Improving the Skin Prick Allergy Test: Exploring the Effect of Operator Dependent Factors of Test Variability

by

Ria Agrawal, Neda Rehan, Darius Walker, and Coco Zhang

Undergraduate Department of Biomedical Engineering

Word Count: 3105

Number of Figures: 3

Number of Tables: 1

Number of Equations: 3

Number of Supplements: 0

Number of References: 15

ADVISOR

Daniel Abebayehu, Department of Biomedical Engineering

Improving the Skin Prick Allergy Test

Ria Agrawal[a]; Neda Rehan[a]; Darius Walker[a]; Coco Zhang[a]; Daniel Abebayehu, PhD[b]

[a] Department of Biomedical Engineering, University of Virginia
[b] Abebayehu Lab, University of Virginia

Abstract

Food allergy is estimated to affect 1 in 10 adults and 1 in 12 children, costing patients in the United States $5.5 billion out-of-pocket annually. Skin prick tests (SPTs) are a quick, inexpensive allergy test to detect type-1 allergens by inserting a small amount of allergen underneath the skin using a lancet. SPTs suffer from high variation between operators, with factors such as depth of penetration, penetration force, and distance between test sites, with an average coefficient variation between tests of 32.25%. The goal of the paper was to explore the effects of operator dependent factors like force applied and angle of application on the effect of allergen delivered. The first results were a comparative study between three common skin prick devices to determine the least variable design, with a null hypothesis that there is no difference between the mean coefficient of variation of devices. While the null hypothesis was unable to be rejected (ANOVA Test, $p = 0.32$), the Duotip was selected as the design with the lowest overall variability. Next, the Duotip was used to explore the effect of angle (35°, 45°, 55°) and force of application (7 mN, 36 mN, and 86 mN) on the amount of allergen delivered, with the null hypothesis that there will no difference in mean between the levels of each factor. The null hypothesis could not be rejected for force (ANOVA Test, $p=0.096$) or angle (ANOVA Test, $p=0.0503$). Although neither factor could statistically explain the variability within the test, the paper goes on to suggest assistive tools to control these factors. More testing must be done to discover the root cause of variability within the SPT. However, this paper leads the charge in establishing consistent, inexpensive and reliable allergy testing for all.

Keywords: Skin Prick Test, Allergy Testing, Diagnostic Accuracy, Diagnostic Devices, Clinical Diagnostics, Medical Device Design, Food Allergy

---

Introduction

Allergy prevalence has risen over decades worldwide, with the CDC reporting a 50% increase in food allergies since the 1990s[1]. Allergies are caused by pathological activation of mast cells, leading to inflammation and possible anaphylaxis[2,3]. Food allergy is estimated to affect 1 in 10 adults and 1 in 12 children, costing patients in the United States $5.5 billion out-of-pocket annually[4,5]. This increase has been considered as an epidemic, highlighting the importance of safe, accessible, and accurate allergy diagnostics for further timely and proper management and treatment. Overdiagnosis of allergy can cause malnutrition, whereas underdiagnosis can result in avoidable allergic reactions and even deaths[4].

Skin prick tests (SPTs) are a quick, inexpensive allergy test used to detect suspected type-1 allergens such as foods, pollen, or dust mites, by inserting a small amount of allergen underneath the epithelial layer of the skin using a lancet[6]. When using SPTs for diagnostics, an allergic reaction is represented by a red welt and flare at the testing site and is quantified by the wheal size, which is the raised, white-edged area surrounding the red center. However, SPTs face a high extent of variation between operators, with factors such as depth of penetration, penetration force[7], and distance between test sites[8]. One study found that the measured sizes from 4 single-spot SPTs had an average coefficient of variation (CV) of 32.25%[9]. While clear standards have not been established, The Childhood Asthma Management Program deemed a coefficient of variation less than 30% to be proficient[8]. Due to the high degree of variability with SPT and the lack of standardization, patients will opt for blood tests for more quantitative allergy diagnostics, which can be more costly and time-consuming. Additionally, allergy tests can have a high false-positive rate, exhibited in a study of 44 children who were avoiding foods due to positive skin or blood allergy tests. After undergoing an oral food challenge, it was found that 93% of avoided foods could, in fact, be tolerated, and that patients were on an overly restrictive diet[10].

