkpatrick level	Type of study	Outcome methodology	Conclusions
Using interactive video to add physical assessment data to computer based patient simulations [22]	Interactive video in patient simulations improves the learning experience [Not formally Stated in paper]	Basic Postgraduate Training	Overall Clinical Care	Learner Behaviour (level 3)	Comparative	Chart Review of interview and physical examination description of actual patients versus simulation performance	Good to fair agreement on overall comparison (kappa=0.72) and physical examinations (kappa=0.7)
The development of shared cognition in paediatric residents analysing a patient video versus a paper patient case [23]	Will supplementing a written case vignette by a PVC as opposed to an equivalent paper case increase shared cognition as measured by the frequency of collaborative concept link formation in the context of collaborative learning about movement disorders. ...?	Medical Students	Knowledge Gain	Leaner Knowledge (level 2b)	Cluster Randomised Control Trial	Identification and frequency of collaborative and individual concept links	The video group showed a significant increase ( $p < 0.01$ ) in collaborative concept links but not in individual concept links after watching the video
Enhancing diagnostic accuracy among non-experts through use of video cases [24]	(1) How does the level of diagnostic accuracy evolve through an interactive teamwork approach using PVCs?  (2) Does the level of diagnostic accuracy differ between non-experts and experts?	Specialist Postgraduate Training	Knowledge Gain	Learner Knowledge (level 2b)	Repeated Measures Design	Analysis of frequency and of new diagnoses and new clinical reasoning processes as new information presented during review of PVC case.	i) New clinical reasoning processes were most frequent at first review of the PVC.  ii) Frequency of new relevant diagnoses were stable at each step whereas less relevant diagnoses decreased.  iii) Relevant clinical reasoning was significantly higher amongst non-experts compared with experts at the small group discussion and think aloud procedure with content expert.
Introduction of patient video clips into computer-based testing: Effects on item statistics and reliability estimates [25]	To compare the basic characteristics and reliability of questions using video-based vignettes to questions using analogous text-based vignettes.	Medical Students	Testing Methods	Learner Knowledge (level 2b)	Parallel test questions with cross-over of video and text descriptors.	Median Item statistics and reliability estimates for test items	Overall, video-based questions had comparable difficulty and discrimination compared to analogous text-based questions.
Video-based test questions: A novel means of evaluation [26]	Video-Based Test items are supported by students [Not formally stated by paper]	Undergraduate professionals allied to Medicine	Testing Methods	Learner reaction (level 1)	(1) Questionnaire utilising repeated measures ANOVAs	Students preference between video-based and multiple choice questions	Students thought video based questions deepened understanding and recommended video-based questions be used in future exams.

**Table 3 Identification of health care settings in which educational patient video clips have been utilised (Continued)**

Video-based test questions: A novel means of evaluation [26]	Unclear	Undergraduate professionals allied to Medicine	Testing Methods	Learner Knowledge (level 2b)	(2) Observational comparative study (One group informed about video questions the other not)	Exam scores in 12 video-based items	Students informed about video clips correctly answered more video based items
A comparison of critical thinking in groups of third-year medical students in text, video, and virtual PBL case modalities [27]	Critical Thinking, as exemplified by the discourse among students during group discussion, differs among groups receiving the same case with the same facilitator in one of three formats.	Medical Students	Knowledge Gain	Learner Knowledge (level 2b)	Three way comparative study (face-2-face with text, face-2-face with video, virtual with video)	Critical thinking discourse analysis	The virtual groups had the highest critical-thinking ratio. Except for the problem-identification stage, the video groups had higher ratios that the text groups did.
Comparison of text and video cases in a postgraduate problem-based learning format [28]	The addition of a video case to written information would lead to a greater increase in the frequency of data exploration, theory building and evaluation and metareasoning than would be achieved by a paper case.	Medical Students	Knowledge Gain	Learner Knowledge (level 2b)	Randomised comparative study of video versus text cases	Frequency of pre-defined clause categories	The verbal interaction showed statistically significant improvements in data exploration, theory building and theory evaluation after the video case
Use of animation-enhanced video clips for teaching abnormal breathing patterns [13]	To gather feedback regarding the Animated Breathing Pattern Videotape	3 <sup>rd</sup> year <b>Veterinary</b> Students, House Officers and Faculty, Qualified Veterinarians	Patient Examination Skills	Learner reaction (level 1)	Questionnaire	Usefulness and Satisfaction	Uniformly positive responses
Using web-based video to enhance physical examination skills in medical students [14]	To measure changes in first year students' performance of physical examinations on standardized [sic] patients after implementation of a web-based curriculum	Medical Students	Patient Examination Skills	Learner Behaviour (level 3)	Before and After Cohort Outcome study	(i) Percent correct score in physical exam item checklist (ii) Mean score on physical exam process instrument	Students on Web-based curriculum had higher level of competency and reduction in poor performance levels
Teaching the plantar reflex [15]	To test to efficacy of video-tape in the evaluation of the planter response	Medical Students	Patient Examination Skills	Learner Behaviour (level 3)	Solomon Four Group Design – Two experimental and control groups (with and without entrance test)	Correct judgement of graded presence of clinical sign	Small non-significant difference between experimental and control groups [evidence of sampling error]. If analysis was restricted to students who performed an entrance test there was a statistical significance in favour of the video group.

