How valid are commercially available medical simulators?

JJ Stunt¹
PH Wulms²
GM Kerkhoffs¹
J Dankelman²
CN van Dijk¹
GJM Tuijthof¹,²

¹Orthopedic Research Center Amsterdam, Department of Orthopedic Surgery, Academic Medical Centre, Amsterdam, the Netherlands; ²Department of Biomechanical Engineering, Faculty of Mechanical, Materials and Maritime Engineering, Delft University of Technology, Delft, the Netherlands

Background: Since simulators offer important advantages, they are increasingly used in medical education and medical skills training that require physical actions. A wide variety of simulators have become commercially available. It is of high importance that evidence is provided that training on these simulators can actually improve clinical performance on live patients. Therefore, the aim of this review is to determine the availability of different types of simulators and the evidence of their validation, to offer insight regarding which simulators are suitable to use in the clinical setting as a training modality.

Summary: Four hundred and thirty-three commercially available simulators were found, from which 405 (94%) were physical models. One hundred and thirty validation studies evaluated 35 (8%) commercially available medical simulators for levels of validity ranging from face to predictive validity. Solely simulators that are used for surgical skills training were validated for the highest validity level (predictive validity). Twenty-four (37%) simulators that give objective feedback had been validated. Studies that tested more powerful levels of validity (concurrent and predictive validity) were methodologically stronger than studies that tested more elementary levels of validity (face, content, and construct validity).

Conclusion: Ninety-three point five percent of the commercially available simulators are not known to be tested for validity. Although the importance of (a high level of) validation depends on the difficulty level of skills training and possible consequences when skills are insufficient, it is advisable for medical professionals, trainees, medical educators, and companies who manufacture medical simulators to critically judge the available medical simulators for proper validation. This way adequate, safe, and affordable medical psychomotor skills training can be achieved.

Keywords: validity level, training modality, medical education, validation studies, medical skills training

Introduction

Simulators for medical training have been used for centuries. More primitive forms of physical models were used before plastic mannequins and virtual systems (VS) were available.¹ Since then, simulation in medical education has been deployed for a variety of actions, such as assessment skills, injections, trauma and cardiac life support, anesthesia, intubation, and surgical skills (SuS).²⁻³ These actions require psychomotor skills, physical movements that are associated with cognitive processes.⁴⁻⁵ Among these psychomotor skills are skills that require (hand–eye) coordination, manipulation, dexterity, grace, strength, and speed. Studies show that medical skills training which requires physical actions can be optimally performed by actual practice in performing these actions, eg, instrument handling.⁶ This is explained by the fact that when learning...
psychomotor skills, the brain and body co-adapt to improve the manual (instrument) handling. This way, the trainee learns which actions are correct and which are not.5

Four main reasons to use simulators instead of traditional training in the operating room have been described.6 Firstly, improved educational experience; when simulators are placed in an easily accessible location, they are available continuously. This overcomes the problem of dependency on the availability of an actual patient case. Simulators also allow easy access to a wide variety of clinical scenarios, eg, complications.6 Secondly, patient safety; simulators allow the trainee to make mistakes, which can equip the resident with a basic skills level that would not compromise patient safety when continuing training in the operating room.7–14 Thirdly, cost efficiency; the costs of setting up a simulation center are in the end often less than the costs of instructors’ training time, and resources required as part of the training.6 Moreover, the increased efficiency of trainees when performing a procedure adds to the return on investment achieved by medical simulators, as Frost and Sullivan demonstrated.15 Lastly, simulators offer the opportunity to measure performance and training progress objectively by integrated sensors that can measure, eg, task time, path length, and forces.6,9,16–18

With the increased developments and experiences in research settings, a wide variety of simulators have become commercially available. The pressing question is whether improvements in performance on medical simulators actually translates into improved clinical performance on live patients. Commercially available simulators in other industries, such as aerospace, the military, business management, transportation, and nuclear power, have been demonstrated to be valuable for performance in real life situations.19–21 Similarly, it is of high importance that medical simulators allow for the correct training of medical skills to improve real life performances. Lack of proper validation could imply that the simulator at hand does not improve skills or worse, could cause incorrect skills training.24,25