We hypothesize that different applied forces and or different angles of application affect the amount of allergen delivered to the skin and increase the variability of SPT results. Therefore, the overall aim was to explore the effect of these operator dependent factors on the

reproducibility of the SPT and create a tool to reduce such variability. This goal has multiple benefits. First, it develops a methodology that assesses and quantifies the effects of different operator-dependent variables for a given SPT device type. Second, a new SPT device prototype will be designed with criteria that inherently alleviates and reduces the test result variability of the various operator-dependent factors to produce optimal, accurate, and reproducible SPT results across individuals who administer the test. Lastly, combining the methodology and new SPT device prototype, a protocol and standardization can be established for SPTs, leading to increased accuracy and reproducible results across test performers. In addition, the standardization can allow for efficient and effective learning and training results for healthcare professionals to produce safe, accurate SPT results. The improved consistency and reliability in SPTs can further promote its accessibility and incentive for the general public to receive allergy testing so that a proper treatment and management plan can be created. The project will contribute to a more safe, accurate, and accessible SPT device and applicable guidelines to achieve standardization and encourage more people to receive allergy testing.

This goal was achieved through 3 phases. A thorough description of the experimental design and materials used can be found in the methods and materials section below. Phase 1 contained comparative testing between market-available skin prick devices, with the aim to confirm and explore the current state of affairs through in vitro testing. This phase A) began the brainstorming process for a new design and B) provided a device with the lowest coefficient of variation. Phase 2 used this device and explored the impact of force of application and angle of application on allergen delivered. The test aimed to eliminate all other potential factors of variation. The goal of this phase was to determine which of the factors contributes to the most variability, such that the designed tool can control for that factor. Phase 3 leverages the results from Phase 2 to suggest possible designs that could reduce the variability of the SPT.

Results

Phase I: Determining variability of current devices

Phase I aimed to identify which device had the lowest variability in the amount of allergen oil delivered. The Coefficient of Variation (CV) of the three devices is summarized in Table 1.

Table 1. Coefficient of Variation (CV) of devices (*n = 4*)

Device	Mean of CV	St. Error of CV
Stallerpoint	0.60	0.23
Duotip	0.32	0.08
Greer Pick	0.78	0.25

Analysis of Variance (ANOVA) testing with the number of groups k=3 and the total number of subjects N=12 computes an F-statistic of 1.31 and a p-value of 0.32. The data, therefore, does not support any significant difference (at $\alpha = 0.05$) between the three devices tested. Without doing a statistical test, however, the Duotip appeared to have the lowest CV compared to the other devices. The Duotip, therefore, was selected for Phase II testing. By calculating the t statistic using an Independent Samples T-test, the confidence level with which this selection was made:

$$t = \frac{\bar{x}_a - \bar{x}_b}{\sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}}}$$

Equation 1a: Formula for a t-statistic using the Independent Samples T-Test

$$t = \frac{0.60 - 0.32}{\sqrt{\frac{0.47^2}{4} + \frac{0.17^2}{4}}}$$

$$t = 1.1$$

$$p\,(t = 1.1,\ dF = 3) = \boxed{0.17}$$

Equation 1b: Calculation of confidence level for the hypothesis that the Duotip has a lower CV of allergen oil delivered than the Stallerpoint

$$t = \frac{0.78 - 0.32}{\sqrt{\frac{0.49^2}{4} + \frac{0.17^2}{4}}}$$

$$t = 1.8$$

$$p\,(t = 1.8,\ dF = 3) = \boxed{0.09}$$

Equation 1c: Calculation of confidence level for the hypothesis that the Duotip has a lower CV of allergen oil delivered than the Greerprick

The Independent Samples T-test evaluated the null hypothesis that the Duotip and Stallerpoint deliver the allergen oil with equal CVs, yielding a p-value of 0.17. For the comparison between the Duotip and Greerpick, the p-value was 0.09. These numbers are higher than the standard of $\alpha \leq 0.05$, making the data insignificant. Calculating these probabilities, however, provides the confidence levels at which the Duotip was selected as the device with the lowest CV to move onto Phase II.

