CHAPTER 3

RESEARCH METHODOLOGY

A. Method and Design of Research

A quasi experimental through pretest and posttest design (Creswell, 2008) were conducted as method of research. Quasi-experimental designs do not include the use of random assignment. In this design there are two groups that were observed, one group served as experiment group and utilize STEM-based instructional material and another group served as control group and utilize non STEM-based instructional material. A set of pretest and posttest were conducted to measure the effectivity of each instructional material on students' learning.

O	X_1	O
O	X_2	O

Figure 3.1 Quasi experiment pretest-postest design

In this design, two already existing, or intact, groups are used. The dashed line indicates that the two groups being compared are already formed that is, the subjects are not randomly assigned to the two groups. X_1 symbolizes the experimental treatment (STEM based instructional material), X_2 symbolizes the control treatment (non-STEM based instructional material), while O refers to observation (measurement) of the dependent variable. The placement of the symbols from left to right indicates the order in time of X and O. Both of group are measured or observed not only after being exposed to a treatment of some sort, but also before. In analyzing the data, each individual's pretest score is subtracted from his or her posttest score, thus permitting analysis of "gain" or "change." The amount of gain often depends on initial performance; that is, the group scoring higher on the pretest is likely to improve more (or in some cases less), and thus subject characteristics still remains somewhat of a threat. Further, administering a pretest raises the possibility of a testing threat.

B. Population and Sample

1. Population

Population of this study comprises of all 8th grader students consist of 521 students distributed among eleven classrooms at a secondary (junior high) school in Bandung.

2. Sample

Sample of this study consist of two classes of 8th grader at a secondary (junior high) school in Bandung. A sample of students was selected from population by selecting clusters of students as classroom groups that has same characteristic.

C. Research Instruments

1. Students' Conceptual Understanding Test

Students' conceptual understanding test in form of multiple choice questions were administered before and after implementation of the instructional material either in control group that was used current science book from National Ministry of Education or experimental group that was used STEM-based instructional material. Conceptual understanding test was aim to capture students' learning gain, a comprehensive multiple-choice examination were administered at the first class meeting and the same instrument were administered as a posttest at the last session. Each question item of conceptual understanding was developed based on indicators that were formulated from basic competence of National Curriculum.

The development of students' conceptual understanding test were done through four steps consist of instrument validation, analysis of test item, selection of test item and revision of test item. Analysis of students' conceptual understanding instrument was conducted to produce a proper instrument. Students' conceptual understanding instrument was validated and analyzed through *Anates V4* program before being implemented in classroom to measure validity, reliability, level of difficulty and power of discrimination of test item.

a) Validity of test

The validity of a test represents the extent to which a test measures what it purports to measures (Tuckman, 1978). Validity concerned with the specific use to be made of the results and with the truthfulness of proposed interpretation. Content validity was conducted through judgment and correction by research supervisor. The revised conceptual understanding test then administered to students and its validity being analyzed through *Anates V4 program* for multiple choice questions. The result of Pearson Product Moment Correlations Coefficient (r_{xy}) that was calculated through *Anates V4 program* was interpreted based on test validity criterion as follow.

Table 3.1 Criteria of test validity

r_{xy}	**Category**
0,80 - 1,00	Very high
0,60 - 0,80	High
0,40 - 0,60	Moderate
0,20 - 0,40	Low
0,00 – 0,20	Very low

(Surapranata, 2009)

The results of validity test through *Anates V4* program shows that among the validity of 20 test items there are 9 test items were categorized as very high, 4 test items were categorized as high, 3 test items were categorized as moderate, 2 test items were categorized as very low and 2 test items were not valid. Based on the analysis of test item validity, there are 15 test items that were appropriate and 5 test items that were not appropriate to be used. A blueprint of conceptual understanding test before being validated is provided in appendix (Appendix B.1, page 170). Recapitulation of conceptual understanding instrument validation was provided on Table 3.2.

Table 3.2 Recapitulation of conceptual understanding instrument validation

Item Number	Discrimination Power (%)	Level of Difficulty		Validity		Reliability	Significance of Correlation	Conclusion	New item number
		P (%)	Category	r_{xy}	Category				
1	63.64	76.19	Easy	0.648	High	0.89 (Very high)	Significant	Accepted	1
2	72.73	78.57	Easy	0.670	High		Significant	Accepted	2
3	0.00	100	Very Easy	NAN	Not valid		NAN	Rejected	-
4	100	57.14	Moderate	0.917	Very high		Very significant	Accepted	3
5	100	42.86	Moderate	0.858	Very high		Very significant	Accepted	4
6	100	57.14	Moderate	0.917	Very high		Very significant	Accepted	5
7	100	45.71	Moderate	0.817	Very high		Very significant	Accepted	6
8	36.36	9.52	Very difficult	0.407	Moderate		Not significant	Rejected	-
9	36.36	52.38	Moderate	0.156	Very low		Not significant	Rejected	-
10	81.82	54.29	Moderate	0.791	High		Significant	Accepted	7
11	0.00	100	Very easy	NAN	Not valid		NAN		-
12	45.45	85.71	Very easy	0.524	Moderate		Significant	Accepted	8
13	90.91	59.05	Moderate	0.805	Very high		Very significant	Accepted	9
14	100.00	54.29	Moderate	0.886	Very high		Very significant	Accepted	10
15	100.00	59.05	Moderate	0.832	Very high		Very significant	Accepted	11
16	100.00	56.67	Moderate	0.827	Very high		Very significant	Accepted	12
17	63.64	59.52	Moderate	0.591	Moderate		Significant	Accepted	13
18	100.00	45.24	Moderate	0.650	High		Significant	Accepted	14
19	100.00	47.62	Moderate	0.878	Very high		Very significant	Accepted	15
20	36.36	52.38	Moderate	0.122	Very low		Not significant	Rejected	-

Review from each aspect of cognitive level shows that there are 2 test items that represent C1 (remember) consist of item number 4 and 8; there are 5 test items that represent C2 (understand) consist of item number 1, 2, 9, 12 and 14; there are 3 test items that represent C3 (apply) consist of item number 11, 13, 17; there are 3 test items that represent C4 (analyze) consist of item number 3, 5 and 10; there are 2 test items that represent C5 (evaluate) consist of item number 6 and 15. A blueprint of students' conceptual understanding test after being validated is provided on Table 3.3.

