ARTICLE

The Development of Moral Reasoning in the Law Curriculum - An Exploration of Various Teaching Activities

Special issue on Education in (Professional) Legal Ethics, Emanuel van Dongen & Jet Tigchelaar (eds.)

*Emanuel van Dongen & Steven Raaijmakers**

Abstract

Developing the capacity for moral judgment is an essential professional competence for lawyers. What teaching and learning activities in the law curriculum can be used in order to contribute to students' moral reasoning and moral judgment? Four teaching methods were tried out in the period 2019 to 2021 at the Utrecht University School of Law: teaching methods that either work with (hypothetical) dilemmas (I); in-class reflection papers (II); experiential learning based on own experiences in a simulation situation (III); or clinical teaching in a real law firm (IV). The effects of these methods on the development of moral reasoning were measured using the Defining Issues Test (DIT). Additional information on the effectiveness and utility of the method was gathered using semi-structured interviews with teachers. The DIT results were compared at the beginning and at the end of the courses and related to the teaching methods. This article presents the outcomes of this study and formulates some recommendations for further research on the topic.

Keywords: moral reasoning, legal education, scholarship of teaching and learning, defining issues test.

1. Introduction

It is important that law students gain insight into professional ethical problems and dilemmas in legal practice from an early stage in their career so that their later practice of the law can be approached in a more reasoned, critical and responsible

* Dr. Emanuel van Dongen is Associated Professor Private Law at the Molengraaff Institute for Private Law, researcher at the Utrecht Centre for Accountability and Liability Law and the Montaigne Centre for Rule of Law and Administration of Justice, Utrecht School of Law. Dr. Steven Raaijmakers is an educational consultant with Educational Consultancy & Professional Development, Faculty of Social and Behavioural Sciences, Utrecht University.

Emanuel van Dongen & Steven Raaijmakers

manner. Developing the capacity of moral judgment, as part of decision-making on legal matters, is an essential professional competence for lawyers (see Van Dongen & Tigchelaar, 2021). However, this competence does not develop by itself, and (educational) support is needed. Introducing (professional) ethics courses during academic study can be a powerful catalyst for the development of moral reasoning. Moral reasoning is needed for moral judgment. Students' moral reasoning can be stimulated both by making it an explicit part of the curriculum taught throughout their studies and by ensuring that moral reasoning is part of and appreciated throughout the educational process (Chapman, 2002). Research also shows that well-designed curricula can contribute significantly to improving moral reasoning (Rhode, 1992) – especially given that moral values and strategies change in early adulthood (Rhode, 2007). It has, however, been argued in Anglo-American literature that law students become cynical and uncritical because of their experiences at university (see Chapman, 2002). Sheldon and Krieger (2004) found that law students' endorsement of intrinsic values declined over the course of their first year, specifically moving away from community service values and towards appearance and image values. These studies are in line with earlier studies showing declines in law students' preference for 'altruistic law practice' (Landsman & McNeel, 2004). From these studies it might be argued that law school has a desensitizing effect on law students. However, these studies did not deal with Dutch law students. According to Chapman (2002), there are indications that legal ethics courses may help overcome the desensitizing effect of law school.[1] It is, however, unclear what elements of legal ethics courses cause this desensitizing effect.

Inspired by the international literature on legal ethics, various teaching methods and learning activities have been tried out in the law curriculum at Utrecht University School of Law. In the current study, the effects of four different teaching methods are explored to ascertain whether these methods contribute to students' moral reasoning. Besides the development of the moral reasoning capacity of students, the approach to teaching and the role of the teacher were also considered.

In this article, first, a discussion of the relevant educational literature will be provided, i.e. on moral reasoning and the effects and possibilities of teaching professional ethics. Second, the teaching background and context as well as the five pilots where the four teaching methods were put into practice, i.e. what the pilots looked like in practice, will be explained. Third, the research methods will be discussed, including the reasons for choosing the particular research instruments, followed by the results and evaluation. Finally, a discussion and conclusion follows with some recommendations for best practices and lessons concerning the respective contributions to students' moral reasoning and attitude, including the role of the teacher and moral reflection.

---

1   This brief summary is based on Van Dongen and Tigchelaar (2021). The present study is a continuation of that contribution in the sense that it has evaluated and measured the pilots described there. The content of these pilots is also briefly described in this contribution.

2. Educational Context

Lawyers, when confronted with professional ethical problems and dilemmas in legal practice, must be able to act in a morally acceptable manner. Morality, however, is a complex and multifaceted concept (Landsman & McNeel, 2004). Rest distinguishes four interrelated and interacting components in the question of what must happen before someone engages in moral behaviour: moral sensitivity, moral judgment, moral motivation and implementation skills (to execute and implement what one ought to do; Rest, 1984). The moral judgment component is the most researched; this component also fits well with the educational context (Landsman & McNeel, 2004) as moral judgment capacity is a central element in ethics education (van Dongen & Tigchelaar, 2021). The major function of this component is to formulate what a moral course of action would be. For assessing and providing insight into the development of moral judgment in law students, the six stages of moral development as described by the cognitive-developmental theory of moral development by Kohlberg (1984) provide a frame of reference.