Phase II: Analyzing potential factors contributing to variability

Proceeding with the Duotip, Phase II aimed to quantify how different operator-dependent factors, specifically force and angle applied in SPTs, will affect the amount of allergen delivered. The null hypothesis for the force test states that different forces will result in the same

mean amount of allergen delivered. The amount of allergen deposited onto the decellularized skin hydrogel is quantified by the absorbance of fluorescence-conjugated ovalbumin. The force test was performed with forces of 7 mN, 36 mN, and 86 mN, with the same angle controlled and applied at 45º, as clinically informed and instructed by the Duotip Directions for use[11]. After running the force test with a sample size of N=5, absorbance values were collected from the fluorometer, and the absorbance reading of a blank well in the 3D-printed, customized well plate was subtracted from each data point to account for normalization (Appendix A & Figure 1). Figure 1 shows the mean absorbance from each of the force conditions, and the error bars represent the standard errors.

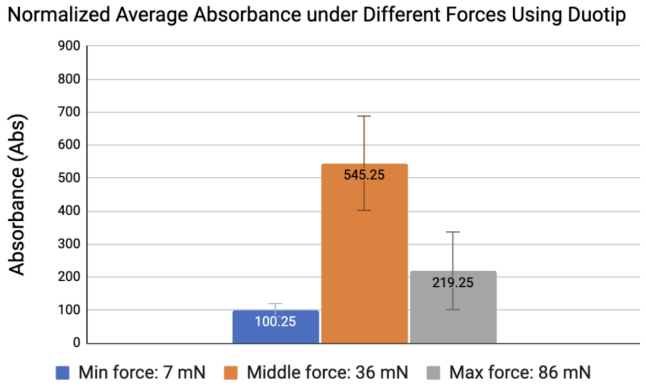

Figure 1. Tests applied under different forces (7, 36, 86 mN) at the same angle (n=5). Mean absorbance ± standard error (error bars).

An ANOVA test was performed after removing outliers, and the p value was calculated to be 0.096 ($p > 0.05$). Therefore, the null hypothesis cannot be rejected to conclude that different forces will result in different amounts of allergen deposited. There was no significant difference between any of the two conditions according to the subsequent Tukey test.

When assessing the impact of varying angles on allergen delivery, the null hypothesis states that the mean amount of allergen delivered remains consistent across all angles. The angle test was performed at angles of 35º, 45º, and 55º, all with controlled forces at 36 mN and a sample size of N=5. After completing the angle test, absorbance readings were collected by using the fluorometer and normalized by subtracting the absorbance value of a blank well from all the original absorbance values (Appendix A & Figure 2). Figure 2 demonstrates the mean absorbance from the different angle conditions, with the error bars representing the standard errors. From the ANOVA test after removing outliers, the p value = 0.0503 ($p > 0.05$). Thus, the null hypothesis is not rejected to conclude that

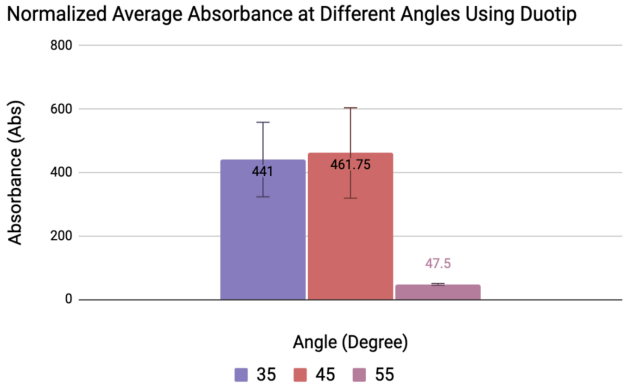

Figure 2. Tests applied at different angles (35, 45, 55 degrees) under the same force of 36 mN (n=5). Mean absorbance ± standard error (error bars).

there is a significant difference in the amount of allergen deposited due to the angle applied.

Additionally, given the extremely low mean amount of allergen delivered in the minimum force condition with 7 mN and in the highest angle condition with 55º, and that both conditions were limited by complications during the experiments, the validity of the data needed further confirmation from future studies. Therefore, t-tests were performed for the rest of the two conditions in the force and angle tests. For the force test, $p = 0.14$ ($p > 0.05$), indicating that the null hypothesis cannot be rejected; thus, there is no significant difference in allergen delivery between the two force conditions (36 mN and 86 mN). For the angle test, $p = 0.41$ ($p > 0.05$), indicating that the null hypothesis cannot be rejected for this test as well; therefore, there is no evidence that the different angles (35º and 45º) affect the amount of allergen deposited.