Table 3.3 Blueprint of students' conceptual understanding test

No	Indicator	Item number distribution					Total item
		C1	**C2**	**C3**	**C4**	**C5**	
1.	Students are able to estimate the position of fulcrum/load/force of lever system to achieve balance.	4	1, 9				**3**
2.	Students are able to explain the meaning of mechanical advantage of lever system.	8					**1**
3.	Students are able to compare mechanical advantage of lever system.		2				**1**
4.	Students are able to associate the principle of lever system in human body with simple tool in daily life.		12, 14				**2**
5.	Students are able to calculate the balance of lever system mathematically.			17, 11			**2**
6.	Students are able to apply the principle of lever system to their daily life.			13			**1**
7.	Students are able to analyze mechanical advantage of lever system.				5, 3		**2**
8.	Students are able to analyze the principle of lever system in human body.				10		**1**
9.	Students are able to conclude work principle of lever system.					6, 15	**2**
Total		2	5	3	3	2	**15**

b) Reliability of test

Reliability deals with the consistency of the result. That is how consistent test scores or other evaluation results are from one measurement to the other. If a test is reliable, then a students' score on it when compared to the scores of his classmates, should be similar to his relative score on the other test measuring the same information. Moskal and Leydens (2000) stated that reliability refers to the degree to which an assessment tool produces stable and consistent results. Reliability measures provide an estimate of how much variation that might expect under different conditions. If the test is reliable, it indicates that the first test and the next test are on the same measure. The result of reliability measurement through *Anates V4* program were interpreted based on test reliability criteria were provided on Table 3.4.

Tabel 3.4. Reliability criteria of test

Reliability score	Category
0,81 - 1,00	Very high
0,61 - 0,80	High
0,41 - 0,60	Moderate
0,21 - 0,40	Low
0,00 – 0,20	Very low

(Surapranata, 2009)

Based on the result of test reliability, the measured reliability of conceptual understanding instrument through *Anates V4* program is 0.89 that was interpreted as very high. This result shows that conceptual understanding test has very high reliability which indicates that the first test and the next test are on the same measure.

c) Difficulty level of test item

A good test is a test which is not too easy or vice versa too difficult to students. The difficulty of the test item is indicated by the percentage of students who get the item right. The more difficult items, the fewer will be the students who select the correct option. And the easier the items are the more will be the students who select the correct one. Very easy items are to build in some affective feelings of "success" among lower ability students and to serve as warm up items, and very difficult items can provide a challenge to the highest-ability students (Brown, 2005). The number that shows the level difficulty of a test can be said as difficulty index (Arikunto, 2006). In this index there are minimum and maximum scores. The lower index of a test, the more difficult the test is. And vice versa, the higher the test, the easier it is. The result of difficulty level measurement through *Anates V4* program were interpreted based on test item difficulty level criteria were provided on Table 3.5.

Tabel 3.5. Difficulty level of test items

Percentages (%)	Category
81 - 100	Very easy
61 - 80	Easy
41 - 60	Moderate
21 - 40	Difficult
0 – 20	Very difficult

Based on the result of test item difficulty level of conceptual understanding test, there are 3 items were categorized as very easy, 2 items were categorized as easy, 14 items were categorized as moderate, and 1 item was categorized as very difficult.

d) Discrimination power of test item

Item discrimination or discrimination power explains how well the items perform in separating the better students from the poorer ones. Discrimination is important because if the test-items can discriminate more, they will be more reliable (Hughes, 2005). It can be defined also as the ability of a test to separate master students and nonmaster students (Arikunto, 2006). A master student is a student with higher scores of test, and a non-master student is a student with lower scores on the test given.

The same as the term of difficulty level, discrimination has discrimination index. It is an indicator of how well an item discriminates between weak candidates and strong candidates (Hughes, 2005). This index is used to measure to the ability of a test in discriminating the upper and lower group of students. Upper students are students who answer with true answer, and lower group are students with false answer. In this index, it has negative point. Different from difficulty index, the negative index of discrimination power shows that the questions identify high group students as poor students and low group students as smart students. A good question is a question that can be answered by upper group and cannot be answered with true answer by lower group. This is the statement underlying the index of discrimination. The results of discrimination index calculation through *Anates V4* program were interpreted based on discrimination index of test item criterion on Table 3.6.

Tabel 3.6. Discrimination power index criteria

Discrimination power (%)	Category
70 - 100	Very good
41 - 70	Good
21 - 40	Sufficient
0 – 20	Poor
Negative	Test item is not good

Based on the result of discrimination index calculation of conceptual understanding test, there are 2 items were categorized as poor discrimination power, 3 items were categorized as sufficient discrimination power, 3 items were categorized as good discrimination power, and 12 items were categorized as very good discrimination power.

2. Students' Engineering Design Behavior

Informed Design Learning and Teaching Matrix (Crismond and Adams, 2012) was adapted and used to characterize observable patterns of student engineering design behaviors within engineering design process during the utilizing of instructional materials. The Informed Design Learning and Teaching Matrix contains nine engineering design strategies and associated patterns that contrast beginning versus informed design behaviors, with links to engineering design process consists of *Pikir* (think), *Desain* (design), *Buat* (Create) dan *Uji* (Test) that aim to support students in developing their engineering design abilities. The Informed Design Learning and Teaching Matrix is provided on Table 3.7.