A criticism of Kohlberg's work and the concept of moral reasoning is that moral reasoning cannot be assumed to translate into good actions. Of course, moral reasoning may support decision-making and moral action, but there is evidence that moral reasoning is not a good predictor of moral action, known as the 'gappiness problem' (see, for instance, Darnell et al., 2019) – the gap between moral judgement and action. Nevertheless, moral reasoning can be seen as a necessary – yet insufficient – condition for moral action.

Research[2] shows that a pedagogy based on contextual, rich, emotionally engaging, role-based problem-solving, coupled with an ongoing reflective discourse, is likely to significantly improve the effective involvement of law students in ethics and the mastery of the role of an ethical practitioner (Lerner, 2004). Research on student learning and moral development advocates an experiential approach. Ideally, courses on professional responsibility should be linked to on-the-job placements, and ethical issues integrated into the curriculum (Rhode, 2009). By using didactics that combine live client or simulated learning with regular opportunities for critical reflective conversation and writing on ethical issues, university teaching can lead students to think more deeply about ethical dilemmas (Lerman, 1998). Discussions about moral dilemmas appear to be an effective intervention for the level of moral reasoning (and ethical judgment). Just giving lectures about this has little effect. It is about what students do that is important for achieving such an effect (Ferris, 2015), which is in line with the theory of Biggs and Tang (2011).

Measuring moral development has been successfully conducted by the Defining Issues Test (DIT) (Rest et al., 1974;[3] see also Section 4) to a variety of different colleges, universities and graduate students in various professional areas, including law (Landsman & McNeel, 2004). Landsman and McNeel studied law students over

---

2   This paragraph is taken from Van Dongen and Tigchelaar (2021).
3   The DIT was originally based on Kohlberg's theory. The renewed version, the DIT-2, is only loosely based on Kohlberg – in addition, it has been influenced by schema theory (Rest et al., 1999a). See Section 4.3.

a period of three years. They concluded that law school did not have a significant effect on the moral judgment of law students (Landsman & McNeel, 2004). However, according to Hartwell (1995), moral judgment can be improved by interventions such as small intensive seminars or clinics. Interesting, but not very hopeful, is one of the findings of Willging and Dunn, who conclude that a third-year required course in legal ethics did not stimulate statistically significant change in the DIT measures of moral reasoning among the students (Willging & Dunn, 1981). Most literature, however, is Anglo-American (and American Law Schools are more vocational, while law study at Dutch universities has a more academic character) – no study on this topic has yet been conducted on Dutch law students. What is their level of moral judgment? Does this change throughout their university education? What effect do various teaching methods on legal or professional ethics have on moral reasoning?

3. Background of the Teaching Environment

Generally, in the international literature on legal ethics, the various teaching methods intended for stimulating moral reasoning that are discussed can be divided into learning activities involving conversations about moral dilemmas; methods based on hypothetical, narrative or real examples; and experiential learning combined with reflection on one's own experiences. During the years 2019 to 2021, four of these teaching methods were tried out. See Table 1, in which the main teaching method used is mentioned.[4]

Table 1   *Pilots: Teaching methods*

	Working with Dilemmas	In-Class Reflection Papers	Experiential Learning: Simulation	Experiential Learning: Clinics and Reflection	English (EN) or Dutch (NL)	Course Level
Pilot I (honours)	x				EN	MA
Pilot II		x			EN	MA
Pilot III (honours)*	x				NL	BA
Pilot IV				x	NL	BA3/MA
Pilot V			x		NL	BA3

* Pilot three was not intended at the start of this project and was not primarily organized by the teaching staff (although frequent consultation took place with an honours teacher). However, as it focuses on legal ethics, and both legal scholars and practitioners gave lectures on this course, with little preparatory work for students, it was interesting to study what effects this had on the development of moral reasoning of students.

4   Three of the five pilots were taught by two teachers, of whom at least one also had practical experience (I, III, IV and V), as a lawyer, for instance. The other pilot was taught by a lecturer who has both practical and theoretical experience and know-how (II). In three of the pilots, one of the lecturers was very experienced in legal theory and ethics (II, III and V).

The first pilot consisted of two lectures, or more precisely, a mix between lecture and seminar, delivered within two weeks, requiring preparatory reading and an obligatory (ungraded) assignment. At the first meeting, the various ethical schools of thought were introduced. An exercise was conducted using ethical dilemmas taken from practice that were first discussed in subgroups of about two or three students and later presented in a plenary session of about 23 students. In addition, a presentation was given by the lecturers on professional ethics, including examples from the teacher's own experience. Students then had to read about these schools of thought before coming up with an ethical problem and had to apply an ethical framework and present it as a subgroup to all the students.

The second pilot was a short course of five meetings at the master's level, where students had to write reflections based on hypothetical or narrative examples. This so-called 'caput', *Philosophy and Ethics of International Law*, started with a discussion of the major ethical schools of thought. Students then had to write an opinion piece on current events in international law, taking a philosophical and/or ethical perspective. They also had to think about their desired future field of employment and the position they would like to hold. They then had to describe a potential legal-ethical dilemma they might be confronted with in that field and in that position, discuss it with peers and come up with a solution. Finally, the course consisted of discussions during the meetings with short reflection papers, which were discussed afterwards. The course aimed to raise awareness of an extra ethical layer in addition to the letter of the law, which also has implications. This element emphasizes the normative context of the law.