Since validation of a simulator is required to guarantee proper simulator training, the aim of this review is to determine the availability of medical simulators and whether they are validated or not, and to discuss their appropriateness. This review is distinctive as it categorizes simulators based on simulator type and validation level. In this way, it provides a complete overview of all sorts of available simulators and their degree of validation. This will offer hospitals and medical educators, who are considering the implementation of simulation training in their curriculum, guidelines on the suitability of various simulators to fulfill their needs and demands.

**Methods**

The approach to achieve the study goal was set as follows. Firstly, an inventory was made of all commercially available simulators that allow medical psychomotor skills training. Secondly, categories that represent medical psychomotor skills were identified and each simulator was placed in one of those categories. Each category will be discussed and illustrated with some representative simulators. Thirdly, validity levels for all available simulators were determined. Lastly, study designs of the validity studies were evaluated in order to determine the reliability of the results of the validity studies.

**Inventory of medical simulators**

The inventory of commercially available medical simulators was performed by searching the Internet using search engines Google and Yahoo, and the websites of professional associations of medical education (Table 1). The search terms were split up in categories to find relevant synonyms (Table 2). Combinations of these categorized keywords were used as search strategy. For each Internet search engine, a large number of “hits” were found. Relevant websites were selected using the following inclusion criteria: the website needs to be from the company that actually manufactures and sells the product; the simulator should be intended for psychomotor skills training in the medical field (this implies that the models, mannequins or software packages that only offer knowledge training or visualization were excluded); if the company’s website provided additional medical simulators, all products that fulfil the criteria were included separately; the website should have had its latest “update” after January 2009, so that it can be expected that the company

| Table 1 List of societies and associations concerning medical education and simulation |
|---------------------------------|----------|
| **Abbreviation** | **Society** |
| SSSH | Society for Simulation in Healthcare |
| SESAM | Society in Europe of Simulation Applied Medicine |
| DSSH | Dutch Society for Simulation in Healthcare |
| INACSL | International Nursing Association for Clinical Nursing Simulation and Learning |
| NLN | National League for Nursing |
| ASSH | Australian Society for Simulation in Healthcare |
| SIRC | Simulation Innovation Resource Center |
| AMEE | An International Association For Medical Education |

**Note:** Some of these societies promote commercially available simulators, which were included in our inventory.
is still actively involved in commercial activities in the field of medical simulators.

**Categorization of simulator type**

For our study purpose, medical simulators were categorized based on their distinct characteristics: VS and physical mannequins with or without sensors (Figure 1). VS are software based simulators. The software simulates the clinical environment that allows practicing individual clinical psychomotor skills. Most of these simulators have a physical interface and provide objective feedback to the user about their performance with task time as the most commonly used performance parameter. The physical mannequins are mostly plastic phantoms simulating (parts of) the human body. The advantage of physical models is that the sense of touch is inherently present, which can provide a very realistic training environment. Most models do not provide integrated sensors and real-time feedback. These models require an experienced professional supervising the skills training. Some physical models have integrated sensors and computer software which allow for an objective performance assessment. As these simulators take over part of the