Table 3.7 Informed design learning and teaching matrix

Phase of Designing	Design Strategies Indicator	Engineering Design Behavior		Rating/scale label (circle)
		What beginning designers do	What informed designers do	
Think (*Pikir*): Learners will define a problem, recognize the need for a new product, establish criteria for success in which the specifications a design solution must meet to be considered successful, gather pertinent information about product functions and features among other things, generating new ideas that may solve the problem and analyze solutions and then decide which solution is best suited for implementation.	Understanding the design challenge	**Problem solving:** Do not grasp the basics of design task, or treat it as a well-defined, straightforward problem that they prematurely attempt to solve.	**Problem framing:** Understand basics of design problem, and then delay making design decisions in order to explore, comprehend and frame the problem better.	1 - 2 - 3 - 4
	Build knowledge and do research	**Skipping research**: Skip doing research and instead pose or build solutions immediately.	**Doing research**: Do investigations and research to learn about the problem, relevant cases and how the system works. thinking, brainstorming, etc.	1 - 2 - 3 - 4
Design (*Desain*): Perform several types of analysis on each design. Analysis that may need to be considered consists of functional analysis, ergonomics, strength analysis, testability, product safety and liability, economic and market analysis. Learners then design a product based on the result of analysis.	Generate ideas	**Idea scarcity**: Work with few or just one idea, which they can get fixated or stuck on, and may not want to discard, add to, or revise.	**Idea fluency**: Practice idea fluency in order to work with lots of ideas by doing divergent	1 - 2 - 3 - 4
	Represent Ideas	**Surface drawing and modeling**: Propose superficial ideas that do not support deep inquiry of a system, and that would not work if built.	**Deep drawing and modeling**: Use multiple representations to explore and investigate design ideas & support deeper inquiry into how system works.	1 - 2 - 3 - 4
	Weigh Options	**Ignore benefits and tradeoffs**:	**Balance benefits and tradeoffs**:	1 - 2 - 3 - 4

	& Make Decisions	Make design decisions without weighing all options, or attend only to pros of favored ideas, and cons of lesser approaches.	Use words and graphics to display and weigh both benefits and tradeoffs of all ideas before picking a design.	
Create (*Buat*): Learners prepare tools and materials, determine the procedures, and build a prototype of the product-the first fully operational production of the complete design solution. A prototype is not fully tested and may not work or operate as intended.	Conduct Experiments	**Confounded tests and experiments**: Do few or no experiments on prototypes, or run confounded tests by changing multiple variables in a single test.	**Valid tests and experiments**: Conduct valid experiments to learn about materials, key design variables and the system work.	1 - 2 - 3 - 4
Test (*Uji*): Testing and verification are important parts of the design process. At all steps in the process, it may find that potential solution is flawed and have to back up to a previous step to get a workable solution. Test the prototype extensively under real conditions in order to identify the part that would have to be redesigned and the process completed until a satisfactory solution was reached.	Troubleshoot	**Unfocused troubleshooting**: Use an unfocused, non-analytical way of viewing prototypes during testing and troubleshooting ideas.	**Diagnostic troubleshooting**: Focus attention on problematic areas and subsystems when troubleshooting devices and proposing ways to fix them.	1 - 2 - 3 - 4
	Revise/Iterate	**Haphazard or Linear designing**: Design in haphazard ways where little learning gets done, or do design steps once in linear order.	**Managed & Iterative Designing**: Do design in a managed way, where ideas are improved iteratively via feedback. Strategies get used as many times as needed, in any order.	1 - 2 - 3 - 4
	Reflect on Process	**Tacit design thinking**: Do tacit designing with little self-monitoring while working or reflecting on	**Reflective Design Thinking**: Practice reflective thinking by keeping tabs on design strategies and thinking.	1 - 2 - 3 - 4

		process.		

There are nine indicators that were observed on this research consist of understand the challenge, build knowledge and do research, generate ideas, represent ideas, weigh options and make decisions, conduct experiments, troubleshoots, revise/iterate and reflect on process. Rubric of students' engineering design behavior by Crismond and Adams (2012) was adapted by adjusting each indicator of students' observable engineering design behaviors during instructional process with engineering design behaviors that were stated on the rubric. Each of indicators was rated from 1 to 4 where each rating represents the level/category of students engineering design behavior. The categorization of students engineering design behavior was presented in Table 3.8 and rubric of students' engineering design behavior was presented in table 3.9.

Tabel 3.8 Categorization of students engineering design behavior

Scale label	**Category**
1	Beginning designer
2	Emerged designer
3	Developing designer
4	Informed designer

Tabel 3.9 Rubric of students' engineering design behavior

Category	Scale label			
	Beginning designer 1	**Emerged designer 2**	**Developing designer 3**	**Informed designer 4**
Understand the challenge: Problem Solving Vs. Problem Framing	• Do not grasp the basics of design problem or treat it as a well-defined • Act prematurely and attempt to solve design challenge immediately • Believing there is a single correct answer without research	• A little bit grasp the basics of design problem and treat it as a well-defined • Understanding the design challenge is straightforward • Delay making design decision only to frame the problem • Believing there is more than a single correct answer without research	• Almost understand basic of design problem and treat it as a well-defined • Understand the challenge as best they can • Delay making design decisions in order to explore, comprehend and frame the problem better. • Believing there is more than a single correct answer and learn through a little bit research	• Understand basics of design problem • Understand the challenge as best they can • Delay making design decisions in order to explore, comprehend and frame the problem better. • Believing there is more than a single correct answer and set out to learn through research, brainstorming, and doing technological investigations.
Build knowledge and do research: Skipping Vs. Doing Research	Skip doing research in favor of generating solutions immediately	• Start doing research/ look for outside help before attempting a solution to learn about the problem • Use nearby objects as the main source of inspiration for design solutions	• Do investigations and research to learn about the problem and how the system works • Use nearby objects as the main source of inspiration for design solutions, while at times yielding creative and effective solutions	• Do research on users, write product histories, and collect information on manufacturing methods, materials, and product standards to build understandings of the problem and potential solutions.
Generate ideas: Idea Scarcity Vs. Idea Fluency	• Work with just one idea • Get fixated or stuck on with just one idea	• work with just one idea • Get fixated or stuck on with one idea	• Work with few ideas as alternatives from which to choose	• Practice idea fluency in order to work with lots of ideas