In the third pilot, a student and lecturer of the bachelor's honours programme (the selective (demanding) three-year programme aimed at excellent law students called 'Utrecht Law College') organized a series of seven lectures on professional ethics, in which guest speakers, academics and legal practitioners highlighted moral issues from the legal professional practice. It was a voluntary series of lectures, for which they did not receive a grade. Lecturing was done partly on the basis of personal experience, where moral dilemmas were linked in part to ethical theories. Core values, linked to professional roles, were also discussed as well as more general values and virtues associated with good governance and citizenship.

The fourth pilot stands out as being really different. It concerned a one-year, selective fellowship at the Utrecht Law Clinic. As fellows, students provided legal advice on several occasions to existing (Dutch) companies in the Utrecht region under the supervision of a lawyer. Students were guided intensively in this process by academic staff from Utrecht University and the law firm Van Benthem & Keulen. The Law Clinic's 2020-2021 Fellow Programme consisted of two parts: training and legal advice. The 'Fellow Programme' associated with the clinic offers training in legal skills in various areas (advisory skills, dealing with clients and other legal aid providers, commercial skills, disciplinary law, etc.). Legal ethics is one of the topics of the programme.

In the fifth pilot, bachelor honours students took part in a simulation game (i.e. the start-up of a Corona app) – a complex, socially relevant, legal simulation game about a start-up, involving various legal roles. Students were explicitly asked to take a legal and ethical approach to their role in their professional product (e.g.

Emanuel van Dongen & Steven Raaijmakers

legal advice) and to reflect on this. In the simulation game, the students were confronted with conflicts between different ethical theories and types of values and standards. Part of this process also included three meetings in smaller groups to discuss (legal) ethics. At the ethics meetings, not only the personal, institutional and professional values but also constitutional ones and values as a citizen were discussed. The moral aspects in the case of the simulation game were scrutinized as well as how ethical theories can help determine a position both in the case of the simulation game (i.e. the start-up of a Corona app) and outside the case, as a lawyer, using illustrations from practice. Finally, the discussion about (professional) ethics and legal philosophy was placed in a broader context, after which the students were presented cases with major inherent dilemmas. One of these was about the scarcity of intensive care beds during the covid-19 pandemic. Suppose someone went – against the government's warnings – 'partying' across the national border, where less strict Corona measures applied, got infected and turned out to need such a bed. However, for the allocation for such a bed, a statement from a housemate is required, stating that this person had complied with the measures. This leads to an ethical dilemma for the housemate: should he produce such a statement?

4. Research Methods

4.1. Research Question and Anticipated Outcomes

The remainder of this article will present the outcomes of the systematic study of the results of the five pilots on moral reasoning of law students. The aim of the study was to compare the four teaching methods used in these five pilots (see Section 3), using the Defining Issues Test (DIT-2), to determine which method(s) contribute(s) to students' capacity for moral reasoning. The research questions were as follows: I. Do the four teaching methods have a positive effect on the development of moral reasoning of law students? II. What do teachers believe to be the strong points of the pilots in relation to their influence on students' moral reasoning, and what conditions are necessary to come to fruition?

We expected that students' moral reasoning would improve when students were actively involved in solving issues or spent more time on the issues compared with learning activities where students were only passively involved or briefly working on ethical issues. Furthermore, we expected that the nature of the task (mandatory or not) and the role of the teacher were important for the way in which students engaged with these matters. Therefore, we expected that the approach taken by the teachers would be an important factor. It should be noted that no explicit assignment was given to the teachers beforehand about which approach to take when teaching legal ethics. Based on these criteria, we anticipated that the fourth and fifth pilot (examples of experiential learning over a longer period, with more time for tasks compared with the other pilots) would score highest when it came to students' moral development. The fourth pilot, in particular (clinics), entailed both 'taking responsibility for others in society' as well as 'making non-hypothetical, irreversible moral choices' – both criteria for moral reasoning (Kohlberg – see Chapman, 2002, p. 83). The qualitative data will confirm whether or not lecturers agree with these expectations.

4.2. Methods and Data Collection

Multiple methods were used in combination in this study. First of all, the DIT was used to measure the development of moral reasoning in bachelor's and master's law students. By means of paired t-tests (pre- and post-measurements – on the level of schemas, patterns of thought, see Section 4.4, and on overall level), we tried to identify which learning activity yields the most gain in moral reasoning (five groups, pre- and post-measurements) in Statistical Package for the Social Sciences (SPSS). Although the pilots started with larger groups of students, we were unable to collect both pre- and post-measurements for all the students. Therefore, we report only on the data of students who have both pre- and post-measurements. Additional information about the effectiveness and utility of the method was gathered using semi-structured interviews with the lecturers (e.g. do you think the method contributed to the moral development of students? Was the method easy to implement?). Their experiences and the reflections of the lecturers will be used to interpret the results: what exactly were the elements that proved to be effective (in the quantitative part of our study)? Our intention was to link the quantitative results (based on the DIT) to the qualitative results from the interviews. However, owing to a low response rate on the questionnaire, we have decided to analyse these results independently. Furthermore, the interviews were analysed together instead of separately. Therefore, our conclusions from the interviews should be related to the teaching of moral reasoning as a whole, not to any one method individually.