**Table 3 Identification of health care settings in which educational patient video clips have been utilised (Continued)**

A videotape-based training method for improving the detection of depression in residents of long-term care facilities [29]	Does a training programme involving video based scenarios improve nursing staffs' detection of depression within long-term care facilities? [Not formally stated by the paper]	Professionals allied to Medicine	Overall Clinical Care	Learner Knowledge (level 2b) and Learner Satisfaction (level 3)	Parallel group delayed intervention design.	(i) Videotape vignette test (ii) Written Test (iii) Course evaluation questionnaire	Significant increase in performance in the intervention group which was maintained for at follow up for both vignette and written test.  Good levels of satisfaction on questionnaire
Advantages of video trigger in problem-base learning [30]	The reasons behind preferences for video triggers or paper cases in students and facilitators who are accustomed to paper cases.	Medical Students	Knowledge Gain	Learner Reaction (level 1)	Questionnaire	Usefulness and Satisfaction	Video triggers were preferred by both students and facilitators over paper cases in Problem Based Learning
A triangulated approach to the assessment of teaching in childhood epilepsy [31]	Evaluation of participant perceptions of learning	Medical Students and Basic Postgraduate Training	Knowledge Gain	Learner Reaction (level 1)	Triangulation Outcome Analysis	Participant assessment (rating scales, open ended questions and focus groups), Lecturer reflection and peer observations	Videos identified as the most useful and interesting teaching tool. Results cross-validated by lecturer and peer observations
How video cases should be used as authentic stimuli in problem-based medical education [32]	To examine students views on the value of video cases compared to text based cases.	Medical Students	Knowledge Gain	Learner Reaction (level 1)	Focus Groups	Thematic Analysis	Video generally valuable but benefit dependant on certain conditions.
Visual expertise in paediatric neurology [33]	To investigate visual attention and cognitive processes of clinicians of varying degrees of experience diagnosing authentic paediatric video case	Medical Students, Basic Postgraduate Training and Consultant CPD	Knowledge Gain	Learner Knowledge (level 2b) and Behaviour (Level 3)	Observational study	Eye-tracking data were analysed with verbal recordings.	More experienced clinicians were more accurate in visual diagnosis and spent more of their time looking at relevant areas
An evaluation of the effectiveness of a videotape programme on inter-observer reliability in outcome assessment for osteoarthritis [34]	Whether interobserver variability in senior medical students could be reduced in a group of patients with OA using only a single viewing of an instructional videotape.	Medical Students	Patient Examinations Skills	Learner Behaviour (level 3)	Before and After Study <b>[Video intervention poorly described]</b>	Change in mean values of previously described observer dependant measures per participant  Reliability coefficients for the group	Pre-standardization reliability coefficients were <0.80 for seven measures. Coefficients for the performance of knee goniometry were uniformly low. Following the intervention, all but four reliability coefficients were >= 0.93.
An evaluation of the effectiveness of a videotape programme on inter-observer reliability in outcome assessment for fibromyalgia [16]	Whether interobserver variability in senior medical students could be reduced in a group of patients with fibromyalgia using only a single viewing of an	Medical Students	Patient Examinations Skills	Learner Behaviour (level 3)	Before and After Study <b>[Video intervention poorly described]</b>	Change in mean values of previously described observer dependant measures per participant	Pre-standardization reliability coefficients were <0.80 for 8 measures. Following standardization all reliability coefficients, but one, approximated or exceeded 0.80

**Table 3 Identification of health care settings in which educational patient video clips have been utilised (Continued)**