	• Do not want to discard, add to, or revise the idea	• Want to revise the idea	• Do not stuck on with one idea • Explore possible design solution through brainstorming	• Explore possible design solution through brainstorming • Doing divergent thinking (doing idea sketching and visual recall, incubating ideas by stopping work on a problem for a while, and generating personal and direct analogies)
Represent ideas: Surface Vs. Deep Drawing & Modeling	• Propose superficial ideas that do not support deep inquiry of a system • Propose superficial ideas that would not work if built.	• Propose superficial ideas that do not support deep inquiry into how system works • Propose superficial ideas that would work if built.	• Use multiple representations (use multiple sides drawing, words, and artifacts) to support deeper inquiry into how system works and communicate their design plans • Propose ideas that would work if built	• Use multiple representations (use gestures, words, and artifacts) to support deeper inquiry into how system works and communicate their design plans • Make drawings, construct physical prototypes, and create virtual models that help them develop deeper understandings of how their designs function.
Weigh options & make decisions: Ignore Vs. Balance Benefits & Tradeoffs	• Ignore attention to design criteria and constraints (decisions may be made based on criteria that are unstated) • Focus only on positive or negative aspects of design ideas without thinking of associated benefits and trade-offs	• Pay a little attention to design criteria and constraints • Make design decisions based on both positive and negative aspects of design idea.	• Weigh both benefits and tradeoffs of all ideas before picking a design. • Consider various plans, make design decisions, and justify them	• Weigh both benefits and tradeoffs of all ideas before picking a design. • Consider various plans, make design decisions, and justify them. • Drawbacks of ideas that they are about to select or reject and look for potential downsides even with the most promising ideas

Conduct experiment: Confounded Vs. Valid Tests & Experiments	• Do few or no experiments on prototypes • Run confounded tests by changing multiple variables in a single test which yields little understanding about potential solutions.	• Do few experiments on prototypes • Run valid test but still get little understanding about potential solutions.	• Do more experiments on prototypes • Conduct valid experiments only to understand how things work	• Run valid tests as part of technological investigations that help them to learn quickly about design variables, users, and materials, to understand how things work, and to optimize the performance of the prototypes they decide to develop.
Troubleshoot: Unfocused Vs. Diagnostic Troubleshooting	• Have an unfocused, non-analytical way of viewing the plans and performance tests of prototypes when troubleshooting their designs. • Do not actively looking out for worrisome patterns when testing prototype performance.	• Have an unfocused, non-analytical way of viewing the plans and performance tests of prototypes when troubleshooting their designs. • Start observing/looking at a product's overall performance during early prototype testing in order to detect unexpected or out-of-range behaviors. • Proposing ways to fix them	• Focus attention on problematic areas and subsystems when troubleshooting devices • Start observing /looking at a product's overall performance during early prototype testing in order to detect unexpected or out-of-range behaviors. • Start conduct diagnosis of the problem, where the designer gives a name to the problems noticed in the product's performance. • Proposing ways to fix them	• Focus attention on problematic areas and subsystems when troubleshooting devices • Start observing /looking at a product's overall performance during early prototype testing in order to detect unexpected or out-of-range behaviors. • Start conduct diagnosis of the problem, where the designer gives a name to the problems noticed in the product's performance. • Give explanation of why those behaviors occur and detect flaws that can inspire ideas for simple fixes, additional features, or entirely new and unimagined systems. • Proposing ways to fix them.
Revise/iterate: Haphazard or Linear Vs. Managed &	• Design in haphazard ways where little learning gets done • Treat design as a set of	• Design in haphazard ways where more learning gets done. • Do design steps twice	• Do design in a managed way where ideas are improved iteratively via	• Do design in a managed way, where ideas are improved iteratively via feedback.

Iterative Designing	strategies to be done once in linear order.	in linear order.	feedback. • Do design steps more than twice in linear order or once in any order • Improving ideas and prototypes based on feedback and cycling back to upgrade their understanding of the problem.	• Improving ideas and prototypes based on feedback and cycling back to upgrade their understanding of the problem. • Manage time and resources strategically • Strategies get used as many times as needed, in any order.
Reflect on process: Tacit Vs. Reflective Design Thinking	• Do tacit designing when they think and act with little or no self-reflection • Do little monitoring of their own or others' actions • Do not articulate what knowledge they know or need to know to further their investigations, and pay scant attention to the progress they make, obstacles they encounter, or design values that influence their decisions.	• Do tacit designing when they think and act with more self-reflection • Do more monitoring of their own or others' actions • Articulate what knowledge they know or need to know to further their investigations, and pay scant attention to the progress they make, obstacles they encounter, or design values that influence their decisions.	• Practice reflective thinking by keeping tabs on their own and others' design work in a metacognitive way • Articulate what knowledge they know or need to know to further their investigations, and pay scant attention to the progress they make, obstacles they encounter, or design values that influence their decisions.	• Practice reflective thinking by keeping tabs on their own and others' design work in a metacognitive way • Articulate what knowledge they know or need to know to further their investigations, and pay scant attention to the progress they make, obstacles they encounter, or design values that influence their decisions. • Reviewing their processes and products once they have completed their work.

3. Students' Teamwork Skills

Students' teamwork skills were observed by using Comprehensive Assessment of Team Member Effectiveness (CATME) that was developed by Loughry, Ohland, and Moore (2007). The CATME collects self-ratings, peer ratings and teacher ratings on five dimensions of team-member contributions that were developed based on the teamwork literature and original empirical research consists of thirty items that load onto five indicators, there are contributing to the team's work, interacting with teammates, keeping the team on track, expecting quality, and having relevant knowledge, skills, and abilities. The CATME likert-short version is provided on Table 3.10 below.