4.3. Choice of Measurement Instrument

We chose one of the most commonly used tools for measuring moral reasoning – derived from theory and empirically validated – namely, the DIT (Rest et al., 1999b).

The DIT has been shown to be sensitive to educational interventions (Rest et al., 1999b, p. 647), even small short-term educational interventions (e.g. Roche & Thoma 2017). The development from 'lower' to 'higher' levels is empirically supported (see e.g. Landsman & McNeel, 2004). Educational interventions are useful and are reflected in the P-score (moral reasoning level; Schlaefi, Rest & Thoma, 1985). Standards for the DIT-2 are present, and research is being conducted with the use of the DIT at Utrecht University. An interesting aspect of the DIT is whether the scenarios should be context specific or not (Doyle, Frecknall-Hughes & Summers, 2009). Moral reasoning in a client-lawyer context might work differently than in a more general context. Additionally, law students might perceive the issues differently from other students. This might decrease the validity of the DIT. At this moment, there are no indications in the literature to assume that this is the case.

The DIT has been effectively used to assess moral reasoning, i.e. the 'psychological construct that characterizes the process by which people determine that one course of action in a particular situation is morally right and another course of action is wrong' (Rest et al., 1997a). The DIT, originally based on Kohlberg's theory, is a probabilistic stage model: the probability of reasoning according to a higher stage of moral judgment increases when the ability of moral

Emanuel van Dongen & Steven Raaijmakers

judgment increases (van den Enden et al., 2019, p. 423). By using a statistical model (*item response theory*), van den Enden et al. (2019) showed that the ordering of the stages fitted the ordering in the underlying stage model well. Furthermore, their findings are compatible with the notion of one latent moral developmental dimension and support the renewed DIT-2. We used the DIT-2 to measure the pre- and post-level of moral reasoning of the law students.

The DIT-2[5] is only loosely based on Kohlberg.[6] In addition, it has been influenced by schema theory (Rest et al., 1999a). Coming from Kohlberg's ideas and extending these Kohlbergian understandings of 'developmentally sequenced and structured patterns of thought', neo-Kohlbergians call for a theoretically based framework that comprises three schemas (Mayhew et al., 2015, pp. 379-380): personal interest schema, maintaining norms schema and post-conventional schema. DIT items cluster around these three general moral schemas, i.e. developmental constructs: arguments that appeal to personal interests (Personal Interest), to maintaining social laws and norms (Maintaining Norms) or appeal to moral ideals and/or theoretical frameworks for resolving complex moral issues (Post-conventional P-score). Besides the older ways of scoring, a new way of scoring was added to the DIT-2 – the N2-score/index – which, according to Rest et al. (1997b), outperforms the P-index.[7] The N2-score has two parts: the degree to which post-conventional items are prioritized plus the degree to which personal interest items (lower stage items) receive lower ratings than the ratings given to post-conventional items (higher stage items; Rest et al., 1997b).

The DIT-2 questionnaires began with a general introduction to how the questionnaire works, followed by five stories (scenarios) about social problems with questions (which action to take; rating of various issues in terms of importance; ranking of these issues in the order of importance), i.e. the Famine dilemma and the Reporter, School Board, Cancer and Demonstration stories, and concluded with some demographic questions. Respondents had to indicate what they thought should be done in that situation (make an action choice). They then had to rate 12 items representing different issues related to the dilemma in terms of their importance in deciding about the social problem (5-point scale). Finally, they had to indicate which four of those 12 aspects they ranked most (and least) important. After analysing the results, students were given a P-value that reflected the degree to which they used higher order moral reasoning and an N2-score that also reflected the extent to which they rejected ideas because they were simplistic or biased.[8]

---

5   https://ethicaldevelopment.ua.edu/about-the-dit.html.
6   The development of the DIT-2 has to do with the criticisms of Kohlberg's model of moral reasoning. One such criticism was that under Kohlberg's model lay a set of moral values influenced by the work of John Rawls and Immanuel Kant (Rest et al., 1999). So 'better moral reasoning' would depend on these values (Graham et al., 2011). Others claim that there is no link between the levels and specific moral theories (Thoma, Bebeau & Narvaez, 2016). A test that is less dependent on value systems is the Moral Foundations Questionnaire (MFQ; based on Moral Foundations Theory). The MFQ maps out which dimensions people consider important when making moral choices. This questionnaire therefore does not use any scenarios or dilemmas.
7   The new way of scoring does not change the fact that the DIT is still based on the stage-typed instruments like the DIT-1 (Rest et al., 1999b; van den Enden et al., 2019).
8   See also www.liberalarts.wabash.edu.