---

**Table 2 Search terms**

<table>
<thead>
<tr>
<th>Simulator</th>
<th>Medical field</th>
<th>Educational</th>
<th>Commercially available</th>
<th>Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulator#</td>
<td>Medic#</td>
<td>Education</td>
<td>Product</td>
<td>Skill#</td>
</tr>
<tr>
<td>Trainer#</td>
<td>Health#</td>
<td>Learning</td>
<td>Company OR firm OR business</td>
<td>Psychomotor</td>
</tr>
<tr>
<td>“Virtual reality” OR VR</td>
<td>Clinical</td>
<td>Teaching</td>
<td>Commer#</td>
<td>Dexterity</td>
</tr>
<tr>
<td>“Skills trainer”</td>
<td>Surg#</td>
<td>Training</td>
<td>Purchase OR buy OR offer</td>
<td>Handiness</td>
</tr>
<tr>
<td>Model</td>
<td>Nurs#</td>
<td></td>
<td></td>
<td>Eye–hand</td>
</tr>
<tr>
<td>Simulation</td>
<td></td>
<td></td>
<td></td>
<td>Coordination</td>
</tr>
<tr>
<td>Phantom</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dummy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mannequin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manikin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mock-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1** Schematic overview of the number of simulators per skills category (in brackets) and the number of simulators per simulator type (in brackets).

**Abbreviations:** VS, virtual systems; PM, physical model; MES, manual patient examination skills; IPIS, injections, needle punctures, and intravenous catheterization skills; BLSS, basic life support skills; SuS, surgical skills.
assessment of training progress, it might be expected that they are validated in a different manner. Therefore, a distinction was made between simulators that provide feedback and simulators that do not.

**Categorization of medical psychomotor skills**

Skills were categorized in the following categories as they are the most distinct psychomotor skills medical professionals will learn during their education starting at BSc level: 1) manual patient examination skills (MES): an evaluation of the human body and its functions that requires direct physical contact between physician and patient; 2) injections, needle punctures, and intravenous catheterization (peripheral and central) skills (IPIS): the manual process of insertion of a needle into human skin tissue for different purposes such as taking blood samples, lumbar or epidural punctures, injections or vaccinations, or the insertion of a catheter into a vein; 3) basic life support skills (BLSS).\(^{29,30}\) BLSS refers to maintaining an open airway and supporting breathing and circulation, which can be further divided into the following psychomotor skills: continued circulation, executed by chest compression and cardiac massage; opening the airway, executed by manually tilting the head and lifting the chin; continued breathing, executed by closing the nose, removal of visible obstructions, mouth-to-mouth ventilation, and feeling for breathing;\(^{29}\) 4) SuS: indirect tissue manipulation for diagnostic or therapeutic treatment by means of medical instruments, eg, scalpels, forceps, clamps, and scissors. Surgical procedures can cause broken skin, contact with mucosa or internal body cavities beyond a natural or artificial body orifice, and are subdivided into minimally invasive and open procedures.

**Inventory of validation and study design quality assessment**

The brand name of all retrieved simulators added to the keyword “simulator” was used to search PubMed for scientific evidence on validity of that particular simulator. After scanning the abstract, validation studies were included and the level of validation of that particular simulator was noted.\(^{32}\) Studies were scored for face validity,\(^{24,32,33}\) (the most elementary level), construct validity,\(^{73}\) concurrent validity,\(^{24,32}\) and the most powerful level, predictive validity.\(^{24,32,33}\)

The validation studies were evaluated for their study design using Issenberg’s guidelines for educational studies involving simulators (Table 3).\(^{34}\) Each study was scored for several aspects concerning the research question, participants, methodology, outcome measures, and manner of scoring (Table 4). An outcome measure is considered appropriate when it is clearly defined and measured objectively.

The validation studies demonstrated substantial heterogeneity in study design, therefore, analysis of the data was performed qualitatively and trends were highlighted.

**Results**

**Inventory and categorization of medical simulators**

In total, 433 commercially available simulators were found (see Supplementary material), offered by 24 different companies. From these simulators, 405 (93.5%) are physical models and 28 (6.5%) are virtual simulators (Figure 1). An almost equal distribution of simulators is available for each of the four defined skills categories (Figure 1), with the SuS category containing the noticeably highest portion of virtual reality simulators (86%). Objective feedback was provided by the simulator itself in 65 cases (15%).

Simulators for patient examination (MES) training provide the possibility for physical care training, eg, respiratory gas exchange, intubation, and anesthesia delivery.\(^{28,35–38}\) The typical simulators in this category predominantly consist of (full body) mannequins that have anatomical structures and simulate physiological functions such as respiration, and peripheral pulses (eg, Supplementary material: simulators 3 and 21).