	instructional videotape.					Reliability coefficients for the group	
An evaluation of the effectiveness of a videotape programme on inter-observer reliability in outcome assessment for ankylosing spondylitis [18]	Whether interobserver variability in senior medical students could be reduced in a group of patients with ankylosing spondylitis using only a single viewing of an instructional videotape.	Medical Students	Patient Examinations Skills	Learner Behaviour (level 3)	Before and After Study [Video intervention poorly described]	Change in mean values of previously described observer dependant measures per participant  Reliability coefficients for the group	Pre-standardization reliability coefficients were < 0.80 for three measures. Following standardization 12 reliability coefficients exceeded 0.80. For the majority of measures pre-standardization reliability coefficients were high and no further improvement in reliability could be demonstrated
An evaluation of the effectiveness of a videotape programme on inter-observer reliability in outcome assessment for rheumatoid arthritis [17]	Whether interobserver variability in senior medical students could be reduced in a group of patients with rheumatoid arthritis using only a single viewing of an instructional videotape.	Medical Students	Patient Examinations Skills	Learner Behaviour (level 3)	Before and After Study [Video intervention poorly described]	Change in mean values of previously described observer dependant measures per participant  Reliability coefficients for the group	Pre-standardization reliability coefficients were >0.80 for all measures and remained above 0.80 following standardization except for one measure
Osteoarthritis antirheumatic drug trials: Effects of a standardized instructional videotape on the reliability of observer-dependent outcome measures [35]	Whether interobserver variability in consultants could be reduced in a group of patients with OA using only a single viewing of an instructional videotape.	Consultant CPD	Patient Examination Skills	Learner Behaviour (level 3)	Before and After Study [Video intervention poorly described]	Change in mean values of previously described observer dependant measures per participant  Reliability coefficients for the group	Prestandardization reliability coefficients were >0.80 for all measures and remained above 0.80 following the intervention
Fibromyalgia antirheumatic drug trials: Effects of a standardized instructional videotape on the reliability of observer-dependent outcome measures [19]	Whether interobserver variability in consultants could be reduced in a group of patients with fibromyalgia using only a single viewing of an instructional videotape.	Consultant CPD	Patient Examinations Skills	Learner Behaviour (level 3)	Before and After Study [Video intervention poorly described]	Change in mean values of previously described observer dependant measures per participant  Reliability coefficients for the group	Prestandardization reliability coefficients were <0.80 for 8 measures. Following standardization all reliability coefficients approximated to or exceeded 0.80.
Rheumatoid arthritis antirheumatic drug trials: Effects of a standardized instructional videotape on the reliability of observer-dependent outcome measures [20]	Whether interobserver variability in consultants could be reduced in a group of patients with rheumatoid arthritis using only a single viewing of an instructional videotape.	Consultant CPD	Patient Examinations Skills	Learner Behaviour (level 3)	Before and After Study [Video intervention poorly described]	Change in mean values of previously described observer dependant measures per participant  Reliability coefficients for the group	Prestandardization reliability coefficients were >0.80 for all measures and remained above 0.80 following standardization
		Consultant CPD					

**Table 3 Identification of health care settings in which educational patient video clips have been utilised (Continued)**

Ankylosing spondylitis antirheumatic drug trials: Effects of a standardized instructional videotape on the reliability of observer-dependent outcome measures [21]	Whether interobserver variability in consultants could be reduced in a group of patients with ankylosing spondylitis using only a single viewing of an instructional videotape.	Patient Examinations Skills	Learner Behaviour (level 3)	Before and After Study <b>[Video intervention poorly described]</b>	Change in mean values of previously described observer dependant measures per participant  Reliability coefficients for the group	Prestandardization reliability coefficients were <0.80 for three measures. Following standardization 12 reliability coefficients exceeded 0.80
--	---	-----------------------------	-----------------------------	--	---	--

trained doctors continuing professional development. Noting the reasons given above the total number of articles evaluated was 17 (which involved 18 distinct studies).

Table 3 contains the descriptive results for the reviewed articles and Table 4 contains the overall judgement on each of the articles. The analysis of the validity of the outcome measures can be found in the Additional file 1: Appendix.

### Discussion

This review examined the evidence on how to measure outcomes when Patient Video Cases (PVCs) are used in healthcare settings. This evidence was small, extremely heterogeneous and there was insufficient evidence to specify the best outcomes to use. The heterogeneity in the articles was created by the diversity of involved health care professionals, varying educational purposes, different types of intervention, a wide range of outcome methodologies, different internal and construct validities and a variety of results. Each of these is examined in turn.

#### Type of healthcare professional

The preponderance of projects in undergraduate education is likely related to the large number of medical education academics at these institutions, the access to a 'captive group' of subjects and the greater ease of assessing undergraduate outcomes. Further investigation into the use of PVCs at postgraduate level and in other healthcare professionals is clearly warranted. For all health care professionals it is also reasonable to attribute the lack of studies to the difficulties in designing [36] and funding studies evaluating PVCs.

#### Educational purposes and types of intervention

Given the small number of studies, it is difficult to identify clear trends in educational purpose or type of intervention. Learner satisfaction and knowledge gain are the easiest of the Kirkpatrick training outcomes to measure as they do not require external observation or intervention. However these domains are the lowest in the hierarchy of evidence needed to confirm that a training

process has been truly effective [37]. No study looked at organisational change, which is in keeping with previous literature. A review aiming to identify methods used to measure change in the clinical practices of health professionals found only 17.6% looked at changes at an organisation level [38]. Also in this review only one study attempted to look at more than one level of training outcome. A systematic review of evaluation in formal continuing medical education [39] noted 28% of studies reviewed looked at two levels and only 6% looked at three.