4.4. Ethics and Data

Before the data collection began, the Ethical Committee of Utrecht University's Faculty of Law, Economics and Governance approved our research proposal, under the condition that we would set up an informed consent providing students with information about the study.[9] The storage of (personal) data was done in a digital safe managed by Utrecht University, to which only the two researchers who participated in this research had access. Yoda (drive) was chosen as the most secure option for storage, where the data would be stored for a period of ten years. The data collected via Qualtrics was exported to the SPSS, anonymized and sent to the Center for the Study of Ethical Development of the University of Alabama, which converted it into score reports. Data transfer was done via a shared drop box and deleted immediately afterwards. Further data analysis was subsequently carried out based on the new (converted) SPSS files. The data (in word (interviews) and surveys (raw anonymous data in Excel files and analysis in score reports (pdfs and SPSS files)) is stored in Yoda.

5. Results

5.1. Results Based on the Defining Issues Test

Here we report the results of our analyses, investigating the effects of four different teaching methods on students' moral reasoning. For analysis, only students who participated in both pre- and post-test were included (see Table 2).

Table 2 *Participants' information: Total number of students in pre-test or post-test (N (before exclusion)), number of students in both pre- and post-test (N (after exclusion)), mean age and gender percentage of number of students in both pre- and post-test (N (after exclusion))*

Pilot	N (Before Exclusion)	N (After Exclusion)	Mean Age	Gender (% Female)
I	20	6	24.3	83
II	2I	6	24.8	33
III	28	I7	20.4	88
IV	8	6	22.8	83
V	37	8	20.5	50

Table 3 shows the starting point of development for students, the means in pre-measurement of their post-conventional scores (P-score) and their N2-score, i.e. the degree to which post-conventional items are prioritized plus the degree to which personal interest items (lower stage items) receive lower ratings than the ratings given to post-conventional items (higher stage items). The N2-scores and P-scores are highly intercorrelated (Bebeau & Thoma, 2003). Our P and N2 results

---

9 The suggestion to translate the surveys from an American to a Dutch context was not accepted, since a section of the respondents were non-Dutch, English-speaking or followed a course or programme taught in English.

Emanuel van Dongen & Steven Raaijmakers

are in line with the DIT scores (US) college students have (in the 40s; Bebeau & Thoma, 2003, p. 8). Dong (2011) reports an average P-score of 35.09 for US undergraduates (SD = 15.21) and 41.06 for US graduates (SD = 15.22) and an average N2-score of 34.76 for US undergraduates (SD = 15.45) and 41.33 for US graduates (SD = 14.57).

Table 3   *Initial means of levels of moral reasoning, measured in post-conventional scores (P) and N2-scores, amount per group before ethical courses (pilots) started*

Pilot	P-score	N2-score
I	40	44
II	48.3	48
III	47.2	47
IV	42.2	44
V	43	43.4

By means of a paired t-test we tested whether the pre- and post-measurement (N2, stages 2-3, stage 4 and P-scores) were equal. A significant result would indicate a difference in pre- and post-measurement. The N2-scores provide insight into the overall level of moral reasoning, and the stage 2/3, stage 4, and P-scores are meant to differentiate between the different developmental schemas. The results are shown in Table 4. Given the small sample sizes, we examined the normality. Three times the assumption of normality was violated (Shapiro Wilk test), and in those instances we performed the nonparametric Wilcoxon signed rank test. In none of the cases did we find a result other than that coming from the parametric tests.

Table 4   *Mean difference (rounded) between post- and pre-measurements (in N2, stage 2/3, stage 4 and P-scores); amounts per group (n), mean (M) with standard deviation (SD). Significant results are indicated by an asterisk (*).*

		M	SD
Pilot I (n = 6)	N2-score	4.33	11.22
	Stage 2/3	0.667	8.27
	Stage 4	-5.33	9.09
	P-score	7.33	15.88
Pilot II (n = 6)	N2-score	8.64 *	5.27
	Stage 2/3	-1.00	6.16
	Stage 4	-12.67 *	10.78
	P-score	10.67 *	7.12
Pilot III (n = 17)	N2-score	1.33	9.14
	Stage 2/3	1.41	10.07

Table 4   (continued)

		M	SD
	Stage 4	-0.706	14.70
	P-score	0.235	14.76
Pilot IV (n = 6)	N2-score	2.88	16.11
	Stage 2/3	-1.63	11.03
	Stage 4	3.34	22.80
	P-score	0.102	22.81
Pilot V (n = 8)	N2-score	1.78	8.38
	Stage 2/3	8.75	17.04
	Stage 4	-13.00 *	8.94
	P-score	5.50	10.52

Unfortunately, in all pilots, the remaining groups, i.e. students that participated in both the pre- and post-measurements, were quite small. Conclusions therefore need to be taken with caution. Nevertheless, an interesting overall observation is that stage 4 results, which focus on maintaining social laws and norms decline (in all pilots, except for pilot 4) – even significantly in pilots 2 and 5 (with a decrease of 12.67 respectively 13.00), while the post-conventional levels, i.e. a person's sense of morality as defined in terms of more abstract principles and values, increase. Also, when all pilots are taken together, looking at changes in scores after ethics teaching, we see an overall trend/rise in N2, i.e. the degree to which post-conventional items are prioritized plus the degree to which personal interest items (lower stage items) receive lower ratings than the ratings given to post-conventional items (higher stage items). This gives us an indication that our teaching (methods) on legal ethics had a positive effect. Although some significant differences in the pre- and post-scores were found, these should be interpreted with utmost care. However, for future research it can be interesting to explore these results further.[10] These results might indicate that students' sense of morality is defined in terms of more abstract principles and values increased, their focus on maintaining social laws and norms declined, and they prioritized post-conventional items more in combination with giving lower ratings to personal interest items than the ratings given to post-conventional items. This preference for higher stage items, which was significant although for a small group, can also be seen as a (positive) trend across all the pilots taken together. Further research could investigate these results further.