Table 3.10 CATME likert-short observation sheet

	Your name	**Your teammate's name**	**Your teammates's name**	**CATME Likert -Short Section Number _______ Team Number _______** 🡐 Write the first and last names of the people on your team including your own name. This self and peer evaluation asks about how you and each of your teammates contributed to the team during the time period you are evaluating. Please read each item that describes a way of contributing. Then confidentially rate yourself and your teammates using the following scale: 1 Strongly Disagree; 2 Disagree; 3 Neither Agree Nor Disagree; 4 Agree; 5 Strongly Agree
Contributing to the Team's Work				Did a fair share of the team's work.
				Fulfilled responsibilities to the team.
				Completed work in a timely manner.
				Did work that was complete and accurate.
				Made important contributions to the team's final product.
				Kept trying when faced with difficult situations.
				Offered to help teammates when it was appropriate.
Interact with team				Communicated effectively.
				Facilitated effective communication in the team.
				Exchanged information with teammates in a timely manner.
				Provided encouragement to other team members.
				Expressed enthusiasm about working as a team.

				Heard what teammates had to say about issues that affected the team.
				Got team input on important matters before going ahead.
				Accepted feedback about strengths and weaknesses from teammates.
				Used teammates' feedback to improve performance.
				Make sure that teammates make a good improvement.
				Let other team members help when it was necessary.
Keeping the Team on Track				Stayed aware of fellow team members' progress
				Assessed whether the team was making progress as expected.
				Stayed aware of external factors that influenced team performance.
				Provided constructive feedback to others on the team.
				Motivated others on the team to do their best.
				Made sure that everyone on the team understood important information.
				Helped the team to plan and organize its work.
Expecti ng Quality				Expected the team to succeed.
				Believed that the team could produce high-quality work.
Having Relevant Knowledge, Skills, and Abilities				Had the skills and abilities that were necessary to do a good job.
				Had enough knowledge of teammates' jobs to be able to fill in if necessary.
				Knew how to do the jobs of other team members.

4. Instructional Material

a) Non-STEM based Instructional Material

Science instructional material that served as non-STEM based instructional material in this research was science book for junior secondary school published by National Ministry of Science Education in 2008. The title of this book was "*Ilmu Pengetahuan Alam 2: SMP/MTs Kelas VIII*" written by Wasis and Sugeng Yuli Irianto. The topic of lever system was a sub-chapter that contained in chapter four of this book which is explained about force. Science concepts in non-STEM based instructional material were not presented in integrated way. The concept of lever system only explained based on physics point of view. The content of non-STEM based instructional material can be seen at a glance in Figure 3.2 below.

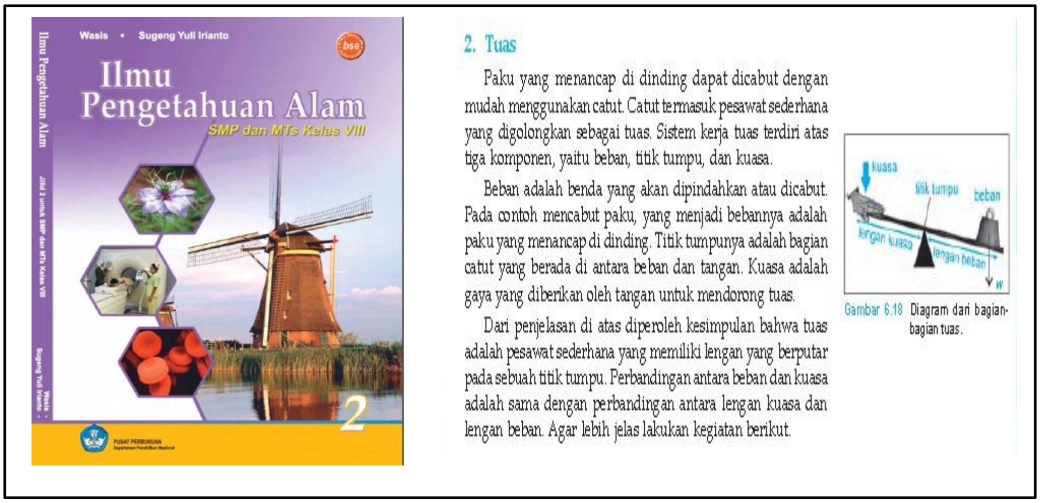


Figure 3.2 Non-STEM based instructional material

The engineering activity for students was prepared and constructed by teacher due to unavailability of students' engineering activity on non-STEM based instructional material. The activity consists of constructing a simple *mangonel* catapult based on work principle of the real *mangonel* catapult. The worksheet allowed students to think about the problem they encountered, design the solution, determine tools and materials and analyze the work principle of their simple *mangonel* catapult that have been designed and constructed by themselves.

b) STEM based Instructional Material

The STEM based Instructional material was developed based on the newest Indonesia curriculum (*Kurikulum 2013*). Instructional material contains STEM concepts in simple machines topic for 8th grader of junior secondary school students. There are science section that talks about the relationship of human body movement with simple machines, technology section that talks about product of science which is related with simple machines, engineering section that challenge students to conduct STEM activity based on engineering design process, and mathematics section that talks about mechanical advantages of simple machines. For assessment purposes, instructional material contains a series of quizzes, students' activity worksheet, pretests and posttests that measure students learning growth before and after reading STEM based instructional material and their response towards STEM based instructional material, respectively. The content of STEM based instructional material can be seen at a glance in figure 3.3.