#### Methods for determining and assessing outcome measures

Reflecting the wide range of different types of studies performed, the validity of the outcome measures used was variable. This represents the difficulties of examining interventions related to education and training. In clinical practice a gold standard approach in assessing the effectiveness of medication is the randomised controlled trial. The primary outcome measure being an objective endpoint such as a defined reduction or gain in a physiological parameter. In training interventions, a single endpoint as an outcome requires a lot of interpretation, and potential criticism. For example, learner satisfaction does not necessarily equate to knowledge change, neither does it have a direct correlation with change in practice. The absence of a gold standard measure to assess training interventions may have led researchers to be opportunistic in their use of outcome measures. In this review seven studies gave no justification for the outcome measure used [13,15,25,26,29,30]. In addition comments by the authors themselves on limitations to the outcome measures were absent in five of the studies [13,26,30,31].

Only one study looked at more than one discrete domain in the Kirkpatrick training evaluation framework [29]. In this work both learner knowledge and learner satisfaction were assessed by different measures (a video test, a written test and a course evaluation). Three other studies [14,25,31] had more than one outcome measure, although these were all subtle variations on a theme

**Table 4 Review of methodological quality of studies using outcome measures to assess the impact of PVCs**

Paper	1 How is the choice of outcome measure justified?	2 Will this choice determine the results the study aims to investigate?	3 To what extent are the writers aware of the disadvantages as well as the advantages of the outcome measures chosen?	4 How do they overcome the disadvantages?	Internal validity?	Construct validity?
Using interactive video to add physical assessment data to computer based patient simulations [22]	Used to justify criterion validity of the intervention used.	Study aims not clear. If presumed to be to elucidate whether the simulations are an effective learning experience the outcome used partially confirms the simulation represent normal practice not that the simulation improved performance or was an useful education tool	Author notes that in actual practice chart noting is done under time pressure whereas with this simulation there was more time available to make case-notes more complete.	No comment made on this. The absence of details on whether a specific proforma for extracting information from the case-notes was used makes it difficult to assess how comparisons were made.	No	No
The development of shared cognition in paediatric residents analysing a patient video versus a paper patient case [23]	Process of capturing concept link formation described with the reason for using verbal protocol analysis supported by published evidence.  Methodology of using simulated recall in individuals following the group discussions not well supported.	Yes	Authors note a small number of participants and only one case so reliability may be questioned.  They also note the ability to accurately recall and record all concept links is not established.  They were aware the interviewers were not blinded to the intervention group of the participants.	The positioning of the simulated recall exercise straight after the group work limits loss of content due to degradation of memories.  Authors comment although not blinded the interviewers were not recording cognitive processes just the thoughts that lead to them.	Yes	Yes
Enhancing diagnostic accuracy among non-experts through use of video cases [24]	Previous work by the author has shown improved cognitive processes when PVCs utilised. Improved diagnostic accuracy is the natural conclusion of relevant or improved cognitive processing.	Yes	Study acknowledges the diagnostic accuracy as an outcome is only a short term variable of learning.  Authors touch on, but don't specifically note, the outcome measure is not directly related to the intervention rather the group discussion following the intervention.	Authors argue increase in diagnostic reasoning in non-experts in may promote further literature study and learning.	Yes	Yes
Introduction of patient video clips into computer-based testing: Effects on item statistics and reliability estimates [25]	No justification for answer analysis is given although standard methodology applied.	Yes	Problems with questions with low discrimination values identified and subject to supplemental analysis.	Items with RPB values of zero or less than 0.2 removed although no explanation of why these values were chosen.	Not Applicable	Not Applicable
Video-based test questions: A novel means of evaluation [26]	No justification for questionnaire methodology given although this format is an accepted primary approach to gathering information on satisfaction with a process.	Yes although reliability of results must be treated with caution.	No comments made	No comments made	Not Applicable	Not Applicable