10   Significant differences in the pre- and post-scores were found for pilot 2 (except for the differences in stages 2/3): a significant increase for the N2-scores at the beginning (M = 47.76, SD = 8.67) and the end of the course (M = 56.40, SD = 8.93); t(5) = -4.014, p = 0.010; a significant decrease for the stage 4 scores at the beginning (M = 28.33, SD = 17.18) and the end of the course (M = 15.67, SD = 9.158); t(5) = 2.877, p = 0.035; a significant increase for the P-scores at the beginning (M = 48.33, SD = 9.416) and the end of the course (M = 59.00, SD = 10.020); t(5) = -3.671, p = 0.014.

5.2. *Experiences of the Teachers or Coordinators Involved*

We interviewed at least one teacher per pilot, and interviews took between 30 and 45 minutes. Owing to the coronavirus pandemic, interviews were conducted through MS Teams. The primary research question in these semi-structured interviews was 'What are [according to the teachers] the effective elements of their courses?' First, teachers were asked to give a brief overview of the course they taught. Next, they were asked to elaborate on the (in their view) effective elements of the course. At the end of the interview, teachers were asked to comment on what made the effective element hard to implement or what could be certain important preconditions for the element to be effective. The interviews were analysed using a process of open-coding. This resulted in seven effective elements (strengths; Table 5) that were mentioned at least by two teachers.

The semi-structured interviews (N = 7) can be summarized in the following five main strengths and difficulties experienced by the teachers (see Table 5).

Table 5   *Strengths and difficulties mentioned by the teachers in the semi-structured interviews*

Strengths	Difficulties
Giving examples from practice, from own experience (ethical conflicts), story telling	Course should be taught by or should involve teachers with experience in practice
Teaching students to recognize problematic situations	Students should discuss dilemmas or experience them in practice
Interacting and reflecting	Teachers should ask students their moral views and their chosen way of action; preparation of students
Acting as a role model, showing steps in teacher's thinking	Teachers should openly discuss their own experience as practitioners
Connecting theory with practice and practice with theory	Although practitioners who (co-)teach a course often put less emphasis on didactics, teachers should link examples from practice with (ethical) theory

6.   Discussion and Conclusion

We began this study with the statement that most literature on (the effectiveness of) teaching methods for moral education of university students is Anglo-American and that as yet no study of this topic had been conducted on Dutch law students. Therefore, their level of moral judgment and whether this changes throughout their university education was unknown. In particular, our research questions were as follows: I. Do the four teaching methods at hand have a positive effect on the development of moral reasoning of law students? II. What perceptions do teachers have concerning the strong points of the pilots in terms of their influence on students' moral thinking, and what conditions are necessary to come to fruition? We discuss these matters now.

Not much is known about what types of moral education for law students are effective. So measuring the effects of moral education on Dutch law students, as we have done using the DIT, is an innovative step. It is not often used, particularly in

the Netherlands, let alone by law lecturers. Our expectation that students' moral reasoning would improve when actively involved in solving issues and more time-on tasks than when passively and/or briefly working on ethical issues could be true considering that the pilots did show a numerical improvement in the desired direction, i.e. higher stage ways of moral thinking. However, owing to the low number of respondents per pilot – it is a pity that, even after issuing reminders, only a few students filled in the questionnaire twice – (and the substantial SD), it is not possible to generalize the results. Consequently, our conclusions can only be tentative and must be interpreted with due caution. Therefore, the statistical tests, in general, lacked the power to discern differences between the pre- and post-measurements.

Our results show that Utrecht School of Law students' N2 results were in line with the DIT scores of (US) college students (i.e. in the 40s). According to Bebeau and Thoma (2003), in heterogeneous samples the level of formal education (junior high, senior high, college, graduate) accounts for 30% to 50% of the variance in DIT scores. In our study, the highest means were found by master's students in an international classroom, i.e. a class with students from various school systems and countries of origin (pilot II) and a bachelor's honours (senior) year elective module (pilot III). The overall scores in our study are higher than the mean levels reported for graduate level under US citizens with English as their primary language by Dong (2011). In Anglo-American literature, it has been argued that law students become cynical and uncritical because of their experiences at university – of course, the students measured in our pilots were not 'average', in the sense that three pilots were followed by honours students, one pilot to a selective clinic programme and one pilot to an international master. The scores in our pilot were relatively high and challenge this assumption. Furthermore, the fact that the scores measured were well above the average norm for higher order and more sophisticated moral thinking, even at the undergraduate level, might justify the tentative and suggestive conclusion that a ceiling effect could have made it difficult to increase the scores any further.