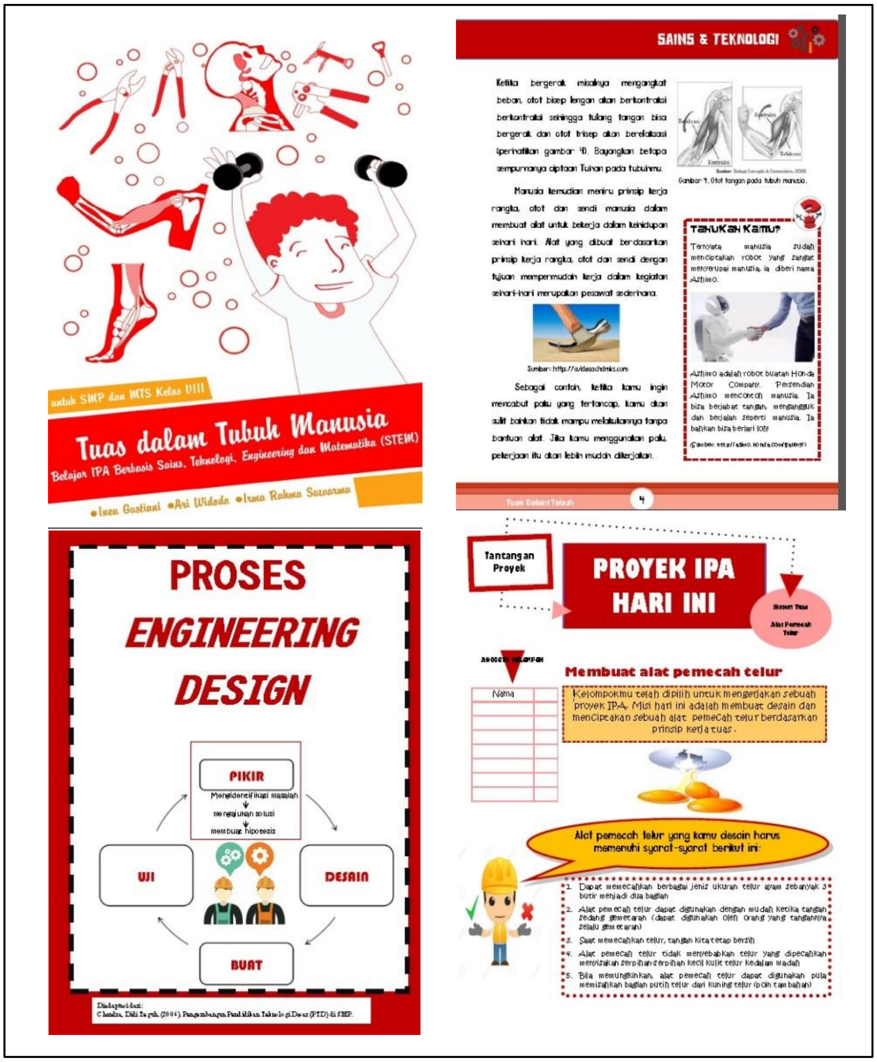


Figure 3.3 STEM based instructional material

In the implementation phase, the developed instructional materials were delivered to a pilot group consist of 8th grader of junior secondary school and the result will be provided in the appendix (appendix C.1, page 204). To measure student learning, a pre-test and post-test were given before and after implementation of the STEM based instructional material. In addition, a questionnaire of readability was given to examine the clarity and difficulty level of each page of instructional material. A questionnaire of students' response towards instructional material given to students and teachers at the end of instructional material reading session to measure layout aspect, content aspect and utility aspect of instructional material for being used in the junior secondary school classroom setting.

The result of STEM based instructional material validation shown that readability aspect of STEM based instructional material is categorized as very high. Students' response towards STEM based instructional material also shows that layout, content and utility of the instructional material achieve very high percentage, pretest and posttest responses revealed that students retained significant amounts information upon completion of the STEM instructional material. Student overall learning gain is 0.67 which is categorized as moderate through a measured by the pretest and posttest results. In essence, the instructional material is valid enough to be used for conducting STEM education in classroom setting. It is recommended for academic/research community for use, and or for further refinement.

D. Research Procedures

This research was conducted through 3 stages which are defined below:

1. Preparation Stage

Activities that were conducted in this stage consist of:

a) Preliminary study and need analysis that consists of an extensive literature study on the subject of implementation as well as appropriate literature related to STEM education, instructional material, engineering design behaviors and teamwork skills, analyzing science curriculum which is used at school and science lesson material for 8th grader of secondary school.

b) Determining research subject and lesson material that will be used in this research.

c) Designing STEM-based instructional material. In this step, the sequences in which the objective will be met are determined. Sequencing the objectives helps to create the outline of the instructional material. After the sequence has been determined, instructional content (information) and activities for each objective identified are selected. The objectives, information, and descriptions of activities are then transferred to storyboards. An outline of the content included identifying different sections of the instructional material and descriptions of the topics list in each section. Once the content outline is developed, an instructional flow is drawn. As a part of this phase, developed materials evaluated to ensure they are correctly designed as intended, and delivered to a pilot group to make sure that the instructional material can be implemented through instructional process in a class.

d) Designing research instrument.

e) Revising research instrument.

f) Preparing research license

2. Implementation Stage

Activities that were conducted in this stage consist of:

a) Conducting pretest in both control group and experiment group to assess students' prior science conceptual understanding about levers in human body topic.

b) Conducting instructional process and implementing STEM based instructional material for experimental group and science book published by National Ministry of Science Education for control group. Both of control group and experiment group were instructed to construct a simple lever system. Students in control group were making a simple *mangonel* catapult and fill the worksheet that was provided by teacher. On the other hand, students in experiment group were making an egg cracker based on the instructions that were provided on STEM based instructional material which leads students to do engineering design process during the activity.

c) Observing students' activity during instructional process. In this process, students engineering design behavior and teamwork skills of each group were observed when generated at least one design solutions addressing a design challenge. Informed Design Learning and Teaching Matrix that were developed by Crismond and Adams (2012) were used to characterize the observable patterns of engineering design behavior. Students conducted self and peer rating based on Comprehensive Assessment of Team Member Effectiveness (CATME) rating scale instrument that was developed by Loughry, Ohland, and Moore (2007) to assess their teamwork skills. During the group activities, students were instructed to rate themselves and their teammates based on teamwork performance, at the same time teacher also rate students' teamwork performance during group activity.