**Table 4 Review of methodological quality of studies using outcome measures to assess the impact of PVCs (Continued)**

Video-based test questions: A novel means of evaluation [26]	No justification given	Aims of the study not clear	No comments made	No comments made	Not Applicable	Not Applicable
A comparison of critical thinking in groups of third-year medical students in text, video, and virtual PBL case modalities [27]	There is a theoretical association between Problem Based Learning and critical thinking. A sound research framework exists to analyse discourse and code for content. Therefore a process, discourse analysis , exists to examine the outcomes of PBL in respect to critical thinking.	Yes	The outcome measure depends on the validity and reliability of the coder.  Although the coder is blind to group type it is possible for this to be suggested by the dialogue.  Time pressures on face-to-face groups may limit opportunities to refine critical thinking compared to the virtual groups.	Only one author did all the coding but coding agreement was checked using a sample of transcripts with two others (one not involved in the study).	Yes	Yes
Comparison of text and video cases in a postgraduate problem-based learning format [28]	A coding system for cognitive and metacognitive thinking has been established. It is theoretically plausible a video case would improve thinking processes.	Yes. Although the actual reason for improved educational outcome in PBL has yet to be defined and the coding schema chosen is only one way of evaluating cognitive and metacognitive processes.	The outcome measure itself is not examined although the article notes the use of one author for both groups and the low numbers of residents in both groups mean caution is required in interpretation.	The use of clause frequency enables variability in group sizes to be addressed.	Yes	Yes
Use of animation-enhanced video clips for teaching abnormal breathing patterns [13]	No justification given for questionnaire methodology although this format is an accepted primary approach to gathering information on satisfaction with a process.	Yes although the reliability of results must be treated with caution.	No comment made. The authors note that some respondents gave inconsistent written comments when compared with their agreement with statements and comment it was likely they had mis-understood the question.	No comments made	Yes	No
Using web-based video to enhance physical examination skills in medical students [14]	Summative clinical skills assessment has been utilised (and presumably validated although this is not stated) in the learning institution the study is taking place in.	Yes, although dependant on the reliability of the Clinical Skills Assessment.	The authors comment they did not track the utilisation of the video clip website by students and note a prospective, randomised controlled study would have been more accurate.	No comments made	Potentially Not	Yes
Teaching the plantar reflex [15]	No justification given. Process for assessing performance described although the standardised rating scale was not demonstrated.	Yes but only if the assessment system is valid.	The authors acknowledge the effect of the entrance test in providing education in itself.	No comment made	Potentially Not	Yes
A videotape-based training method for improving the detection of depression in residents	Outcome measures well described but no comment on the reason for using them.	As patient outcome not measured methodology can only assess how the training programme improves performance	Authors not a large sample size is needed to counter affects of attrition.	Authors tested after a control period and withheld feedback to participants about their test performances.	Potentially Not	Yes

**Table 4 Review of methodological quality of studies using outcome measures to assess the impact of PVCs (Continued)**

of long-term care facilities [29]		in the outcomes tested.	They also not the vignette video test may improve practice in its own right.  Staff were allowed to choose timing and type of session according to their needs with no control for group or individual sessions.				
Advantages of video trigger in problem-base learning [30]	No justification for questionnaire methodology given although this format is an accepted primary approach to gathering information on satisfaction with a process.	Yes as long as sample of participants valid.	No comments made	No comments made	Yes	Yes	
A triangulated approach to the assessment of teaching in childhood epilepsy [31]	Triangulation used as a more complete and robust measure to validate findings.	Yes although must be employed in a methodological fashion. However the use of video clips was not the sole purpose of the study and questions not posed to determine this.	Very little attention paid to confounding influences and the fact that the cross – validation was not particular well demonstrated.	No comments made	No		Not applicable
How video cases should be used as authentic stimuli in problem-based medical education [32]	Focus groups a well refined qualitative tool which all deep analysis of concepts presented.	Yes	Clear acknowledgement of the problems with individuals dominating or evading group discussion.	Very experienced facilitator used	Yes	Yes	
Visual expertise in paediatric neurology [33]	Important differences in perception between experts and novices studying dynamic stimuli has been documented . Authors note this field is underexplored in the medical domain but use a high quality eye tracking machine and linked to spoken cognitive processes	Yes	Note that outcome method was novel and made efforts to triangulate findings to gold standard outcomes (such as correct diagnosis)	Used variety of experience in subject population	Yes	Yes	
An evaluation of the effectiveness of a videotape programme on inter-observer reliability in outcome assessment for osteoarthritis [34]	Outcome measure used in previous studies to assess performance in musculo-skeletal examination.	Yes	A larger matrix and more observers and patients may have been used to improve reliability.	The authors own previous work has indicated the 6x6 is pragmatic and representative.	No	Yes	
Osteoarthritis antirheumatic drug trials: Effects of a standardized instructional videotape on the reliability of observer-dependent dependent outcome measures [35]	Outcome measure used in previous studies to assess performance in musculo-skeletal examination.	Yes	A larger matrix and more observers and patients may have been used to improve reliability.	The authors own previous work has indicated the 6x6 is pragmatic and representative.	No	Yes	

such as scores in different types of clinical examination in the same test.

Only two of the studies [27,33] satisfied all domains when deciding on whether internal and construct validity had been achieved. Three other papers [15,23,29] had minor concerns, generally relating to the extent which the outcome measure itself affected the results. Questionnaire studies reflecting learner satisfaction tended not to perform well as control between groups was not possible and confounding factors were very difficult to assess.

### Results of the interventions

Nearly all papers were positive regarding the use of PVCs (regardless of whether the analysis above had revealed concerns over the validity of the outcome measure). The medical student studies regarding critical analysis and thinking showed strong results in favour of the use of PVCs. The underlying hypotheses of these studies [23,24,27,28,32] were plausible and the methodologies used rigorous. A researcher independent of these groups has also recently shown students prefer this use of PVCs to current problem based learning techniques [30] so triangulation has in some respects been achieved in this field. A recent paper demonstrating experts are more focused on the relevant clinical features within patient video clips has been further supported by, as yet unpublished evidence, that eye movement modelling may improve diagnostic reasoning. This methodology, where the minute movements of the eye are tracked while observing dynamic images, has strong construct validity. It is felt the cognitive 'load' of dynamic video clips may encourage cognitive processing [40] and therefore methodologies to explore the extent of this load created by PVCs are welcome. Future research must be cognisant of the fact that under- or over-load may occur depending on the capacity of the individual engaging in the activity. Extraneous cognitive [41] load may be able to be controlled to some extent by investigators and this will aid determination of its impact on the outcome of the intervention.