We started with various teaching methods involving conversations about moral dilemmas, methods based on hypothetical, narrative or real examples, and experiential learning combined with reflection on one's own (role played/ hypothetical) situation or actual experiences. Although pilot II included in-class reflection papers, it also worked with dilemmas. Therefore, it is not possible to give a conclusive answer as to whether one of them would work sufficiently, even though pilot I only worked with dilemmas. The groups are too small. Certainly because the literature has stated that discussions about moral dilemmas appear to be an effective intervention for the level of moral reasoning (and ethical judgment), further research is needed in this regard. The two pilots that included experiential learning (III and IV), unfortunately, did not give statistically significant and decisive results. This does not mean that this is not a potentially effective intervention but that research on a larger group of students is needed. Nevertheless, our expectation that the third and fourth pilots (examples of experiential learning over a longer period, with – compared with the other pilots – more time-on tasks)

would have the highest score in students' moral development could not be answered based on the data we collected.

Nevertheless, pilot II has the largest positive increase in N2-score (8.64) and P-score (10.67). Owing to this large increase, which was significant (even though this group was also very small), this pilot is interesting. Its result is remarkable as it is opposite to the results of Landsman and McNeel (2004) and Willging and Dunn (1981). Pilot II was an intensive course that was held over three consecutive weeks. This affirms the statement of Bebeau and Thoma (2003) that DIT scores show significant gains owing to moral educational programmes of more than three weeks (if one considers the preparation of the course it did slightly exceed the three weeks; otherwise, their statement is even applicable in our situation for a course of three weeks). Another remarkable difference, compared with the other pilots, is the international classroom where pilot II was held (various (ideological) insights) and the person of the teacher, who had both an ethical/theoretical background and practical experience as a lawyer. Finally, this is the only pilot for which the deliverables (opinion piece, application paper and participation, for which in-class reflection papers are used), including ethical reflections and writing, were part of the assessment. It is common knowledge that assessment is a powerful initiator for learning. One of the exercises was a kind of role-based problem-solving exercise that was coupled with ongoing reflective discourse. Students had to think about their desired future field of employment and position and describe a potential legal-ethical dilemma they could be confronted with in that field and in that position, discuss it with peers and come up with a solution. Lerner (2004) mentions that this is likely to significantly improve the effective involvement of law students in ethics and the mastery of the role of an ethical practitioner.

Our expectation that students' moral reasoning would improve when actively involved in solving issues and more time-on tasks than when passively and/or briefly working on ethical issues could be true considering the fact that pilot II was an intensive course. Furthermore, we suspected that the nature of the task (mandatory or not) and the role of the teacher were important for the way in which students engaged with these matters. This point is tentatively affirmed. The approach taken by the teachers was regarded as important: providing examples from practice, from personal experience (ethical conflicts), teachers should encourage discussions (between students) and should proceed step by step towards the difficult point of a dilemma. Teachers should ask students their moral views and their chosen way of action, making this preparation for students; teachers should openly discuss their own experiences as a practitioner and link examples from practice with (ethical) theory.

In conclusion, further research is needed to provide more insight into what types of moral education (possibly combined) are most effective in developing moral reasoning, preferably through various interventions over a longer period, stretching from the start of the bachelor's degree in law until (post-doctoral training in) practice. However, our study of the pilots has shown that both the design of effective moral education and the measurement of progress by means of DIT are feasible. The next step is to investigate differences between the types of

moral education with larger groups of students in order to be able to reach firm conclusions.

A final, positive, though still tentative, result of these pilots is that there is no sign of cynical or amoral reasoning among (the small group of) Utrecht University law students. Our study might tentatively indicate that the (legal ethics) teaching at Utrecht University is already preventing students from becoming cynical and uncritical. This finding is at odds with the US findings. The reasons for this tentative finding, which might stem from the context of learning and/or cultural differences (e.g. on attitudes towards following rules), as well as the question of how ethical decision-making has developed over the years at law schools in the Netherlands, and also in other countries in Europe, might be an interesting avenue for future research.

References

Bebeau, M. J. & Thoma, S. J. (2003). *Guide for DIT-2. A guide for using the defining issues test, version 2 ('DIT-2') and the scoring service of the center for the study of ethical development*. University of Minnesota: Center for the Study of Ethical Development.

Biggs, J. & Tang, C. (2011). *Teaching for Quality Learning at University* (4th ed.). Maidenhead: Open University Press/McGraw-Hill.

Chapman, J. (2002). Why teach legal ethics to undergraduates? *Legal Ethics, 5*(1-2), 68-89.

Darnell, C., Gulliford, L., Kristjánsson, K. & Paris, P. (2019). Phronesis and the knowledge-action gap in moral psychology and moral education: A new synthesis? *Human Development, 62*(3), 101-129. doi:10.1159/000496136.

Dong, Y. (2011). *Norms for DIT2: From 2005-2009*. Center for the Study of Ethical Development. http://ethicaldevelopment.ua.edu/.