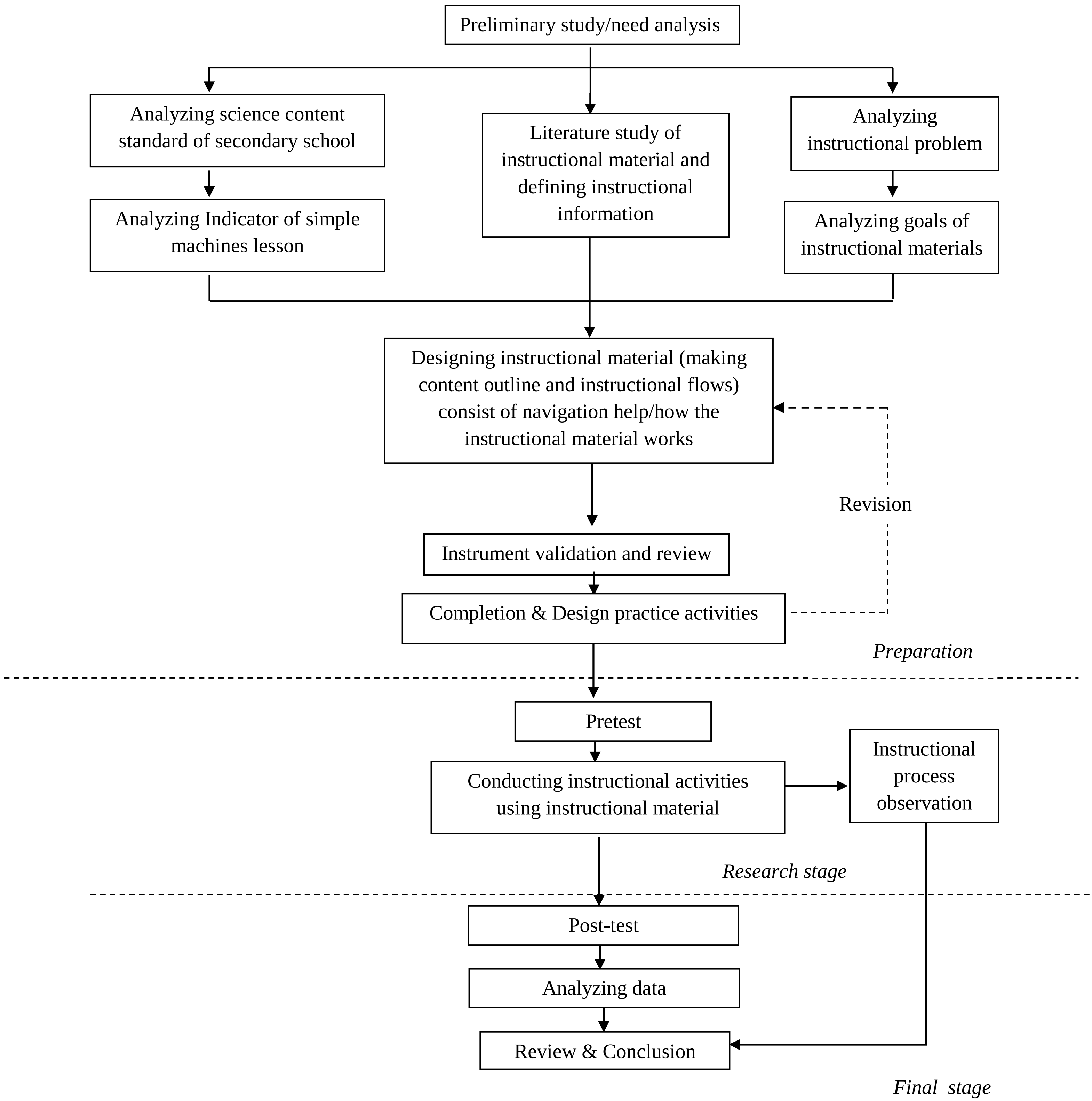
d) Conducting posttest in both control group and experiment group to assess students' conceptual understanding about levers in human body topic after learning through STEM based instructional material for experimental group and science book published by National Ministry of Science Education for control group.

3. Final Stage

Activities that were conducted in this stage consist of:

a) Collecting and processing data
b) Analyzing data and taking conclusion
c) Arrange and presenting thesis
d) Revising thesis

E. Scheme of Research



F. Data Processing Technique and Analysis

Data that were collected through research consist of quantitative data and qualitative data. Quantitative data comprises of pretest and posttest result, meanwhile qualitative data comprises of observation result of students' engineering design behavior and teamwork skills. Quantitative data analysis was intended to identify the improvement of students' conceptual understanding before and after the implementation of instructional material. Data processing technique and analysis will be explained more detail as follow.

1. Students' Conceptual Understanding Data Processing Technique and Analysis

Pretest and posttest results that were collected through students' conceptual understanding test were analyzed descriptively and inferentially to test the hypotheses. The analysis of students' conceptual understanding data were conducted through the following steps.

a) Scoring of pretest and posttest results

All of students' pretest and posttest results were examined and scored. Each of right answer was given score 1 (one) and each of wrong answer was given score 0 (zero). The scoring pretest and posttest results were calculated through the following formula:

$$S = \sum R$$

Where:

S = Students' score

R = Students' correct answer

Students' conceptual understanding score on pretest and posttest were converted into score with scale 0-100. The score then served the purpose to test the hypotheses. The score was calculated through the following formula:

$$Score = \frac{\Sigma\, amount\ of\ gained\ scores}{\Sigma amount\ of\ total\ scores}\ x\ 100$$

b) Calculating N-gain

The improvement of students' conceptual understanding could be identified through the comparison of N-gain scores. The calculation of N-gain score was conducted through the formula that was developed by Hake (1999) as follow.

$$\text{N – Gain} = \frac{\text{posttest score} - \text{pretest score}}{\text{perfect score} - \text{pretest score}}$$

The results of N-gain scores were interpreted based on Table 3.11 as follow.