Studies concerning testing methods and clinical examination showed no obvious differences between PVCs and current assessment methods. The potential difficulty and cost of placing video clips into examinations (whether formative or summative) may have limited the number of validation studies in this area. In studies of clinical examination technique which aimed to show improvement following a PVC intervention, there was supportive evidence although initial skill sets tended to be relatively high. The importance of controlling for this was demonstrated by the use of the Solomon Four Group design on a video intervention to improve examination of the plantar reflex [15]. In this study an effect was only seen when pre-intervention performance was assessed.

The video-based training method for improving the detection of depression in residents of long term care facilities demonstrated an increase in performance of the intervention group in both knowledge assessments [29]. Direct patient benefit was not assessed so an improvement in clinical care as a use of PVC cannot be claimed. However given the good levels of satisfaction on questionnaire testing it is likely that participants would not have been averse to incorporating newly acquired learning into their day-to-day practice.

### Limitations

The heterogeneity of the current published evidence made a robust narrative review extremely difficult. Apart from the work on how PVCs encourage discourse and critical thinking, there were no common themes in which to be able to extract information and analyse composite outcomes. This may represent difficulty in undertaking research in the field (the cost of production of video clips), the difficulty in defining valid outcome measures or publication bias due to a paucity of positive outcomes. This exemplifies the challenge that much medical education research is Action Research, research based on the instructors' own practice.

Publication bias is unlikely to be significant as there is literature in which research is positive [42] regarding the use of video and online technologies but there are also negative [43] publications in existence. It would seem unlikely a particular modality of online or audiovisual learning would be subject to a different research agenda.

The main limitation of this study is the low number of articles that were found. The search strategy used was expansive although "Patient Video Clip" or similar terms are not used by all researchers in the field. It is possible terms other than those searched have been used although the number of papers missed is likely to be very small. Extraction of data was performed by a sole reviewer so it is possible so errors of typology were made although the small number of final articles has allowed extensive examination of the papers by all the authors.

### Conclusion

This review process has demonstrated the diverse nature of research in determining the effectiveness of PVCs in education. Medical education occurs in a variety of environments and the complicated interplay of confounding variables makes interpretation of outcomes difficult. The following recommendations would enable the production of a standard conceptual framework to guide future research in the area.

- Studies should classify which facet of training or educational outcome the study is aiming to explore.

- Studies should aim to validate a particular outcome measure, preferably by reproducing previous work rather than adopting new methods.
- A description of the validity of the chosen outcome measure should be included in study protocol.
- Although control groups are useful for demonstrating the benefit of a PVC intervention, more evidence is needed on whether the outcome measure demonstrates construct validity.
- Studies on PVCs should take account of cognitive theory with the cognitive processing enhancement, demonstrated in a number of the medical student papers, tested at a postgraduate level. Although pragmatic outcome measures are easier to achieve explanatory trials are needed.
- Prior-knowledge and behaviour testing is vital to demonstrate improvement.

## Additional file

**Additional file 1: Appendix.** Review of studies to assist in evaluation of internal and construct validity based on framework via Farrington.

## Competing interests

The author(s) declare that they have no competing interests. This report is independent research arising from a Doctoral Research Fellowship supported by the National Institute for Health Research. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

## Authors' contributions

DR proposed the original research concept, carried out the literature review and prepared a first draft. DM and TC made significant contributions to the design and methodology of the review as well as to subsequent drafting of the article. All authors read and approved the final manuscript.

## Acknowledgements

The authors wish to acknowledge the help given by Sarah Sutton, senior medical librarian at the University Hospitals of Leicester NHS Trust, in the preparation of the literature search.

## Author details

<sup>1</sup>Emergency Medicine Academic Group, Emergency Department secretaries c/o Elizabeth Cadman-Moore, Leicester Royal Infirmary, Leicester LE1 5WW, UK. <sup>2</sup>Room B94C Medical School, Queens Medical Centre, Nottingham NG7 2UH, UK.