Doyle, E., Frecknall-Hughes, J. & Summers, B. (2009). Research methods in taxation ethics: Developing the defining issues test (DIT) for a tax-specific scenario. *Journal of Business Ethics, 88*, 35-52. doi:10.1007/s10551-009-0101-5.

Ferris, G. (2015). *Uses of value in legal education*. Cambridge: Intersentia.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S. & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology. 101*(2), 366-385. doi:10.1037/a0021847.

Hartwell, S. (1995). Promoting moral development through experiential teaching. *Clinical Law Review, 1*(3), 505-540.

Kohlberg, L. (1981/1984). *Essays on moral development*. San Francisco, CA: Harper & Row.

Landsman, M. & McNeel, S. P. (2004). Moral judgment of law students across three years: Influences of gender, political ideology and interest in altruistic law practice. *South Texas Law Review, 45*, 891-919.

Lerman, L. G. (1998). Teaching moral perception and moral judgment in legal ethics courses: A dialogue about goals. *William & Mary Law Review, 39*(2), 457-488.

Lerner, A. M. (2004). Using our brains: What cognitive science and social psychology teach us about teaching law students to make ethical, professionally responsible choices. *Quinnipiac Law Review, 23*(3), 643-706.

Mayhew, M. J., Pascarella, E., Trolian, T. J., & Selznick, B. S. (2015). Measurements matter: Taking the DIT-2 multiple times and college students' moral reasoning development. *Research in Higher Education, 56*(4), 378-396. doi:10.1007/s11162-014-9348-5.

Rest, J. R. (1984). The major components of morality. In W. Kurtines & J. Gewirtz (Eds.), *Morality, moral development and moral behavior* (pp. 24-38). New York: John Wiley.

Emanuel van Dongen & Steven Raaijmakers

Rest, J. R., Cooper, D., Coder, R., Masanz, J. & Anderson, D. (1974). Judging the important issues in moral dilemmas: An objective measure of development. *Developmental Psychology, 10*(4), 491-501. doi:10.1037/h0036598.

Rest, J., Narvaez, D., Bebeau, M. & Thoma, S. (1999a). A Neo-Kohlbergian approach: The DIT and schema theory. *Educational Psychology Review, 11*(4), 291-324. doi:10.1023/A:1022053215271.

Rest, J., Narvaez, D., Thoma, S. & Bebeau, M. (1999b). DIT2: Devising and testing a revised instrument of moral judgment. *Journal of Educational Psychology, 91*(4), 644-659. doi:10.1037/0022-0663.91.4.644.

Rest, J., Thoma, S. & Edwards, L. (1997a). Designing and validating a measure of moral judgment: Stage preference and stage consistency approaches. *Journal of Educational Psychology, 89*(1), 5-28. doi:10.1037/0022-0663.89.1.5.

Rest, J., Thoma, S. J., Narvaez, D. & Bebeau, M. J. (1997b). Alchemy and beyond: Indexing the defining issues test. *Journal of Educational Psychology, 89*(3), 498-507. doi:10.1037/0022-0663.89.3.498.

Rhode, D. L. (1992). Ethics by the pervasive method. *Journal of Legal Education, 42*(1), 31-56. doi:10.5840/profethics199211/21.

Rhode, D. L. (2007). Teaching legal ethics. *Saint Louis University Law Journal, 51*(4), 1043-1058.

Rhode, D. L. (2009). Legal ethics in legal education. *Clinical Law Review, 16*, 43-56.

Roche, C. & Thoma, S. (2017). Insights from the defining issues test on moral reasoning competencies development in community pharmacists. *American Journal of Pharmaceutical Education, 81*(8), 21-32. doi:10.5688/ajpe5913.

Schlaefi, A., Rest, J. R. & Thoma, S. J. (1985). Does moral education improve moral judgment? A meta-analysis of intervention studies using the defining issues test. *Review of Educational Research, 55*(3), 319-352. doi:10.3102/00346543055003319.

Sheldon, K. M. & Krieger, L. S. (2004). Does legal education have undermining effects on law students? Evaluating changes in motivation, values, and well-being. *Behavioral Sciences and the Law, 22*, 261-286. doi:10.1002/bsl.582.

Thoma, S. J., Bebeau, M. J. & Narvaez, D. (2016). How not to evaluate a psychological measure: Rebuttal to criticism of the defining issues test of moral judgment development by Curzer and colleagues. *Theory and Research in Education, 14*(2), 241-249. doi:10.1177/1477878516635365.

Willging, T. E. & Dunn, T. G. (1981). The moral development of the law student theory and data on legal education. *Journal of Legal Education, 31*, 306-358.

van den Enden, T., Boom, J., Brugman, D. & Thoma, S. (2019). Stages of moral judgment development: Applying item response theory to defining issues test data. *Journal of Moral Education, 48*(4), 423-438. doi:10.1080/03057240.2018.1540973.

van Dongen, E. & Tigchelaar, J. (2021). Professionele ethiek in het academisch juridisch onderwijs. Enige inhoudelijke en didactische aanknopingspunten. *Law and Method*, 1-25. doi:10.5553/REM/.000000.