Discussion

Phase I: Determining variability of current devices

The data from Phase I was inconclusive because no significant differences (at $\alpha = 0.05$) were established, as shown in the Results section. Additional trials should be conducted to increase the sample size (n) and determine if establishing significance is possible. If increasing trials indeed established that the Duotip had a lower CV of amount of allergen delivered than the other SPT devices (at $\alpha = 0.05$), it is possible that the Duotip's unique shape allows for more consistent allergen oil pickup. The Duotip's two prongs are parallel and close together and may potentially pick up allergen oil due to capillary action, whereas the other two devices tested pick up allergen oil

due to simple adherence. To test this hypothesis, devices could be weighed when dry and then again after being dipped in allergen oil. The consistency in weights, or lack thereof, can be related to consistency in allergen oil pickup. The device with the most consistent allergen oil pickup should inform the design of the new device, which should aim to use similar methods for picking up allergen oil, i.e. capillary action.

In the clinic, SPTs are conducted by nurses who are trained to use SPT devices. To improve clinical relevance of this experiment, therefore, trials should be done with nurses rather than engineering students. It is possible that the engineering students, who were not trained in SPT administration, may have produced different results than nurses themselves. Because nurses may produce more consistent results using their device of preference, it is important to control for nurse SPT device preference and experience by encouraging nurses from different clinics and skill levels to conduct the test in future studies.

Phase II: Analyzing potential factors contributing to variability

A major realization during Phase II was the limitations from the experimental designs, partly due to the Instron machine settings. These restraints imply that the validity of some of the data needs to be confirmed either through the repetition of more tests or modification of the experimental designs. Specifically, for the force test, the different values were initially determined as 294 mN, 441 mN, and 589 mN, calculated and referenced from literature[12]. However, these forces exceeded the upper bound allowed in the settings of the Instron machine, as the velocity parameter was used to control for the forces as the independent variable as inputs for the Instron protocol. To address this, a lower set of forces was selected: 7 mN, 36 mN, and 86 mN. However, given the extremely and consistently low absorbance reading from the 7 mN condition, it was likely that the force exerted was so small that there was not enough allergen delivered to be detectable.

Similarly, for the angle test, the 55º condition presented difficulties when using the Instron, because the well plate was raised at an angle so large that it collided with the Instron before the Duotip was long enough to contact the surface of the decellularized skin hydrogel. Therefore, the experimental setup was improvised to extend the Duotip by taping it to a marker so that it could reach the hydrogel before any collision. However, this may have contributed to structural instability in the simulated SPT setup, which was reflected by the extremely low absorbance reading and suggested that insufficient allergen was deposited for it to be accurately detected. In the future, an Instron machine that is able to reach faster speeds should be used, and an Instron attachment that can administer the test without colliding into the well plate should be fabricated, potentially via 3-D printing.

Due to time constraints and limited amounts of available decellularized skin hydrogel kits, the team was unable to repeat more trials with any attempts to modify the experimental process. An alternative platform for testing could have been a cheaper hydrogel, for example alginate, and/or a hydrogel with longer stability at room temperature. Based on these limitations inherent in the experimental design, it was proposed amongst the team that in future studies, the experimental process could be designed without an Instron due to its upper bound constraints and its inability to directly control forces. By addressing the complications that arose from the Phase II experiments, the validity of the absorbance data can be ensured to help draw stronger conclusions of the impacts on allergen delivery due to different operator-dependent factors.

Future Steps – Phase III: Designing an SPT accessory

The next steps of this project are to design an accessory to help improve the application of the SPT for more consistent results. Phase III was conducted in parallel with Phase II to stay on schedule with the project timeline. Although the data was not conclusive enough to determine whether angle or force of application had the greatest impact on the results, this phase proceeded under the assumption that both factors significantly influenced the outcome.

To control for the angle of application, two designs were modeled with CAD to limit the angle at which an SPT device can approach the skin, as shown in Figure 3.

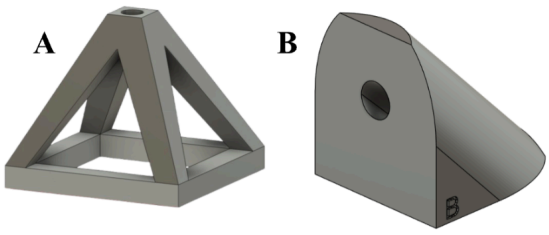

Figure 3. CAD models of potential SPT accessories designed to limit the angle of application to **A.** 90 degrees and **B.** 45 degrees

Both designs feature a predetermined path that a device would follow, ensuring it pricks the skin at the desired angle. These models would contain an adhesive at the bottom, or contain a feature similar to a bandaid strap, to prevent movement of the accessory once on the skin, and would be designed to be removed and reapplied with each application of the device. Although no specific designs were modeled to control for the force of application, inspiration was drawn from the mechanism of a clicker pen. Just as a pen uses a spring mechanism to require a specific amount of force to operate, a similar accessory could be designed with a spring, or flexible plastic, that applies a controlled amount of force to the skin.