Received: 30 March 2012 Accepted: 12 December 2012

Published: 21 December 2012

## References

1. McNulty JA, Hoyt A, Gruener G, Chandrasekhar A, Espiritu B, Price R Jr, Naheedy R: **An analysis of lecture video utilization in undergraduate medical education: associations with performance in the courses.** *BMC Med Educ* 2009, **9**:6.
2. Leone TA, Rich W, Finer NN: **Educational Perspectives.** *NeoReviews* 2006, **7**(12):e597–e601.
3. Romanov K, Nevgi A: **Do medical students watch video clips in eLearning and do these facilitate learning?** *Med Teach* 2007, **29**(5):484–488.
4. Roland D, Wahl H, Lakhanpaul M, Blackwell N, Davies F: **Education by video.** *BMJ (Careers)*; 2011. available at <http://careers.bmj.com/careers/advice/view-article.html?id=20001865> (website last accessed 14<sup>th</sup> December 2012).
5. (National Patient Safety Agency) NPSA: **Review of patient safety for children and young people**; 2009. available via <http://www.nrls.npsa.nhs.uk/EasySiteWeb/getresource.axd?AssetID=60227&type=full&servicetype=Attachment> (website last accessed 14<sup>th</sup> December 2012).
6. Bates R: **A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence.** *Eval Program Plann* 2004, **27**(3):341–347.
7. Kirkpatrick D, Kirkpatrick J: **Evaluating Training Programs: The Four Levels: 3<sup>rd</sup> ed.** San Francisco: Berrett-Koehler Publishers, Inc.; 2006.
8. Arthur W Jr, Bennett W Jr, Edens PS, Bell ST: **Effectiveness of training in organizations: a meta-analysis of design and evaluation features.** *J Appl Psychol* 2003, **88**(2):234–245.
9. Barr H, Freeth D, Hammick M, Koppel I, Reeves S: **Evaluations of interprofessional education: a United Kingdom review for health and social care.** 1st edition. London: London: Centre for the Advancement of Interprofessional Education; 2000.
10. Cook T, Campbell D: **Quasi-experimentation: Design and analysis issues for field settings: First ed.** Boston, MA: Houghton Mifflin Company; 1979.
11. Shadish W, Cook T, Campbell D: **Experimental and quasi-experimental designs for generalized causal inference: First ed.** Boston, MA: Houghton Mifflin Company; 2002.
12. Farrington DP: **Methodological Quality Standards for Evaluation Research.** *Ann Am Acad Pol Soc Sci* 2003, **587**(1):49–68.
13. Hawkins EC, Hansen B, Bunch BL: **Use of animation-enhanced video clips for teaching abnormal breathing patterns.** *J Vet Med Educ* 2003, **30**(1):73–77.
14. Orientale E Jr, Kosowicz L, Alerte A, Pfeiffer C, Harrington K, Palley J, Brown S, Sapieha-Yanchak T: **Using web-based video to enhance physical examination skills in medical students.** *Fam Med* 2008, **40**(7):471–476.
15. Raijmakers PG, Cabezas MC, Smal JA, van Gijn J: **Teaching the plantar reflex.** *Clin Neurol Neurosurg* 1991, **93**(3):201–204.
16. Bellamy N, Anjema C, Chhina T, Dudek N, Hurley B, Landesman B, Probyn L, Hill J, Campbell J: **An evaluation of the effectiveness of a videotape programme on interobserver reliability in outcome assessment for fibromyalgia.** *Inflammopharmacology* 1999, **7**(2):119–130.
17. Bellamy N, Anjema C, Frankovic TS, Horowitz N, Mills GCR, Shulman S, Smith C, Hill J, Campbell J: **An evaluation of the effectiveness of a videotape programme on interobserver reliability in outcome assessment for rheumatoid arthritis.** *Inflammopharmacology* 1999, **7**(2):131–141.
18. Bellamy N, Anjema C, Dhanoa D, Joglekar A, Mills GC, Nesrallah G, Smith C, Ucar C, Hill J, Campbell J: **An evaluation of the effectiveness of a videotape programme on interobserver reliability in outcome assessment for ankylosing spondylitis.** *Inflammopharmacology* 1999, **7**(2):107–117.
19. Bellamy N, Buchbinder R, Hall S, Lowthian P, Patrick M, Ryan P, Stockman A, Grant G, Green S, Campbell J: **Fibromyalgia antirheumatic drug trials: Effects of a standardized instructional videotape on the reliability of observer-dependent outcome measures.** *Inflammopharmacology* 1997, **5**(3):261–272.
20. Bellamy N, Muirden K, Bendrups A, Boyden K, McColl G, Moran H, Stockman A, Tellus M, Travers R, Campbell J: **Rheumatoid arthritis antirheumatic drug trials: Effects of a standardized instructional videotape on the reliability of observer-dependent outcome measures.** *Inflammopharmacology* 1997, **5**(3):273–284.
21. Bellamy N, Muirden K, Boyden K, McColl G, Moran H, Stockman A, Tellus M, Wicks I, Campbell J: **Ankylosing spondylitis antirheumatic drug trials: Effects of a standardized instructional videotape on the reliability of observer-dependent outcome measures.** *Inflammopharmacology* 1997, **5**(3):297–308.
22. White JE: **Using interactive video to add physical assessment data to computer-based patient simulations in nursing.** *Comput Nurs* 1995, **13**(5):233–235.
23. Balslev T, de Grave W, Muijtjens AMM, Eika B, Scherpbier AJJA: **The development of shared cognition in paediatric residents analysing a patient video versus a paper patient case.** *Adv Health Sci Educ* 2009, **14**(4):557–565.
24. Balslev T, De Grave WS, Muijtjens AMM, Scherpbier AJJA: **Enhancing diagnostic accuracy among nonexperts through use of video cases.** *Pediatrics* 2010, **125**(3).
25. Lieberman SA, Frye AW, Litwins SD, Rasmusson KA, Boulet JR: **Introduction of patient video clips into computer-based testing: Effects on item statistics and reliability estimates.** *Acad Med* 2003, **78**(10 SUPPL).
26. Hertenstein MJ, Wayand JF: **Video-based test questions: A novel means of evaluation.** *J Instr Psychol* 2008, **35**(2):188–191.
27. Kamin C, O'Sullivan P, Deterding R, Younger M: **A comparison of critical thinking in groups of third-year medical students in text, video, and virtual PBL case modalities.** *Acad Med* 2003, **78**(2):204–211.