IPIS simulators provide training on needle punctures and catheterization. Such simulators usually consist of a mimicked body part, eg, an arm or a torso. An example is the Lumbar Puncture simulator (Kyoto Kagaku Co., Kitane-Voycho Fushimi-ku Kyoto, Japan).\(^{39,40}\) This simulator consists of a life-like lower torso with a removable “skin” that does not show the marks caused by previous needle punctures. Integral to the simulator is a replaceable “puncture block”, which can represent different types of patients (eg, “normal”, “obese”, “elderly”), and which is inserted under the “skin”.\(^{41}\)

BLSS simulators allow for emergency care skills training, such as correct head tilt and chin lift, application of cervical collars, splints, and traction or application to spine board.\(^{42}\) These simulators predominantly consist of full body mannequins having primary features such as anatomically correct landmarks, articulated body parts to manipulate the full range of motion, removable mouthpieces and airways, permitting the performance of chest compressions, oral or nasal intubation, and simulated carotid pulse (eg, Supplementary material: simulators 243, 245, and 270). SuS simulators are used for skills training required when
performing open or minimally invasive surgery, like knot tying, suturing, instrument and tissue handling, dissection, simple and complex wound closure. Both physical and virtual simulators form part of this category. A representative example of a physical simulator is the life-sized human torso with thoracic and abdominal cavities and neck/trachea. Such a model is suited to provide training on a whole open surgery procedure, including preparing the operative area, (local) anesthesia, tube insertion, and closure (eg, Supplementary material: 355 and 357). The torso is covered with a polymer that mimics the skin and contains red fluid that mimics blood. Virtual reality systems start to take an important place in minimally invasive surgical procedure training, especially for hand–eye co-ordination training. The VR provide instrument handles with or without a phantom limb and a computer screen in which a virtual scene is presented (eg, the Symbionix simulators 316–322 [Symbionix, Cleveland, OH, USA] and the Simendo simulators 425–426 [Simendo B.V., Rotterdam, the Netherlands] in the Supplementary material). Software provides a “plug-and-play” connection to a personal computer via a USB port.33,44

Inventory of validation and study design quality assessment

One hundred and thirty validation studies evaluated 35 commercially available medical simulators for levels of validity ranging from face to predictive validity (Figure 2). From these 35 simulators, two (5.7%) simulators were tested for face validity, four (11.4%) for content validity, seven (20%) for construct validity, 14 (40%) for concurrent validity and 8 (22.9%) for predictive validity (Figure 2). References of the validated simulators are shown in the Supplementary material (between brackets). Twenty-four (37%) simulators that provide objective feedback have been validated, from which six occurred in MES, one in IPIS and 17 in SuS (Figure 1).

The numbers of validated simulators per category were substantially different, as was the level of validity (Figure 2). SuS simulators were most validated (62.9%), and most frequently for the highest validity level (Figure 2, predictive validity). MES simulators were primarily tested for content and concurrent validity. The proportion of validated IPIS and BLSS simulators was small (Figure 2).

The quality of the study designs was verified for ten important aspects. Although all studies clearly described the researched question, study population, and outcome measures, few studies met all other criteria on the checklist. Most studies did not perform a power analysis to guarantee a correct number of participants before inclusion. Twelve percent of the 130 studies used a standardized assessment system or performed blind assessment. The majority of the studies (111) performed a correct selection of subjects: either based on experience level or with a randomly selected control group. However, 20 studies did not select their control group randomly or had no control group at all (Table 3) (37 studies tested face or content validity).55–48

Each study used proper outcome measures to test the efficacy of the simulator, which indicated psychomotor skills performance. The most commonly used performance measures are depicted in Table 4. To assess performance data objectively the following standardized scoring methods were used: team leadership-interpersonal skills (TLIS) and emergency clinical care scales (ECCS),49 objective structural clinical examination (OSCE),27,50 objective structured assessment of technical skills (OSAT),39,51–55 and global rating scale (GRS).56–59 All other studies used assessment methods that were developed specifically for that study.