Moving forward, the next steps will involve refining these design concepts through prototyping and testing to evaluate their effectiveness in achieving consistent results. Continued research will be essential to optimize both the force and angle control mechanisms, ensuring reliable and repeatable applications of the SPT device

General Reflections

This project had multiple working, often concurrent, parts. One particularly difficult component of experimentation was the hydrogel, which had a short shelf-life and needed to be manufactured within a few hours of the test. This shelf-life meant that experimentation was limited to days when the team had a fully working day available. Furthermore, this shelf-life may have contributed to inconsistent results, especially towards the ends of experiments when the gel began to degrade. Another difficulty of this project was that Phase II was dependent on the results of Phase I. Although the data from Phase I was inconclusive, Phase II needed to be started to finish the project on schedule. Therefore, the Duotip was selected for Phase II despite there being no significant data to support its selection.

To continue this project, the first step would be to design more experiments that isolate operator dependent factors that may cause variation. This can include the amount of allergen picked up on the device or duration of penetration. Future tests can explore more skin prick devices for a better scope of design options. Secondly, Phase I can be repeated to compare the standard Duotip to the Duotip with designed prototypes. This will allow us to continuously test whether the prototypes are improving the variability of the test. Lastly, the designs can be taken into the field to gather user feedback and begin a clinical trial. One key benefit of the skin prick test is it can quickly be performed by nurses. We must ensure that the user population is not encumbered by a new device. Furthermore, by implementing our designs into clinical practice, we can receive more information of the root of variation within the test and make larger strides to improving the reliability of skin prick test.

Materials and Methods

Materials

To conduct this experiment, three skin prick devices were used: 1) the QUINTIP® (also called the Stallerpoint) manufactured by HollisterStier Laboratories LLC,[13] 2) the Greer®Pick® manufactured by GREER® Laboratories, Inc.[14] and 3) the Duotip-Test®, manufactured by Lincoln Diagnostics, Inc.[11] All devices were obtained from the University of Virginia Hospital. These devices were used on a decellularized skin ECM hydrogel from Xylyx Bio[15] to simulate the properties of skin, which was prepared using the recommended instructions at 6 mg/mL. Fluorescence-conjugated ovalbumin was used to simulate allergen oil, and the concentration of fluorescence present in wells was calculated using a fluorometer.

Custom 48-well plates with 45° angled openings were 3D printed and used for devices that required angled application. Traditional 96-well plates were used at all other instances. Custom angled stands were printed to accommodate the application angles of 35, 45, and 55 degrees in Phase II. All 3D printed parts were designed using Fusion360 and printed in either Stacey Hall or Clemons Library at the University of Virginia. An Instron machine was used to control the speed and depth of application during Phase II. All tests were performed in Dr. Daniel Abebayehu's lab, or in IDEAS Lab in Thornton Hall at the University of Virginia.

Methods

In Phase I, the three skin prick devices were tested to determine which of them had the least variability between operators. Decellularized skin ECM hydrogel was prepared in both the 96-well and custom 48-well plates. Three different operators followed the protocols as described in the instructions with the three different devices, dipping the tip of the device into the ovalbumin and then pricking the hydrogel. The plate was then inserted into a fluorometer and the absorbance values were read. The coefficient of variation (CV) was calculated for each of three trials, with each trial consisting of all three operators' attempts to account for the variability between operators.

In Phase II, the Duotip was used to test the effect of angle and force of application on the volume of allergen delivered into the skin. Decellularized skin ECM hydrogel was prepared in the custom 48-well plate. For all tests in Phase II, the Duotip was dipped into the ovalbumin and then attached to the clasp of the instron. For the angle test, in which the force was held constant while varying the angle, the three angled stands were placed under the well plates to provide the appropriate angle. The Instron machine was used to lower the Duotip to a specific depth into the hydrogel at a constant speed. Four trials were conducted for the angle test. A supplemental set up using a marker as an extension was required for the test at 55° due to length restraints and collision prevention. For the force test, in which the angle was held constant while varying the force, the Instron machine was used to adjust the application speed, as it could not directly control the application force. A calculated speed was selected to correspond to the desired force level. Five trials were conducted for the force test. The plates were inserted into a fluorometer and the absorbance values were read. The volume of ovalbumin deposited into the hydrogel was calculated.

