Appendix S1.

MATERIALS AND METHODS

General concept

The structure of the present study is shown in Fig. S1[1]. The aims of the first part of the study (pre-eczema school studies) were to ensure that the R&L domains were perceived as adequate by experts and consumers, and to train and evaluate participating nurses. In the second part (eczema school studies), the R&L performance was evaluated by using it on children in our regular eczema school. The objective SCORAD was used as gold standard (16). The study was performed in parallel at the Departments of Dermatology, Karolinska University Hospital, Solna and Skåne University Hospital, Malmö, Sweden. The study was approved by the Regional Ethical Review Board in Lund (number 2012/417). Informed consent was obtained from all participants.

Pre-eczema school studies

Content validity. To investigate whether the R&L scale covers adequate domains, i.e. content validity, a questionnaire with 15 different items of eczema was distributed to 45 persons; 11 senior dermatologists and 9 dermatology nurses, 9 adults (25–56 years) and 6 children (8–16 years) with eczema and 10 parents of children aged 0.5–11 years, with eczema. These items (STable I) were chosen based on the literature (17) and the authors' personal experience. Respondents were asked to grade the importance of each item on a Likert scale (0=unimportant, 1=rather unimportant, 2=neither–nor, 3=important, 4=very important). In addition, they were asked to select and

STable I. *Questionnaire to explore the content validity of different eczema score items*

Please consider the importance of the items below, in order to follow the severity of atopic eczema

Completed by: ❑ Patient ❑ Parent ❑ Nurse ❑ Dermatologist
Date…………………………
Age of the patient………………

1	Skin erythema	❑ Very important ❑ Important ❑ Neither-nor ❑ Rather unimportant ❑ Unimportant
2	Eczema oozing	as above
3	Skin scaling	as above
4	Eczema extent	as above
5	Eczema course	as above
6	Itch intensity	as above
7	Skin dryness	as above
8	Skin swelling	as above
9	Skin pain	as above
10	Skin odour	as above
11	Duration of treatment	as above
12	Eczema localization	as above
13	How visible the eczema is for others	as above
14	Number of scratches	as above
15	Number of ulcers	as above

Please write down the five most important items of the fifteen above and rank them.

1. ....
2. ....
3. ....
4. ....
5. ....

rank the 5 most important items. A median rating of "important" or "very important" was required to rate a domain or item as adequate (10).

Training. All evaluators had many years of experience of professional management of AD in children. To ensure that all evaluators reached consensus on how to use the R&L and the objective SCORAD scales a discussion, followed by a training session on 2 children with AD, was held with 2 dermatologists and 5 dermatology nurses participating.

Inter- and intra-observer reliability. After the training session the inter-observer reliability was investigated on 7 children (age range 1.5–12 years) with AD. This was done independently by the 4 nurses who were to run the eczema school, and by 1 dermatologist. Intra-observer reliability was studied using a test-retest procedure, where each nurse separately scored 9 children with AD with R&L twice, at least 30 min apart. The intraclass correlation (ICC) was computed as a measure of inter- and intra-observer reliability. ICC values were interpreted as follows: $<0·20$ (poor), $0·21–0·40$ (fair), $0·41–0·60$ (moderate), $0·61–0·80$ (good), and $0·81–1·00$ (very good) (18).

Eczema school studies

Eczema school. For many years both departments have been running eczema school for children (0–16 years of age) with AD. Therefore we had the opportunity to evaluate the R&L eczema severity score with objective SCORAD as gold standard. In addition, HRQoL instruments (Infants' Dermatitis Quality of Life Index (IDQoL) (19) and Children's Dermatology Life Quality Index (CDLQI) (20)), were used, so that changes in eczema scores could be supported by the corresponding changes in HRQoL. These 2 HRQoL instruments had previously been used in both departments (5, 21). The questionnaires cover the preceding 7 days and have 10 questions, each scoring 0–3, giving a maximum score per questionnaire of 30; the higher the score, the more the QoL is impaired. In addition, the IDQoL includes a question scored separately from the QoL index, dealing with dermatitis severity (0–4) as perceived by the carer.

The inclusion criteria for this part of the study were checked at an introductory visit to a dermatologist; these criteria were: age 2–16 years, fulfilment of the UK Working Party's Diagnostic Criteria for AD (22), willingness to participate, being in command of the Swedish language, and absence of mental or cognitive disability. Exclusion criteria were: eczema in remission and having attended the eczema school within the previous 6 months.

After inclusion (Nov 2012–Nov 2013), the patients were evaluated by the nurses during 3 visits (0 month=baseline, 2 months and 4 months). A total of 104 children (median age 5 years, range 2–15 years, 38 boys) were included, of whom 87 completed the study.

Construct validity. To investigate whether the R&L scale measures the construct as it should, i.e. eczema severity, the convergent and divergent construct validities were calculated. The convergent construct validity describes whether 2 outcome measurements that are presumed to measure the same latent construct are related. The divergent construct validity measures whether 2 outcome measurements that are presumed to measure different constructs are unrelated. The correlation coefficient for an adequate convergence was set at >0.7 and for an adequate divergence at <0.7 (10).

Internal consistency. To investigate the internal consistency of the R&L scale, i.e. the degree to which responses are consistent across the items within a measure, Cronbach's alpha was calculated (23). Cronbach's alpha for an adequate or acceptable internal consistency on a group level was set at ≥0.7 and 0.6–0.69, respectively (10, 23).

Sensitivity to change. The correlation of changes in R&L and objective SCORAD from baseline to 4 months was calculated to explore sensitivity to change. A correlation between 0.6 and 0.8 was acceptable, and a correlation > 0.8 was judged as adequate (10). Moreover, paired t-test was used for analyses of changes from 0 to 4 months in R&L and objective SCORAD scores, respectively. For the R&L scale, a sensitivity to detect a change of 0.8 scale step units was considered adequate and clinically relevant and this value was also used in the power calculation.

Time consumption. The time needed to score with R&L and SCORAD was measured with chronometers and expressed in minutes. A scoring duration < 3 min was rated as adequate and between 3 and 5 min as acceptable in everyday clinical practice and < 7 min and 7–10 min, respectively, for clinical trials (10).

To assess whether there was a difference in time spent on R&L and SCORAD a paired *t*-test was performed.

Statistical analysis

Prior to the study a power calculation (with statistical power=0.9, significance level=0.05, SD=1.6, clinical relevant decrease in R&L score=0.8) showed that 85 patients were required in the eczema school study. The statistical methods for the different quality items are presented adjacent to the description of each variable (see above). Intra-class correlation was computed with ANOVA (18). The changes in HRQoL scores between 0 and 4 months were analysed with paired *t*-tests. $p < 0.05$ were considered statistically significant. Stata 12 (StataCorp LP, College Station, TX, USA) statistical software was used.