Methodologically speaking, the studies that tested concurrent and predictive validity outperformed the studies that tested face, content, and construct validity.

**Discussion**

This study reviewed the availability of medical simulators, their validity level, and the reliability of the study designs. Four hundred and thirty-three commercially available simulators were found, of which 405 (94%) were physical models. Evidence of validation was found for 35 (6.5%) simulators (Figure 2). Mainly in category two and three, the number of validated simulators was marginal. Solely SuS simulators were validated for the highest validity level. Sixty-three percent of the 65 simulators that provide feedback on
Table 3 Checklist for the evaluation of validation study, using Issenberg's guidelines for educational studies involving simulators

<table>
<thead>
<tr>
<th>Guidelines for educational studies involving simulators</th>
<th>No of studies</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Clear statement of the research question</td>
<td>130</td>
<td>56,64,65,67,73,75,95,103,132,135</td>
</tr>
<tr>
<td>2. Clear specification of participants</td>
<td>130</td>
<td>1,5,27,28,39,43,51,52–57,64,66,67,74,75,78,83,85,87,88,90–97,100,101,105,106,125,132,139,148–150</td>
</tr>
<tr>
<td>4. Power analysis</td>
<td>10</td>
<td>27,39,49–54,56–59,77,102,128</td>
</tr>
<tr>
<td>5. Random selection of subjects</td>
<td>41</td>
<td>49,53,56,74,87,89,91,96,97,125,127,128,134,135,139,147,148</td>
</tr>
<tr>
<td>7. Is the outcome measure the proper one for the study?</td>
<td>130</td>
<td>11,26</td>
</tr>
<tr>
<td>8. Standardized scoring of performance data</td>
<td>15</td>
<td>60,61</td>
</tr>
<tr>
<td>10. Pre-intervention measurement: yes/no</td>
<td>20</td>
<td>45,46</td>
</tr>
</tbody>
</table>

Notes: In the first column, the ten important aspects the studies were evaluated for are stated. The second column shows the number of studies that met the criteria. The third column shows the references of the concerned studies. Data from Wolpert et al.7

These findings can partly be explained: the necessity of a high level validation and the extent to which simulators need to mimic reality is firstly dependent on the type of skills training, and secondly on the possible consequences for patients when medical psychomotor skills are insufficient. This could especially be the case for SuS skills, because minimally invasive SuS are presumably most distinct from daily use of psychomotor skills, and as a result not well developed.

Table 4 Outcome measures to test the efficacy of the simulator (the parameters indicate psychomotor skills performance)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>No of studies</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Path length</td>
<td>15</td>
<td>44,62,63,105,111–113,121,124,128,131,132,134–136</td>
</tr>
<tr>
<td>Smoothness</td>
<td>9</td>
<td>128,129,132–136,145,148</td>
</tr>
<tr>
<td>Number and economy of movement</td>
<td>21</td>
<td>39,43,52,58,72,76,77,82,83,87,105,110–112,119,120,126,127,128,129,133</td>
</tr>
<tr>
<td>Number of targets reached or missed</td>
<td>10</td>
<td>2.68–70,72,81,82,124,125,150</td>
</tr>
<tr>
<td>Tissue handling</td>
<td>7</td>
<td>39,43,52,58,87,110,127</td>
</tr>
<tr>
<td>Technical skills</td>
<td>9</td>
<td>39,54,57,79,102,119,120,122,145</td>
</tr>
<tr>
<td>Number of errors or instrument collisions</td>
<td>21</td>
<td>22,44,58,62,63,66,79,82,89,90,105,109–112,119,121–124,126</td>
</tr>
<tr>
<td>Accuracy</td>
<td>15</td>
<td>74,91,109,110,117,119,128,129,132–136,146,147</td>
</tr>
</tbody>
</table>

In addition, when these skills are taught incorrectly, it can have serious consequences for the patient, eg, if a large hemorrhage occurs as a result of an incorrect incision. To guarantee patient safety, it is important that simulators designed for this type of training demonstrate high levels of validity.60,61 For other types of skills, such as patient examination, a lower validity level can be acceptable, because these skills are closer related to everyday use of psychomotor skills, and solely require a basic level of training on a simulator, which can be quickly adapted in a real-life situation.45,46 Moreover, it requires less extensive methodology to determine face validity than to determine predictive validity.