End Matter

Acknowledgement

We would like to thank the Abebayehu Lab and the University of Virginia Undergraduate Biomedical Engineering Capstone Team for their support.

References

1. Digging Up the Roots of Food Allergies | NIH Intramural Research Program. Accessed November 14, 2024. https://irp.nih.gov/blog/post/2023/05/digging-up-the-roots-of-food-allergies

2. P K, D EAK, M M, F S, M M. Understanding human mast cells: lesson from therapies for allergic and non-allergic diseases. *Nat Rev Immunol*. 2022;22(5). doi:10.1038/s41577-021-00622-y

3. Food Allergies | Pediatrics In Review | American Academy of Pediatrics. Accessed October 6, 2024. https://publications.aap.org/pediatricsinreview/article/41/8/403/35426/Food-Allergies?autologincheck=redirected

4. Lk T, P D. Food allergy in the World Health Organization's International Classification of Diseases (ICD)-11. *Pediatr Allergy Immunol Off Publ Eur Soc Pediatr Allergy Immunol*. 2022;33(11). doi:10.1111/pai.13882

5. Gupta R, Holdford D, Bilaver L, Dyer A, Holl JL, Meltzer D. The Economic Impact of Childhood Food Allergy in the United States. *JAMA Pediatr*. 2013;167(11):1026-1031. doi:10.1001/jamapediatrics.2013.2376

6. Frati F, Incorvaia C, Cavaliere C, et al. The skin prick test. *J Biol Regul Homeost Agents*. 2018;32(1 Suppl. 1):19-24.

7. Andersen HH, Lundgaard AC, Petersen AS, et al. The Lancet Weight Determines Wheal Diameter in Response to Skin Prick Testing with Histamine. *PLOS ONE*. 2016;11(5):e0156211. doi:10.1371/journal.pone.0156211

8. Fatteh S, Rekkerth DJ, Hadley JA. Skin prick/puncture testing in North America: a call for standards and consistency. *Allergy Asthma Clin Immunol*. 2014;10(1):44. doi:10.1186/1710-1492-10-44

9. Carr WW, Martin B, Howard RS, Cox L, Borish L, the Immunotherapy Committee of the American Academy of Allergy A and I. Comparison of test devices for skin prick testing. *J Allergy Clin Immunol*. 2005;116(2):341-346. doi:10.1016/j.jaci.2005.03.035

10. Fleischer DM, Bock SA, Spears GC, et al. Oral Food Challenges in Children with a Diagnosis of Food Allergy. J Pediatr. 2011;158(4):578-583.e1. doi:10.1016/j.jpeds.2010.09.027

11. Lincoln Diagnostics, Inc. Duotip Directions for Use. Published online March 2020.

12. Chiaranairungroj M, Chatchatee P, Srituravanich W. The effect of applied force and device design on skin prick test performance. *Ann Allergy Asthma Immunol*. 2023;130(3):312-316. doi:10.1016/j.anai.2022.11.014

13. HollisterStier. QuinTIp Directions for Use. Published online 2/09. www.hsallergy.com

14. Greer Laboratories, Inc. Greer Pick Directions for Use. Published online February 2016.

15. XYLYX Bio. XYLYX Skin dECM Fact Sheet.

Appendix A: Raw data from Phase II

Force test (applied with consistent angle of 45º):

Force (mN)	Raw Absorbance (Abs)					Mean (Abs)	Standard Deviation (Abs)	Standard Error (Abs)
7	63	91	91	156	342	100.25	39.44	19.72
36	124	714	646	644	177	545.25	285.46	142.73
86	467	576	90	73	138	219.25	235.09	117.55

Angle test (applied with consistent force of 36 mN):

Angle (Degrees)	Absorbance (Abs)					Mean (Abs)	Standard Deviation (Abs)	Standard Error (Abs)
35	416	239	606	184	735	441.00	235.06	117.53
45	149	522	835	317	173	461.75	285.27	142.64
55	18	41	44	53	52	47.50	5.92	2.96