Tabel 3.11 Classification of N-Gain interpretation

Score	Interpretation
<g> > 0.7	High
0.3 ≤ <g> ≤ 0,7	Moderate
<g> < 0,3	Low

(Hake, 1999)

c) Analysis of pretest

Pretest data of students' conceptual understanding were analyzed using descriptive statistics to analyze whether both of control group and experiment group has an equal initial condition. The result of pretest analysis shows that there was no significance difference of students' pretest score between control group and experiment group before treatment. With the result that testing hypotheses to examine whether the mean differences of students' conceptual understanding improvement between experiment group and control group has significance difference or the other way was using posttest data.

d) Statistical test

Hypotheses testing were conducted to examine the significance differences of students' conceptual understanding improvement between experiment group and control group. Hypotheses testing were conducted by calculating posttest data because pretest data shows that there is no significance difference of students' conceptual understanding between control group and experiment group. The hypotheses testing were conducted through these following steps:

1) Normality test

A normality test should be conducted as prerequisite testing before conducting hypotheses testing of students' conceptual understanding posttest data. Saphiro-Wilk test through *IBM SPSS 22* software were conducted in order to test the assumption of normality. The normality test was intended to examine distribution spread of data. The criteria of testing the assumption of normality is *sig* > α (α = 0.05), therefore the data was normally distributed. The following hypotheses were stated to conduct normality test.

H_0 : Data were originated from normally distributed population.

H_1 : Data were not originated from normally distributed population.

2) Homogeneity of variance test

The assumption of homogeneity of variance was tested through the Levene's F test for equality of variances by using *IBM SPSS 22* software. Homogeneity of variance test was intended to examine the variance equality data between control class and experiment class. The criteria of testing the assumption of homogeneity is *sig* > α (α = 0.05), therefore the data was homogeny. The following hypotheses were stated to conduct homogeneity of variance test.

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

Where:

σ_1^2 = data variance of experiment group

σ_2^2 = data variance of control group

3) Statistical test

Due to the result of test of normality and homogeneity of variance that met the assumption (data were normally distributed and with the same variance) and the data are independent of each other, independent-sample t-test in 95% confidence interval will be conducted to test the hypotheses and see the mean difference of students' conceptual understanding improvement between control group and experiment group. The criteria of independent-sample t-test is *sig* > α (α = 0.05). Hypotheses testing will be conducted through *IBM SPSS 22* software. Hypotheses formulation was stated as follow.

Statistical hypotheses:

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Where:

H_0 = there is no significance difference of students' conceptual understanding improvement between control group and experiment group.

H_1 = there is significance difference of students' conceptual understanding improvement between control group and experiment group.

μ_1 = average posttest or N-gain score of students' conceptual understanding in experiment group.

μ_2 = average posttest or N-gain score of students' conceptual understanding in control group.

2. Students' Engineering Design Behavior Data Processing Technique and Analysis

Students' engineering design behavior was analyzed descriptively based on the data that were collected through observation during engineering design process that consist of *Pikir* (think), *Desain* (design), *Buat* (Create) and *Uji* (Test). Engineering design behavior of each group were observed twice (first phase and second phase) during engineering design process activity to see which category of engineering design behavior that students possess in each phase and whether each group made progress from the first phase until the second phase. Rubric of engineering design behavior will be served as basis of observation to identify the level of students' engineering design behavior of control group and experiment group for each indicator from the first phase until the second phase.

students' engineering design behaviors profile were analyzed by calculating and contrasting the percentages of students that were categorized in each level of engineering design behavior between control group and experiment group in first phase and second phase. Meanwhile, students' engineering design behavior development of each indicator were analyzed by plotting a diagram that shows engineering design behavior level development from first phase until second phase between control group and experiment group.

3. Students' Teamwork Skills Data Processing Technique and Analysis

Students' teamwork skills were analyzed based on Comprehensive Assessment of Team Member Effectiveness (CATME) likert scale data that were collected through observation during instructional process. Students' teamwork skills of each group were observed based on three kinds of rating consists of self-rating where each students rate themselves from scale 1 until 5 regarding to their performance in teamwork based on CATME observation sheet, peer-rating where each students rate each of their teammates and teacher rating where teacher rate teamwork performance of each students based on CATME observation sheet. The result of students' teamwork skills rating-scale were interpreted through students' teamwork skills level category based on Team Assessment Report (2015) that were provided on Table 3.12.

Tabel 3.12 Students' teamwork skills category

Average rate	Category
<average rate> < 3.2	Low
3.3 ≤ <average rate> ≤ 3.6	Medium
<average rate> > 3.7	High

The total average rating of students' teamwork skills between control group and experiment group were compared to see the overall result of students' teamwork skills difference between control group who learnt science through non-STEM based instructional material and experiment group who learnt science through STEM based instructional material. Analysis of students' teamwork skills for each indicator in more detail will be conducted through calculating and comparing the average rating of students' teamwork skills between control group and experiment group for each indicator. To measure Inter-Rater Reliability of students' teamwork skills, two-way random types of the Intraclass Correlation Coefficient (ICC) analysis (Appendix D, page 246) was used because this model assumes that rater which is involved comes from research population and each rater rates the same subject. Landis and Koch (1977) propose classifying reliability based on the magnitude of a reliability coefficient ($\bar{r}$) in Table 3.13.

Tabel 3.13 Students' ICC category

Intraclass coefficient value ($\bar{r}$)	Category
($\bar{r}$) = 0	Non-existing
0 < ($\bar{r}$)< 0.2	Slight
0.2 < ($\bar{r}$) < 0.4	Fair
0.4 < ($\bar{r}$) < 0.6	Moderate
0.6 < ($\bar{r}$) < 0.8	Substantial
0.8 < ($\bar{r}$) < 1.0	Almost Perfect
($\bar{r}$) = 1	Perfect