Certain factors made it difficult to score all validity studies on equal terms; substantial heterogeneity exists among the studies. However, in general, it can be stated that a substantial part of the validation studies showed methodological flaws. For example, many studies did not describe a power analysis, so it was difficult to judge whether these studies included the correct number of participants. Furthermore, only 15 of 130 studies used standardized assessment methods and blinded assessors. Unvalidated assessment methods and unblinded ratings are less objective, which affects reliability and validity of the test.26 This raises the question whether the presented studies were adequate enough to determine the validity level of a certain simulator. Future validity studies should focus on a proper study design, in order to increase the reliability of the results.

There are several limitations to our study. Firstly, our inventory of commercially available medical simulators was performed solely by searching the Internet. We did not complement our search by contacting manufacturers or by visiting conferences. This might implicate that our list of available simulators is not complete. Secondly, the available...
level of validity for the simulators was also determined by searching public scientific databases. Quite possibly, manufacturers have performed validity tests with a small group of experts, but refrained from publishing the results. It is also possible that studies have been rejected for publication or have not been published yet. Therefore, the total number of simulators and number of validated simulators that was found, might be underestimated. However, this does not undermine the fact that few simulators were validated. Especially high levels of validation are scanty.

Our results should firstly make medical trainers aware of the fact that a low number of simulators are actually tested, while validation is truly important. Although it is possible that unvalidated simulators provide proper training, validity of a device is a condition to guarantee proper acquisition of psychomotor skills and lack of validity brings the risk of acquisition of improper skills. Secondly, a simulator that provides feedback independent of a professional supervisor, should have been validated to guarantee that the provided feedback is adequate and appropriate in real-life settings. Thirdly, for reliable results of validity studies, proper study design is required. Well conducted studies have shown to be limited so far. Lastly, it is necessary to determine the type of skills educators will offer to their trainees with a simulator and the level of validity that is required to guarantee adequate training.

Our plea is for researchers to collaborate with manufacturers to develop questionnaires and protocols to test newly developed simulators. Simulators from the same category can be tested simultaneously with a large group of relevant participants. When objective evidence for basic levels of validity is obtained, it is important to publish the results so that this information is at the disposal of medical trainers. Before introducing a simulator in the training curriculum, it is recommended to first consider which skills training is needed, and the complexity and possible clinical consequences of executing those skills incorrectly. Subsequently, the minimum required level of validity should be determined for the simulator that allows for that type of skills training. The qualitative results support the concept that the level of validation depends on the difficulty level of skills training and the unforeseen consequences when skills are insufficient or lead to erroneous actions. This combination of selection criteria should guide medical trainers in the proper selection of a simulator for safe and adequate training.

**Conclusion**

For correct medical psychomotor skills training and to provide objective and correct feedback it is essential to have a realistic training environment. Scientific testing of simulators is an important way to prove and validate the training method. This review shows that 93.5% of the commercially available simulators are not known to be tested for validity, which implies that no evidence is available that they actually improve individual medical psychomotor skills. From the validity studies that were done for 35 simulators, many show some methodological flaws, which weaken the reliability of the results. It is also advisable for companies that manufacture medical simulators to validate their products and provide scientific evidence to their customers. This way, a quality system becomes available, which contributes to providing adequate, safe, and affordable medical psychomotor skills training.

**Disclosure**

This research was funded by the Marti-Kuening Eckhart Foundation, Lunteren, the Netherlands. The authors report no conflict of interest.

**References**