28. Balslev T, de Grave WS, Muijtjens AM, Scherpbier AJ: **Comparison of text and video cases in a postgraduate problem-based learning format.** *Med Educ* 2005, **39**(11):1086–1092.
29. Wood S, Cummings JL, Schnelle B, Stephens M: **A videotape-based training method for improving the detection of depression in residents of long-term care facilities.** *Gerontologist* 2002, **42**(1):114–121.
30. Chan LK, Patil NG, Chen JY, Lam JC, Lau CS, Ip MS: **Advantages of video trigger in problem-based learning.** *Med Teach* 2010, **32**(9):760–765.
31. Bye AM, Connolly AM, Netherton C, Looker P, Burgess A, Lonergan A: **A triangulated approach to the assessment of teaching in childhood epilepsy.** *Med Teach* 2007, **29**(2–3):255–257.
32. de Leng B, Dolmans D, van de Wiel M, Muijtjens A, van der Vleuten C: **How video cases should be used as authentic stimuli in problem-based medical education.** *Med Educ* 2007, **41**(2):181–188.
33. Balslev T, Jarodzka H, Holmqvist K, de Grave W, Muijtjens AM, Eika B, van Merriënboer J, Scherpbier AJ: **Visual expertise in paediatric neurology.** *Eur J Paediatr Neurol* 2012, **16**(2):161–166.
34. Bellamy N, Anjema C, Alikhan N, Chhina T, Dhanoa D, Edelist D, Esufali Z, Ismail F, Hill J, Campbell J: **An evaluation of the effectiveness of a videotape programme on interobserver reliability in outcome assessment for osteoarthritis.** *Inflammopharmacology* 1999, **7**(2):143–154.
35. Bellamy N, Bachmeier C, Brooks P, Browne C, Cohen M, March L, Conaghan P, Day R, Campbell J: **Osteoarthritis antirheumatic drug trials: Effects of a standardized instructional videotape on the reliability of observer-dependent outcome measures.** *Inflammopharmacology* 1997, **5**(3):285–295.
36. Norman G: **RCT = results confounded and trivial: the perils of grand educational experiments.** *Med Educ* 2003, **37**(7):582–584.
37. Kirkpatrick D: **Evaluation of training.** In *Training and development handbook: A guide to human resource development*. 1st edition. Edited by Craig RL. New York: McGraw-Hill; 1976:317.
38. Hakkenes S, Green S: **Measures for assessing practice change in medical practitioners.** *Implement Sci* 2006, **1**:29.
39. Tian J, Atkinson N, Portnoy B, Gold R: **A systematic review of evaluation in formal continuing medical education.** *J Contin Educ Health Prof* 2007, **27**(1):16–27.
40. Paas F, Renkl A, Sweller J: **Cognitive Load Theory and Instructional Design: Recent Developments.** *Educ Psychol* 2003, **38**(1):1–4.
41. Chandler P, Sweller J: **Cognitive Load Theory and the Format of Instruction.** *Cognit Instruct* 1991, **8**(4):293–332.
42. Hadley J, Kulier R, Zamora J, Coppus SF, Weinbrenner S, Meyerrose B, Decsi T, Horvath AR, Nagy E, Emparanza JI, Arvanitis TN, Burls A, Cabello JB, Kaczor M, Zanrei G, Pierer K, Kunz R, Wilkie V, Wall D, Mol BW, Khan KS: **Effectiveness of an e-learning course in evidence-based medicine for foundation (internship) training.** *J R Soc Med* 2010, **103**(7):288–294.
43. Bakker RJ, van de Putte EM, Kuis W, Sinnema G: **Effects of an educational video film in fatigued children and adolescents: a randomised controlled trial.** *Arch Dis Child* 2011, **96**(5):457–460.

doi:10.1186/1472-6920-12-125

**Cite this article as:** Roland et al.: Towards a conceptual framework demonstrating the effectiveness of audiovisual patient descriptions (patient video cases): a review of the current literature. *BMC Medical Education* 2012 **12**:125.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

